

LA INTELIGENCIA ARTIFICIAL EN LOS PROCESOS DE EVALUACIÓN EN EDUCACIÓN SUPERIOR: UN ANÁLISIS BIBLIOMÉTRICO (2014-2024)

Artificial intelligence in assessment processes in higher education: a bibliometric analysis (2014-2024)

ALBA GALÁN-ÍÑIGO¹, JUDIT RUIZ-LÁZARO² Y EVA JIMÉNEZ-GARCÍA¹

¹ Universidad Europea de Madrid (España)

² Universidad Nacional de Educación a Distancia (España)

DOI: 10.13042/Bordon.2025.107797

Fecha de recepción: 14/6/2024 • Fecha de aceptación: 12/11/2024

Autora de contacto / *Corresponding author*: Alba Galán-Íñigo. E-mail: alba.galan@universidadeuropea.es

Cómo citar este artículo: Galán-Íñigo, A., Ruiz-Lázaro, J. y Jiménez-García, E. (2025). La inteligencia artificial en los procesos de evaluación en educación superior: un análisis bibliométrico (2014-2024). *Bordón, Revista de Pedagogía*, 77(3), 131-154. <https://doi.org/10.13042/Bordon.2025.107797>

INTRODUCCIÓN. Para la educación superior, la inteligencia artificial (IA) representa una innovación significativa que está transformando la enseñanza, el aprendizaje y la evaluación académica. Esta última es esencial para proporcionar retroalimentación al alumnado, evaluar su comprensión y favorecer la mejora continua del aprendizaje. El objetivo del presente estudio se centra en analizar la producción científica relativa a la IA en la evaluación en educación superior durante la última década. Este análisis abarca la productividad cronológica y geográfica, así como un estudio detallado de las fuentes, palabras clave y el número de citas de los artículos más destacados en este ámbito. **MÉTODO.** Para ello, se realizó una revisión bibliométrica y sistemática siguiendo las indicaciones de la Declaración PRISMA. **RESULTADOS.** Inicialmente se identificaron 2.275 estudios en la base de datos *Web of Science* y tras aplicar los criterios de elegibilidad se seleccionaron 130 estudios para su posterior análisis. **DISCUSIÓN.** Las investigaciones revisadas concluyen que las principales formas de integración de la IA en el proceso de evaluación incluyen el *feedback* automatizado, la predicción del rendimiento académico mediante análisis de datos basados en IA, el uso de modelos de lenguaje como ChatGPT y la consideración de cuestiones éticas asociadas. Se observa un notable incremento en el volumen de publicaciones durante el período 2023-2024, con Estados Unidos liderando la producción científica y la revista *Education Sciences* siendo la principal fuente de difusión. El análisis de co-palabras revela una falta de uniformidad terminológica, lo que sugiere la necesidad de estandarizar el lenguaje para mejorar la claridad en el campo.

Palabras clave: *Inteligencia artificial, Evaluación, Evaluación educativa, Educación superior.*

Introducción

La integración de la inteligencia artificial (en adelante, IA) en el ámbito educativo está transformando los métodos de enseñanza y evaluación (Jiménez-García *et al.*, 2024; Ruiz-Lázaro, 2024). Según Fischer *et al.* (2020) y Ruiz-Lázaro *et al.* (2024), la evolución tecnológica y la creciente disponibilidad de datos han permitido el desarrollo de sistemas inteligentes capaces de analizar y generar información relevante para la toma de decisiones en el ámbito educativo. En particular, en la educación superior, la IA ha demostrado ser una herramienta muy útil que ayuda al alumnado a mejorar los resultados de aprendizaje y permite a los docentes optimizar sus estrategias de enseñanza (Chu *et al.*, 2022). La integración de la IA abre la puerta a formas de evaluación más dinámicas, adaptativas y precisas con el propósito de mejorar tanto la evaluación formativa como la sumativa en el ámbito universitario.

Los métodos tradicionales de evaluación pueden presentar limitaciones en términos de personalización, eficiencia y adaptabilidad a las necesidades individuales de los estudiantes (Huang *et al.*, 2021). Sin embargo, la IA ofrece una solución al personalizar la evaluación según el perfil y desempeño de cada estudiante. De esta manera, los sistemas de evaluación basados en IA pueden ajustar la dificultad y el contenido de las preguntas en función de los conocimientos previos y las habilidades de cada estudiante, optimizando el tiempo dedicado a la evaluación y proporcionando una experiencia más significativa (Luckin, 2017).

Además, las tecnologías de IA integradas en el proceso de evaluación están siendo desarrolladas con diversos propósitos. Por un lado, la IA puede mejorar la precisión y relevancia de las evaluaciones al analizar múltiples fuentes de datos, como el progreso académico previo, las preferencias de aprendizaje o los patrones de participación e interacción de los estudiantes. Por otro lado, estudios recientes, como los de Sailer *et al.* (2023) y Bauer *et al.* (2023), demuestran la capacidad de la IA para proporcionar orientación y retroalimentación inmediatas que favorecen el aprendizaje.

La necesidad de realizar un análisis bibliométrico sobre este tema deriva de la creciente importancia que la IA ha adquirido en el ámbito educativo. La tendencia observada en los últimos tres años muestra un aumento sustancial en la producción académica sobre este tema. Este fenómeno resalta la urgencia de evaluar críticamente la calidad y la dirección de la investigación existente, así como de identificar las áreas de oportunidad y los desafíos que conlleva esta transformación. En esta línea, se observa la necesidad de realizar un estudio que permita analizar aspectos como la productividad cronológica, el país de origen, las revistas que más han publicado, los autores destacados, las palabras clave más utilizadas y las temáticas más recurrentes en la literatura científica.

Por ello, el objetivo general del presente estudio se centra en analizar bibliométricamente la producción científica existente sobre la integración de la IA en la evaluación en educación superior en la última década. Los objetivos específicos son:

- Analizar la productividad cronológica de los estudios de la última década (2014-2024).
- Evaluar la producción científica en función del país de publicación, la revista, el autor y las citas recibidas.
- Determinar las fuentes y autores más influyentes en el campo de la IA en la evaluación en educación superior.
- Identificar las tendencias temáticas y las palabras clave más utilizadas mediante el análisis de co-ocurrencia de palabras.

Método

Esta revisión sistemática y bibliométrica se ha llevado a cabo siguiendo las directrices establecidas en la Declaración PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses, Urrútia y Bonfill, 2010) para asegurar un enfoque estructurado, riguroso y transparente en el proceso de investigación (Uman, 2011). Específicamente, se han seguido las etapas propuestas por Velt *et al.* (2020), garantizando un proceso metódico y exhaustivo. Estas etapas incluyen: (1) la formulación de los objetivos de investigación, (2) la identificación de datos relevantes a través de bases de datos bibliográficas, (3) la selección de datos utilizando criterios específicos para evaluar la pertinencia y validez de la información recopilada, (4) la confirmación de los datos mediante la revisión y verificación de la información seleccionada, y (5) el análisis de datos aplicando técnicas estadísticas y herramientas informáticas para interpretar y presentar los resultados.

La base de datos seleccionada para la búsqueda fue Web of Science (WoS) debido a su confiabilidad y autoridad. Para optimizar la relevancia, este estudio utilizó varios grupos de palabras clave para buscar artículos sobre IA en la evaluación de la educación superior, utilizando la frase de búsqueda que se observa en la Tabla 1.

TABLA 1. Frase de búsqueda

Base de datos	Frase de búsqueda
Web of Science (Fecha de búsqueda: mayo 2024)	(“artificial intelligence” OR “machine intelligence” OR “machine learning” OR “AIEd”) AND (“higher education” OR “graduate” OR “undergraduate” OR “university”) AND (“evaluation” OR “assessment” OR “feedback”)

Fuente: Elaboración propia.

La ventana temporal seleccionada para este análisis bibliométrico abarca desde 2014 hasta 2024, asegurando así la actualidad y precisión de la producción científica. Este enfoque temporal contextualiza históricamente el avance tecnológico ocurrido en torno a la integración de la IA en el ámbito educativo.

Para ser incluido en esta revisión sistemática, cada estudio debía cumplir con los criterios indicados en la Tabla 2.

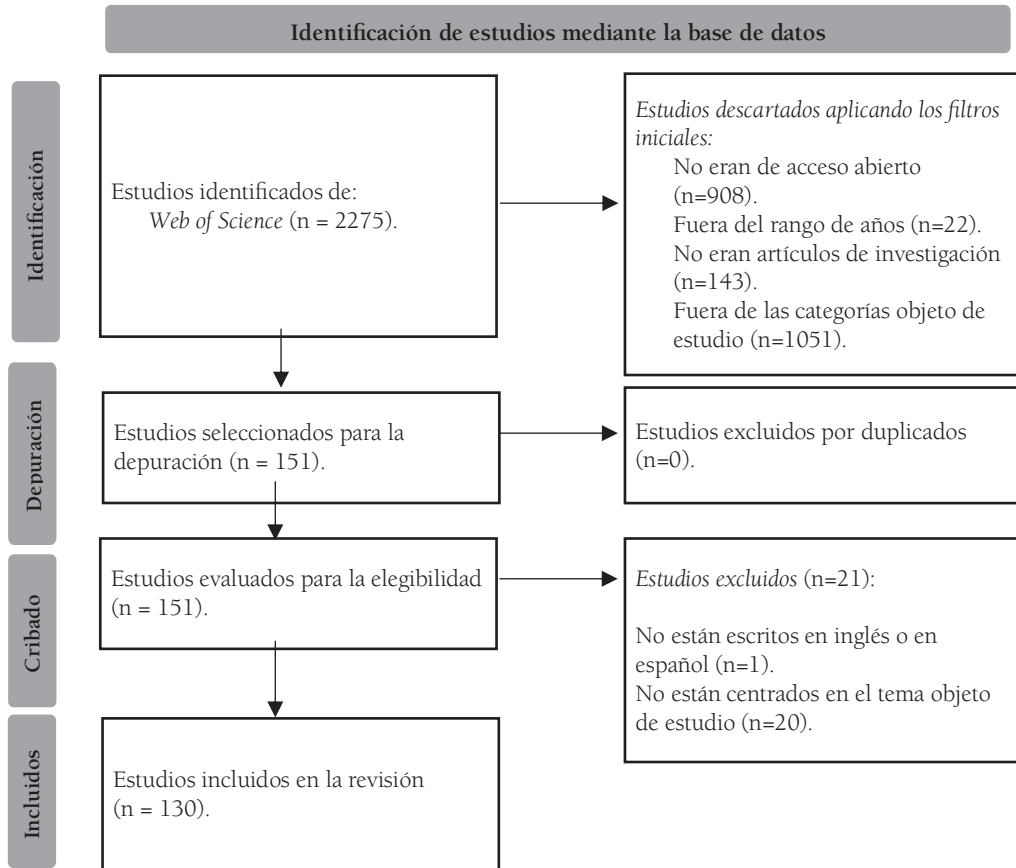
TABLA 2. Criterios de inclusión y exclusión

