



AIESAD

Ried

**Revista Iberoamericana de
Educación a Distancia**

La Revista Iberoamericana de la Educación Digital

VOL. 29 N° 1 ENERO, 2026
ISSN: 1138-2783



RIED. Revista Iberoamericana de Educación a Distancia

La Revista Iberoamericana de la Educación Digital

Depósito legal: M- 36.279-1997

ISSN: 1138-2783 / E-ISSN: 1390-3306

1º semestre, enero, 2026

RIED

Esta publicación de periodicidad semestral está dirigida a los estudiosos e investigadores del ámbito educativo, docentes universitarios y público interesado en su objeto de estudio. La RIED centra su atención en la difusión de ensayos, trabajos de carácter científico y experiencias innovadoras dentro del ámbito de la educación a distancia en cualesquiera de sus formulaciones y de las tecnologías aplicadas a la educación.

La RIED se gestiona íntegramente a través del Open Journal System (OJS), tanto para la edición como para la relación con los autores y revisores, así como para la difusión electrónica en abierto.

La RIED, además de su formato impreso, se publica en formato electrónico en dos sedes: OJS en UNED de España: <http://revistas.uned.es/index.php/ried>

INTERCAMBIOS y SUSCRIPCIONES:

RIED. Revista Iberoamericana de Educación a Distancia.

UTPL – SAN CAYETANO ALTO, s/n

Loja (Ecuador)

ried@utpl.edu.ec

Consejo Directivo de AIESAD (Asesor en RIED)

- **Presidente:** Jaime Leal Afanador, Rector Magfco. Universidad Nacional Abierta y a Distancia UNAD, Colombia
- **Directora Ejecutiva:** Constanza Abadía García, Vicerrectora. Universidad Nacional Abierta y a Distancia – UNAD. Colombia.
- **Vicepresidente Primero:** Ricardo Mairal Usón, Rector Magfco. Universidad Nacional de Educación a Distancia UNED. España
- **Vicepresidente Segundo:** Rodrigo Arias Camacho, Rector Magfco. Universidad Estatal a Distancia UNED. Costa Rica
- **Vicepresidente Tercero:** Santiago Acosta Aide, Rector Magfco. Universidad Técnica Particular de Loja. Ecuador
- **Vicepresidente Cuarto:** David Mejía Rodríguez, Rector Magnífico. Universidad Internacional de La Rioja – UNIR. México.
- **Vocales:**
 - Alfredo Alonso, Rector Magfco. Universidad Nacional de Quilmes UNQ. Argentina
 - Carla Padrel de Oliveira, Rectora Magfca. Universidad Aberta UAB. Portugal
 - Alejandra Hernández Acosta, Rectora Magnífica. Universidad Abierta para Adultos – UAPA. República Dominicana
 - Fray Álvaro José Arango Restrepo, O.P., Rector Magnífico. Universidad Santo Tomás – USTA. Colombia
- **Secretaría permanente y Tesorería:** Laura Alba Juez, Vicerrectora de Internacionalización. Universidad Nacional de Educación a Distancia – UNED. España.

Director/Editor (Director/Editor-in-Chief)

- Dr. Lorenzo García Aretio, Universidad Nacional de Educación a Distancia (UNED), España

Directora/Editora Adjunta (Deputy Director/Editor)

- Dra. Cristina Díaz de la Cruz, UTPL (Ecuador)

Consejo Editorial y Científico (Editorial and Scientific Board)

- Jordi Adell Segura. Univ. Jaime I (España).
- José Ignacio Aguaded Gómez. Univ. de Huelva (España). Editor “Comunicar”
- Luisa Aires. Univ. Aberta (Portugal).
- Terry Anderson. Athabasca Univ. (Canadá). Editor Emeritus IRRODL.
- Manuel Area Moreira. Univ. de La Laguna (España).
- Elena Barberá Gregori. UOC (España).
- Elena Bárcena Madera. UNED (España)
- Antonio Bartolomé Pina. Univ. de Barcelona (España).
- Julio Cabero Almenara. Univ. de Sevilla (España). Editor Pixelbit.
- Isabel Cantón Mayo. Univ. de León (España).
- Linda Castañeda. Univ. de Murcia (España).
- Manuel Castro Gil. UNED (España).
- M. Elena Chan Núñez. Univ. de Guadalajara (México).
- Cristóbal Cobo. Univ. of Oxford (R. Unido). The World Bank.
- César Coll Salvador. Univ. de Barcelona (España).
- Grainne Conole. e4innovation (R.Unido).
- Laura Czerniewicz. Univ. of Cape Town (Sudáfrica). CILT.
- Carlos Delgado Kloos. Univ. Carlos III (España). Cátedra UNESCO Educación Digital.
- Frida Díaz Barriga. UNAM (México).
- M. Esther del Moral Pérez. Univ. de Oviedo (España).
- Pierre Dillenbourg. EPFL (Suiza). Coordinator CHILI Lab.
- Josep M. Duarte. UOC (España). Editor ETHE Journal. EDEN President.
- Rubén Edel Navarro. Univ. Veracruzana (México).
- María Jesús Gallego-Arrufat. Univ. de Granada (España).
- Francisco J. García Peñalvo. Univ. de Salamanca (España). Editor EKS.

- Ana García-Valcárcel Muñoz-Repiso. Univ. de Salamanca (España).
- Mercè Gisbert Cervera. Univ. Rovira i Virgili (España).
- Carina Soledad González González. Univ. de La Laguna (España).
- Mercedes González Sanmamed. Univ. de A Coruña (España).
- Jaime Leal Afanador. UNAD (Colombia). Presidente AIESAD.
- M. del Carmen Llorente-Cejudo. Univ. de Sevilla (España).
- Ricardo Mairal Usón. Rector UNED (España). EADTU Ex-President.
- Carlos Marcelo García. Univ. de Sevilla (España).
- João Mattar. Pontificia Univ. Católica de São Paulo (Brasil).
- Rory McGreal. Athabasca Univ. Editor de IRRODL (Canadá). UNESCO Chair in OER.
- Daniel Mill. Univ. Federal de São Carlos (Brasil). Grupo Horizonte.
- António Moreira Teixeira. Univ. Aberta (Portugal). EDEN Ex-President.
- Lina Morgado. Univ. Aberta (Portugal).
- Jaime Muñoz Arteaga. Univ. Autónoma Aguascalientes (México).
- Sara Osuna-Acedo. UNED (España).
- Adolfinia Pérez Garcías. Univ. Islas Baleares (España).
- Mar Pérez-Sanagustín. Univ. Paul Sabatier (Francia).
- Teresa Pessoa. Univ. de Coimbra (Portugal).
- M. Paz Prendes Espinosa. Univ. de Murcia (España).
- Claudio Rama. IESAL/UNESCO (Venezuela).
- M. Soledad Ramírez Montoya. TEC Monterrey (México). Cátedra UNESCO-ICDE.
- Rosabel Roig. Univ. de Alicante (España). Editora NAER.
- Jesús Salinas Ibáñez. Univ. Islas Baleares (España). Editor EDUTEC.
- Ángeles Sánchez-Elvira. UNED (España) Cátedra UNESCO EaD.
- Juana Sancho Gil. Univ. de Barcelona (España).
- Albert Sangrá. UOC (España). Cátedra UNESCO E-learning.
- Alan W. Tait. The OU (R.Unido). EDEN Ex-President. Ex-Editor Open Learning.
- Hernán Thomas. Univ. Nacional de Quilmes (Argentina).
- Javier Tourón Figueroa. UNIR (España).
- Martín James Weller. The OU (R. Unido). Director OER Hub. Editor JIME.
- Miguel Zapata Ros. Univ. de Murcia (España). Editor RED.

Delegación de AIESAD en RIED (AIESAD's Delegate in RIED)

- Santiago Acosta Aide, Univ. Técnica Particular de Loja (UTPL), Ecuador
- Cristina Díaz de la Cruz. UTPL, Ecuador

Secretaría Técnica (Technical Secretariat)

- José Luis García Boyé, AIESAD, España

Consejo de Redacción (Editing Board)

- Martha Vanessa Agila-Palacios , UTPL (Ecuador).
- Laura T. Alonso Díaz, Univ. Extremadura (España).
- Julia Ángel Osorio, UNAD (Colombia).
- Analía Chiecher, Univ. Río Cuarto (Argentina).
- Cristina Díaz de la Cruz, UTPL (Ecuador)
- Paz Díez Arcón, UPM (España)
- Anna Engel, Univ. de Barcelona (España).
- Francesc Esteve Mon, Univ. Jaime I (España).
- Alicia García Holgado, Univ. Salamanca (España)
- Montse Guitert Catasús, UOC (España).
- M. José Hernández Serrano, Univ. Salamanca (España)
- Agustín Lagunes Domínguez, Univ. Veracruzana (México).
- Andrea López Ingelmo, Univ. Salamanca (España)
- Verónica Marín Díaz, Univ. Córdoba (España).
- Victoria I. Marín, Univ. Lleida (España)
- Elena M. Martín Monje, UNED (España).
- Rafael Morales Gamboa, Univ. Guadalajara (México).
- João Paz, Univ. Aberta (Portugal).
- Neuza Pedro, Univ. Lisboa (Portugal)
- Iliana Ramírez Asanza, UTPL (Ecuador).
- Enrique Ruiz Velasco, UNAM, (México).
- José M. Sáez López, UNED (España).
- Cassio Santos, Univ. Lisboa (Portugal).
- Carolina Schmitt Nunes, Univ. Sta. Catarina (Brasil).
- Paola Sarango-Lapo, UTPL (Ecuador)
- Juan José Sosa Alonso, Univ. La Laguna (España).
- Romero Tori, Univ. Sao Paulo (Brasil).
- Gemma Tur Ferrer, Univ. Islas Baleares (España).
- José Carlos Vázquez Parra, TEC Monterrey (México).

.Redes Sociales (Social Networks)

- José Luis García Boyé
- Carla Netto
- Carolina Schmitt

Apoyo Técnico (Technical Assistance)

- Alexis Moreno-Pulido, Responsable de Biblioteca (UNED)

Soporte OJS y Publicación digital

- Servicio Publicación y Difusión Digital - BIBLIOTECA, UNED

LA REVISTA IBEROAMERICANA DE EDUCACIÓN A DISTANCIA (RIED) SE ENCUENTRA INDIZADA ACTUALMENTE EN LAS SIGUIENTES BASES DE DATOS Y CATÁLOGOS:

BASES DE DATOS Y PLATAFORMAS DE EVALUACIÓN

- BASE. Bielefeld Academic Search Engine
- CAPES
- CARHUS Plus+
- CCHS-CSIC
- CEDAL (Instituto Latinoamericano de Comunicación Educativa (ILCE) de México)
- CIRC (Clasificación Integrada de Revistas Científicas)
- CiteFactor – Academic Scientific Journals
- CREDI- OEI (Centro de Recursos de la OEI)
- Crossref (Metadata Search)
- Dialnet (Alertas de Literatura Científica Hispana)
- DICE (Difusión y Calidad Editorial de Revistas)
- EI Compendex
- EBSCO. Fuente Académica Premier
- ERA. Educational Research Abstracts
- ERIH-Plus. European Reference Index for the Humanities and Social Sciences.
- EZB-Electronic Journals Library Genamics JournalSeek
- HEDBIB (International Bibliographic Database on Higher Education)
- IN-RECS (Índice de Impacto de Revistas Españolas de Ciencias Sociales)
- IRESIE (Índice de Revistas de Educación Superior e Investigación Educativa)
- ISOC (CSIC/CINDOC)
- JournalTOCs
- MIAR (Matriz para Evaluación de Revistas)
- ProQuest-CSA
- Psycodoc
- REDIB. Red Iberoamericana de Innovación y Conocimiento Científico
- REDALYC. Red de Revistas Científicas de América Latina y el Caribe, España y Portugal
- REDINED. Red de Información Educativa
- RESH - Revistas Españolas de Ciencias Sociales (CSIC/CINDOC)
- ResearchBib. Academic Resource Index
- Scopus
- Web of Science (SSCI)
- WEBQUALIS

DIRECTORIOS Y BUSCADORES

- DOAJ
- Dulcinea
- Google Scholar
- LATINDEX (Publicaciones Científicas Seriadadas de América, España y Portugal)
- Recolecta
- Sherpa/Romeo
- Scirus
- Ulrich's Periodicals (CSA)

PORTALES Y REPOSITARIOS ESPECIALIZADOS

- Actualidad Iberoamericana
- Asociación Internacional de Estudios en comunicación social
- CLARISE - Comunidad Latinoamericana Abierta Regional de Investigación Social y Educativa
- Educ.ar
- Enlaces educativos en español de la Universitat de València
- e-sPacio-UNED. Repositorio institucional de la UNED
- Institut Français de L'éducation
- Plataforma de revistas 360º
- Red Iberoamericana de Revistas de Comunicación y Cultura
- REDIAL & CEISAL
- Universia. Biblioteca de recursos

CATÁLOGOS DESTACADOS DE BIBLIOTECA

- 360grados
- British Library
- Buz
- Catàleg Col·lectiu de les Universitats de Catalunya
- Catálogo Colectivo de Publicaciones
- Periódicas Español CCPP
- Catálogo de la Biblioteca de Educación (Ministerio de Educación, Cultura y Deporte)
- Catálogo del CSIC (CIRBIC)
- CENDOC
- CIDE
- CISNE
- COMPLUDOC
- COPAC (Reino Unido)
- ICDL
- INRP
- IOE (Institute of Education. University of London)
- Library of Congress (LC)
- KINGS
- MIGUEL DE CERVANTES
- Observatorio de revistas científicas de Ciencias Sociales
- REBIUN
- SUDOC (Francia)
- UBUCAT
- UIB
- WORDLCAT (OCLC)
- ZDB (Alemania)

La Asociación Iberoamericana de Educación Superior a Distancia (AIESAD) es una entidad sin ánimo de lucro, constituida por universidades o instituciones de educación superior que imparten sus ofertas educativas en esta modalidad de enseñanza y promueve el estudio e investigación del modelo de enseñanza superior a distancia. *RIED. Revista Iberoamericana de Educación a Distancia* es el instrumento de la AIESAD para la difusión internacional de los avances en la investigación e innovación dentro del ámbito de la enseñanza y aprendizaje abiertos y a distancia.



RIED. Revista Iberoamericana de Educación a Distancia es una publicación científica que se edita semestralmente los meses de enero y julio. Promueve el intercambio institucional con otras revistas de carácter científico. La *RIED* no se hará responsable de las ideas y opiniones expresadas en los trabajos publicados. La responsabilidad plena será de los autores de los mismos.



"Los textos publicados en esta revista están sujetos a una licencia "Reconocimiento-No comercial 3.0" de Creative Commons. Puede copiarlos, distribuirlos, comunicarlos públicamente, siempre que reconozca los créditos de la obra (autor, nombre de la revista, instituciones editoras) de la manera especificada en la revista."