Criterios de inclusión	Criterios de exclusión
Aplicación de la IA en la evaluación de Educación superior.	No relacionados con la inteligencia artificial en la evaluación de la educación superior.
Artículos de investigación publicados en revistas científicas.	Tesis doctorales, libros o artículos sin una sección sobre metodología.
Artículos escritos en inglés o español.	Artículos escritos en otro idioma diferente al inglés o español.
Desde el 2014 hasta la actualidad (2024).	Artículos publicados antes del 2014.

Fuente: Elaboración propia.

Esta búsqueda arrojó un total de 2.275 resultados, y tras aplicar los criterios de selección, se incluyeron 130 artículos, tal y como queda reflejado en la Figura 1.

FIGURA 1. Diagrama de flujo



Fuente. Elaboración propia adaptado de Page et al. (2021).

Conocido el total de estudios incluidos en la revisión (Anexo 1) se llevó a cabo un análisis bibliométrico apoyado en técnicas estadísticas, ofreciendo objetividad, precisión y rigor. Este análisis ha utilizado datos bibliográficos tales como el número de citas, las publicaciones o los términos más usados para evaluar la importancia y el impacto de trabajos específicos en este campo. Para ello, se ha utilizado el programa VOSviewer aplicando el método Association Strength para medir la fuerza de asociación basada en el peso de la cita. Además, se han creado mapas de red de co-palabras donde las líneas o conexiones entre los nodos son de mayor grosor cuanto más fuertes son las relaciones de co-ocurrencia entre ellos.

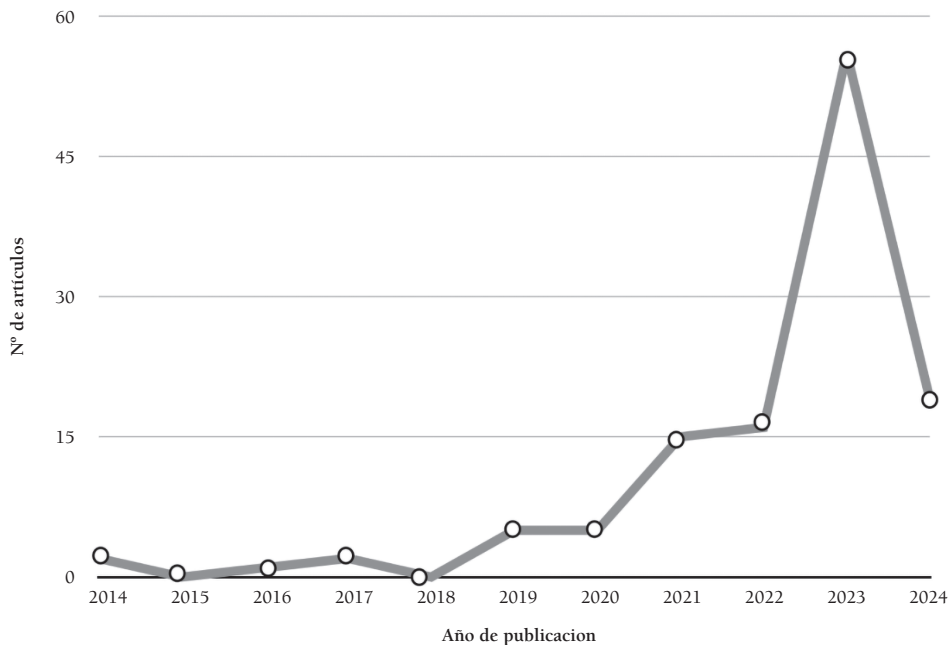
Resultados

A continuación, se presentan los resultados para dar respuesta tanto al objetivo general como a los objetivos específicos.

Productividad cronológica

En la Figura 2 se presenta la productividad científica entre los años 2014 y 2024, destacando dos años clave (2021 y 2023) en los que se observa un incremento significativo de publicaciones con respecto al año anterior. En el 2021 vemos que se publicó el 11.5% de la producción total de documentos, duplicando la producción del año anterior (5.3%). En 2023 la producción experimentó un notable aumento pasando de un 16% al 43.8%. Entre enero y mayo de 2024, se han publicado 20 artículos y se espera que esta cifra continúe creciendo a lo largo del año.

FIGURA 2. Productividad cronológica de publicaciones

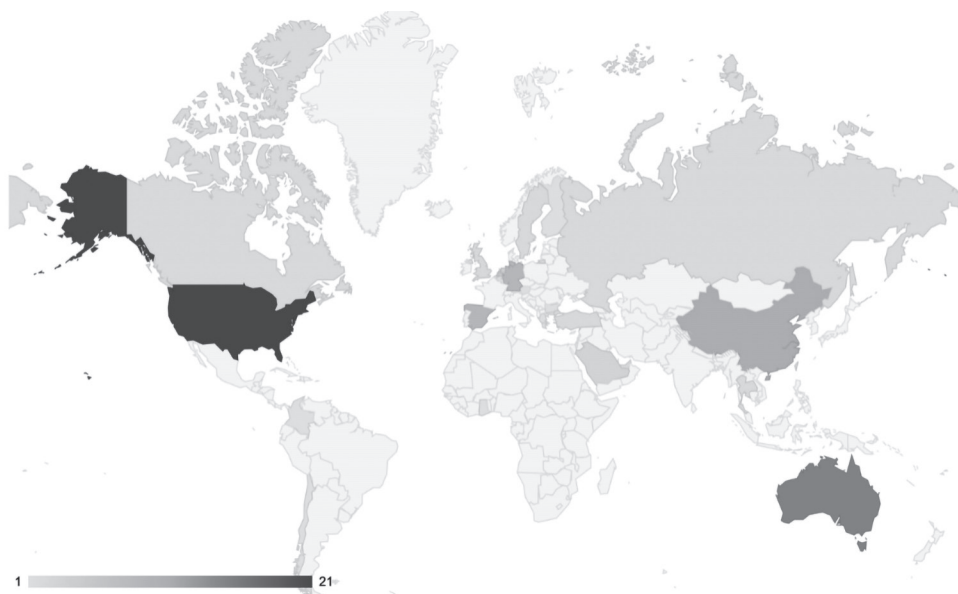


Fuente: Elaboración propia.

Producción geográfica

En la Figura 3, se muestra la productividad científica en función de la ubicación geográfica, destacando Estados Unidos con 24 publicaciones, seguido de Australia con 22, Alemania con 14, Inglaterra con 13, España con 12, China con 9, Canadá con 8 y Singapur con 4.

FIGURA 3. Productividad en función de la ubicación geográfica de las publicaciones



Fuente: Elaboración propia.

La productividad abarca un total de 51 países, de los cuales 43 tienen tres o menos publicaciones; es decir, un 84% de la productividad está concentrada en los ocho países mencionados anteriormente.

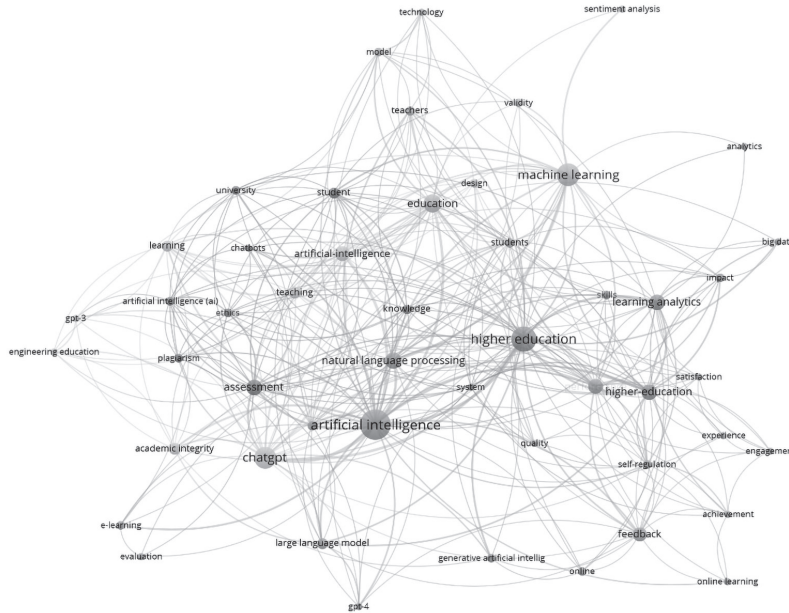
Análisis de co-términos

En este análisis se presentan las palabras clave extraídas de los estudios evaluados con el fin de identificar los principales temas de investigación sobre la IA en la evaluación en el ámbito de la educación superior.

Se analizaron las asociaciones de co-ocurrencia entre cada par de palabras clave. Inicialmente, se trabajó con un conjunto de 680 términos, de los cuales 50 cumplieron con el umbral predefinido de relevancia (al menos tres apariciones). Como se observa en la Figura 4, aparecen cinco nodos principales que destacan por su alta frecuencia de aparición en los estudios. Estos términos se presentan, a continuación, siguiendo un orden descendente de relevancia, y habiendo sido traducidos al castellano: “inteligencia artificial” (48 citas), “educación superior” (31 citas), “chat GPT” (24 citas), “aprendizaje automático” (27 citas) y “educación” (16 citas).

Los resultados arrojados en este análisis generaron siete grupos de nodos, también conocidos como clústeres. En la Tabla 3 se presentan las diferentes redes generadas, proporcionando información específica sobre la categoría temática de cada nodo, las palabras clave incluidas en cada clúster y la palabra más frecuente en cada uno.

FIGURA 4. Mapa de VosViewer de relaciones de palabras clave



Fuente: Elaboración propia.

TABLA 3. Agrupaciones de palabras clave en función de la co-ocurrencia (al menos tres)

Clúster	Categoría temática	Palabras claves	Palabra más citada
Clúster 1	Evaluación y ética	Artificial intelligence (ai), assessment, chatbots, ethics, knowledge, plagiarism, student, system, university	Assessment (13 veces)
Clúster 2	Retroalimentación del rendimiento académico	Achievement, engagement, experience, feedback, online learning, performance, quality satisfaction, self-regulation	Performance (12 veces)
Clúster 3	Análisis de datos en Educación superior	Analytics, big data, generative artificial intelligence, higher education, higher-education, impact, learning analytics, online, skills	Higher education (31 veces)
Clúster 4	Integridad con ChatGPT	Academic integrity, artificial- intelligence, ChatGPT, design, engineering education, GPT-3, learning, teaching	ChatGPT (24 veces)
Clúster 5	Inteligencia artificial y PLN	Artificial intelligence, e-learning, evaluation, GPT-4, large language model, medical education, natural language processing	Artificial intelligence (48 veces)
Clúster 6	Aprendizaje automático en la educación	Education, machine learning, sentiment analysis, validity	Machine learning (27 veces)
Clúster 7	Educación y tecnología	Model, students, teachers, technology	Student (5 veces), teachers (5 veces)

Fuente: Elaboración propia.

Este análisis refleja la relación entre las palabras clave empleadas por diversos autores. No obstante, se observa una discrepancia actual en la denominación de los términos que ha afectado los resultados obtenidos. Por un lado, algunas palabras clave han sido descartadas por no contar con tres o más coincidencias, a pesar de referirse al mismo concepto. Por otro lado, algunos términos que inicialmente tenían pocas co-ocurrencias mostraron una frecuencia significativamente mayor al unificarse los conceptos. Dado que la investigación en IA aplicada a la educación es un campo relativamente nuevo que ha experimentado un crecimiento exponencial en los últimos dos años, se hace necesario unificar los 680 términos de los 130 artículos para obtener resultados más consistentes y realistas.

En la Tabla 4 se presenta el término general junto con todos los términos similares que aluden al mismo concepto, aunque con algunos matices, así como la frecuencia con la que se repiten.

TABLA 4. Palabras clave utilizadas para definir el mismo término

Término general	Términos relacionados	Total
Artificial intelligence (Inteligencia artificial)	Artificial intelligence (48), artificial-intelligence (12), Artificial intelligence in education, ai (2), artificial intelligence in education (aied), artificial intelligence education, artificial intelligence (ai), explainable artificial intelligence	67
Assessment (Evaluación)	Assessment (13), Assessments, evaluation (2), automated assessment, automated assessment of understanding, performance assessment, computer-aided assessment, course evaluation, performance evaluation, reflective assessment, student evaluation, summative assessment, formative assessment (4), computer aided assessment., assessment for learning, authentic assessment, ongoing assessment task, course evaluation, short answer assessment, teaching and learning assessment, exams, online tests, educational evaluation, student evaluation, evaluation system, assess, evaluation (3), assessment for learning	46
Higher education (Educación superior)	Higher education (10), tertiary education, university education, higher-education (13), higher education institutions, higher-education institutions, universities, university teaching, online higher education, university (4), undergraduate education (2), higher institutions (he is), university education	38
Chagpt	chatgpt (24), chat gpt, gpt-3 (3), gpt- 3.5, gpt 4 (3), openai, chatgpt chatgpt3, gps	33
Machine learning (Aprendizaje automático)	Machine learning (27), Machine learning models, machine learning application, machine learning (ml), machine learning algorithms, machine_learning	32
NLP (Procesamiento de lenguaje natural)	Large language model (6), natural language processing (14), natural language processing (nlp) (2), ai-powered language tools	23
E-learning (Aprendizaje en línea)	E-learning (4), Online learning (3), digital learning, distance education, mobile learning, virtual learning environments, virtual classroom, distance learning, e-learning systems, online learning spaces, interactive learning environments, multiuser virtual environment, telelearning, online (4)	22
Feedback (Retroalimentación)	feedback (10), automated feedback (2), informative feedback, interactions with feedback, machine feedback, adaptive feedback, feedback literacy, ai-generated feedback, feedback sources, state-feedback control.	20

TABLA 4. Palabras clave utilizadas para definir el mismo término (cont.)

Término general	Términos relacionados	Total
E-learning (Aprendizaje en línea)	E-learning (4), Online learning (3), digital learning, distance education, mobile learning, virtual learning environments, virtual classroom, distance learning, e-learning systems, online learning spaces, interactive learning environments, multiuser virtual environment, telelearning, online (4)	22
Performance (Rendimiento)	Performance (12), student academic performance, performance assessment, performance evaluation, student performance, students' performance, academic performance, work performance, academic-performance	20
Data analysis (Análisis de datos)	Data expedition, data science, trace data, clickstream Data, data mining (3), education data mining (2), educational data mining (2), text mining (2), data visualization, analytic rubrics, analytics (3), automated analysis, task analysis, data mining techniques,	21
Ethics (Ética)	Academic dishonesty, ethics (4), ethics in ai, dishonest practices, academic integrity (7), academic misconduct, ethical IA	16

Fuente: Elaboración propia.

Ley zonal de Bradford

Para identificar las revistas con mayor producción de actividad científica relacionada con la IA integrada en el proceso de evaluación en educación superior, se utiliza la división por zonas de Bradford. Según Bradford, es posible organizar las revistas en orden decreciente en función de la cantidad de artículos que estas han publicado sobre un tema determinado (Ardanuy, 2012). Tras la distribución equitativa de los artículos en diferentes zonas, se evidencia un conjunto de revistas “núcleo” que concentran la mayor cantidad de publicaciones, junto con varias zonas que contienen un número similar de documentos producidos que siguen una progresión matemática aproximada a la forma (Bradford, 1934). Esta ley prueba que el número de revistas ubicadas fuera del núcleo tienden a aumentar y, a su vez, producen una cantidad menor de publicaciones a medida que se alejan de la Zona 1. Actualmente, según investigaciones recientes, este continúa siendo un método destacado de clasificación de artículos de revistas en las ciencias sociales (Palomo *et al.*, 2017).

En este caso, con la finalidad de observar la distribución de las revistas productoras, se ha utilizado la clasificación zonal de Bradford, optando por un total de tres zonas. La Tabla 5 muestra que la Zona 1 o “núcleo” representa el 33.8% de los artículos producidos (42), que pertenecen a cinco revistas. También encontramos una segunda zona, que representa el 31.5% de los artículos publicados (39) en 16 revistas, y una tercera zona que representa el 34.6% de las publicaciones (43), que pertenecen a 44 revistas. En este caso, la progresión obtenida ($n=2.98$) se encuentra lejos de la propuesta del modelo original de Bradford. Es decir, podemos concluir que no se sigue una progresión geométrica cuadrática ($n=2$) como la propuesta por Bradford. Las revistas que forman parte del núcleo incluyen: *International Journal of Educational Technology in Higher Education*, *Education and Information Technologies*, *International Journal of Emerging Technologies in Learning y Education Sciences*.

TABLA 5. Reparto de publicaciones según la ley de Bradford

Zonas	Revista (total)	Revistas (%)	Artículos (total)	Artículos (%)	Multiplicador de Bradford
Zona 1 (Núcleo)	5	7.7 %	42	33.8 %	-
Zona 2	16	24.6 %	39	31.5 %	3.2
Zona 3	44	67.7 %	43	34.6 %	2.75
Total	65	100 %	77	100 %	n=2.98
Revistas que forman parte del núcleo					
	Revista	Nº de artículos			
	<i>Education Sciences</i>	11			
	<i>International Journal of Educational Technology in Higher Education</i>	10			
	<i>Education and information Technologies</i>	9			
	<i>Australasian Journal of Educational Technology</i>	6			
	<i>Cogent education</i>	6			

Fuente: Elaboración propia.

A continuación, se exponen las temáticas principales y el factor de impacto (de ahora en adelante FI) de las tres revistas más relevantes que conforman la Zona 1 o núcleo. *Education Sciences* es una revista revisada por pares cuyo objetivo es difundir hallazgos de investigación de alta calidad en una amplia gama de temas educativos; su FI, según *Journal's Impact*, es de 3.0. *International Journal of Educational Technology in Higher Education* es una revista académica revisada por pares que busca contribuir al avance del conocimiento científico sobre tecnología en la educación superior; su FI, según *Journal's Impact*, es de 7.66. *Education and Information Technologies* es una revista que explora la relación entre las tecnologías de la información y la comunicación y la educación, con un enfoque que va desde aplicaciones específicas en el aula hasta preocupaciones a nivel nacional, abarcando diversos niveles educativos; su FI según *Journal's Impact* es de 5.5.