AIESAD

Ried

Revista Iberoamericana de
Educación a Distancia

VOL. 29 N° 1

Enero, 2026

Índice

Inteligencia artificial en la educación superior: diseño, competencias, evaluación y desafíos

Estudios e investigaciones

Inteligencia artificial generativa y autonomía educativa: metáforas históricas y principios éticos para la transformación pedagógica

Marc Alier-Forment, María José Casañ-Guerrero, Juan Antonio Pereira-Varela, Francisco José García-Peñalvo, Faraón Llorens-Largo 9

Inteligencia artificial e inteligencia colectiva en la educación superior digital: estudio cuasi-experimental con la plataforma Kampal

Alicia Martínez-De la muela, Juan Pablo Ruiz-Fuentes, José L. González-Geraldo 29

Impacto de la IA generativa en competencias digitales universitarias: evidencia experimental basada en el marco DigComp

Concepción González García, Nina Pallarés 53

AI-ED-SAT: diseño y validación de un cuestionario para autoevaluar competencias docentes en IA educativa

Andresa Sartor-Harada, Juliana Azevedo Gomes 79

Autopercepción y utilidad de la inteligencia artificial generativa en docentes en formación

Ana María Pinto-Llorente, Vanessa Izquierdo-Álvarez, Marta M. Dolcet-Negre 111

Docentes y divulgación científica en redes sociales: desarrollo y validación inicial de instrumentos de evaluación

Dieter Reynaldo Fuentes-Cancell, Odiel Estrada-Molina, Mónica Gutiérrez-Ortega 133

Tecnologías inmersivas en la universidad: perfiles del profesorado y obstáculos para su integración

Julio Cabero-Almenara, Antonio Palacios-Rodríguez, Julio Barroso-Osuna, Carmen Siles-Rojas .. 161

Desarrollo y validación de la escala de Aprendizaje Autorregulado con IA (AI-SRL)

Mehmet Ali Ayaz, Fatih Karataş, Erkan Yüce 185

Influencia de la motivación hacia el aprendizaje y la procrastinación en la dependencia a ChatGPT

Héctor Galindo-Domínguez, Martín Sainz-de-la-Maza, Lucía Campo, Daniel Losada-Iglesias 217

Retroalimentación de aprendizajes con inteligencia artificial generativa en estudiantes universitarios

María Verónica Leiva-Guerrero, Ignacio Araya Zamorano, Rafael Escobar Collins, Francisca Silva Castro 241

Aplicación de PathRAG en aprendizaje adaptativo con IA generativa para una educación inclusiva y sostenible Rubén Juárez Cádiz	267
Nivelación inteligente: estrategia de aprendizaje adaptativo con IA en educación superior Patricia Aldape-Valdes, Elvira G. Rincon-Flores, Leticia Castano, Sadie Guerrero	299
Recomendador de evaluación para preguntas cortas utilizando modelos de lenguaje en propiedad intelectual David Bañeres Besora, Ana-Elena Guerrero Roldán, M. Elena Rodríguez González	321
Medición de la habilidad escrita en español como lengua extranjera con inteligencia artificial generativa María-Victoria Cantero Romero, María-Teresa Martín-Valdivia, Ana María Ortiz-Colón, Salud María Jiménez-Zafra	353
Metodología participativa para la creación de recursos digitales de microaprendizaje con inteligencia artificial generativa Bell Manrique-Losada, Sandra Isabel Arango-Vásquez, Mariutsi Alexandra Osorio-Sanabria	381
Aprendizaje activo en la educación superior de ingeniería: un decálogo para diseñar quizzes online post-clase Dubravka Novkovic, Ricardo Torres Kompen, Josep Petchamé	411

*Estudios e
investigaciones*

Inteligencia artificial generativa y autonomía educativa: metáforas históricas y principios éticos para la transformación pedagógica

Generative artificial intelligence and educational autonomy: historical metaphors and ethical principles for pedagogical transformation



-  Marc Alier-Forment - *Universitat Politècnica de Catalunya, UPC (España)*
-  María José Casañ-Guerrero - *Universitat Politècnica de Catalunya, UPC (España)*
-  Juan Antonio Pereira-Varela - *Universidad del País Vasco, UPV/EHU (España)*
-  Francisco José García-Peñalvo - *Universidad de Salamanca, USAL (España)*
-  Faraón Llorens-Largo - *Universidad de Alicante, UA (España)*

RESUMEN

Este artículo analiza la integración de la inteligencia artificial generativa en educación desde una perspectiva crítica, histórica y ética. Se identifica una creciente preocupación por la opacidad de las herramientas de inteligencia artificial actuales, especialmente en sistemas de aprendizaje. El trabajo utiliza un enfoque basado en metáforas para entender cómo las narrativas tecnológicas influyen en la adopción de innovaciones educativas. Se revisan metáforas históricas en las tecnologías aplicadas a la educación, desde Multivac y Matrix hasta el Bazar del software libre y la App Store, y se proponen nuevas imágenes conceptuales que podrían aplicarse al contexto actual en el que la inteligencia artificial irrumpe en la educación. A partir de este análisis metafórico, se plantean siete principios éticos para una adopción segura de la inteligencia artificial generativa en educación, centrados en la privacidad, la alineación pedagógica, la supervisión humana y la transparencia tecnológica. Estos principios se ejemplifican con el entorno LAMB (Learning Assistant Manager and Builder), un marco de código abierto que permite diseñar asistentes de aprendizaje basados en inteligencia artificial de forma ética y contextualizada. Se presentan casos reales de aplicación de LAMB en educación superior, incluyendo una experiencia controlada con estudiantes que muestran mejoras significativas en autonomía y coherencia pedagógica. Finalmente, se destaca cómo LAMB encarna los principios éticos propuestos y responde a las metáforas críticas identificadas, ofreciendo un modelo de integración tecnológica centrado en la autonomía de los docentes, la alineación con los principios y prácticas de la institución educativa y el aprendizaje significativo de los estudiantes.

Palabras clave: tecnología educativa; inteligencia artificial generativa; aprendizaje adaptativo; asistente de aprendizaje; ética; diseño instruccional.

ABSTRACT

This article examines the integration of generative artificial intelligence in education from a critical, historical, and ethical perspective. It highlights growing concerns about the opacity of current artificial intelligence tools, particularly in learning systems. The study adopts a metaphor-based approach to explore how technological narratives influence the adoption of educational innovations. It reviews historical metaphors used to describe educational technologies, from Multivac and Matrix to the free software Bazaar and the App Store, and proposes new conceptual frameworks that may better reflect the current context in which artificial intelligence is entering the educational sphere. Based on this metaphorical analysis, the article outlines seven fundamental ethical principles for the safe adoption of generative artificial intelligence in education, focusing on privacy, pedagogical alignment, human oversight, and technological transparency. These principles are illustrated through a practical application: the LAMB (Learning Assistant Manager and Builder) environment, an open-source software framework that enables the ethical and contextualized design of artificial intelligence-based learning assistants. The article presents real-world cases of LAMB implementation in higher education, including a controlled experience with students that demonstrates significant improvements in student autonomy and pedagogical coherence. Finally, it emphasizes how LAMB embodies the proposed ethical principles and responds to the identified critical metaphors, offering a model for technology integration centered on teacher autonomy, alignment with institutional values and practices, and meaningful student learning that prioritizes pedagogical control over technological determinism.

Keywords: educational technology; generative artificial intelligence; adaptive learning; learning assistant; ethics; instructional design.

INTRODUCCIÓN

En septiembre de 2024, se mantuvieron conversaciones con diez expertos en innovación educativa y tecnología sobre el “Manifiesto por una IA segura en la educación” (Alier, García-Peñalvo, Casañ et al., 2024). Estas discusiones giraron en torno a una propuesta de marco de evaluación para tecnologías educativas basadas en inteligencia artificial generativa (GenAI). Este marco dio como resultado un cuestionario para la evaluación de iniciativas educativas basadas en GenAI. El cuestionario asociado ponía el foco en aspectos clave: la privacidad de los datos, el control de acceso dentro de las instituciones y, sobre todo, la capacidad de las herramientas para alinearse con los objetivos pedagógicos. De forma sorprendente, más de la mitad de los expertos consideraban que ninguna solución basada en GenAI disponible actualmente cumpliría estos criterios o que fuera posible cumplirlos.

De estos intercambios surgió una observación relevante: muchos tecnólogos con amplia experiencia en el ámbito educativo no ven la GenAI, y en particular los sistemas de aprendizaje adaptativo basados en IA que prometen personalizar la experiencia del estudiante, como herramientas claras y moldeables, sino como artefactos potentes pero opacos, difíciles de entender o controlar. Este fenómeno nos llevó a proponer la metáfora del *Palantír* (Tolkien, 1954-1955).

En el universo narrativo de J.R.R. Tolkien, el *Palantír* es una esfera mágica que permite ver a través del tiempo y el espacio y comunicarse con usuarios de otros *Palantíres*. No obstante, el uso del *Palantír* también entraña riesgos significativos, como se ilustra con la corrupción del mago Saruman, que Sauron ejerce a través de este objeto. Esta metáfora, aplicada a la IA generativa en educación, captura dos sentimientos muy presentes: la fascinación ante su aparente poder de personalización y predicción, y el temor o escepticismo sobre quién la controla y con qué fines.

Llamaremos a esto la metáfora del *Palantír*: la GenAI entendida como una herramienta centralizada, poderosa, inevitable, pero en última instancia fuera del control de sus usuarios reales (profesores, estudiantes e instituciones). Esta imagen ayuda a explicar una actitud cada vez más común en el ámbito educativo: aceptar la implementación de sistemas de IA adaptativos tal como vienen, sin conocer en profundidad cómo se entrenan, qué sesgos contienen o cómo se toman las decisiones que afectan al aprendizaje individual.

Para efectos de este trabajo, utilizamos el término “asistentes de aprendizaje basados en GenAI” para referirnos específicamente a herramientas conversacionales que proporcionan apoyo personalizado al estudiante mediante procesamiento de lenguaje natural, distinguiéndolas de los sistemas adaptativos tradicionales que ajustan automáticamente rutas de aprendizaje basándose en análisis de rendimiento.

La idea de tecnología vista como “magia”, ya ha sido ampliamente discutida en el pasado. Por ejemplo, Arthur C. Clarke explicó en 1962 que una tecnología suficientemente avanzada podía convertirse en algo similar a la magia (Clarke, 1962).

Sin embargo, es necesario matizar. Aunque la metáfora del Palantír es útil para describir ciertas narrativas actuales, también tiene limitaciones. La inteligencia artificial no es magia, y describirla como tal puede reforzar una visión determinista de la tecnología, que inhibe la capacidad crítica y de comprensión de la misma.

Las metáforas, cuando se explicitan con claridad, son herramientas analíticas valiosas. Reflejan e influyen en cómo se imaginan, diseñan, implementan y, sobre todo, cómo se difunden las tecnologías. Como señalan Weller y Hofstadter, las metáforas y analogías no son solo recursos estilísticos, sino marcos de pensamiento que orientan nuestras decisiones (Weller, 2022; Hofstadter y Sander, 2018).

Por ejemplo, la metáfora de los “nativos digitales” moldeó políticas educativas durante años, muchas veces con resultados erróneos. De forma similar, la idea de una “Uber de la educación” inspiró modelos de negocio que ignoraban la complejidad del aprendizaje (Adell-Segura et al., 2018).

En la actualidad, se percibe una presión creciente para integrar la GenAI en los sistemas educativos. Ante esta situación, este trabajo presenta en el apartado dos, un recorrido crítico por la historia de la tecnología educativa, con el objetivo de analizar cómo distintas innovaciones han sido introducidas, qué desafíos han planteado y qué marcos conceptuales, a menudo en forma de metáforas, han guiado su comprensión y uso. A partir de este análisis histórico y conceptual, en el apartado tres se identifican principios éticos fundamentales que los autores consideran necesarios para una incorporación responsable de la GenAI en contextos educativos. A continuación, en la sección cuatro se propone una aplicación práctica que traduce estos principios en orientaciones concretas para el diseño y uso de sistemas de aprendizaje adaptativo basados en inteligencia artificial. Finalmente, se exponen las conclusiones del artículo.

MARCO ANALÍTICO: EVOLUCIÓN METAFÓRICA DE LAS TECNOLOGÍAS EDUCATIVAS

A lo largo de la historia, la introducción de nuevas tecnologías en el ámbito educativo no ha sido guiada únicamente por avances técnicos, sino también por narrativas, metáforas e imaginarios que han dado sentido a su adopción y modelado sus aplicaciones. En este apartado se explora cómo determinadas metáforas han moldeado la evolución de la tecnología educativa en las últimas décadas y qué lecciones se pueden extraer de esta historia para acabar explicando el enfoque actual respecto a la GenAI (Llorens-Largo, 2019). Como recuerda Weller (2022), “La tecnología suele venir envuelta en sus propias metáforas” (p. 59), y esas metáforas iniciales pueden tener profundas consecuencias sobre la forma en que dicha tecnología se desarrolla y se aplica. En esta línea, Mason (2018) demuestra empíricamente cómo las metáforas subyacentes estructuran el discurso de la tecnología educativa, sugiriendo que aunque es necesario usar metáforas para comprender nuevos fenómenos digitales, algunas de las metáforas específicas comúnmente empleadas pueden impedir

enfoques más reflexivos para conceptualizar e implementar nuevas tecnologías en el ámbito educativo.

Se ha dividido este apartado en cuatro momentos clave de la evolución de las tecnologías educativas, analizando las metáforas dominantes que acompañaron cada etapa y sus implicaciones en las formas de aprendizaje, especialmente en relación con la personalización y adaptación.

De Multivac al mainframe educativo: la metáfora de la educación como sistema centralizado

Isaac Asimov, en sus relatos sobre Multivac, imaginó una supercomputadora capaz de tomar decisiones cruciales para toda la humanidad. Esta visión refleja una idea poderosa: la tecnología como entidad centralizada que recopila datos, procesa información masiva y dirige las acciones de la sociedad (Asimov, 1955-1976).

Esta visión centralizada y autoritaria del ordenador también se reflejó en las primeras propuestas de la Enseñanza Asistida por Ordenador (EAO) o Instrucción Asistida por ordenadores. B. F. Skinner, figura clave del conductismo, planteó en su obra *The Technology of Teaching* (Skinner, 1968) que los ordenadores podían optimizar el aprendizaje mediante la aplicación sistemática de principios conductistas. Estos sistemas no concebían el aprendizaje como un proceso abierto o exploratorio, sino como una serie de trayectorias programadas y controladas desde un único centro.

Aplicado a los sistemas de aprendizaje basados en GenAI, Multivac representa una visión donde el sistema informático dicta el ritmo y el contenido que se presenta al alumno, relegando la adaptabilidad a un modelo de supervisión rígido. Se priorizaba la optimización de resultados predefinidos, no la personalización genuina basada en intereses, contextos o motivaciones individuales.

Esta metáfora sigue siendo relevante hoy cuando ciertos sistemas tienden a priorizar la eficiencia de la instrucción sobre la autonomía del estudiante, reproduciendo una lógica de control centralizado más que de aprendizaje autodirigido.

De Matrix a las píldoras educativas: la metáfora del aprendizaje como transferencia instantánea de conocimiento

En la película *The Matrix* (Wachowski y Wachowski, 1999), los personajes adquieren habilidades complejas, como pilotar helicópteros, mediante la descarga inmediata de programas en su cerebro. Esta metáfora de la “transferencia instantánea de conocimiento” se hizo muy influyente durante la era del CD-ROM y las primeras plataformas e-learning durante la segunda mitad de los 1990s.

En educación, se tradujo en el desarrollo de unidades de aprendizaje cerradas y estandarizadas, las llamadas “píldoras educativas”, que pretendían ofrecer formación

rápida, modular y autónoma. Bajo esta lógica, aprender se convirtió en “consumir” paquetes de información discretos, no en construir conocimiento activamente.

La relación con las herramientas de IA generativa contemporáneas es clara: muchos enfoques basados en IA siguen asumiendo que el aprendizaje es una cuestión de administrar la dosis adecuada de contenido, ajustada a la velocidad o nivel del alumno, pero sin necesariamente considerar la dimensión crítica, reflexiva o social del aprendizaje.

La metáfora de Matrix alimenta la promesa de la GenIA como solución mágica, pero también oculta los riesgos de simplificar el aprendizaje a un proceso meramente receptivo, pasivo y descontextualizado.

Del Bazar al Campus Virtual: la metáfora de la construcción colaborativa de saberes

La siguiente revolución en la tecnología educativa llegó de la mano de la web, y la web no puede entenderse sin la cultura del software libre. Este movimiento nació en los laboratorios del MIT, cuando Richard Stallman, en respuesta a las restricciones impuestas por las licencias de usuario de software (EULA), que anteriormente se compartían libremente entre programadores, publicó el *Manifiesto GNU* (Stallman, 1985) y fundó la *Free Software Foundation (FSF)* en 1985. La FSF proporcionó un marco legal fundamental: la licencia GPL, que permitía a los desarrolladores compartir su software garantizando cuatro libertades esenciales a los usuarios: usar, estudiar, compartir y mejorar el software. Esta filosofía, resumida en el lema “libre como en libertad, no como en cerveza gratis” (“*free as in freedom not as in free beer*”), facilitó la colaboración y el intercambio de código fuente a una escala sin precedentes (Ceruzzi, 2003).

En contraste con las visiones centralizadas anteriores, el movimiento del software libre, se puede sintetizar en la metáfora del Bazar (Raymond, 2001), que introdujo un nuevo paradigma en los sistemas de aprendizaje: sistemas abiertos, donde múltiples actores colaboran, innovan y adaptan soluciones de manera descentralizada y democrática.

En su influyente ensayo *The Cathedral and the Bazaar* (Raymond, 2001), Eric S. Raymond observa que las comunidades de software libre, donde el conocimiento y el código se comparten y los miembros ganan influencia en función de sus contribuciones más que de su estatus o posición, ofrecen una solución a la “crisis endémica del software” descrita por Pressman en *Software Engineering: A Practitioner’s Approach* (Pressman, 1991). Esta aplicación práctica de los principios del software libre genera una escisión en el movimiento con la aparición de la *Open Source Initiative (OSI)*. La OSI propuso el término “software de código abierto” (*Open Source Software*) como alternativa al “software libre”, introduciendo licencias más flexibles (como Apache o MIT) que resultaban más atractivas para las grandes empresas.

Esta filosofía inspiró la creación de entornos como Moodle y Sakai, plataformas de gestión de aprendizaje que, al ser de código abierto, permitieron a docentes, estudiantes, instituciones y desarrolladores adaptar las herramientas a sus necesidades específicas. No se trataba solo de consumir contenidos, sino de construir entornos de aprendizaje dinámicos, compartidos y configurables.

En este contexto, la metáfora del Bazar señala la posibilidad de diseñar sistemas que no sean cajas negras, sino espacios abiertos donde las adaptaciones se comprendan, negocien y reconfiguren colectivamente. Esta perspectiva refuerza la idea de que la adaptabilidad no debe ser solo algorítmica, sino también pedagógica y social.

De la App Store a la nube educativa: la metáfora de la dependencia invisible

Con la irrupción de los teléfonos inteligentes y de las tiendas de aplicaciones —la App Store de Apple y Google Play— se afianzó la metáfora de que “hay una app para todo”. La promesa era contar con soluciones inmediatas, personalizadas y accesibles, pero a cambio de una dependencia casi total de infraestructuras propietarias y opacas —plataformas— que venden seguridad y conveniencia al precio del control. Como recuerda Ben Thompson al explicar el verdadero motor de esas dinámicas: “la realidad es que las plataformas no son un problema huevogallina: lo primero que cuenta son los usuarios, que atraen a los desarrolladores a pesar de los obstáculos” (Thompson, 2025). En otras palabras, más que cada app individual, la plataforma misma es el centro de poder.