Análisis de citación de referencias

Los estudios con el mayor número de citas y, por lo tanto, de mayor relevancia en este dominio dentro de la comunidad científica se reflejan en la Tabla 3. En ella se presenta el título de los artículos, el autor, el año de publicación y el número de citas, fijando el mínimo en 40.

En el artículo escrito por Crawford *et al.* (2019), con 58 citas, se plantean los riesgos y oportunidades que presenta la herramienta de ChatGPT. A pesar de mencionar la preocupación por el plagio, el estudio resalta los beneficios de esta herramienta para mejorar los resultados académicos y facilitar la evaluación auténtica, que busca que los estudiantes apliquen sus conocimientos en situaciones o problemas del mundo real.

El artículo de Williams (2019), con 52 citas, analiza cómo la tecnología *blockchain* está transformando la educación superior y los métodos de evaluación. Este estudio propone una estrategia en la que el énfasis de la evaluación no se base exclusivamente en la disciplina académica, sino en la formación integral del individuo para que lleguen preparados a su futuro profesional.

El artículo escrito por Fergus *et al.* (2021), con 50 citas, destaca la capacidad de ChatGPT para responder a las preguntas de química en un programa de ciencias farmacéuticas. Se encontró que ChatGPT puede generar respuestas para preguntas que requieren conocimiento y comprensión, pero tiene limitaciones para preguntas que necesitan aplicar conocimientos e interpretar información no textual.

TABLA 6. Análisis de citación referencias

Artículo	Autores	Año	Citas
Leadership is needed for ethical ChatGPT: Character, assessment, and learning using artificial intelligence (AI)	Crawford, Cowling, Allen	2023	58
Does competency-based education with blockchain signal a new mission for universities	Williams	2019	52
Evaluating Academic Answers Generated Using ChatGPT	Fergus, Botha, Ostovar	2023	50
Chatbot to improve learning punctuation in Spanish and to enhance open and flexible learning environments	Vázquez-Cano, Esteban; Mengual-Andrés, Santiago; López-Meneses, Eloy	2021	45
Experiment-Based Approach to Teach Optimization Techniques	Precup, Hedrea, Bojam-Dragos	2021	42

Fuente: Elaboración propia.

Temáticas destacadas

A partir del análisis de los 130 estudios seleccionados, se realizó una clasificación por temáticas para determinar los asuntos de mayor relevancia en la integración de la IA en el proceso de evaluación en educación superior. A continuación, se presentan los temas más relevantes junto al porcentaje de estudios que los abordan.

TABLA 7. Temáticas destacadas

Temática	% art
Feedback automatizado (técnicas PLN)	14.6%
Predicción del rendimiento académico a través del análisis de datos sobre la interacción y comportamiento de los estudiantes	14.6%
ChatGPT en la evaluación	13 %
Aspectos éticos en la evaluación con IA	10 %
Detección de las dificultades académicas	6.9%
Evaluación del desempeño docente	6.9%
Chatbot/asistente virtual/STI	6.9%
Calificación automática	4.6%
Itinerarios de aprendizaje personalizado en las evaluaciones	4.6%

Fuente: Elaboración propia.

El porcentaje restante de temáticas no se encuentra en la Tabla 6 por no ser temas que hayan sido objeto de estudio en más de dos artículos. Entre ellos se encuentran estudios relacionados con la simulación y realidad virtual en la evaluación (Liaw *et al.*, 2023; Sailer *et al.*, 2023), la evaluación por pares mediante tecnología de aprendizaje automático (Bauer *et al.*, 2023; Rico-Juan *et al.*, 2019) o la vigilancia en la evaluación a través de sistemas de IA (Sefcick *et al.*, 2022), entre otros.

Discusión y conclusiones

El objetivo general del presente estudio se centra en analizar bibliométricamente la producción científica sobre la integración de la IA en la evaluación en educación superior desde 2014 hasta 2024.

Para ello, se ha realizado el análisis de la productividad cronológica, que ha arrojado resultados reveladores en cuanto a la alta tasa de publicaciones en el 2023, siendo el año en el que se ha publicado casi la mitad de la producción total de documentos. De acuerdo con las afirmaciones de varios autores como Banerjee *et al.* (2023), el lanzamiento público de ChatGPT en noviembre del 2022 impulsó este aumento, al generar un mayor interés en la investigación sobre la IA en la evaluación educativa. Este fenómeno ha motivado a algunos investigadores a diseñar modelos y sistemas de evaluación basados en IA con el propósito de mejorar el proceso de evaluación y aprendizaje, en contraste con enfoques centrados exclusivamente en los aspectos asociados al plagio, a la ética o al temor a que la IA reemplace a los docentes (Cotton *et al.*, 2023; Huallpa, 2023).

En cuanto a la localización geográfica, Estados Unidos es el país con mayor producción de artículos sobre el tema. La productividad analizada abarca un total de 51 países, lo que refleja una importante diversidad geográfica. La disparidad en la producción científica sugiere discrepancias en el nivel de compromiso y enfoque investigativo entre naciones. Este panorama evidencia un interés global en torno a la evaluación en educación superior con IA. Sin embargo, se hace necesaria la colaboración internacional para abordar los desafíos y maximizar los beneficios que esta sinergia puede aportar al campo educativo a nivel mundial.

Tras el análisis de los co-términos se identificaron los términos más recurrentes, entre los que destacan: “inteligencia artificial”, “aprendizaje automático”, “educación superior” y “procesamiento de lenguaje natural”. No obstante, los resultados pueden generar cierta ambigüedad debido a la falta de consenso entre los diferentes autores a la hora de referirse a algunos conceptos. Si estos términos se hubieran empleado de manera uniforme, los resultados obtenidos habrían sido distintos, ya que algunos han sido expresados de manera tan dispar que no han sido considerados en el análisis al no cumplir con el requisito mínimo de contar con al menos tres coincidencias. Este fenómeno se observa particularmente en aquellos términos relacionados con la predicción del rendimiento, la evaluación docente o la satisfacción del estudiante. Un término relevante es “minería de datos”, que inicialmente apareció con tres coincidencias; sin embargo, al consolidar todos sus sinónimos, se alcanzaron 21 menciones. Una situación similar se observó en términos como “educación superior”, que incrementó de 10 a 37, o “ética”, que aumentó de 7 a 16 menciones. En cuanto a “evaluación”, se evidenció una diferencia significativa, apareciendo 13 veces en el análisis original y aumentando a 46 menciones tras la unificación de términos. Esta discrepancia en la definición de algunos términos sugiere la necesidad de establecer consensos

con el fin de unificar las temáticas de estudio, así como analizar las palabras clave a través de la función Keywords Plus para arrojar términos más descriptivos y precisos que las palabras clave aportadas por los autores (Zhang *et al.*, 2015). Asimismo, se sugiere que futuros estudios puedan aplicar métodos cualitativos para desambiguar el significado de las palabras clave en investigaciones bibliométricas similares.

El análisis del reparto de publicaciones siguiendo la ley zonal de Bradford ha identificado las cuatro revistas que actualmente publican más sobre el objeto de estudio. En este caso, la progresión obtenida (n=2.98) se encuentra lejos de la propuesta del modelo original de Bradford (n=2). Esto puede explicarse por el auge de monográficos que algunas revistas han publicado sobre la integración de la IA en educación entre los años 2021-2024. A la luz de los resultados, se hace notable el alto número de revistas que publican en torno a esta temática buscando promover e impulsar la investigación en este ámbito en constante expansión. Sin embargo, se sugiere que en futuras investigaciones pueda debatirse una revisión y ajuste de la distribución de las zonas para mejorar la precisión y adaptabilidad del modelo, teniendo en cuenta las características específicas del área de estudio.

En cuanto al análisis de citación de referencias, se han identificado los tres artículos más citados que han sido objeto de estudio en el marco de esta investigación. Estos artículos destacan la necesidad de replantear y adaptar los métodos de evaluación para responder a las capacidades y limitaciones de herramientas de IA como ChatGPT. Para ello, será fundamental que la educación superior ponga énfasis en el desarrollo de competencias integrales y habilidades de pensamiento superior como el análisis crítico o la resolución de problemas. La alta frecuencia de citaciones de estos artículos en comparación con sus competidores posiciona el uso de ChatGPT en la evaluación como una de las tendencias temáticas destacadas, lo que ha quedado reflejado en este análisis. Asimismo, se ha concluido que el *feedback* automatizado con técnicas del Procesamiento de Lenguaje Natural (Afzaal *et al.*, 2023; Banihashem *et al.*, 2024) y la predicción del rendimiento académico a través del análisis de la interacción y participación de los estudiantes (Fahd *et al.*, 2022; Liu *et al.*, 2023) han sido las temáticas más recurrentes. Los estudios más recientes de 2024 se centran principalmente en el papel de las herramientas de IA generativa, como ChatGPT, en la evaluación educativa (Saleem *et al.*, 2024; Tarisayi, 2024), así como en las cuestiones éticas relacionadas con la integridad académica y el plagio (Lesage *et al.*, 2024). No obstante, diferentes autores señalan las oportunidades y el potencial que tienen estas herramientas en el futuro de la educación (Williams, 2024).

Para concluir, se recomienda que futuras investigaciones puedan ampliar o reajustar los criterios de selección con la finalidad de observar otros resultados —afines o no a la presente investigación— y así reflejar la naturaleza multidimensional del campo de estudio. Por último, se sugiere ampliar la búsqueda en otras bases de datos de prestigio como Scopus, Dialnet o Scielo y llevar a cabo actualizaciones periódicas que permitan evidenciar y difundir las nuevas investigaciones en torno a esta temática.

Los hallazgos de esta investigación ponen de manifiesto el rápido crecimiento de la producción científica sobre el uso de la IA en la evaluación en educación superior, lo que refleja no solo el interés académico, sino también el impacto que puede llegar a tener el uso de la IA en los procesos evaluativos en educación superior.