En educación, se dio una proliferación de apps educativas cuya información y gestión ya no se da en sistemas controlados por las instituciones, sino en “la nube” (una nueva metáfora), ofreciendo experiencias de aprendizaje a medida, pero a cambio de una pérdida de control sobre los datos, los procesos y los criterios de personalización. Los algoritmos deciden qué contenidos recomendar, qué rutas seguir, qué evaluaciones aplicar, casi siempre sin transparencia ni participación real de educadores o instituciones. Obsérvese como “la nube” y el “algoritmo” se convierten en los sujetos de análisis y crítica, cuando en realidad son instrumentos operados por las empresas que las diseñan y controlan.

Esta metáfora nos invita a preguntarnos: ¿quién adapta para quién? ¿Qué intereses comerciales, qué sesgos culturales o técnicos modelan los caminos que recorren los estudiantes?

La metáfora de la plataforma —el App Store— advierte que el aprendizaje personalizado puede convertirse en una personalización simulada, predefinida por intereses corporativos, si no se establecen mecanismos de soberanía tecnológica y pedagógica.

Metáforas alternativas para el aprendizaje adaptativo impulsado por inteligencia artificial

Si bien la metáfora del Palantír resulta útil para representar la fascinación y el temor que rodean el uso de la inteligencia artificial en contextos educativos, necesitamos marcos que nos permitan pensar en formas más críticas y pedagógicamente útiles de integrar estas tecnologías. Para ello los autores proponen dos metáforas alternativas que pueden ayudar a imaginar un sistema de apoyo al aprendizaje con GenAI más abierto, controlado por las comunidades educativas y centrado en el aprendizaje significativo: el conjunto de bloques de Lego y el dilema de Prometeo.

La metáfora de bloques de Lego se puede utilizar para explicar que, en lugar de concebir la IA generativa como un artefacto cerrado y mágico, la metáfora de Lego plantea una visión más accesible, modular y comprensible. Bajo este enfoque, la inteligencia artificial no es un “oráculo” que ofrece rutas de aprendizaje automáticas e incuestionables, sino un conjunto de piezas reutilizables que pueden combinarse de múltiples formas para diseñar experiencias de aprendizaje personalizadas, contextuales y con fundamentos pedagógicos.

Aplicado a las herramientas de aprendizaje basados en GenAI, esto implica que:

- Las herramientas basadas en IA generativa pueden ser configuradas por docentes para responder a las necesidades reales de sus estudiantes.
- El “algoritmo” no sustituye al educador, sino que le ofrece materiales con los que construir rutas alternativas.
- Las adaptaciones no se producen solo en función del rendimiento o la velocidad, sino también de intereses, lenguajes, motivaciones y trayectorias diversas.

Este enfoque conecta con una visión constructivista y participativa del aprendizaje, donde la tecnología no reemplaza las bases pedagógicas, sino que las amplifica. En este sentido, puede vincularse con la metáfora de la *construcción*, propuesta por Weller (2022), que entiende la tecnología educativa como un proceso en el que los usuarios (docentes y estudiantes) construyen activamente sus herramientas y entornos de aprendizaje, promoviendo la autonomía y la personalización. Además, permite imaginar futuros en los que los sistemas sean transparentes, auditables y creados por comunidades educativas, evitando así la dependencia de plataformas opacas o decisiones automatizadas que no pueden ser explicadas.

Una segunda metáfora relevante es la del dilema prometeico. En la mitología griega, Prometeo roba el fuego de los dioses y lo entrega a los humanos, un acto que representa tanto un avance para la civilización como un castigo potencial para la humanidad. “En el tallo de una caña escondí la chispa, el manantial del fuego robado, que para los mortales es maestra de todas las artes; y por ese delito pago ahora la pena, clavado con cadenas bajo el éter.” (Jonas, 1984).

Las herramientas impulsadas por GenIA pueden entenderse como sistemas con capacidad para ofrecer recomendaciones “personalizadas”, adaptar contenidos en tiempo real y predecir trayectorias educativas. Todo esto representa una forma de “fuego”, poderoso y transformador, pero también un riesgo si se usa sin tener en cuenta la ética, y la transparencia.

Esta metáfora alerta sobre:

- La posibilidad de delegar en exceso el juicio educativo a algoritmos automatizados, desplazando el papel del docente.
- El riesgo de generar trayectorias limitadas o cerradas para los estudiantes, guiados por patrones históricos de datos, en lugar de abrir caminos nuevos.
- Las implicaciones de reducir el aprendizaje a patrones de consumo o eficiencia, ignorando dimensiones más complejas como la creatividad, el conflicto o la exploración.

Así, mientras la metáfora del Lego apunta a la posibilidad de uso emancipador, la metáfora prometeica nos recuerda que toda tecnología poderosa conlleva tensiones éticas y políticas, especialmente cuando se despliega a gran escala en sistemas educativos.

PRINCIPIOS ÉTICOS PARA LA INTRODUCCIÓN SEGURA DE LA GENAI EN EDUCACIÓN

La creciente presión para integrar la GenAI en los sistemas educativos plantea no solo oportunidades, sino también desafíos éticos, legales y pedagógicos de gran calado. A nivel regulatorio, en la Unión Europea, los sistemas de GenAI deben garantizar una adopción segura, respetuosa con las normativas de privacidad de la Unión Europea (European Parliament & Council of the European Union, 2016). Además, su irrupción reciente ha impulsado el desarrollo de legislaciones específicas sobre inteligencia artificial en diversas regiones, como la Ley de IA de la Unión Europea (European Parliament, 2024) o las regulaciones emergentes en China (Cyberspace Administration of China et al., 2023).

Para garantizar una adopción segura, respetuosa con las normativas de privacidad y alineada con los valores, estrategias y prácticas institucionales, los autores proponen en esta sección una serie de principios que orienten la evaluación y despliegue de aplicaciones de GenAI en educación. Estos principios buscan asegurar que las tecnologías basadas en GenAI: 1) Se alineen con los objetivos educativos de las instituciones. 2) Mantengan niveles adecuados de seguridad, precisión e integridad ética. 3) Favorezcan un aprendizaje de calidad y equitativo, minimizando riesgos asociados al uso indebido, la privacidad o la desinformación.

Siguiendo estos principios, las instituciones educativas podrán aprovechar las oportunidades que ofrece la GenAI al tiempo que mitigan los riesgos asociados a su implementación en contextos de aprendizaje adaptativo.

Principios de GenAI segura en educación

Más allá del cumplimiento legal, la integración ética de la inteligencia artificial requiere principios prácticos y accionables que permitan evaluar tanto la adecuación pedagógica como la seguridad tecnológica de las soluciones adoptadas.

A partir de un análisis de los marcos legales y de la naturaleza técnica de la GenAI, Alier, García-Peñalvo y Camba (2024) han propuesto una serie de principios que, posteriormente, se han ampliado a siete directrices específicas para entornos educativos (García-Peñalvo et al., 2024). Estos principios permiten realizar una evaluación técnica, pedagógica y ética de las estrategias de integración de la GenAI en educación. A continuación, se exponen dichos principios.

- **(SAIE1) Garantía de confidencialidad:** los sistemas de GenAI deben proteger estrictamente los datos de los estudiantes, asegurando la privacidad de sus identidades, expedientes e interacciones educativas.
- **(SAIE2) Alineación con las estrategias educativas:** las herramientas de GenAI deben apoyar los objetivos institucionales y adherirse a las políticas de gobernanza tecnológica, facilitando el aprendizaje y la creación de conocimiento sin fomentar malas prácticas como el plagio o la elusión de controles académicos.
- **(SAIE3) Ajuste a las prácticas didácticas:** las aplicaciones de GenAI deben actuar bajo parámetros pedagógicos definidos. No deben resolver automáticamente problemas o tareas, sino guiar al estudiante en procesos de razonamiento alineados con el diseño instruccional.
- **(SAIE4) Precisión y minimización de errores:** los sistemas de IA deben priorizar la veracidad y relevancia de sus respuestas, especialmente para minimizar los riesgos de información errónea derivada de sesgos o alucinaciones de los modelos.
- **(SAIE5) Interfaz comprensible y comportamiento adecuado:** la GenAI debe presentar sus funcionalidades y limitaciones de manera clara para docentes y estudiantes, evitando falsas expectativas o interpretaciones erróneas de su nivel de fiabilidad.
- **(SAIE6) Supervisión humana y responsabilidad:** las decisiones basadas en GenAI deben ser supervisadas por personas. El profesorado debe mantener el control sobre los procesos educativos, asegurando la posibilidad de apelación y revisión de las decisiones automatizadas.
- **(SAIE7) Entrenamiento ético y transparencia:** los modelos de GenAI deben entrenarse de manera ética, garantizando transparencia sobre sus fuentes

de datos y metodologías, y minimizando los sesgos para asegurar resultados justos y explicables.

La aplicación de estos principios tiene importantes consecuencias prácticas para las instituciones educativas:

- **Confidencialidad (SAIE1):** las instituciones deben garantizar el control sobre las plataformas de IA utilizadas, evitando la obligatoriedad de registros en herramientas externas como ChatGPT o Gemini, lo cual podría vulnerar la privacidad estudiantil.
- **Alineación estratégica (SAIE2):** herramientas de propósito general, como los modelos de lenguaje de gran tamaño o Large Language Model en inglés (LLM), presentan dificultades de integración institucional debido a su complejidad de uso (Willison, 2023) y los riesgos asociados a la integridad académica (González-Geraldo y Ortega-López, 2024). Además, agregar complejidad al proceso de aprendizaje no es una buena práctica pedagógica, ya que aumenta la carga cognitiva del estudiantado (Chen et al., 2023).
- **Adaptación didáctica (SAIE3):** las aplicaciones deben integrarse de forma coherente con las metodologías educativas. Por ejemplo, en entornos como la formación médica (Hwang et al., 2024), los sistemas de GenAI deben actuar como asistentes en el razonamiento clínico, no como sustitutos automáticos de los procesos analíticos de los estudiantes.
- **Precisión y minimización de errores (SAIE4):** en entornos educativos, resulta fundamental diseñar sistemas de IA adaptativos capaces de citar fuentes y proporcionar mecanismos de validación de información (Tonmoy et al., 2024).
- **Interfaz y transparencia (SAIE5):** el diseño de las interfaces debe favorecer la comprensión del alcance y limitaciones de la GenAI, evitando la presentación de respuestas especulativas como verdades absolutas.
- **Supervisión y responsabilidad (SAIE6-7):** la supervisión humana no solo asegura la calidad educativa, sino que protege la equidad en el aprendizaje y garantiza que los sistemas adaptativos no reproduzcan sesgos o exclusiones sistemáticas.

Estos principios han dado lugar al desarrollo del Safe AI Manifesto (Alier, García-Peñalvo, Casañ et al., 2024), un documento vivo que recoge los compromisos necesarios para una integración ética y responsable de la GenAI en educación, disponible para consulta y suscripción en <https://manifesto.safeaieducation.org>. Además, los principios anteriores permiten realizar una evaluación técnica, pedagógica y ética de las estrategias de integración de la GenAI en educación, tal como se desarrolla en García-Peñalvo et al. (2024).

Relación de los principios de GenAI segura en educación y las metáforas educativas

Cada uno de los principios de GenAI segura que se proponen responden, de alguna manera, a los riesgos o aspiraciones representados por las metáforas históricas analizadas en el apartado anterior:

- **(SAIE1) Garantía de confidencialidad** combate el riesgo de la App Store y la nube educativa: La metáfora de la App Store muestra cómo la dependencia de plataformas opacas pone en riesgo la privacidad. La confidencialidad exige revertir esa lógica, asegurando que los datos de los estudiantes estén protegidos y bajo control institucional.
- **(SAIE2) Alineación con las estrategias educativas** responde al peligro de Matrix. Frente a la visión del aprendizaje como descarga pasiva de información predefinida, alinear la GenAI con las estrategias educativas permite construir sistemas que fomenten procesos activos de aprendizaje, en vez de automatizar la transmisión de contenidos.
- **(SAIE3) Ajuste a las prácticas didácticas** conecta con la metáfora del Bazar. Como en el Bazar del software libre, el principio busca que las herramientas de GenAI sean configurables y adaptables por los docentes a las prácticas pedagógicas específicas, no productos cerrados que impongan metodologías estándar.
- **(SAIE4) Precisión y minimización de errores** combate las ilusiones de la visión prometeica. El dilema de Prometeo nos recuerda que el “fuego” de la GenAI puede ser peligroso si no se gestiona adecuadamente. Minimizar errores y alucinaciones ayuda a controlar ese riesgo inherente a sistemas tan potentes como frágiles.

LAMB: UNA PROPUESTA PRÁCTICA PARA INTEGRAR LA GENIA SEGURA EN EDUCACIÓN

Tras analizar la evolución de las metáforas que han acompañado a las tecnologías educativas y definir los principios de GenAI segura (SAIE), resulta necesario mostrar ejemplos concretos que incluyan estos principios en su diseño y aplicación. En esta sección se presenta LAMB (Learning Assistant Manager and Builder), un marco de software diseñado específicamente para crear asistentes de aprendizaje basados en GenAI de manera segura, controlada y pedagógicamente alineada con los principios de la institución educativa. En primer lugar, se explica el concepto de asistente de aprendizaje basado en GenAI. A continuación, se define LAMB y como este marco permite crear asistentes de aprendizaje basados en GenAI. Después de presentar casos de uso de LAMB, como este marco está alineado con los principios presentados en la sección 3 y, finalmente, su relación con las metáforas educativas.

Asistentes de aprendizaje basados en GenIA

Un asistente de aprendizaje basado en GenIA es una herramienta conversacional que interactúa con los estudiantes para ofrecer apoyo educativo personalizado. A diferencia de un chatbot genérico, un asistente de aprendizaje está diseñado para cumplir funciones específicas dentro de un contexto educativo (Kochmar et al., 2020; Wollny et al., 2021). La aparición de herramientas como ChatGPT ha marcado un punto de inflexión en este campo, transformando las posibilidades de personalización e interacción educativa (Bettayeb et al., 2024; Chan y Lee, 2023):

- Responder a preguntas sobre el contenido de un curso.
- Guiar en la resolución de problemas aplicando métodos pedagógicos definidos.
- Sugerir recursos de apoyo pertinentes.
- Promover el razonamiento y la autoexplicación en los estudiantes.
- Acompañar procesos de evaluación formativa proporcionando retroalimentación ajustada.

Estos asistentes no buscan reemplazar al profesorado, sino amplificar su capacidad de acompañamiento y retroalimentación, especialmente en entornos de gran número de estudiantes o en modalidades híbridas y a distancia.

LAMB

LAMB es un entorno de software que permite a instituciones educativas y docentes crear asistentes de aprendizaje propios, adaptados a cada asignatura, curso o necesidad concreta. Este marco también permite integrar estos asistentes de manera sencilla en plataformas de gestión de aprendizaje (LMS) como Moodle, a través del estándar IMS LTI (Learning Tools Interoperability), de forma que los estudiantes acceden a los asistentes de aprendizaje a través del LMS de la institución (IMS Global Learning Consortium [IMS GLC], 2014). Además, LAMB permite controlar y configurar el conocimiento base que alimenta las respuestas del asistente, asegurando la alineación pedagógica con la materia. Adicionalmente, este marco permite gestionar la seguridad, privacidad y supervisión de los asistentes creados.

Alier, Pereira et al. (2024) presentan la descripción técnica, así como la arquitectura detallada de LAMB. De forma resumida y desde una perspectiva técnica, LAMB combina:

- Modelos de lenguaje de gran tamaño (LLMs) que generan respuestas en lenguaje natural.

- Técnicas de generación aumentada por recuperación o Retrieval-Augmented Generation en inglés (RAG), que limitan las respuestas del asistente a información verificada proveniente de fuentes controladas por la institución.
- Un sistema de plantillas de interacción y configuraciones pedagógicas que el docente define previamente.

De esta forma, LAMB permite aprovechar la potencia de la IA generativa, pero bajo condiciones de control, transparencia y responsabilidad humana.

Ejemplos de asistentes creados con LAMB

Para ilustrar la aplicación real de LAMB en contextos educativos, se presentan algunas de las experiencias realizadas.

En primer lugar, hay que destacar la experiencia desarrollada en la asignatura de Economía y Entorno económico, una asignatura obligatoria de segundo curso en el Grado de Ingeniería en Informática de la Facultad de Informática de Barcelona.

Antes de abordar experiencias piloto con estudiantes, es importante contextualizar el proceso de desarrollo de LAMB a través del asistente llamado “Macroeconomics Study Coach”, que funcionó como prototipo para el diseño y refinamiento del marco LAMB. Este asistente no fue sometido a evaluación formal con estudiantes, sino que sirvió como banco de pruebas técnico y pedagógico durante las iteraciones de desarrollo del sistema LAMB.

Este asistente fue creado a partir de 30 vídeos de conferencias del curso. Su funcionalidad principal consistió en responder preguntas de los estudiantes basándose en las transcripciones de los vídeos y documentos PDF complementarios, proporcionando citas específicas y enlaces a los vídeos con instantes precisos donde se abordaban las consultas realizadas.

Una vez consolidada la estabilidad del framework a través de este caso, se procedió a diseñar e implementar experiencias controladas con estudiantes, siendo el asistente PESTLE la primera de estas validaciones empíricas.