Notas

Este estudio forma parte de una tesis doctoral que investiga la integración de la inteligencia artificial en la educación superior, específicamente en el área de Educación de la Universidad Europea de Madrid.

Referencias bibliográficas

- Afzaal, M., Zia, A., Nouri, J., y Fors, U. (2023). Informative Feedback and Explainable AI-Based Recommendations to Support Students' Self-regulation. *Technology, Knowledge and Learning*, 29(1), 331–354. <https://doi.org/10.1007/s10758-023-09650-0>
- Ardanuy, J. (2012). *Breve introducción a la bibliometría*. Universitat de Barcelona. Recuperado de: <http://diposit.ub.edu/dspace/bitstream/2445/30962/1/breve%20introduccion%20bibliometria.pdf>
- Banerjee, P., Srivastava, A., Adjeroh, D., Reddy, Y. R., y Karimian, N. (2023). *Understanding Chat-GPT: Impact Analysis and Path Forward for Teaching Computer Science and Engineering*. TechRxiv. <https://doi.org/10.36227/techrxiv.22639705.v1>
- Banihashem, S., Kerman, N., Noroozi, O., Moon, J., y Drachsler, H. (2024). Feedback sources in essay writing: Peer-generated or AI-generated feedback? *International Journal of Educational Technology in Higher Education*, 21(23), 1–15. <https://doi.org/10.1186/s41239-024-00455-4>
- Bauer, E., Greisel, M., Kuznetsov, I., Berndt, M., Kollar, I., Dresel, M., Fischer, M. R. R., y Fischer, F. (2023). Using natural language processing to support peer-feedback in the age of artificial intelligence: A cross-disciplinary framework and a research agenda. *British Journal of Educational Technology*, 54, (5) 1222–1245 <https://doi.org/10.1111/bjet.13336>
- Bradford, S. C. (1934). Sources of information on specific subjects. *Engineering*, 137, 85-86.
- Chu, H. C., Hwang, G. H., Tu, Y. F., y Yang, K. H. (2022). Roles and research trends of artificial intelligence in higher education: A systematic review of the top 50 most-cited articles. *Australasian Journal of Educational Technology*, 38(3), 22–42. <https://doi.org/10.14742/ajet.7526>
- Cotton, D. R., Cotton, P. A., y Shipway, J. R. (2023). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 61(2), 228–239. <https://doi.org/10.1080/14703297.2023.2190148>
- Crawford, J., Cowling, M., y Allen, K. A. (2023). Leadership is needed for ethical ChatGPT: Character, assessment, and learning using artificial intelligence (AI). *Journal of University Teaching & Learning Practice*, 20(3), 1–21. <https://doi.org/10.53761/1.20.3.02>
- Fahd, K., Venkatraman, S., Miah, S., y Ahmed, K. (2022). Application of machine learning in higher education to assess student academic performance, at-risk, and attrition: A meta-analysis of literature. *Education and Information Technologies*, 27(3), 3743–3775. <https://doi.org/10.1007/s10639-021-10741-7>
- Fergus, S., Botha, M., y Ostovar, M. (2023). Evaluating academic answers generated using ChatGPT. *Journal of Chemical Education*, 100(4), 1672–1675. <https://doi.org/10.1021/acs.jchemed.3c00087>
- Fischer, C., Pardos, Z. A., Baker, R. S., Williams, J. J., Smyth, P., Yu, R., y Warschauer, M. (2020). Mining big data in education: Affordances and challenges. *Review of Research in Education*, 44(1), 130–160. <https://doi.org/10.3102/0091732X20903304>

- Huallpa, J. J. (2023). Exploring the ethical considerations of using ChatGPT in university education. *Periodicals of Engineering and Natural Sciences*, 11(4), 105–115. <http://dx.doi.org/10.21533/pen.v11.i4.200>
- Huang, J., Saleh, S., y Liu, Y. (2021). A review on artificial intelligence in education. *Academic Journal of Interdisciplinary Studies*, 10(3), 206–217. <https://doi.org/10.36941/ajis-2021-0077>
- Jiménez-García, E., Orenes-Martínez, N., y López-Fraile, L. A. (2024). Rueda de la Pedagogía para la Inteligencia Artificial: adaptación de la Rueda de Carrington. *RIED-Revista Iberoamericana de Educación a Distancia*, 27(1), 87–113. <https://doi.org/10.5944/ried.27.1.37622>
- Lesage, J., Brennan, R., Eaton, S., Moya, B., Mcdermott, B., Wiens, J., y Herrero, K. (2024). Exploring natural language processing in mechanical engineering education: Implications for academic integrity. *International Journal of Mechanical Engineering Education*, 52(1), 88–105. <https://doi.org/10.1177/03064190231166665>
- Liaw, S. Y., Tan, J. Z., Lim, S., Zhou, W., Yap, J., Ratan, R., Ooi, S. L., Wong, S. J., Seah, B., y Chua, W. L. (2023). Artificial intelligence in virtual reality simulation for interprofessional communication training: Mixed method study. *Nurse Education Today*, 122, 105718. <https://doi.org/10.1016/j.nedt.2023.105718>
- Liu, Y., Fan, S., Xu, S., Sajjanhar, A., Yeom, S., y Wei, Y. (2023). Predicting Student Performance Using Clickstream Data and Machine Learning. *Education Sciences*, 13(1), 17. <https://doi.org/10.3390/educsci13010017>
- Luckin, R. (2017). Towards artificial intelligence-based assessment systems. *Nature Human Behaviour*, 1(3), 0028. <https://doi.org/10.1038/s41562-016-0028>
- Palomo, J., Figueroa-Domecq, C., y Laguna, P. (2017). Women, peace and security state-of-art: a bibliometric analysis in social sciences based on SCOPUS database. *Scientometrics*, 113(1), 123–148. <https://doi.org/10.1007/s11192-017-2484-x>
- Rico-Juan, J. R., Gallego, A.-J., y Calvo-Zaragoza, J. (2019). Automatic detection of inconsistencies between numerical scores and textual feedback in peer-assessment processes with machine learning. *Computers & Education*, 140, 103609, 1–16. <https://doi.org/10.1016/j.compedu.2019.103609>
- Ruiz-Lázaro, J., Jiménez-García, E., y Huetos-Domínguez, M. (2024). Revisión sistemática sobre el uso de la tecnología en educación y el compromiso de los estudiantes en la última década. *Campus Virtuales*, 14(1), 139–152. <https://doi.org/10.54988/cv.2025.1.1318>
- Ruiz-Lázaro, J. (2024). El aliado invisible: la inteligencia artificial revoluciona la preparación de la evaluación para el acceso a la universidad (EvAU). *Padres y Maestros / Journal of Parents and Teachers*, (398), 19–26. <https://doi.org/10.14422/pym.i398.y2024.003>
- Sailer, M., Bauer, E., Hofmann, R., Kiesewetter, J., Glas, J., Gurevych, I., y Fischer, F. (2023). Adaptive feedback from artificial neural networks facilitates pre-service teachers' diagnostic reasoning in simulation-based learning. *Learning and Instruction*, 83, 101620, 1–10. <https://doi.org/10.1016/j.learninstruc.2022.101620>
- Saleem, N., Mufti, T., Sohail, S., y Madsen, D. (2024). ChatGPT as an innovative heutagogical tool in medical education. *Cogent Education*, 11(1), 1–9. <https://doi.org/10.1080/2331186X.2024.2332850>
- Sefcik, L. T., Veeran-Colton, T., Baird, M., Price, C., y Steyn, S. (2022). An examination of student user experience (UX) and perceptions of remote invigilation during online assessment. *Australasian Journal of Educational Technology*, 38(2), 49–69. <https://doi.org/10.14742/ajet.6871>
- Tarisayi, K. (2024). ChatGPT use in universities in South Africa through a socio-technical lens. *Cogent Education*, 11(1), 1–10. <https://doi.org/10.1080/2331186X.2023.2295654>

- Uman, L. S. (2011). Systematic Reviews and Meta-Analyses. *Journal of the Canadian Academy of Child & Adolescent Psychiatry*, 20(1), 57–59.
- Urrútia, G., y Bonfill, X. (2010). Declaración PRISMA: una propuesta para mejorar la publicación de revisiones sistemáticas y metaanálisis. *Medicina Clínica*, 135(11), 507-511.
- Velt, H., Torkkeli, L., y Laine, I. (2020). Entrepreneurial ecosystem research: Bibliometric mapping of the domain. *Journal of Business Ecosystems (JBE)*, 1(2), 43-83. <https://doi.org/10.4018/JBE.20200701-oa1>
- Williams, R. (2024). The ethical implications of using generative chatbots in higher education. *Frontiers in Education*, 8, 1–8. <https://doi.org/10.3389/educ.2023.1331607>
- Zhang, J., Yu, Q., Zheng, F., Long, C., Lu, Z., y Duan, Z. (2016). Comparing keywords plus of WOS and author keywords: A case study of patient adherence research. *Journal of the Association for Information Science and Technology*, 67(4), 967–972. <https://doi.org/10.1002/asi.23437>