El asistente de aprendizaje del caso de PESTLE fue diseñado como recurso para la asignatura Fundamentos de Administración y Dirección de Empresas, impartida en el grado de Ingeniería en Informática en la EPSEVG (Escuela Politécnica Superior de Ingeniería de Vilanova y la Geltrú). El asistente creado ayudaba a los estudiantes a trabajar en un caso de evaluación de un proyecto empresarial impulsado por tecnología. El asistente había sido dotado de una base de conocimiento específica para el caso. Los estudiantes, organizados en equipos, debían analizar el caso utilizando la metodología PESTLE (Política, Economía, Sociedad, Tecnología, Legal y Medioambiente), lo cual implicaba que debían plantear preguntas desde varias dimensiones antes de redactar su informe final. Estas preguntas las plantearon al asistente de aprendizaje creado con LAMB. En cursos anteriores, los estudiantes

simulaban ser expertos con ayuda de motores de búsqueda, documentos específicos e incluso ChatGPT en el otoño de 2023 (Casañ et al., 2024).

La experiencia con el asistente para ayudar a los estudiantes en el caso de PESTLE se desarrolló durante el curso 2023-2024. La actividad involucró un total de 47 estudiantes de segundo curso del grado organizados en grupos de 6-8 miembros.

La experiencia se estructuró siguiendo una metodología de estudio de caso centrada en el análisis del robot humanoide Optimus de Tesla, utilizando el framework PESTLE (Political, Economic, Social, Technological, Legal, Environmental). La actividad se desarrolló durante dos sesiones de clase de 2 horas cada una, con una semana de intervalo entre ellas para permitir a los estudiantes continuar el análisis de forma autónoma.

En la primera sesión se realizó la presentación del caso y explicación de la metodología PESTLE, el trabajo colaborativo en grupos para analizar las seis dimensiones del PESTLE, así como la utilización del asistente para obtener información de expertos. Durante la semana los estudiantes tuvieron tiempo para seguir con el análisis utilizando el asistente desde casa. En la segunda sesión se introdujo el análisis SWOT integrado con PESTLE. A continuación, se pidió a los estudiantes que categorizaran los elementos o aspectos que habían encontrado en cada dimensión del PESTLE como “altamente relevante”, “relevante” o “poco relevante”. Para acabar los estudiantes elaboraron un informe final con conclusiones.

La metodología de recogida de datos para evaluar la efectividad del asistente, se administró mediante un cuestionario no obligatorio de 5 preguntas utilizando una escala Likert de 5 puntos (1=totalmente en desacuerdo, 5=totalmente de acuerdo). El cuestionario obtuvo **27 respuestas** de un total de 47 estudiantes (tasa de respuesta del 57,4 %). Las preguntas del cuestionario fueron las siguientes:

1. El asistente me ha ayudado a encontrar información relevante para el caso más rápidamente que si hubiera tenido que hacerlo yo mismo utilizando internet.
2. Las preguntas adicionales sugeridas por el sistema son buenas incitaciones para buscar más información sobre el caso.
3. Las respuestas a las preguntas sugeridas por el sistema proporcionan información útil.
4. Poder consultar las fuentes de datos ha sido útil.
5. La interfaz del asistente es sencilla y fácil de usar.

Los resultados muestran una recepción mayoritariamente positiva del asistente por parte de los estudiantes:

- **Pregunta 1** (Eficiencia en la búsqueda de información): 83,3 % de los estudiantes puntuaron con 4 o 5 (17 respuestas favorables de 27 total).
- **Pregunta 2** (Calidad de las preguntas sugeridas): 77,7 % puntuaron con 4 o 5.
- **Pregunta 3** (Utilidad de las respuestas): 85,2 % puntuaron con 4 o 5.

- **Pregunta 4** (Acceso a fuentes): 81,5 % puntuaron con 4 o 5.
- **Pregunta 5** (Usabilidad de la interfaz): 92,6 % puntuaron con 4 o 5.

Los resultados indican que ningún estudiante puntuó con 1 o 2 en las preguntas 1, 2 y 4, y solo un pequeño porcentaje lo hizo en las preguntas 3 y 5, evidenciando una ausencia significativa de valoraciones negativas.

Los resultados cualitativos y cuantitativos confirman varios beneficios pedagógicos:

En general, como retroalimentación positiva en la experiencia, las respuestas de los estudiantes destacaron beneficios como la consistencia y alineación. Se destacó la capacidad de los asistentes para ceñirse a la base de conocimiento proporcionada, sin generar información irrelevante o alucinaciones. En cuanto a la reducción de carga de trabajo, se valoró la utilidad de los asistentes para responder consultas repetitivas sobre procedimientos del curso. En el caso específico del asistente para ayudar a los estudiantes a resolver una evaluación de un proyecto empresarial, se observó una mejora en la calidad de los informes estudiantiles gracias al acceso a conocimiento experto.

Entre los retos y limitaciones cabe destacar la dependencia excesiva en este tipo de herramientas, ya que se advirtió sobre el riesgo de que los estudiantes usen el asistente como atajo, sin profundizar en los materiales del curso proporcionados por el profesorado. También se propusieron mejoras futuras de LAMB para una integración más fluida con el LMS, concretamente para permitir enviar consultas al docente cuando el asistente no tiene respuesta. Finalmente, se señaló la necesidad de apoyar lenguas minoritarias como el catalán o el euskera. En cursos anteriores, los estudiantes simulaban ser expertos utilizando motores de búsqueda, documentos específicos y, en el trimestre de otoño de 2023, ChatGPT de forma no controlada. La implementación del asistente LAMB proporcionó un entorno más controlado y pedagógicamente alineado, con información verificada y contextualizada específicamente para el caso de estudio.

Relación entre LAMB y los principios de GenIA segura en educación

El diseño de LAMB refleja de forma directa el cumplimiento de los siete principios de IA segura planteados en el apartado tres:

- **SAIE1 (Confidencialidad):** LAMB garantiza la privacidad de los datos estudiantiles al operar sobre servidores de la institución o mediante APIs controladas. La información sensible no se expone a plataformas externas no autorizadas.
- **SAIE2 (Alineación con estrategias educativas):** los asistentes creados con LAMB solo utilizan fuentes aprobadas por la institución, asegurando que

el contenido y las interacciones estén alineados con los objetivos y valores educativos institucionales.

- **SAIE3 (Ajuste a prácticas didácticas):** LAMB permite que los docentes definan plantillas de interacción y bases de conocimiento específicas para cada curso o actividad, integrándose de manera coherente con los diseños instruccionales, en lugar de imponer interacciones genéricas.
- **SAIE4 (Precisión y minimización de errores):** mediante técnicas de generación aumentada por recuperación (RAG), los asistentes basan sus respuestas en documentos de referencia fiables, reduciendo el riesgo de alucinaciones o errores típicos de los LLM generalistas.
- **SAIE5 (Interfaz comprensible y comportamiento adecuado):** LAMB proporciona asistentes con comportamientos predecibles y delimitados, aclarando su función y sus limitaciones, evitando la ilusión de infalibilidad que suele acompañar a muchos sistemas de IA.
- **SAIE6 (Supervisión humana y responsabilidad):** el profesorado diseña y supervisa los contenidos, estrategias y comportamientos del asistente, preservando su rol central como mediador pedagógico. Esta aproximación es consistente con experiencias prácticas donde los LLMs actúan como asistentes que automatizan tareas repetitivas (revisión de código, seguimiento de proyectos), pero manteniendo al docente en el rol de guía y supervisor del proceso educativo (Pereira et al., 2025).
- **SAIE7 (Entrenamiento ético y transparencia):** aunque los modelos subyacentes son de terceros, LAMB permite a las instituciones construir bases de interacción propias, que en el futuro podrán utilizarse para ajustar o entrenar modelos adaptados éticamente a sus contextos.

LAMB y las metáforas en educación

La implementación práctica de LAMB y los asistentes de aprendizaje en contextos reales ha revelado dificultades que, de manera significativa, validan las advertencias contenidas en el análisis metafórico desarrollado en la primera parte de este artículo. Esta convergencia entre teoría crítica y experiencia empírica refuerza la importancia de mantener una vigilancia activa sobre las narrativas que acompañan la integración tecnológica en educación.

Una de las preocupaciones más evidentes observadas en las experiencias con LAMB es el riesgo de dependencia excesiva por parte de los estudiantes, quienes pueden utilizar el asistente como un atajo que les evite profundizar en los materiales del curso proporcionados por el profesorado. Esta dificultad práctica materializa precisamente los peligros anticipados en la metáfora del Palántir.

Esta observación no invalida la utilidad de los asistentes, sino que refuerza la necesidad de marcos pedagógicos claros que orienten su uso. Como señaló uno de los expertos consultados: *“debe quedar claramente definido que su ‘conocimiento’*

no va más allá del contenido correspondiente y no trata contenidos más allá de las necesidades del profesor”. La metáfora del Palantír nos recuerda que el poder de la tecnología debe ir acompañado de estructuras de control y reflexión ética.

Las dificultades de integración con los sistemas de gestión del aprendizaje (LMS) observadas en las experiencias prácticas reflejan la tensión entre las metáforas del Bazar y la App Store analizadas anteriormente. Los estudiantes han solicitado mejoras para una integración más fluida con la plataforma Moodle (el LMS utilizado en la institución educativa), especialmente la posibilidad de realizar consultas al docente cuando el asistente no tenga respuesta adecuada.

Esta demanda ilustra la limitación inherente a los sistemas cerrados (metáfora de la App Store). Cuando las herramientas tecnológicas operan de manera aislada, sin capacidad de interconexión y colaboración, se genera frustración y se limita el potencial pedagógico. En contraste, la metáfora del Bazar sugiere la necesidad de sistemas abiertos, modulares y colaborativos donde diferentes herramientas y actores (estudiantes, profesores, sistemas) puedan interactuar de manera fluida.

LAMB, por diseño, intenta responder a esta tensión mediante su arquitectura modular y su integración con estándares abiertos como LTI. Sin embargo, las dificultades prácticas observadas subrayan que la apertura técnica no es suficiente: se requiere también una apertura pedagógica que permita la conexión entre la inteligencia artificial y la inteligencia humana del profesorado.

La necesidad identificada de soporte para lenguas minoritarias como el catalán y el euskera constituye una manifestación concreta de las limitaciones de la metáfora de Matrix discutida en el marco conceptual. La visión de Matrix sugiere que el conocimiento puede “descargarse” de manera instantánea y universal, ignorando las realidades contextuales, culturales y lingüísticas específicas.

La experiencia práctica con LAMB demuestra que esta universalidad aparente es ilusoria. Los modelos de lenguaje, por muy avanzados que sean, reflejan los sesgos y limitaciones de sus datos de entrenamiento, que tienden a favorecer las lenguas mayoritarias.

Desde una perspectiva conceptual, LAMB, rechaza la lógica de la App Store y del Palantír, donde los usuarios dependen de sistemas cerrados y opacos. En cambio, se alinea con la metáfora de Lego: los asistentes son herramientas modulares, personalizables y transparentes, adaptables.

En resumen, LAMB ofrece un ejemplo realista de cómo las instituciones pueden reapropiarse de la GenIA educativa como un recurso didáctico y no como una amenaza a su autonomía o a la agencia pedagógica.

CONCLUSIONES

Este artículo ha planteado una mirada crítica sobre la incorporación de la GenAI en educación, combinando análisis histórico, ético y conceptual. A través del uso de metáforas, se ha mostrado cómo las visiones sobre la tecnología educativa han

oscilado entre el control centralizado y la autonomía participativa. Metáforas como el Palantír, Matrix o la App Store ayudan a entender los riesgos actuales: opacidad, dependencia tecnológica y desalineación pedagógica. Frente a ello, se han propuesto marcos alternativos, como los bloques de Lego o el dilema prometeico, que invitan a una integración crítica, transparente y controlada por la comunidad educativa.

Sobre esta base, se han formulado siete principios para una GenIA segura, centrados en la confidencialidad, la alineación institucional, la supervisión humana y la transparencia. Estos principios no son teóricos, sino que se han aplicado en la práctica mediante LAMB, un entorno de software que permite a las instituciones diseñar y desplegar asistentes de aprendizaje configurables y auditables. Los casos analizados muestran beneficios claros en coherencia pedagógica, autonomía del alumnado y reducción de tareas repetitivas, pero también señalan retos como la dependencia o la integración con los LMS de las instituciones educativas.

En conjunto, este trabajo reivindica la necesidad de recuperar el control pedagógico sobre la tecnología, para asegurar que la inteligencia artificial en educación esté al servicio del aprendizaje significativo y no de intereses externos u opacos.

REFERENCIAS

- Adell-Segura, J., Castañeda, L. y Esteve-Mon, F. M. (2018). ¿Hacia la ubersidad? Conflictos y contradicciones de la universidad digital. *RIED-Revista Iberoamericana de Educación a Distancia*, 21(2), 51-68. <https://doi.org/10.5944/ried.21.2.20669>
- Alier, M., García-Peñalvo, F. J. y Camba, J. D. (2024). Generative artificial intelligence in education: From deceptive to disruptive. *International Journal of Interactive Multimedia and Artificial Intelligence*, 8(5), 5-14. <https://doi.org/10.9781/ijimai.2024.02.011>
- Alier, M., García-Peñalvo, F. J., Casañ, M. J., Pereira, J. A. y Llorens-Largo, F. (2024). *Safe AI in Education Manifesto (v 0.4.0)*. <https://manifesto.safeaieducation.org>
- Alier, M., Pereira, J., García-Peñalvo, F. J., Casañ, M. J. y Cabré, J. (2024). LAMB: An open-source software framework to create artificial-intelligence assistants deployed and integrated into learning-management systems. *Computer Standards & Interfaces*, 92, 103940. <https://doi.org/10.1016/j.csi.2024.103940>
- Asimov, I. (1955–1976). *Historias de Multivac* [Serie de relatos]. Diversas publicaciones.
- Bettayeb, A. M., Abu Talib, M., Sobhe Altayasinah, A. Z. y Dakalbab, F. (2024). Exploring the impact of ChatGPT: Conversational AI in education. *Frontiers in Education*, 9, 1379796. <https://doi.org/10.3389/feduc.2024.1379796>
- Casañ, M. J., Alier, M., Pereira, J. y García-Peñalvo, F. J. (2024). Asistentes de aprendizaje basados en inteligencia artificial: Principios de seguridad y experiencias de implementación en educación superior. En M. Navarro-Granados, J. J. Sánchez Amate, P. Berbel Oller y C. Rodríguez Jiménez (Eds.), *Investigación y conocimientos en la educación actual* (pp. 13-35). Dykinson.
- Ceruzzi, P. E. (2003). *A history of modern computing* (2ª ed.). MIT Press.

- Chan, C. K. Y. y Lee, K. K. W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*, 20(1), 43. <https://doi.org/10.1186/s41239-023-00411-8>
- Chen, O., Paas, F. y Sweller, J. (2023). A cognitive-load theory approach to defining and measuring task complexity through element interactivity. *Educational Psychology Review*, 35(2), 63. <https://doi.org/10.1007/s10648-023-09782-w>
- CyberspaceAdministrationofChina,National Development and Reform Commission, Ministry of Education, Ministry of Science and Technology, Ministry of Industry and Information Technology, Ministry of Public Security, & National Radio and Television Administration. (2023). *Interim measures for the management of generative artificial intelligence services*. Promulgadas el 13 de julio de 2023, en vigor desde el 15 de agosto de 2023. https://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm
- Clarke, A. C. (1962). *Profiles of the future: An inquiry into the limits of the possible*. Harper & Row.
- European Parliament. (2024, 13 de junio). *Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (published in OJ on 12.7.2024)*. <https://data.europa.eu/eli/reg/2024/1689/oj>
- European Parliament, & Council of the European Union. (2016, 27 de abril). *Regulation (EU) 2016/679 of the European Parliament and of the Council on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (OJ L 119, 4.5.2016, pp. 1-88)*. <https://data.europa.eu/eli/reg/2016/679/oj>
- García-Peñalvo, F. J., Alier, M., Pereira, J. y Casañ, M. J. (2024). Safe, transparent, and ethical artificial intelligence: Keys to quality sustainable education (SDG 4). *International Journal of Educational Research and Innovation*, 22, 1-21. <https://doi.org/10.46661/ijeri.11036>
- González-Geraldo, J. L. y Ortega-López, L. (2024). Can AI fool us? University students' ability to detect ChatGPT. *Education in the Knowledge Society*, 25, e31760. <https://doi.org/10.14201/eks.31760>
- Hofstadter, D. R. y Sander, E. (2018). *La analogía: El motor del pensamiento* (R. Musa, Trad.). Tusquets.
- Hwang, G.-J., Tang, K.-Y. y Tu, Y.-F. (2024). How artificial intelligence supports nursing education: Profiling the roles, applications and trends of AI in nursing education research (1993–2020). *Interactive Learning Environments*, 32(1), 373–392. <https://doi.org/10.1080/10494820.2022.2086579>
- IMS Global Learning Consortium. (2014). *IMS Learning Tools Interoperability (LTI) implementation guide v2.0*.
- Jonas, H. (1984). *The imperative of responsibility: In search of an ethics for the technological age*. University of Chicago Press. <https://doi.org/10.7208/chicago/9780226850337.001.0001>
- Kochmar, E., Vu, D. D., Belfer, R., Gupta, V., Serban, I. V. y Pineau, J. (2020). Automated personalized feedback improves learning gains in an intelligent tutoring system. En I. I. Bittencourt, M. Cukurova, K. Muldner, R. Luckin y E. Millán (Eds.), *Artificial Intelligence in Education: 21st International Conference, AIED 2020, Ifrane, Morocco*,

- July 6-10, 2020, *Proceedings, Part II* (Lecture Notes in Computer Science, Vol. 12164, pp. 140-146). Springer. https://doi.org/10.1007/978-3-030-52240-7_26
- Llorens-Largo, F. (2019, 13 de febrero). Las tecnologías en la educación: Características deseables, efectos perversos. *Universidad*, sí. <https://www.universidadsi.es/las-tecnologias-en-la-educacion-caracteristicas-deseables-efectos-perversos/>
- Mason, J. (2018). A critical metaphor analysis of educational technology research in the social studies. *Contemporary Issues in Technology and Teacher Education*, 18(3). <https://citejournal.org/volume-18/issue-3-18/social-studies/a-critical-metaphor-analysis-of-educational-technology-research-in-the-social-studies/>
- Pereira, J., López-Gil, J. M. y Alier, M. (2025). The AI-powered classroom: LLMs as teacher assistants for enhanced software-engineering learning experiences. En R. Molina Carmona, C. J. Villagrà Arnedo, P. Compañ Rosique, F. García-Peñalvo y A. García-Holgado (Eds.), *Proceedings of TEEM 2024: The Twelfth International Conference on Technological Ecosystems for Enhancing Multiculturality* (pp. 115-124). Springer. (Lecture Notes in Educational Technology). https://doi.org/10.1007/978-981-96-5658-5_12
- Pressman, R. S. (1991). *Software engineering: A practitioner's approach* (3ª ed.). McGraw-Hill.
- Raymond, E. S. (2001). *The cathedral & the bazaar*. O'Reilly.
- Skinner, B. F. (1968). *The technology of teaching*. Appleton-Century-Crofts.
- Stallman, R. M. (1985). *The GNU manifesto*. Free Software Foundation. <https://www.gnu.org/gnu/manifesto.html>
- Thompson, B. (2025, 5 de mayo). *Platform power is underrated*. Stratechery. <https://stratechery.com/2025/platform-power-is-underrated/>
- Tolkien, J. R. R. (1954-1955). *The lord of the rings* (Vols. 1-3). George Allen & Unwin.
- Tonmoy, S. M. T. I., Mehedi Zaman, S. M., Jain, V., Rani, A., Rawte, V., Chadha, A. y Das, A. (2024). A comprehensive survey of hallucination-mitigation techniques in large language models. *arXiv preprint arXiv:2401.01313*. <https://doi.org/10.48550/arXiv.2401.01313>
- Wachowski, L. y Wachowski, L. (1999). *The Matrix* [Película]. Warner Bros.
- Weller, M. (2022). *Metaphors of ed tech*. Athabasca University Press. <https://doi.org/10.15215/aupress/9781771993500.01>
- Willison, S. (2023). *Prompt injection attacks: It's only going to get worse*. <https://simonwillison.net/2023/Apr/25/prompt-injection>
- Wollny, S., Schneider, J., Di Mitri, D., Weidlich, J., Rittberger, M. y Drachsler, H. (2021). Are we there yet? A systematic literature review on chatbots in education. *Frontiers in Artificial Intelligence*, 4, 654924. <https://doi.org/10.3389/frai.2021.654924>

Fecha de recepción del artículo: 1 de junio de 2025

Fecha de aceptación del artículo: 27 de agosto de 2025

Fecha de aprobación para maquetación: 26 de septiembre de 2025

Fecha de publicación en OnlineFirst: 16 de octubre de 2025

Fecha de publicación: 1 de enero de 2026

Inteligencia artificial e inteligencia colectiva en la educación superior digital: estudio cuasi-experimental con la plataforma Kampal

Artificial intelligence and collective intelligence in digital higher education: a quasi-experimental study using the Kampal platform



 Alicia Martínez-De la muela - *Universidad Internacional de La Rioja, UNIR (España)*
 Juan Pablo Ruiz-Fuentes - *Universidad Internacional de La Rioja, UNIR (España)*
 José L. González-Geraldo - *Universidad de Castilla-La Mancha, UCLM (España)*

RESUMEN

La irrupción de la Inteligencia Artificial (IA) en educación superior plantea el desafío de integrar el aprendizaje personalizado con aprendizaje colaborativo. La inteligencia colectiva se sitúa en este estudio como una herramienta que permite articular ambos enfoques, al transformar aportaciones individuales en conocimiento compartido mediante tecnologías mediadas por IA. En este marco, se analiza la percepción del uso de la plataforma Kampal como entorno de aprendizaje que favorece simultáneamente la autonomía del estudiante y la construcción colectiva del conocimiento. Se empleó un diseño cuasi-experimental pretest-posttest con una muestra inicial de 399 estudiantes de educación superior a distancia digital y una muestra final emparejada de 25 casos coincidentes. A través de un cuestionario estructurado, se evaluaron tres dimensiones: familiaridad con la herramienta, percepción de eficacia de esta y expectativas de uso o disposición futura a utilizar tecnologías similares. Los resultados muestran un incremento significativo en la familiaridad con Kampal y en la disposición a emplear herramientas de inteligencia colectiva en el futuro, aunque no se hallaron cambios significativos en las expectativas ni en la percepción de eficacia inmediata. El análisis factorial reveló dos dimensiones principales: familiaridad y eficacia/expectativas. Los hallazgos sugieren que estas herramientas fomentan una actitud favorable hacia su uso futuro, aunque sin cambios significativos en la percepción de eficacia. Esto resalta la necesidad de una mediación docente intencionada. Supervisada adecuadamente, la IA potencia entornos híbridos que integran personalización y colaboración, donde la inteligencia colectiva opera como nexo, facilitando dinámicas complejas como la inteligencia de enjambre en educación superior.

Palabras clave: inteligencia artificial; aprendizaje mediado por IA; educación superior; inteligencia colectiva; educación a distancia; Kampal.

ABSTRACT

The emergence of Artificial Intelligence (AI) in higher education poses the challenge of integrating personalized learning with collaborative learning. Collective intelligence is presented in this study as a tool that allows both approaches to be articulated, transforming individual contributions into shared knowledge through AI-mediated technologies. Within this framework, the perception of the Kampal platform as a learning environment that simultaneously promotes student autonomy and collective knowledge construction is analyzed. A pretest-posttest quasi-experimental design was used with an initial sample of 399 digital distance higher education students and a final matched sample of 25 cases. Three dimensions were evaluated through a structured questionnaire: familiarity with the tool, perception of its effectiveness, and expectations of use or future willingness to use similar technologies. The results show a significant increase in familiarity with Kampal and willingness to use collective intelligence tools in the future, although no significant changes were found in expectations or perceptions of immediate effectiveness. Factor analysis revealed two main dimensions: familiarity and effectiveness/expectations. The findings suggest that these tools promote a favorable attitude toward their future use, although without significant changes in perceived effectiveness. This highlights the need for intentional teacher mediation. When properly supervised, AI empowers hybrid environments that integrate personalization and collaboration, where collective intelligence operates as a nexus, facilitating complex dynamics such as swarm intelligence in higher education.

Keywords: artificial intelligence; AI-mediated learning; higher education; collective intelligence; distance education; Kampal.

INTRODUCCIÓN

Hablar de Inteligencia Artificial (IA), en especial dentro del ámbito educativo, se ha convertido peligrosa pero también significativamente en sinónimo de ChatGPT o, en su defecto, de otros grandes modelos de lenguaje (LLM: Large Language Models) como Copilot, Gemini o Claude, por señalar simplemente algunos de los principales LLM, todos ellos privativos, aunque existen alternativas en código abierto. No obstante, como es evidente a poco que nos paremos a observar la evolución del uso e implementación ya no solo de la IA, sino de las Tecnologías de la Información y de la Comunicación (TIC) en educación y en especial en educación a distancia (Bozkurt, 2023), nos daremos cuenta de que, en efecto, la IA –sin tener por ello que resignarnos a la meramente generativa de texto– como también ha ocurrido anteriormente con otros avances tecnológicos –MOOC/NOOC, por ejemplo– nos obliga a replantear los modelos educativos vigentes, como señala García Aretio, tanto de manera urgente como profunda (UNED, 2025).

Estrictamente hablando, el concepto de IA lleva con nosotros casi setenta años, habiendo evolucionado irregularmente –lo que se conoce como los inviernos de la IA– hasta nuestros días (Russell y Norvig, 2020). Sin detallar los distintos subcampos que la conforman (*machine learning*, por ejemplo), así como las distintas IA que existen en función de su especificidad (IA débil-fuerte) entenderemos que bajo el concepto de IA hacemos referencia al uso de algoritmos que permiten realizar y agilizar trabajos inteligentes normalmente asociados a los seres humanos. En este sentido, conviene también señalar que, en educación especialmente, no se trata tanto de ver qué es lo que la IA puede *por* nosotros –sustitución– sino más bien *para* nosotros –provisión– e incluso *con* nosotros –colaboración–.

Asimismo, resulta oportuno considerar la hoja de ruta propuesta por OpenAI hacia la inteligencia artificial general, que establece cinco niveles progresivos (Tabla 1) que orientan el desarrollo actual y futuro de estas tecnologías (Duenas y Ruiz, 2024).

Tabla 1
Marco de cinco niveles de OpenAI para el desarrollo de IA

Nivel	Nombre	Características claves
1	Chatbots	Habilidades en el lenguaje conversacional
2	Razonadoras	Resolución de problemas a nivel humano
3	Agentes	Acción autónoma
4	Innovadores	Apoyo a la invención y la creatividad
5	Organizaciones	Gestión organizacional compleja

Fuente: Duenas y Ruiz, 2024, p. 3

En esta línea se aborda el potencial de la IA en la educación a distancia digital yendo más allá de los LLM o nivel 1, al mismo tiempo que, como veremos, reflexiona sobre algunos de los retos a los que hoy, queramos o no, hemos de enfrentarnos. En este último sentido, es imprescindible tener como pieza clave el Reglamento de la Unión Europea sobre Inteligencia Artificial (Unión Europea, 2024), comúnmente denominado “IA Act”, en el que se establecen las pautas y directrices éticas y responsables necesarias para que estas nuevas plataformas puedan, en efecto, catalizar la mejora educativa y no, simplemente, delegar la toma de responsabilidades que, por supuesto, ha de quedar siempre en manos no solo humanas, sino también expertas, sobre todo en ámbitos sensibles como la justicia, la sanidad y, en el caso que nos ocupa, la educación y en especial la educación superior.

Por estos motivos, plataformas como la que nos ocupa, centradas en el uso de la IA con una finalidad colectiva y pedagógica, nos permitirían actuar como una extensión de la memoria grupal, optimizar la atención colectiva, fortalecer el razonamiento colectivo, realizar una tutoría a escala personalizada y, en consecuencia y de manera paralela, ahorrar tiempo en tareas rutinarias (Riedl y De Cremer, 2025). Estos beneficios, por supuesto, tal como señalan estos mismos y otros autores, no están exentos de riesgos: sesgos, dependencia, brechas digitales, resistencia al cambio, etc. Entre ellos, como veremos con más detalle en la parte final del artículo, la potencial reducción de la diversidad intelectual (Riedl y Bogert, 2024). Por otro lado, el trabajo colaborativo en entornos virtuales favorece el desarrollo de la motivación intrínseca y el fortalecimiento de la autoestima, al contribuir a reducir la sensación de aislamiento y permitir que cada participante perciba su contribución como esencial para el logro colectivo (Marrón Luna, 2021).

El presente artículo se enmarca en esa necesidad crítica, con un enfoque particular en una cuestión clave en el debate pedagógico actual: cómo integrar el aprendizaje personalizado con el aprendizaje colaborativo en entornos mediados por inteligencia artificial, incluyendo un buen equilibrio entre la interacción mediada por la IA y el factor humano en contextos de aprendizaje y la construcción colectiva del conocimiento (Msambwa et al., 2025). Mientras que las aplicaciones actuales tienden a priorizar el aprendizaje individualizado mediante sistemas conversacionales, existe aún un escaso desarrollo de herramientas que promuevan procesos colectivos de construcción del conocimiento. Esta escasez, pese al exponencial avance de todo lo relacionado con la IA, se hace extensivo a las pocas evidencias existentes en la literatura relacionada con plataformas que vayan más allá de los modelos de generación de lenguaje y, en especial, con respecto al concepto de inteligencia colectiva asistida por IA.

La inteligencia colectiva se presenta en este estudio, por tanto, como una herramienta que apoya tanto el aprendizaje personalizado, a través de la revisión de las contribuciones y la monitorización del comportamiento individual, como al aprendizaje colectivo, mostrando un potencial para compartir conocimientos y llegar a resultados conjuntos mediante una plataforma digital. Kampal Collective Learning

(KCL) es una plataforma de inteligencia colectiva desarrollada por Kampal Data Solutions, una empresa derivada de la Universidad de Zaragoza. Su objetivo principal es facilitar la resolución colaborativa de problemas mediante la combinación de aportaciones individuales y colectivas en tiempo real. La plataforma opera a través de fases estructuradas donde los participantes pueden proponer, revisar y modificar ideas, fomentando un aprendizaje dinámico y colaborativo.

Antes de continuar, es imprescindible no solo señalar que Kampal no es un *chatbot*, sino que el uso que esta hace de la IA queda alejado del control y conocimiento de los propios usuarios y, en este caso, también beneficiarios de la misma. Con otras palabras, ni profesores ni estudiantes son conscientes de qué y cómo funcionan los algoritmos que subyacen a Kampal y que, en definitiva, conforman los grupos de trabajo y las redes sobre las que se construye dicha Inteligencia Colectiva. Estamos, por tanto, ante una plataforma que entiende la IA desde el prisma del Big Data (BD) y Machine Learning (ML) y que, como también debatiremos en la parte final de estas líneas, se enfrenta, entre otros asuntos, a la escurridiza trazabilidad de los modelos de IA tipificados de “caja negra”. De esta forma, tanto para complementar y diferenciarnos de la avalancha de estudios sobre la relación entre IA-(LLM) y educación, como tanto para explorar hasta qué punto otras plataformas son aceptadas y cuál puede ser el impacto y potencial uso en nuestros programas educativos, hemos seleccionado como epicentro de este estudio la herramienta Kampal (<https://ic.kampal.com/?locale=es>), cuyo objetivo final consiste, a través del uso de la IA, en posibilitar y potenciar el uso de la Inteligencia Colectiva (IC o, en ocasiones, *Thinkhub*) para la resolución colaborativa de problemas.

Pese a que la literatura con respecto al uso de Kampal no es muy extensa, estamos ante un medio ciertamente adecuado para una educación multitudinaria y cuyos beneficios sobrepasan los estrictamente académicos (Gonzalo et al., 2023) o los meramente informacionales, abordando así dilemas morales y no solo conocimientos (Bautista-Alcaine et al., 2025), algunos de ellos ciertamente relacionados también en su génesis con una implementación deficiente de la IA, como la existencia de las fake news (Cebollero-Salinas et al., 2024). De cualquier forma, como observamos en la metodología de estas mismas investigaciones, adecuados para una educación de calidad, y por tanto superior, que bien puede ser a distancia.

Inteligencia colectiva en el contexto de educación superior a distancia digital

Tal como señala García Aretio (2020), la educación a distancia ha experimentado una transformación semántica y funcional hacia modelos más complejos (presencial, o a distancia o mixta/combinada) pudiendo ser integrados en el concepto Educación a Distancia Digital (EaDD) en el ámbito de educación superior. Este contexto, lejos de limitarse a un modelo de aprendizaje individualista, orienta cada vez más sus esfuerzos hacia el fortalecimiento del aprendizaje personalizado dentro del proceso

educativo. Bayly-Castaneda et al. (2024) señalan, tras una revisión sistemática de la literatura, que el enfoque del uso de la IA se dirige principalmente al ámbito de educación superior y donde predomina el uso de tecnologías de aprendizaje adaptativo orientadas al desarrollo de rutas de aprendizaje personalizadas. Este paradigma busca integrar la personalización del aprendizaje con la construcción de una comunidad de conocimiento, en la que todos sus miembros asumen una responsabilidad compartida en la generación y consolidación del saber. Esta visión se ve reforzada por la idea de utilizar herramientas de visualización para facilitar la interpretación de las contribuciones en entornos de inteligencia colectiva (Ullmann et al., 2019).

Resulta oportuno destacar la contribución de Fidalgo-Blanco et al. (2017), quienes desarrollaron un modelo teórico-práctico de inteligencia colectiva en el aula universitaria presencial, basado en la gestión del conocimiento generado por el alumnado mediante una red social privada. Aunque su experiencia se desarrolla en un entorno presencial, los principios que subyacen y el enfoque de cooperación a partir de la inteligencia individual del colectivo se ve extendido en el uso de plataformas que automatizan y escalan esas interacciones mediante inteligencia artificial. Del análisis de Zawacki-Richter et al. (2019), se advierte que su uso en la educación superior en la mayoría de las aplicaciones desarrolladas hasta la fecha ha relegado a los docentes a un rol periférico. Según los autores, las soluciones basadas en IA se centran en el diseño automático de contenidos, la predicción de resultados o la monitorización del rendimiento, pero descuidan el papel pedagógico del educador como mediador del aprendizaje, diseñador de contextos significativos e impulsor del pensamiento crítico y colectivo.