Anexo 1. Referencias incluidas en la revisión

Autor/es	Año	Título
Abbot, K. L., George, B. C., Sandhu, G., Hardbaugh, C. M., Gauger, P. G., Ötles, E., Matusko, N., y Vu, J. V.	2021	Natural Language Processing to estimate Clinical Competency Committee Ratings
Abid, A., Murugan, A., Banerjee, I., Purkayastha, S., Trivedu, H., y Gichoya, J.	2024	AI Education for Fourth-Year Medical Students: Two-Year Experience of a Web-Based, Self-Guided Curriculum and Mixed Methods Study
Afzaal, M., Zia, A., Nouri, J., y Fors, U.	2023	Informative Feedback and Explainable AI-Based Recommendations to Support Students' Self-regulation
Ahmed, H. M. M., y Sorous, S. E.	2024	Classification-driven intelligent system for automated evaluation of higher education exam paper quality
Al-Alawi, L., Al Shaqsi, J., Tarhini, A., y Al-Busaidi, A. S.	2023	Using machine learning to predict factors affecting academic performance: The case of college students on academic probation
Ali, K., Bahrom, N., Tamimi, F., y Duggal, M.	2024	ChatGPT: A double-edged sword for healthcare education? Implications for assessments of dental students.
Almasre, M	2024	Development and Evaluation of a Custom GPT for the Assessment of Students' Designs in a Typography Course.
Archibald, A., Hudson, C., Heap, T., Thompson, R. R., Lin, L., DeMeritt, J., y Lucke, H.	2023	A Validation of AI-Enabled Discussion Platform Metrics and Relationships to Student Efforts
Banihashem, S. K., Kerman, N. T., Noroozi, O., Moon, J., y Drachsler, H.	2024	Feedback sources in essay writing: peer-generated or AI-generated feedback?
Baragash, R. S., Aldowah, H., y Umar, I. N.	2022	Students' perceptions of E-Learning in Malaysian Universities: Sentiment analysis based Machine Learning Approach
Baran, E., Alzoubi, D., y Morales, A. S.	2023	Design and Implementarion of an Automated Classroom Analytics System:Skateholder Engagement and Mapping Author.
Barret, A., y Pack, A.	2023	Not quite eye to AI: student and teacher perspectives on the use of generative artificial intelligence in the wiring process.
Bauer, E., Greisel, M., Kuznetsov, I., Berndt, M., Kollar, I., Dresel, M., Fischer, M. R. R., y Fischer, F	2023	Using natural language processing to support peer-feedback in the age of artificial intelligence: A cross-disciplinary framework and a research agenda
Bearman, M., Tai, J., Dawson, P., Boud, D., y Ajjawi, R.	2024	Developing evaluative judgement for a time of generative artificial intelligence.
Bekmanova, G., Ongarbayev, Y., Somzhurek, B., y Mukatayev, N.	2021	Personalized training model for organizing blended and lifelong distance learning courses and its effectiveness in Higher Education

Autor/es	Año	Título
Bernard, J., Sonnadara, R., Saraco, A. N., Mitchell, J. P., Bak, A. B., Bayer, I., y Wainman, B. C.	2023	Automated grading of anatomical objective structured practical examinations using decision trees: An artificial intelligence approach.
Berriri, M., Djema, S., Rey, G., y Dartigues-Pallez, C.	2021	Multi-Class Assessment Based on Random Forests.
Bertolini, R., Finch, S. J., y Nehm, R. H.	2023	An application of Bayesian inference to examine student retention and attrition in the STEM classroom
Cebrian-Robles, V., Ruiz-Rey, F. J., Raposo-Rivas, M., y Cebrian-de-la-Serna, M.	2023	Impact of Digital Contexts in the Training of University Education Students
Chan, C. K. Y.	2023	A comprehensive AI policy education framework for university teaching and learning
Chaudhry, I. S., Sarwary, S. A. M., El Refae, G. A., y Chabchoub, H.	2023	Time to Revisit Existing Student's Performance Evaluation Approach in Higher Education Sector in a New Era of ChatGPT - A Case Study
Chiu, T. K. F.	2023	The impact of Generative AI (GenAI) on practices, policies and research direction in education: a case of ChatGPT and Midjourney
Chrysafiadi, K., Virvou, M., Tsihrintzis, G. A., y Hatzilygeroudis, I.	2023	Evaluating the user's experience, adaptivity and learning outcomes of a fuzzy-based intelligent tutoring system for computer programming for academic students in Greece.
Cianciolo, A. T., LaVoie, N., y Parker, J.	2021	Machine Scoring of Medical Students' Written Clinical Reasoning: Initial Validity Evidence
Cowling, M., Crawford, J., Allen, K. A., y Wehmeyer, M.	2023	Using leadership to leverage ChatGPT and artificial intelligence for undergraduate and postgraduate research supervision
Crawford, J., Vallis, C., Yang, J. H., Fitzgerald, R., y O'Dea, C.	2023	Editorial: Artificial Intelligence is Awesome, but Good Teaching Should Always Come First
Crawford, J., Cowling, M., y Allen, K. A.	2023	Leadership is needed for ethical ChatGPT: Character, assessment, and learning using artificial intelligence (AI)
Cunningham, S., Cathcart, A., y Graham, T.	2023	Allegations, Abuse and Discrimination: Using Student Evaluation of Teaching Surveys to Support Student and Educator Wellbeing
Cunningham, S., Laundon, M., Cathcart, A., Abul Bashar, M., y Nayak, R.	2023	First, do no harm: Automated detection of abusive comments in student evaluation of teaching surveys.
Cunningham-Nelson, S., Baktashmotlagh, M., y Boles, W.	2019	Visualizing Student Opinion Through Text Analysis
Daftary, M. N., Jordan, J., Habib, M., Pather, I., y Tofade, T.	2020	Implementing virtual experiences and remote assessments during the COVID-19 pandemic: A college experience
Dake, D. K., y Gyimah, E.	2023	Using sentiment analysis to evaluate qualitative students' responses.
De la Hoz, E., Zuluaga, R., y Mendoza, A.	2021	Assessing and Clarification of academic efficiency in engineering teaching programs.
Deho, O. B., Zhan, C., Li, JY., Liu, J. X., Liu, L., y Le, T. D.	2022	How do the existing fairness metrics and unfairness mitigation algorithms contribute to ethical learning analytics?
Emerson, A., Min, W., Azevedo, R., y Lester, J.	2023	Early prediction of student knowledge in game-based learning with distributed representations of assessment questions
Escalante, J., Pack, A., y Barrett, A.	2023	AI-generated feedback on writing: Insights into efficacy and ENL student preference
Ezeiza, A.	2023	Return to virtual forum as sociodiscursive context for the development of University Academic Writing
Fahd, K., Venkatraman, S., Miah, S. J., y Ahmed, K.	2022	Application of machine learning in higher education to assess student academic performance, at-risk, and attrition: A meta-analysis of literature

Autor/es	Año	Título
Fan, S. S.	2023	Evaluation on Innovation and Development of University Education Management Informatization Construction Under the Background of Big Data
Fan, Z., y Chiong, R.	2023	Identifying digital capabilities in university courses: An automated machine learning approach
Fergus, S., Botha, M., Ostovar, M.	2023	Evaluating Academic Answers Generated Using ChatGPT
Fernández, A. G., Higuera, A. M. G., González, M. A. C., y Llamas, C. F.	2020	Evaluation of students' academic results through the analysis of their use of Version Control Systems
Gallent-Torres, C., Zapata-González, A., y Ortego-Hernando, J.L.	2023	The impact of Generative Artificial Intelligence in higher education: a focus on ethics and academic integrity
Gibson, M. J., Luong, J., Cho, H., Moh, B., Zanin, S., Djatmiko, M., y Aandahl, R. Z.	2022	Where's the harm? Screening student evaluations of teaching for offensive, threatening or distressing comments
Grájeda, A., Burgos, J., Córdova, P., y Sanjinés, A.	2024	Assessing student-perceived impact of using artificial intelligence tools: Construction of a synthetic index of application in higher education
Guerrero-Roldán, A.-E., Elena Rodríguez-González, M., Baneres, D., Elasmri-Ejjaberi, A., y Cortadas, P.	2021	Experiences in the use of an adaptive intelligent system to enhance online learners' performance: A case study in Economics and Business courses
Hackl, V., Müller, A. E., Granitzer, M., y Sailer, M.	2023	Is GPT-4 a reliable rater? Evaluating consistency in GPT-4's text ratings
Hasrod, T., Nuapia, Y. B., y Tutu, H.	2024	ChatGPT Helped Me Build a Chemistry App, and Here's How You Can Make One Also
Huang R. S. T., Lu, K. J. Q., Meaney, C., Kemppainen, J., Punnett, A., y Leung, F. H.	2023	Assessment of Resident and AI Chatbot Performance on the University of Toronto Family Medicine Residency Progress Test: Comparative Study
Ifenthaler, D., Schumacher, C., y Kuzilek, J.	2023	Investigating students' use of self-assessments in higher education using learning analytics
Jacobs, S. M., Lundy, N. N., Issenberg, S. B., y Chandran, L.	2023	Reimagining Core Entrustable Professional Activities for Undergraduate Medical Education in the Era of Artificial Intelligence
Jain, G. P., Gurupur, V. P., Schroeder, J. L., y Faulkenberry, E. D.	2014	Artificial Intelligence-Based Student Learning Evaluation: A Concept Map-Based Approach for Analyzing a Student's Understanding of a Topic
Jescovitch, L. N., Scott, E. E., Cerchiara, J. A., Merrill, J., Urban-Lurain, M., Doherty, J. H., y Haudek, K. C.	2021	Comparison of Machine Learning Performance Using Analytic and Holistic Coding Approaches Across Constructed Response Assessments Aligned to a Science Learning Progression
Kavadella, A., Silva, M. A. D. D., Kaklamanos, E. G., Stamatopoulos, V., y Giannakopoulos, K.	2024	Evaluation of ChatGPT's Real-Life Implementation in Undergraduate Dental Education: Mixed Methods Study
Kelly, A., Sullivan, M., y Strampel, K.	2023	Generative artificial intelligence: University student awareness, experience, and confidence in use across disciplines
Kiryakova, G., y Angelova, N.	2023	ChatGPT-A Challenging Tool for the University Professors in Their Teaching Practice
Kumar, R.	2023	Faculty members' use of artificial intelligence to grade student papers: a case of implications
Langan, A. M., y Harris, W. E.	2019	National student survey metrics: Where is the room for improvement?

Autor/es	Año	Título
Laupichler, M. C., Hadizadeh, D. R., Wintergerst, M. W. M., von der Emde, L., Paech, D., Dick, E. A., y Raupach, T.	2022	Effect of a flipped classroom course to foster medical students' AI literacy with a focus on medical imaging: a single group pre-and post-test study
Laupichler, M. C., Aster, A., Perschewski, J. O., y Schleiss, J.	2023	Evaluating AI Courses: A Valid and Reliable Instrument for Assessing Artificial-Intelligence Learning through Comparative Self-Assessment
Lesage, J., Brennan, R., Eaton, S. E., Moya, B., Mcdermott, B., Wiens, J., y Herrero, K.	2024	Exploring natural language processing in mechanical engineering education: Implications for academic integrity
Liaw, S. Y., Tan, J. Z., Lim, S., Zhou, W. T., Yap, J., Ratan, R., Ooi, S. L., Wong, S. J., Seah, B., y Chua, W. L.	2023	Artificial intelligence in virtual reality simulation for interprofessional communication training: Mixed method study
Liu, Y., Fan, S., Xu, S., Sajjanhar, A., Yeom, S., y Wei, Y.	2022	Predicting Student Performance Using Clickstream Data and Machine Learning
Lünich, M., Keller, B., y Marcinkowski, F.	2023	Fairness of Academic Performance Prediction for the Distribution of Support Measures for Students: Differences in Perceived Fairness of Distributive Justice Norms
Ma, J.	2021	Intelligent Decision System of Higher Educational Resource Data Under Artificial Intelligence Technology
Mahapatra, S.	2024	Impact of ChatGPT on ESL students' academic writing skills: a mixed methods intervention study
Mangera, E., Supratno, H., y Suyatno	2023	Exploring the relationship between Transhumanist and Artificial Intelligence in the Education Context: Particularly Teaching and Learning Process at Tertiary Education
Martin, P. P., y Graulich, N.	2024	Beyond Language Barriers: Allowing Multiple Languages in Postsecondary Chemistry Classes Through Multilingual Machine Learning
Martin-Marchante, B.	2022	The use of ICTs and artificial intelligence in the revision of the writing process in Valencian public universities
Matthews, J., y Volpe, C. R.	2023	Academics' perceptions of ChatGPT-generated written outputs: A practical application of Turing's Imitation Game
Mientus, L., Wulff, P., Nowak, A., y Borowski, A.	2023	Fast-and-frugal means to assess reflection-related reasoning processes in teacher training-Development and evaluation of a scalable machine learning-based metric
Molina, O. E., y Cancell, D. F.	2021	Is it possible to predict academic performance? An analysis from educational technology
Moorhouse, B. L., Yeo, M. A., y Wan, Y. W.	2023	Generative AI tools and assessment: Guidelines of the world's top-ranking universities
Moral-Sánchez, S. N., Rey, F. J. R., y Cebrian-de-la-Serna, M.	2023	Analysis of artificial intelligence chatbots and satisfaction for learning in mathematics education
Morjaria, L., Burns, L., Bracken, K., Ngo, Q. N., Lee, M. R., Levinson, A. J., Smith, J., Thompson, P., y Sibbald, M.	2023	Examining the Threat of ChatGPT to the Validity of Short Answer Assessments in an Undergraduate Medical Program
Muratov, E., Lewis, M., Fourches, D., Tropsha, A., y Cox, W. C.	2017	Computer-Assisted Decision Support for Student Admissions Based on Their Predicted Academic Performance
Murzo, Y., Chuvileva, N., y Schetinina, A.	2022	Using Semantic Patterns in Web Search and Assessment of Professionally Oriented Texts in a Foreign Language for Training Students in Higher Education Institutions of Mineral Resource Profile
Ng, D., Ku, D. R. Y., Xi, Y. M., Ho, D., y Ko, J. M.	2023	School Reform: New Future-Ready Quality Outcomes and Proposed Measures

Autor/es	Año	Título
Nikolic, S., Daniel, S., Haque, R., Belkina, M., Hassan, G. M., Grundy, S., Lyden, S., Neal, P., y Sandison, C	2023	ChatGPT versus engineering education assessment: a multidisciplinary and multi-institutional benchmarking and analysis of this generative artificial intelligence tool to investigate assessment integrity
Okoye, K., Arrona-Palacios, A., Camacho-Zúñiga, C., Achem, J. A. G., Escamilla, J., y Hosseini, S.	2022	Towards teaching analytics: A contextual model for analysis of students' evaluation of teaching through text mining and machine learning classification
Oravec, J. A.	2022	AI, Biometric Analysis, and Emerging Cheating Detection Systems: The Engineering of Academic Integrity?
Ou, A. W., Stöhr, C., y Malmström, H.	2024	Academic communication with AI-powered language tools in higher education: From a post-humanist perspective
Ouyang, F., Wu, M., Zheng, L., Zhang, L., y Jiao, P.	2023	Integration of artificial intelligence performance prediction and learning analytics to improve student learning in online engineering course
Perkins, M., y Roe, J.	2023	Decoding Academic Integrity Policies: A Corpus Linguistics Investigation of AI and Other Technological Threats
Pham, T., Nguyen, T. B. ., Ha, S., y Nguyen Ngoc, N. T.	2023	Digital transformation in engineering education: Exploring the potential of AI-assisted learning
Poonpon, K., Manorom, P., y Chansanam, W.	2023	Exploring effective methods for automated essay scoring of non-native speakers
Precup, R. E., Hedrea, E. L., Roman, R. C., Petriu, E. M., Szedlak-Stinean, A. I., y Bojan-Dragos, C. A.	2021	Experiment-Based Approach to Teach Optimization Techniques
Rahiman, H. U., y Kodikal, R.	2024	Revolutionizing education: Artificial intelligence empowered learning in higher education
Raković, M., Iqbal, S., Li, T., Fan, Y., Singh, S., Surendrannair, S., y Gašević, D.	2023	Harnessing the potential of trace data and linguistic analysis to predict learner performance in a multi-text writing task
Rapanta, C., y Walton, D.	2016	The Use of Argument Maps as an Assessment Tool in Higher Education
Rico-Juan, J. R., Gallego, A.-J., y Calvo-Zaragoza, J.	2019	Automatic detection of inconsistencies between numerical scores and textual feedback in peer-assessment processes with machine learning
Rodríguez, J. A., Santana, M. G., Perera, M. V. A., y Pulido, J. R.	2021	Embodied conversational agents: artificial intelligence for autonomous learning
Romo-Pérez, V., García-Soidán, J. L., Özdemir, A. S., y Leiróos-Rodríguez, R.	2023	ChatGPT has arrived! What do we do now? Creativity, our last refuge
Sailer, M., Bauer, E., Hofmann, R., Kiesewetter, J., Glas, J., Gurevych, I., y Fischer, F.	2023	Adaptive feedback from artificial neural networks facilitates pre-service teachers' diagnostic reasoning in simulation-based learning
Sajja, R., Sermet, Y., Cwiertny, D., y Demir, I.	2023	Platform-independent and curriculum-oriented intelligent assistant for higher education
Saleem, N., Mufti, T., Sohail, S. S., y Madsen, D. O.	2024	ChatGPT as an innovative heutagogical tool in medical education
Salinas-Navarro, D. E., Vilalta-Perdomo, E., Michel-Villarreal, R., y Montesinos, L.	2024	Using Generative Artificial Intelligence Tools to Explain and Enhance Experiential Learning for Authentic Assessment
Sallam, M., y Al-Salahat, K.	2023	Below average ChatGPT performance in medical microbiology exam compared to university students
Schwerter, J., y Brahm, T.	2024	Voluntary E-Learning Exercises Support Students in Mastering Statistics

Autor/es	Año	Título
Sefcik, L. T., Veeran-Colton, T., Baird, M., Price, C., y Steyn, S.	2022	An examination of student user experience (UX) and perceptions of remote invigilation during online assessment
Singh, J., Perera, V., Magana, A. J., Newell, B., Wei-Kocsis, J., Seah, Y. Y., Strimel G. J., y Xie, C.	2022	Using machine learning to predict engineering technology students' success with computer-aided design
Somers, R., Cunningham-Nelson, S., y Boles, W.	2021	Applying natural language processing to automatically assess student conceptual understanding from textual responses
Stadelmann, T., Keuzenkamp, J., Grabner, H., y Würsch, C.	2021	The AI-Atlas: Didactics for Teaching AI and Machine Learning On-Site, Online, and Hybrid
Stojanov, A.	2023	Learning with ChatGPT 3.5 as a more knowledgeable other: an autoethnographic study
Sweeney, S.	2023	Who wrote this? Essay mills and assessment-Considerations regarding contract cheating and AI in higher education
Talebinamvar, M., y Zarrabi, F	2022	Clustering students' writing behaviors using keystroke logging: A learning analytic approach in EFL writing
Tamayo, P. A., Herrero, A., Martin, J., Navarro, C., y Tranchez, J. M.	2020	Design of a chatbot as a distance learning assistant
Tarisayi, K. S.	2024	ChatGPT use in universities in South Africa through a socio-technical lens
Tavares, C., Tallone, L., Oliveira, L., y Ribeiro, S.	2023	The Challenges of Teaching and Assessing Technical Translation in an Era of Neural Machine Translation
Tomas, C., Whitt, E., Lavelle-Hill, R., y Severn, K.	2019	Modeling Holistic Marks With Analytic Rubrics
Tossell, C. C., Tenhundfeld, N. L., Momen, A., Cooley, K., y de Visser, E. J.	2024	Student Perceptions of ChatGPT Use in a College Essay Assignment: Implications for Learning, Grading, and Trust in Artificial Intelligence
Tseng, C. H., Lin, H. C. K., Huang, A. C. W., y Lin, J. R.	2023	Personalized programming education: Using machine learning to boost learning performance based on students' personality traits
Usman, M., Iqbal, M. M., Iqbal, Z. Chaudhry, M. U., Farhan, M., y Ashraf, M.	2017	E-Assessment and Computer-Aided Prediction Methodology for Student Admission Test Score
Uysal, D., y Uysal, A. K.	2022	Automatic classification of EFL learners' self-reported text documents along an affective continuum
Vazquez-Cano, E., Mengual-Andres, S., y Lopez-Meneses, E.	2021	Chatbot to improve learning punctuation in Spanish and to enhance open and flexible learning environments
Veerasamy, A. K., Laakso, M. J., y D'Souza, D.	2022	Formative Assessment Tasks as Indicators of Student Engagement for Predicting At-risk Students in Programming Courses
Wakelam, E., Jefferies, A., Davey, N., y Sun, Y.	2020	The potential for student performance prediction in small cohorts with minimal available attributes
Wambsganss, T., Janson, A., y Leimeister, J. M.	2022	Enhancing argumentative writing with automated feedback and social comparison nudging
Watson, R. A.	2014	Use of a Machine Learning Algorithm to Classify Expertise: Analysis of Hand Motion Patterns During a Simulated Surgical Task
Williams, P.	2023	AI, Analytics and a New Assessment Model for Universities
Williams, P.	2019	Does competency-based education with blockchain signal a new mission for universities?
Williams, R. T.	2024	The ethical implications of using generative chatbots in higher education
Wise, B., Emerson, L., Van Luyn, A., Dyson, B. Bjork, C., y Thomas, S. E.	2024	A scholarly dialogue: writing scholarship, authorship, academic integrity and the challenges of AI