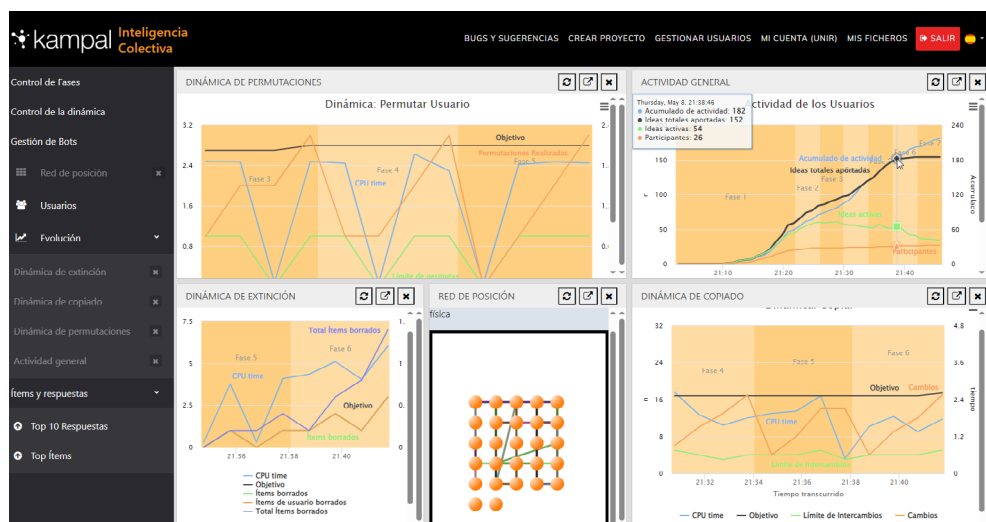
La educación digital debe aspirar a integrar dicho enfoque con la construcción colectiva del conocimiento. En este sentido, herramientas como Kampal representan una propuesta híbrida de notable valor pedagógico ya que permiten combinar la trazabilidad individual del proceso de aprendizaje con la participación activa en experiencias de inteligencia colectiva. Los algoritmos empleados permiten organizar las contribuciones estudiantiles en función de su popularidad relativa, pero también visibilizan los procesos de edición, copia y transformación de ideas, creando así un entorno donde lo individual y lo colectivo se retroalimentan. En términos pedagógicos, la inteligencia colectiva posibilita que el aprendizaje personalizado no se limite a la adaptación de contenidos o ritmos, sino que se amplíe hacia la personalización de las contribuciones cognitivas, las cuales pueden luego ser puestas al servicio del grupo. A su vez, el aprendizaje colaborativo se ve enriquecido con una base más amplia y diversa, donde cada estudiante participa no solo como receptor, sino como productor y evaluador del conocimiento compartido. Esta dinámica se ve fortalecida por la trazabilidad que ofrece la herramienta, permitiendo al docente monitorear la participación, detectar patrones de interacción y tomar decisiones formativas más informadas.

De este modo, la inteligencia colectiva se configura como un punto de encuentro entre la singularidad del pensamiento individual y la potencia transformadora del diálogo grupal, el cual pasa por una serie de fases que permiten editar, eliminar o copiar ideas hasta llegar a una respuesta colectiva. Su implementación, mediada por IA, no solo optimiza los procesos de enseñanza-aprendizaje, también promueve una cultura pedagógica más inclusiva, reflexiva y comprometida con la construcción de conocimiento situado y compartido. Esta solución tecnológica, diseñada específicamente para fomentar procesos de inteligencia colectiva, permite que las aportaciones individuales de los estudiantes, reflejo de sus propios estilos de aprendizaje, conocimientos previos y niveles de desarrollo, sean puestas en común en un entorno digital que favorece el análisis, la síntesis y la evaluación colaborativa de dichas ideas. Mediante algoritmos de análisis basados en IA (BD y ML, principalmente) y visualización de datos, la plataforma genera indicadores como las tres o diez respuestas más populares, facilitando la identificación de los consensos emergentes y las contribuciones más significativas con respecto al agrado de la mayoría.

Kampal es, por tanto, una herramienta *online* que facilita al profesorado, a través de un proceso ordenado por IA, monitorizar el comportamiento individual en el desarrollo de ideas del aprendizaje colectivo de grupos de estudiantes. Según la empresa Kampal Data Solutions “la plataforma genera una dinámica interna usando Inteligencia Artificial, Redes Complejas, etc., para dinamizar la interacción entre los participantes y garantizar la supervivencia de las respuestas más populares de entre todas las emitidas” (p. 5). La Figura 1 muestra un ejemplo de las diferentes dinámicas durante el desarrollo de la actividad, donde se evidencia empíricamente la inteligencia colectiva mediante la convergencia de respuestas individuales hacia una síntesis compartida.

Figura 1

Panel de visualización de métricas dinámicas y estructurales de la plataforma Kampal en tiempo real durante una actividad de inteligencia colectiva



Fuente: Elaboración propia.

Uno de los beneficios que subyacen de esta herramienta es que “todos los individuos parten de la igualdad y cada idea debe valorarse por su contenido y no por quién la propone” (Bautista-Alcaine et al., 2025, p. 2). Por su parte, Orejudo et al. (2022) llevaron a cabo una investigación con 900 estudiantes donde se demuestra el potencial de la inteligencia colectiva para incrementar la colaboración del grupo y cómo a través de las fases se guía a una respuesta de alta calidad.

En conjunto, estas evidencias apuntan a que la inteligencia colectiva, mediada por herramientas tecnológicas como Kampal, representa una evolución cualitativa en los modelos de enseñanza-aprendizaje de la educación superior a distancia digital. No solo responde a las exigencias de personalización cada vez más demandadas en entornos virtuales, sino que permite incorporar, de forma estructurada y medible, dinámicas de colaboración profunda que sitúan al estudiante como co-constructor activo del conocimiento. La inteligencia colectiva emerge como una vía legítima para reequilibrar la dimensión social del aprendizaje en línea, poniendo en valor tanto las aportaciones individuales como las sinergias emergentes del grupo. En este sentido, su implementación constituye no solo una innovación metodológica, sino también una respuesta a los desafíos pedagógicos, tecnológicos y éticos que plantea la educación superior digital del siglo XXI.

Este estudio analiza el impacto del uso de Kampal en la percepción del aprendizaje colaborativo en educación superior digital, evaluando variaciones en familiaridad, eficacia percibida y expectativas sobre tecnologías mediadas por IA.

METODOLOGÍA

De acuerdo con la naturaleza del problema, se plantea como adecuado un diseño cuasi-experimental de corte cuantitativo (Hernández Sampieri et al., 2010) ofreciendo mayores posibilidades de generalización a contextos reales y cotidianos que diseños experimentales, siendo al mismo tiempo la alternativa más viable y ética, que puede proporcionar información relevante sobre el impacto de un tratamiento o la evolución de un proceso de cambio. A tal efecto, se realiza una medición pretest postest, en un solo grupo, tras el uso de la herramienta de inteligencia colectiva Kampal como intervención. Esta estructura metodológica permite observar los efectos en una muestra no aleatorizada, realizando una comparación entre los resultados obtenidos antes y después del tratamiento.

Los alumnos de la asignatura de “Tecnologías para la innovación educativa” del Máster oficial en innovación educativa de la Universidad Internacional de La Rioja (UNIR), participaron en la resolución de una actividad colaborativa, dentro de un foro académico, en la que se utiliza la herramienta de inteligencia colectiva Kampal, para analizar y sintetizar las respuestas del alumnado en torno a un caso práctico relacionado con el uso ético de la inteligencia artificial (IA) en el aula. La primera intervención es la respuesta a tres preguntas claves, las cuales fueron trabajadas de manera síncrona y colectiva a través de Kampal. La actividad de foro colaborativo se desarrolla a lo largo de tres semanas, permitiendo una participación escalonada y reflexiva. El uso de Kampal tuvo una temporalidad de una hora donde los estudiantes pasaron por siete fases generadas por dinámicas automáticas para fomentar la colaboración, pudiendo analizar la información de sus “vecinos”, modificar sus propias respuestas y/o copiar la de uno de sus compañeros. El objetivo de Kampal es generar una respuesta colectiva con base en las aportaciones del alumnado en la primera intervención, que sirvió como base para posteriores reflexiones y mejoras de propuestas. En concreto, la tercera intervención tuvo una relación directa sobre los contenidos que surgieron de la inteligencia colectiva en la sesión presencial virtual. Dado el carácter formativo de la actividad, integrado en el desarrollo curricular de la asignatura y vinculado a una evaluación formativa, no fue posible establecer un grupo de control, lo cual constituye una limitación metodológica a considerar.

En este estudio se consideraron tres dimensiones conceptuales esenciales: familiaridad, eficacia y expectativas de uso. La familiaridad alude al reconocimiento intuitivo y facilidad de uso percibida, determinante en la aceptación inicial de plataformas de IA en entornos colaborativos universitarios (Chan y Hu, 2023). La eficacia se refiere a la calidad y relevancia pedagógica de las contribuciones colaborativas facilitadas por IA (Wang et al., 2025). Las expectativas de uso describen

la disposición futura a emplear tecnologías de IA en actividades académicas, alineadas con modelos de adopción tecnológica que vinculan experiencias positivas con mayor intención de uso (Khlaif et al., 2024).

A partir del enfoque cuasi-experimental pretest-posttest con un solo grupo, se formularon tres hipótesis de investigación que orientaron el análisis de los efectos del uso de Kampal cabiendo esperar:

- Un incremento significativo de la familiaridad de los estudiantes con las tecnologías de inteligencia colectiva y con la propia plataforma Kampal en comparación con su nivel previo a la intervención.
- Una mejora significativa de la percepción de eficacia del trabajo colaborativo y la capacidad de relacionar y mejorar las contribuciones individuales en comparación con las percepciones previas tras el uso de Kampal.
- Un aumento de la disposición futura de los estudiantes a emplear herramientas basadas en inteligencia colectiva en entornos académicos tras la participación en actividades colaborativas mediadas por Kampal.

Estas hipótesis fueron contrastadas mediante análisis estadísticos no paramétricos, adecuados para el tipo de datos y la estructura de la muestra, con el fin de valorar el impacto percibido de la herramienta en un entorno de educación superior a distancia digital.

Contexto

La investigación se desarrolla en la Universidad Internacional de La Rioja (UNIR), que imparte enseñanza superior 100 % *online* con carácter oficial. Reconocida legalmente a través de la Ley 3/2008, de 13 de octubre, de reconocimiento de la Universidad Internacional de La Rioja, el Parlamento de la comunidad autónoma, aprueba su personalidad jurídica propia y forma de sociedad anónima, caracterizada por la enseñanza a distancia y cuyo funcionamiento se asienta en las TIC, sin campus físico, con clases síncronas y asíncronas (UNIR, 2024), ofreciendo enseñanza universitaria, y realizando el servicio público de la educación superior mediante la investigación, la docencia y el estudio.

Población

La muestra para este estudio ($n = 399$) está compuesta por estudiantes del Máster Universitario en Innovación Educativa de UNIR matriculados en la asignatura “Tecnología para la Innovación Educativa”. La invitación garantizaba la confidencialidad de los datos, dejando un total de $n = 152$ sujetos en el momento pretest, y $n = 77$ sujetos en el posttest. Tras depurar los datos finalmente la muestra

en pretest fue $n = 147$ y en postest $n = 62$. A la hora de analizar la percepción en el aprendizaje de los estudiantes universitarios sobre el uso de la aplicación de inteligencia colectiva Kampal entre el pretest y el postest, se identificó una muestra coincidente suficiente para con respecto a los objetivos propuestos ($n = 25$). La pérdida de continuidad en los identificadores anónimos asignados a cada estudiante, junto con el carácter voluntario y la participación en la sesión en directo de la herramienta, han podido influir en esta muestra final.

La mayoría de los participantes en todas las mediciones se identifican como mujeres, manteniéndose esta tendencia tanto en el pretest (65,31 %) como en el postest (66,13 %) y en los casos coincidentes (60 %). La representación masculina es menor en todos los momentos, y las categorías no binarias o no declaradas tienen una presencia muy reducida. En cuanto a la distribución por país (Tabla 2), la mayoría de la muestra deriva de Colombia, tanto en el pretest (76,87 %) como en el postest (70,97 %) y en los casos coincidentes (64 %). Aunque otros países como España, Ecuador y Argentina tienen una presencia menor, se observa un ligero incremento relativo en su representación en el postest y en los casos coincidentes. La participación de otros países es marginal, y algunos como Chile o Estados Unidos solo aparecen en una de las fases o desaparecen en los datos coincidentes.

Tabla 2

Distribución de la muestra según país

País	Pretest		Postest		Pre-Post Coincidentes	
	n	%	n	%	n	%
Colombia	113	76.87	44	70.97	16	64.00
España	9	6,12	7	11.29	4	16.00
Ecuador	5	3.40	5	8.06	2	8.00
Perú	5	3.40	3	4.84	0	0.00
Chile	5	3.40	0	0.00	0	0.00
Argentina	3	2.04	2	3.23	2	8.00
Resto Latinoamérica	2	1.36	1	1.61	1	4.00
Otros	4	2.72	0	0.00	0	0.00
Estados Unidos	1	0.68	0	0.00	0	0.00
TOTAL	147	100	62	100	25	100

Fuente: Elaboración propia.

Respecto a la experiencia docente (Tabla 3), predomina el perfil con una trayectoria superior a 10 años, tanto en el pretest (42,18 %) como en el postest (37,10 %) y en los casos coincidentes (40 %). Sin embargo, se observa una mayor

representación relativa de docentes con menos de 3 años de experiencia en el postest (27.42 %) y en los casos coincidentes (24 %).

Tabla 3

Distribución de la muestra según nivel de experiencia docente

Categoría	Pretest		Postest		Pre-Post Coincidentes	
	n	%	n	%	n	%
Más de 10 años	62	42.18	23	37,10 %	10	40,00 %
De 6 a 10 años	34	23.13	10	16,13 %	6	24,00 %
Menos de 3 años	29	19.73	17	27,42 %	6	24,00 %
Entre 3 y 5 años	16	10.88	9	14,52 %	2	8,00 %
Ninguna experiencia	6	4.08	3	4,84 %	1	4,00 %
TOTAL	147	100.00	62	100,00 %	25	100,00 %

Fuente: Elaboración propia.

Instrumento

Se administró un instrumento de evaluación “Cuestionario para analizar la Percepción en el aprendizaje de los estudiantes universitarios sobre el uso de la aplicación de Inteligencia Colectiva Kampal” (CPIC), previamente validado por un comité de tres expertos con distintos perfiles académicos, todos ellos especialistas en el ámbito pedagógico, antes del inicio de la intervención y migrado a la aplicación *Microsoft Forms* para facilitar la distribución y recogida de datos. Los resultados del índice V de Aiken fueron superiores a .7 en todos los ítems salvo tres, no siendo en ningún caso inferior a .6, obteniendo así una adecuada validez de contenido (Penfield y Giacobbi, 2004).

El cuestionario presenta como primer ítem un consentimiento informado en cumplimiento de las normativas para la aprobación de investigaciones por parte del Comité de Ética de la Investigación (Ref. 037/2025) de la Universidad Internacional de La Rioja (UNIR), siguiendo todos los protocolos en términos de confidencialidad y anonimato (UNIR, 2025).

El instrumento consta de cuatro bloques:

- **Bloque I: Información biográfica.** Compuesta por los ítems del 2 al 5, se incluyeron preguntas para obtener información sobre ID personal, sexo (masculino, femenino, ninguno de los anteriores, prefiero no contestar), país (España, Colombia, Ecuador, Perú, México, Argentina, Chile, Estados Unidos,

Resto Latinoamérica, otros), y nivel de experiencia docente en años (ninguna experiencia, menos de 3 años, entre 3 y 5 años, de 6 a 10 años, más de 10 años).

- Bloque II: Familiaridad. Incluye los ítems del 6 al 8, con la función de extraer la familiaridad de los estudiantes con aplicaciones de inteligencia colectiva, concretando en Kampal.
- Bloque III: Eficacia. Integra los ítems del 9 al 12, con el objetivo de obtener la percepción de eficacia de la inteligencia colectiva en los estudiantes.
- Bloque IV: Expectativas de uso. Registra los ítems de 12 al 14, para obtener las expectativas de usabilidad en términos de aprendizaje, colaboración y disposición de aplicaciones de inteligencia colectiva.

En la Tabla 4 se registran las preguntas de cada bloque según ítem:

Tabla 4
Bloques, ítems y preguntas CPIC

Ítems y preguntas	
Bloque I: Información Biográfica	1. ID Personal. 2. Sexo. 3. País. 4. Indique nivel de experiencia en años.
Bloque II: Familiaridad	5. ¿Crees que el nivel de uso con aplicación de Inteligencia Artificial ha aumentado tras la experiencia? 6. ¿Crees que el nivel de uso con aplicaciones de inteligencia colectiva ha aumentado tras la experiencia? 7. ¿Crees que el nivel de uso con la aplicación de Kampal ha aumentado tras la experiencia?
Bloque III: Eficacia	8. ¿Hasta qué punto crees que el uso de una aplicación de Inteligencia Colectiva podría mejorar la calidad del trabajo colaborativo? 9. ¿El uso de la aplicación de Inteligencia Colectiva te ha ayudado a cambiar, mejorar o descartar tus respuestas iniciales a las preguntas planteadas en la actividad? 10. ¿Crees que las respuestas surgidas de la Inteligencia Colectiva te han ayudado a relacionar contribuciones depositadas en el foro?
Bloque IV: Expectativas de uso	11. Creo que esta aplicación de Inteligencia Colectiva me va a servir en mi proceso de aprendizaje. 12. Creo que esta aplicación de Inteligencia Colectiva me va a servir en términos de colaboración. 13. ¿Tendrías disposición a utilizar aplicaciones de inteligencia colectiva en futuras actividades académicas?

Fuente: Elaboración propia.