Autor/es	Año	Título
Wulff, P., Buschhueter, D., Westphal, A., Nowak, A., Becker, L., Robalino, H., Stede, M., y Borowski, A.	2021	Computer-Based Classification of Preservice Physics Teachers' Written Reflections.
Yang, H., Gao, C., y Shen H.	2024	Learner interaction with, and response to, AI-programmed automated writing evaluation feedback in EFL writing: An exploratory study
Yilmaz, Y., Nunez, A. J., Ariaeinejad, A., Lee, M., Sherbino, J., y Chan, T. M.	2022	Harnessing Natural Language Processing to Support Decisions Around Workplace-Based Assessment: Machine Learning Study of Competency-Based Medical Education
Zaki, N., Turaev, S., Shuaib, K., Krishnan, A., y Mohamed, E.	2023	Automating the mapping of course learning outcomes to program learning outcomes using natural language processing for accurate educational program evaluation
Zeng, Y.	2020	Evaluation of Physical Education Teaching Quality in Colleges Based on the Hybrid Technology of Data Mining and Hidden Markov Model
Zheng, L., Long, M., Chen, B., y Fan, Y.	2023	Promoting knowledge elaboration, socially shared regulation, and group performance in collaborative learning: an automated assessment and feedback approach based on knowledge graphs
Zuluaga, R. and Camelo-Guarín, A., y De La Hoz, E.	2023	Assessing relative impact of Colombian Higher Education Institutions using Fuzzy Data Envelopment Analysis (Fuzzy-DEA) in State Evaluations

Fuente: Elaboración propia.

Abstract

Artificial intelligence in assessment processes in higher education: a bibliometric analysis (2014-2024)

INTRODUCTION. Artificial Intelligence (AI) in higher education represents a significant innovation that is transforming teaching, learning, and academic assessment. The latter is essential for providing feedback to students, evaluating their understanding, and fostering continuous learning improvement. The aim of this study is to analyze the scientific production related to AI in assessment in higher education over the last decade. This analysis covers chronological and geographical productivity, as well as a detailed study of sources, keywords, and citation counts of the most prominent articles in this field. **METHOD.** A bibliometric and systematic review was conducted following the guidelines of the PRISMA Statement. **RESULTS.** Initially, 2,275 studies were identified in the Web of Science database, and after applying eligibility criteria, 130 empirical studies were selected for further analysis. **DISCUSSION.** The research reviewed concludes that the primary methods of integrating AI in the assessment process include automated feedback, prediction of academic performance through AI-based data analysis, the use of language models such as ChatGPT, and the consideration of associated ethical issues. A significant increase in the volume of publication was observed during the 2023-2024 period, with the United States leading scientific production and Education Sciences being the main source of dissemination. The co-word analysis reveals a lack of terminological uniformity, suggesting a need to standardize language to improve clarity in the field.

Keywords: *Artificial intelligence, Assessment, Educational assessment, Higher education.*

Résumé

L'intelligence artificielle dans les processus d'évaluation dans l'enseignement supérieur: une analyse bibliométrique (2014-2024)

INTRODUCTION. L'intelligence artificielle (IA) dans l'enseignement supérieur constitue une innovation majeure qui transforme l'enseignement, l'apprentissage et l'évaluation académique. Cette dernière est essentielle pour fournir un retour d'information aux étudiants, évaluer leur compréhension et favoriser l'amélioration continue de l'apprentissage. L'objectif de cette étude est d'analyser la production scientifique relative à l'IA dans l'évaluation dans l'enseignement supérieur au cours de la dernière décennie. Cette analyse porte sur la productivité chronologique et géographique, ainsi que sur une étude détaillée des sources, des mots-clés et du nombre de citations des articles les plus influents dans ce domaine. **MÉTHODE.** Une revue bibliométrique et systématique a été réalisée en suivant les directives de la déclaration PRISMA. **RÉSULTATS.** Au départ, 2 275 études ont été identifiées dans la base de données Web of Science. Après application des critères d'éligibilité, 130 études empiriques ont été retenues pour une analyse approfondie. **DISCUSSION.** L'analyse conclut que les principales méthodes d'intégration de l'IA dans le processus d'évaluation incluent la rétroaction automatisée, la prédiction de la performance académique par l'analyse de données basée sur l'IA, l'utilisation de modèles linguistiques tels que ChatGPT, ainsi que la prise en compte des questions éthiques associées. Une augmentation significative du volume de publications a été observée sur la période 2023-2024, avec les États-Unis en tête de la production scientifique et les sciences de l'éducation comme principal domaine de diffusion. L'analyse des co-occurrences de mots révèle un manque d'uniformité terminologique, suggérant la nécessité de standardiser le langage afin d'améliorer la clarté dans ce domaine.

Mots-clés : *Intelligence artificielle, Évaluation, Évaluation éducative, Enseignement supérieur.*

Perfil profesional de las autoras

Alba Galán-Íñigo (autora de contacto)

Doctoranda en la Universidad Europea de Madrid. Su tesis doctoral se centra en la integración de la inteligencia artificial en educación superior. Es docente en el Máster U. en Innovación Educativa de la Universidad Europea de Madrid y es miembro del grupo de investigación Innedu-UEM. En su formación académica destaca el Máster Oficial en Educación Universitaria (2021, UEM), el Grado de Educación Primaria (2015, UCM) y el Grado en Comunicación Audiovisual (2012, UCM), en el que realizó una estancia en la Kingston University of London (2011). Es profesora especializada en metodologías emergentes y neuroeducación. Dentro de su experiencia profesional destaca como coordinadora académica de Educación Secundaria y Bachillerato en el Instituto Psicológico Desconnect@ (Madrid) y profesora-tutora en el colegio International Leadership of Texas (EE. UU.). Sus últimos estudios y comunicaciones en congresos se enmarcan en la integración de la IA y en el análisis de espacios innovadores en educación superior.

ORCID: <https://orcid.org/0009-0000-2446-2806>

Correo electrónico de contacto: alba.galan@universidadeuropea.es

Dirección para la correspondencia: C/Tajo, s/n, 28670. Fac. de Ciencias Jurídicas, Educación y Humanidades. Villaviciosa de Odón, Madrid.

Judit Ruiz-Lázaro

Doctora en Educación por la Universidad Complutense de Madrid (2021) con mención “Doctor Internacional”, calificación “Sobresaliente Cum Laude” y “Premio Extraordinario de Doctorado”. Acreditada a Profesora Titular de Universidad por ANECA (2024). Dispone de un sexenio de investigación vivo. Actualmente, Profesora Ayudante Doctora en la Universidad Nacional de Educación a Distancia (UNED, Dpto. de Didáctica, Organización Escolar y Didácticas Especiales). Sus últimos estudios y publicaciones se enmarcan en la evaluación para el acceso a la universidad en el contexto español, el análisis de la formación del profesorado en España y el uso de la inteligencia artificial en el ámbito educativo. Es miembro del grupo de investigación consolidado Medida de Evaluación y Sistemas Educativos (MESE) de la UCM e Innedu-UEM de la UEM. ORCID: <https://orcid.org/0000-0003-2036-0428>
Correo electrónico de contacto: Judit.ruiz@edu.uned.es

Eva Jiménez-García

Doctora Acreditada en Educación con Premio Extraordinario de Doctorado (2016) y licenciada en Pedagogía por la Universidad Complutense de Madrid. Actualmente trabaja como directora de Investigación y directora del Centro de Investigación Educativa (CIE-UE) de la Facultad de Educación de la Universidad Europea de Madrid. Titular en Métodos de Investigación y Diagnóstico en Educación (ANECA). Su actividad investigadora se centra en la medida y evaluación de sistemas educativos. Forma parte del Grupo de Investigación de Medida y Evaluación de Sistemas Educativos, de la Universidad Complutense de Madrid. Miembro del Consejo Asesor de la revista *Tendencias Pedagógicas* y miembro del Consejo Evaluador de dos revistas de impacto. ORCID: <http://orcid.org/0000-0001-6541-3517>
Correo electrónico de contacto: eva.jimenez@universidadeuropea.es