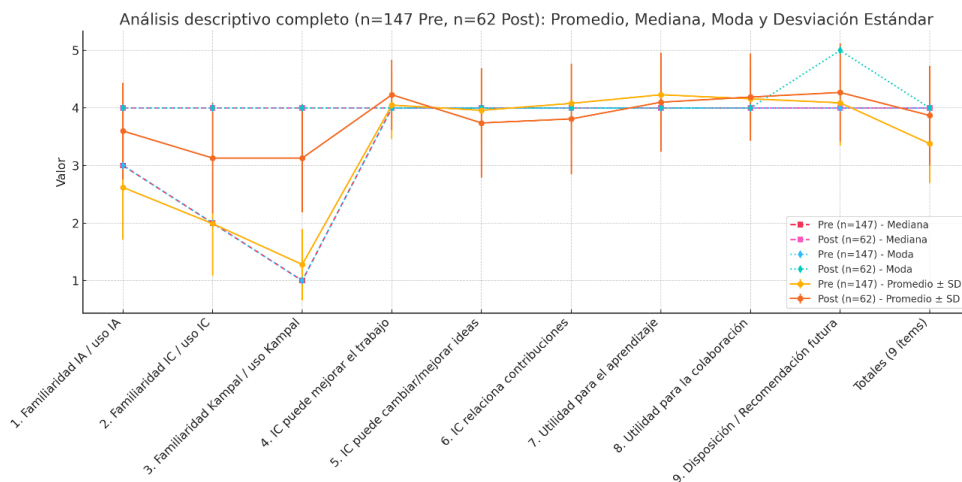
Una vez que los encuestados procedieron a completar el cuestionario, los datos fueron importados al paquete estadístico para las ciencias sociales *IBM SPSS Statistics 25* con el fin de obtener las oportunas propiedades psicométricas, así como para realizar los pertinentes análisis estadísticos.

ANÁLISIS Y RESULTADOS

En el marco de la incorporación de tecnologías colaborativas en el ámbito educativo, se realizó un análisis estadístico descriptivo, así como otro análisis de las propiedades psicométricas y la utilidad percibida de una herramienta de inteligencia colectiva. Desde un enfoque cuantitativo, se aplicaron pruebas de consistencia interna, análisis factorial exploratorio, y técnicas no paramétricas, con el propósito de valorar la adecuación del instrumento para captar percepciones significativas. Los resultados obtenidos ofrecen evidencias relevantes en torno a la validez, fiabilidad y utilidad percibida de una herramienta de inteligencia colectiva aplicada al contexto educativo.

Análisis estadístico descriptivo

Del pre-test al post-test se observa un avance claro en la adopción de las tecnologías (Figura 2): la familiaridad media con la inteligencia artificial registra un incremento de 2.62 a 3.74, con la inteligencia colectiva de 1.99 a 3.60 y con Kampal de 1.28 a 3.13. La creencia de que la IC mejora el trabajo colaborativo también aumenta de 4.05 a 4.23, mientras que las expectativas sobre su impacto en las ideas y en la relación de contribuciones descienden levemente. Las valoraciones de utilidad para el aprendizaje y la colaboración se mantienen en torno a 4.10 y 4.20 y la disposición a recomendar la herramienta crece de 4.09 a 4.27, desplazándose la moda del 4 al 5. En conjunto, la media global de los nueve ítems pasa de 3.38 a 3.87, con desviaciones típicas siempre inferiores a 1.

Figura 2**Resultados análisis estadístico descriptivo**

Fuente: Elaboración propia.

Con todo ello, se registra una mejora general en la familiaridad con la Inteligencia Artificial, la Inteligencia Colectiva y la herramienta Kampal. También aumentan ligeramente las valoraciones sobre su utilidad y la disposición a recomendar su uso, mientras que se mantienen estables las percepciones sobre su contribución al aprendizaje y la colaboración. Las respuestas muestran consistencia, con una baja variabilidad en las puntuaciones.

Análisis estadístico inferencial

De acuerdo con Hair et al. (2019), el instrumento demostró una alta consistencia interna ($\alpha = .856$), lo que respalda su fiabilidad e idoneidad para captar las percepciones de los participantes sobre la herramienta analizada.

En relación con el análisis factorial exploratorio, se utilizó una rotación ortogonal Varimax con los ítems del 5 al 13. La adecuación muestral fue moderada ($KMO = .764$), pero aceptable para llevar a cabo el procedimiento (Field, 2009), respaldado por una prueba de esfericidad de Bartlett significativa ($p < .000$). El análisis identificó dos factores, que consiguen explicar el 59.82 % de la variabilidad del constructo. El primer factor explica el 39.72 % y el segundo el 20.10 %. Por otra parte, para la posterior interpretación de los factores, partimos de la matriz inicial de componente rotado (Tabla 5) que determinó diferentes saturaciones factoriales para la selección de los ítems incluidos en cada uno de los 2 factores, que expresan la magnitud de la correlación entre el ítem y los factores, ordenados por tamaño y

suprimiendo los coeficientes pequeños con un valor absoluto por debajo de .400 basado en la sugerencia de Stevens (2002).

Tabla 5

Matriz de componente rotado

	Componente
Familiaridad IA Inteligencia Artificial	.738
Familiaridad Inteligencia Colectiva	.862
Familiaridad Inteligencia Kampal	.663
¿Hasta qué punto crees que el uso de una aplicación de Inteligencia Colectiva puede mejorar la calidad del trabajo colaborativo?	.828
¿Hasta qué punto crees que la Inteligencia Colectiva puede ayudar a cambiar, mejorar o descartar tu idea inicial a determinados contenidos?	.754
¿Hasta qué punto crees que las respuestas que surjan de una inteligencia colectiva te ayudarán a relacionar contribuciones y mejorarlas?	.762
Creo que esta aplicación de Inteligencia Colectiva me va a servir en mi proceso de aprendizaje.	.795
Creo que esta aplicación de Inteligencia Colectiva me va a servir en términos de colaboración.	.815
¿Tendrías disposición a utilizar aplicaciones de inteligencia colectiva en futuras actividades académicas?	.621

Fuente: Elaboración propia.

La denominación de los factores encontrados se ha determinado a partir de los elementos que lo constituyen y el contenido al que se refiere cada uno de ellos, siendo finalmente nombrados como: factor 1, familiaridad; factor 2, eficacia/expectativas. Pese a que los resultados factoriales son óptimos, cabe señalar que es posible que exista cierta solapación conceptual entre ambos factores al entender que el aumento en la familiaridad pudiera influir en la percepción de la eficacia y las expectativas de uso.

La evaluación de la normalidad de los datos, mediante las pruebas de Kolmogórov-Smirnov y Shapiro-Wilk, evidenció desviaciones significativas respecto a la distribución normal en todas las variables ($p < .05$). Esta condición justificó el uso de pruebas no paramétricas, más apropiadas y robustas en presencia de distribuciones asimétricas o valores atípicos.

A continuación, se realizaron comparaciones por sexo mediante la prueba U de Mann-Whitney (Tabla 6). Los resultados no mostraron diferencias estadísticamente significativas entre hombres y mujeres en ninguna de las dimensiones evaluadas.

Tabla 6
Resultados prueba Mann-Whitney según “sexo”

	pre_ familiaridad_ factor1	post_ familiaridad_ factor1	pre_ eficacia/ expectativas factor2	post_ eficacia/ expectativas factor2
U de Mann-Whitney	53.000	67.000	51.000	52.500
W de Wilcoxon	173.000	187.000	171.000	172.500
Z	-1.253	-.459	-1.354	-1.262
Sig. asint. (bilateral)	.210	.646	.176	.207
Significación exacta [2 sig. unilateral]	.238	.683	.196	.216

Fuente: Elaboración propia.

Posteriormente se exploraron las diferencias en función de la experiencia docente mediante la prueba de Kruskal-Wallis. Los resultados (Tabla 7) no arrojaron diferencias estadísticamente significativas, lo que indica una valoración positiva transversal de la herramienta, independientemente del nivel de experiencia profesional.

Tabla 7
Resultados prueba Kruskal-Wallis según la experiencia docente

	pre_ familiaridad_ factor1	post_ familiaridad_ factor1	pre_ eficacia/ expectativas factor2	post_ eficacia/ expectativas factor2
H de Kruskal-Wallis	1.281	3.655	1.799	.524
gl	4	4	4	4
significación asintótica	.865	.455	.773	.971

Fuente: Elaboración propia.

Finalmente, se llevó a cabo un análisis pre-post mediante la prueba de los rangos con signo de Wilcoxon. Los resultados revelaron mejoras estadísticamente significativas en la familiaridad con la herramienta ($z = -4.32$, $p < .001$, $r = .68$) y en la disposición futura a utilizar tecnologías basadas en inteligencia colectiva ($z = -4.26$, $p < .001$, $r = .67$), ambos con tamaños del efecto grandes. En cambio, no se observaron cambios significativos en las expectativas sobre los resultados ($z = -0.29$, $p = .770$, $r = .05$) ni en la percepción de autoeficacia ($z = -0.75$, $p = .453$, $r = .12$).

DISCUSIÓN Y CONCLUSIONES

En la actualidad, la inteligencia artificial (IA) en el ámbito educativo se ha reducido predominantemente a aplicaciones basadas en modelos de lenguaje de gran escala (LLM), como ChatGPT, que operan según los niveles delimitados por OpenAI, en un nivel 1 conversacional. Este enfoque limitado ha generado preocupaciones respecto a su impacto en los procesos de evaluación y aprendizaje en la educación superior (Consuegra-Fernández et al., 2024). En este estudio se ha trabajado desde un segundo nivel superior de procesamiento gracias al enfoque de inteligencia colectiva que se propone con Kampal y que trascienden de la simple interacción conversacional. Por otro lado, se trabaja desde un nivel 3 (agentes) ya que la IA toma decisiones estructurales en nombre del usuario (formación de grupos, selección de ideas, etc.), aunque sin plena autonomía. Sin embargo, la convergencia más significativa se situaría en los niveles 4 (innovadores) y nivel 5 (organizaciones) ya que Kampal se presenta como una plataforma que permite soluciones colaborativas emergentes, dinámicas complejas y autoorganizadas, estructuras que aprenden y evolucionan.

Aunque el estudio partió de una población definida (399 estudiantes), la reducción progresiva hasta una muestra coincidente de 25 sujetos en la medición pretest-postest introduce ciertas limitaciones respecto a la validez externa. No obstante, como señalan Creswell y Creswell (2017) este tipo de estudios exploratorios con muestras no probabilísticas aportan información relevante en contextos reales, especialmente cuando se busca evaluar la aceptación y percepción de nuevas tecnologías educativas. La muestra final, con predominancia femenina, alta representación geográfica de Colombia y elevada experiencia docente, ofrece una perspectiva relevante dentro del marco de la educación superior digital, aunque sus resultados no sean directamente generalizables a otras poblaciones. Asimismo, el diseño pretest–postest sin grupo control, propio de enfoques cuasi-experimentales, limita la posibilidad de establecer inferencias causales definitivas, al no poder aislar completamente los efectos de variables externas. Sin embargo, tal como señalan Shadish et al. (2002), este tipo de diseños sigue siendo adecuado en etapas iniciales de investigación aplicada, donde el interés reside más en explorar patrones de cambio que en comprobar hipótesis causales con máxima controlabilidad. Desde el punto de vista analítico, la elección de pruebas no paramétricas —Kolmogórov-Smirnov y Shapiro-Wilk para normalidad, Mann-Whitney U, Kruskal-Wallis y Wilcoxon— se justificó por el tamaño reducido de la muestra, la asimetría de las distribuciones y la naturaleza ordinal de los datos tipo Likert (Sullivan y Artino, 2013).

Los resultados obtenidos en este estudio evidencian un impacto positivo en la familiaridad del alumnado con la herramienta Kampal, así como una elevada disposición futura para utilizar tecnologías de inteligencia colectiva en contextos educativos, confirmando las dos primeras hipótesis, lo que hace plantear que el contacto con entornos digitales bien diseñados puede transformar las actitudes del

alumnado hacia formas más colaborativas de aprendizaje (Hogan et al., 2023). Esta disposición positiva es especialmente relevante en entornos de educación superior a distancia digital, donde las oportunidades para la interacción entre pares pueden verse limitadas sin un diseño intencional de dinámicas colectivas. Sin embargo, respecto a la tercera hipótesis no se evidenciaron cambios significativos en las expectativas de resultados ni en la percepción de la eficacia inmediata, lo que sugiere una revisión de su potencial pedagógico y la necesidad de valorar las respuestas surgidas colectivamente como las más populares no siendo necesariamente las más idóneas, complejas o deseadas. Esta ausencia de cambios puede deberse a diversas circunstancias, entre las que podríamos destacar el breve tiempo de exposición a la herramienta o de identificación de los potenciales beneficios, todavía latentes, por descubrir. Desde un punto de vista pedagógico, estas reflexiones nos animan a establecer un mayor seguimiento y acompañamiento no solo durante su uso, sino también al final del proceso, incorporando estrategias como evaluaciones longitudinales, diseños controlados, indicadores objetivos de rendimiento y la réplica del estudio en distintas instituciones y contextos formativos.

Estos hallazgos sugieren que el uso de la herramienta Kampal puede facilitar procesos de aproximación, exploración y apertura hacia su implementación, aunque no necesariamente transforma de forma inmediata las creencias sobre sus beneficios académicos, evidencian su capacidad para favorecer la participación activa, la construcción colectiva del conocimiento y una disposición positiva hacia el uso de tecnologías colaborativas mediadas por IA en contextos educativos similares. Asimismo, el uso de la herramienta presenta ciertas limitaciones ya que requiere de la participación síncrona del estudiantado, lo cual puede restringir su integración en dinámicas más flexibles propias de la educación a distancia. No obstante, los productos generados colectivamente podrían servir como base para la reflexión asíncrona en otros espacios de interacción, como los foros colaborativos. Esto sugiere una doble lectura: por un lado, la intervención directa permite que cada estudiante perciba su contribución como parte constitutiva de la respuesta colectiva; por otro, plantea la necesidad de cuestionar si las soluciones emergentes del proceso de inteligencia colectiva representan efectivamente las respuestas más pertinentes o simplemente las más populares.

En este sentido, desde un punto de vista crítico, es necesario contrastar estos resultados con aquellos que sugieren que el uso de estas y otras herramientas de IA conllevan una reducción de la diversidad intelectual (Riedl y Bogert, 2024), provocando un efecto negativo de homogeneización excesiva al pretender agilizar procesos de toma de decisión donde lo más frecuente no es, quizá, lo que realmente andamos buscando desde el prisma pedagógico. Con otras palabras, Kosmyrna et al. (2025) acuñan el término “deuda cognitiva” al haber observado una disminución significativa de conectividad neuronal entre los que usaron LLM y aquellos que realizaron tareas sin asistencia de IA, corroborando así el peligro de un uso automatizado incorrecto.

En el caso de Kampal, las respuestas que emergen como “colectivas” son aquellas que han alcanzado mayor aceptación o popularidad en el grupo, pero esto no garantiza que sean las más complejas, pertinentes o deseables desde un punto de vista formativo. En relación con la afirmación de Mulgan (2018) “los colectivos humanos y la inteligencia humana deben seguir estando en el centro” (p. 235), a pesar del papel facilitador de la inteligencia artificial, el estudio valora que son las personas las que activan el papel pedagógico de estas herramientas, resaltando así la necesidad de supervisión, no solo humana sino también experta, al trabajar con cualquier tipo de plataforma basada en algoritmos de IA. Se recomienda que los docentes actúen como mediadores críticos, orientando la reflexión y participación en las dinámicas de inteligencia colectiva. En este sentido, y adicionalmente, aunque es cierto que Kampal tiene previstas las suficientes herramientas que permiten la trazabilidad de las interacciones propuestas, no debemos perder de vista que los modelos de IA, sobre todo, los basados en redes neuronales, suelen caer en lo que se denomina como modelos de “caja negra”, siendo imprescindible no solo supervisar el resultado alcanzado sino también, y quizá sobre todo al tratarse de una herramienta pedagógica, el camino seguido hasta el mismo. Uno de los debates centrales que se desprenden de este estudio es el relativo a la opacidad de los sistemas de IA, comúnmente referida como el problema de la “caja negra”.

Como línea futura de investigación se propone la constitución de comités de evaluación que analicen las respuestas seleccionadas algorítmicamente, contrastándolas con criterios pedagógicos, éticos y epistemológicos explícitos. Una validación experta de los productos emergentes de la inteligencia colectiva permitiría comprobar si lo más aceptado coincide con lo más ajustado a los fines formativos definidos, abriendo así un debate necesario sobre qué entendemos por “mejor respuesta” en un contexto de aprendizaje colaborativo. Este enfoque contribuiría a equilibrar la eficiencia técnica con la exigencia educativa, dotando de mayor legitimidad al proceso colectivo automatizado.

Poner el foco en el humano experto plantea una formación hacia las competencias de los facilitadores de la inteligencia colectiva, como plantean Broome y Hogan (2020) a través de un recorrido progresivo por tres niveles: competencias fundacionales, performativas y de maestría. Este enfoque no solo requiere habilidades técnicas, sino también competencias pedagógicas que permitan guiar procesos colaborativos mediante herramientas como Kampal. El facilitador debe ser capaz de orientar al grupo en la generación de nuevas perspectivas, gestionar intervenciones mediadas por tecnologías como los “bots” con IA, y evaluar dinámicas grupales a través del control de la dinámica o la red de posición. En particular, el uso de “bots” con IA actúa como “agentes mediadores” lo que ayudaría al vínculo entre aprendizaje personalizado y aprendizaje colectivo, aunque debemos ser conscientes que este tipo de elementos incrementarían también los riesgos de “caja negra”.

De esta forma, bajo las cautelas previstas y en función de los resultados alcanzados junto con las previsiones de mejora de los instrumentos utilizados, así

como en los de la literatura citada, observamos cómo el caso de Kampal dentro de la educación superior digital puede facilitar la convergencia entre lo individual y lo colectivo, ofreciendo un entorno donde cada estudiante aporta contenidos, desde sus intereses y estilos cognitivos, al tiempo que se incorporan a una dinámica colectiva dirigida a proporcionar un conocimiento compartido. Una línea de trabajo es explorar futuros desarrollos donde la IA no solo analice respuestas, sino que aprenda a facilitar dinámicas similares a las Inteligencia de Enjambre Artificial (ASI, por sus siglas en inglés) “porque permiten que un grupo muy grande de individuos se reúna durante un tiempo muy corto y tome decisiones efectivas y relativamente precisas” (Baltzersen, 2022, p.104). Dentro del contexto educativo, esto significaría el uso responsable y eficiente de una IA con capacidad para orquestar, en tiempo real, procesos de decisión colectiva entre grandes grupos de estudiantes, promoviendo la convergencia rápida de ideas, el consenso y la resolución colaborativa de problemas complejos. Su aplicación responsable en contextos educativos como la educación superior a distancia digital podría permitir generar espacios de deliberación dinámica, resolución de problemas complejos y toma de decisiones compartidas que respondan tanto a criterios de inclusión como de calidad formativa. Siempre que se diseñen experiencias de aprendizaje estructuradas, con mediación docente intencionada, tareas auténticas y mecanismos de trazabilidad que aseguren la transparencia y el valor pedagógico del proceso colectivo.

Declaración de conflictos de interés y transparencia: Los autores declaran que no existen vínculos económicos o institucionales con la empresa propietaria de Kampal.

REFERENCIAS

- Baltzersen, R. K. (2022). *Cultural-historical perspectives on collective intelligence: Patterns in problem solving and innovation*. Cambridge University Press. <https://doi.org/10.1017/9781108981361>
- Bayly-Castaneda, K., Ramirez-Montoya, M.-S. y Morita-Alexander, A. (2024). Crafting personalized learning paths with AI for lifelong learning: A systematic literature review. *Frontiers in Education*, 9, 1-12. <https://doi.org/10.3389/feduc.2024.1424386>
- Bautista-Alcaine, P., Íñiguez-Berrozpe, T. y Fernández-Albás, J. (2025). Collective intelligence online experiments: A learning method and its pedagogical applications through online collaborative work platforms. En V. Goglio y C. Biancalana (Eds.), *Mapping the evolution of platform society* (pp. 13-26). Emerald Publishing Limited. <https://doi.org/10.1108/978-1-83608-028-220251002>
- Bozkurt, A. (2023). Generative artificial intelligence (AI) powered conversational educational agents: The inevitable paradigm shift. *Asian Journal of Distance Education*, 18(1), 198-204. <https://doi.org/10.5281/zenodo.7716416>
- Broome, B. y Hogan, M. (2020). Core competencies of collective intelligence facilitators. *Systems Research and Behavioral Science*, 38, 940-944. <https://doi.org/10.1002/sres.2734>
- Cebollero-Salinas, A., Elboj-Saso, C., Íñiguez-Berrozpe, T. y Bautista-Alcaine, P. (2024). Confronting fake news through

- collective intelligence in adolescents. *International Journal of Human-Computer Interaction*. Advance online publication. <https://doi.org/10.1080/10447318.2024.2405567>
- Chan, C. K. Y. y Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*, 20, 43. <https://doi.org/10.1186/s41239-023-00411-8>
- Consuegra-Fernández, M., Sanz-Aznar, J., Burguera-Serra, J. G. y Caballero-Molina, J. J. (2024). ChatGPT: El dilema sobre la autoría de las actividades evaluables en educación universitaria. *Revista de Investigación Educativa*, 42(2). <https://doi.org/10.6018/rie.565391>
- Creswell, J. W. y Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches* (5ª ed.). SAGE Publications.
- Duenas, T. y Ruiz, D. (2024). *The path to superintelligence: A critical analysis of OpenAI's five levels of AI progression [Preprint]*. ResearchGate. <https://doi.org/10.13140/RG.2.2.33794.70085>
- Fidalgo-Blanco, Á., Sein-Echaluce, M. L. y García-Peñalvo, F. J. (2017, 4-6 octubre). Inteligencia colectiva en el aula. Un paradigma cooperativo [Presentación en conferencia]. *IV Congreso Internacional sobre Aprendizaje, Innovación y Competitividad*, Zaragoza, España. https://doi.org/10.26754/CINAIC.2017.000001_125
- Field, A. (2009). *Discovering statistics using SPSS* (3ª ed.). SAGE Publications.
- García Aretio, L. (2020). Bosque semántico: ¿educación/enseñanza/aprendizaje a distancia, virtual, en línea, digital, eLearning...? *RIED-Revista Iberoamericana de Educación a Distancia*, 23(1), 9-28. <https://doi.org/10.5944/ried.23.1.25495>
- Gonzalo, A., Sanz-García, F., Pelacho, M., Tarancón, A., Rivero, A., Varela, O. y Moreno, A. (2023). Collective intelligence to find solutions to the challenges posed by the Sustainable Development Goals. *Citizen Science: Theory and Practice*, 8(1), 47. <https://doi.org/10.5334/cstp.587>
- Hair, J. F., Black, W. C., Babin, B. J. y Anderson, R. E. (2019). *Multivariate data analysis* (8ª ed.). Cengage.
- Hernández Sampieri, R., Fernández Collado, C. y Baptista Lucio, M. P. (2010). *Metodología de la investigación* (5ª ed.). McGraw-Hill.
- Hogan, M. J., Barton, A., Twiner, A., James, C., Ahmed, F., Casebourne, I., Hamilton, P., Shi, S., Zhao, Y., Harney, O. y Wegerif, R. (2023). Education for collective intelligence. *Irish Educational Studies*, 44(1), 137-166. <https://doi.org/10.1080/03323315.2023.2250309>
- Khlaif, Z. N., Ayyoub, A., Hamamra, B., Bensalem, E., Mitwally, M. A. A., Ayyoub, A., Hattab, M. K. y Shadid, F. (2024). University teachers' views on the adoption and integration of generative AI tools for student assessment in higher education. *Education Sciences*, 14(10), 1090. <https://doi.org/10.3390/educsci14101090>
- Kosmyna, N., Hauptmann, E., Yuan, Y. T., Situ, J., Liao, X.-H., Beresnitzky, A. V., Braunstein, I. y Maes, P. (2025, 10 de junio). *Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing task [Preprint]*. arXiv. <https://doi.org/10.48550/arXiv.2506.08872>
- Ley Orgánica 3/2008, de 13 de octubre, de reconocimiento de la Universidad Internacional de La Rioja. *Boletín Oficial del Estado*, núm. 261, 42805-42806. <https://www.boe.es/eli/es-ri/1/2008/10/13/3>
- Marrón Luna, Y. A. (2021). Trabajo colaborativo: una opción en la

- educación en línea. *Revista SUAyED Psicología*, (2), 26–31. https://www.psicologia.unam.mx/documentos/pdf/publicaciones/revista_suayed/suayed_psicologia_2021_2.pdf
- Msambwa, M. M., Wen, Z. y Daniel, K. (2025). The impact of AI on the personal and collaborative learning environments in higher education. *European Journal of Education*, 60(1). <https://doi.org/10.1111/ejed.12909>
- Mulgan, G. (2018). *Big mind: How collective intelligence can change our world*. Princeton University Press. <https://doi.org/10.1515/9781400888511>
- Orejudo, S., Cano-Escoriaza, J., Cebollero-Salinas, A. B., Bautista, P., Clemente-Gallardo, J., Rivero, A., Rivero, P. y Tarancón, A. (2022). Evolutionary emergence of collective intelligence in large groups of students. *Frontiers in Psychology*, 13, 1-16. <https://doi.org/10.3389/fpsyg.2022.848048>
- Penfield, R. D. y Giacobbi, P. R. (2004). Applying a score confidence interval to Aiken's item content-relevance index. *Measurement in Physical Education and Exercise Science*, 8(4), 213-225. https://doi.org/10.1207/s15327841mpee0804_3
- Riedl, C. y Bogert, E. (2024). Effects of AI feedback on learning, the skill gap, and intellectual diversity [Preprint]. *arXiv*. <https://doi.org/10.48550/arXiv.2409.18660>
- Riedl, C. y De Cremer, D. (2025). AI for collective intelligence. *Collective Intelligence*, 4(2). <https://doi.org/10.1177/26339137251328909>
- Russell, S. J. y Norvig, P. (2020). *Artificial intelligence: A modern approach* (4ª ed.). Pearson.
- Shadish, W. R., Cook, T. D. y Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.
- Stevens, J. P. (2002). *Applied multivariate statistics for the social sciences*. Routledge. <https://doi.org/10.4324/9781410604491>
- Sullivan, G. M. y Artino, A. R. (2013). Analyzing and interpreting data from Likert-type scales. *Journal of Graduate Medical Education*, 5(4), 541-542. <https://doi.org/10.4300/JGME-5-4-18>
- Ullmann, T. D., De Liddo, A. y Bachler, M. (2019). Un panel de visualización para la inteligencia colectiva controvertida: analíticas de aprendizaje para la mejora de creación de ideas en grupos de discusión. *RIED-Revista Iberoamericana de Educación a Distancia*, 22(1), 41-80. <https://doi.org/10.5944/ried.22.1.22294>
- UNED. (2025, 20 de febrero). Lorenzo García Aretio: un gran legado en la educación a distancia. <https://www.uned.es/universidad/inicio/noticias/mostrarnoticia.html?noticia=lorenzo-garcia-aretio-un-gran-legado-en-la-educacion-a-distancia>
- Unión Europea. (2024). Reglamento del Parlamento Europeo y del Consejo de 13 de marzo de 2024 por el que se establecen normas armonizadas en materia de inteligencia artificial y se modifican determinados actos legislativos de la Unión. *Diario Oficial de la Unión Europea*, L 1689, 12 de julio de 2024. <https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX:32024R1689>
- Universidad Internacional de La Rioja. (2024). *Manual de calidad*. UNIR. https://static.unir.net/calidad/Manual_calidad.pdf
- Universidad Internacional de La Rioja. (2025). *Centro de Estudios sobre Innovación (CEI)*. UNIR. <https://www.unir.net/universidad-online/vicerrectorados/vicerrectorado-investigacion/cei/>
- Wang, H., Wang, C., Zhan, C., Liu, F., Bao, C. y Xianlong, X. (2025). Impact of AI-agent-supported collaborative learning on the learning outcomes of university

programming courses. *Education and Information Technologies*, 30(12), 17717-17749. <https://doi.org/10.1007/s10639-025-13487-8>

Zawacki-Richter, O., Marín, V. I., Bond, M. y Gouverneur, F. (2019). Systematic

review of research on artificial intelligence applications in higher education: Where are the educators? *International Journal of Educational Technology in Higher Education*, 16(1), 1-27. <https://doi.org/10.1186/s41239-019-0171-0>

Fecha de recepción del artículo: 1 de junio de 2025

Fecha de aceptación del artículo: 24 de agosto de 2025

Fecha de aprobación para maquetación: 24 de septiembre de 2025

Fecha de publicación en OnlineFirst: 15 de octubre de 2025

Fecha de publicación: 1 de enero de 2026

Impacto de la IA generativa en competencias digitales universitarias: evidencia experimental basada en el marco DigComp

Impact of generative AI on university students' digital competences: experimental evidence based on the DigComp framework



 Concepción González García - *Universidad Católica de Murcia, UCAM (España)*
 Nina Pallarés - *Universidad Católica de Murcia, UCAM (España)*

RESUMEN

Este estudio analiza el impacto del uso formativo de inteligencia artificial (IA) generativa en el desarrollo de competencias digitales en estudiantes universitarios. La intervención se implementó mediante un ensayo controlado aleatorizado. El grupo experimental recibió formación orientada a utilizar estratégicamente modelos de IA generativa para la realización de tareas, mientras que el grupo de control completó las mismas actividades sin orientación específica sobre IA. El impacto se evaluó mediante un modelo de diferencias en diferencias con efectos fijos, basado en cuestionarios pre y postintervención. Las competencias se analizaron según el marco europeo DigComp 2.2, contemplándose cuatro áreas: alfabetización en información y datos, comunicación y colaboración, seguridad y resolución de problemas. Los resultados muestran mejoras estadísticamente significativas en alfabetización en información y datos y en resolución de problemas, tanto en su dimensión funcional como metacognitiva. Asimismo, se identificaron efectos diferenciales según el nivel inicial de competencia digital, siendo más pronunciados entre estudiantes con menor dominio previo, quienes presentan avances significativos en todas las competencias evaluadas. Estos hallazgos sugieren un efecto compensatorio del uso didáctico de la IA, capaz de reducir brechas y promover aprendizajes más equitativos. El estudio respalda la integración guiada de tecnologías emergentes en la educación superior para fortalecer las competencias digitales.

Palabras clave: inteligencia artificial; competencias digitales; educación superior; aprendizaje autónomo; DigComp.

ABSTRACT

This study analyzes the impact of the formative use of generative artificial intelligence (AI) on the development of digital competencies in university students. The intervention was implemented through a randomized controlled trial research design. The experimental group received training aimed at strategically using generative AI models to complete academic tasks, while the control group carried out the same activities without specific AI guidance. The impact was assessed using a difference-in-differences model with fixed effects, based on pre- and post-intervention questionnaires. Competences were analyzed according to the European DigComp 2.2 framework, covering four main competence areas: information and data literacy, communication and collaboration, safety, and problem solving. The results show statistically significant improvements in information and data literacy and in problem solving, both in their functional and metacognitive dimensions. Differential effects were also identified depending on the initial level of digital competence, with more pronounced gains among students with lower prior proficiency, who showed significant progress across all evaluated competencies. These findings suggest a compensatory effect of the didactic use of AI, capable of reducing gaps and promoting more equitable and inclusive learning processes. The study supports the guided integration of emerging technologies in higher education to strengthen digital competencies.

Keywords: artificial intelligence; digital competences; higher education; autonomous learning; DigComp.

INTRODUCCIÓN

La irrupción de herramientas basadas en inteligencia artificial (IA) generativa ha supuesto una transformación significativa en los procesos de enseñanza y aprendizaje en la educación superior (Hwang et al., 2020; Lee et al., 2024). Estas herramientas han reconfigurado las formas en las que los estudiantes acceden, analizan y evalúan la información, pudiendo influir en la mejora de sus competencias digitales. En este contexto, y ante la creciente adopción de estas tecnologías, resulta fundamental analizar sus efectos, particularmente en las competencias digitales, elemento clave para afrontar las exigencias de un mercado laboral en permanente transformación (Van Laar et al., 2020).

Estas competencias comprenden el conjunto de conocimientos y habilidades necesarios para utilizar de manera efectiva las tecnologías digitales (Ferrari y Punie, 2013; Laupichler et al., 2022). Actualmente, incluyen no solo la búsqueda y gestión de información, sino también la capacidad de comprender y evaluar críticamente fuentes digitales (Polizzi, 2020). Estudios previos han señalado su importancia en diversos ámbitos como la educación, el empleo y las relaciones sociales (Bastian et al., 2023; Buchholz et al., 2020; Morgan et al., 2022). A su vez, la Comisión Europea las define como una de las ocho competencias clave para el aprendizaje permanente, subrayando su papel en el uso seguro, crítico y responsable de las tecnologías digitales para el aprendizaje, el trabajo y la participación activa en la sociedad (Comisión Europea, 2019). Por ello, promover su desarrollo se ha consolidado como una prioridad en las agendas educativas y políticas.

Investigaciones recientes han demostrado que la incorporación de tecnologías con fines educativos favorece el desarrollo de las competencias digitales, así como la autorregulación y una percepción positiva del aprendizaje por parte del estudiantado (Blau et al., 2020; Lin et al., 2025). Además, cuando estas tecnologías se integran de forma intencional y con una orientación clara, no solo fortalecen dichas competencias, sino que también pueden contribuir al aprendizaje autónomo (Ting, 2015). En la misma línea, Prior et al. (2016) muestran que un mayor nivel de competencias digitales, junto con actitudes positivas hacia la tecnología, se asocia con niveles más altos de autoeficacia, lo que incide positivamente en el aprendizaje autónomo y la participación.

El uso de herramientas de IA generativa en la educación superior ha sido ampliamente investigado, destacándose tanto sus beneficios como los riesgos asociados (Dimitriadou y Lanitis, 2023; Zhai et al., 2021). García Peñalvo et al. (2024) destacan que, si bien las herramientas de IA generativa ofrecen importantes beneficios, como la personalización del aprendizaje, la automatización de tareas repetitivas, la generación de contenidos educativos y el apoyo al desarrollo del pensamiento crítico, también presentan limitaciones significativas. Asimismo, se señalan desafíos vinculados a la equidad en el acceso, la protección de datos personales, y la necesidad de formar tanto a docentes como a estudiantes en un uso

ético y crítico de estas tecnologías. Popenici y Kerr (2017) señalan que la incorporación de tecnologías basadas en inteligencia artificial está transformando profundamente los procesos de enseñanza y aprendizaje en las universidades, lo que plantea nuevos desafíos para las instituciones y requiere una revisión de los enfoques tradicionales.

Pese al creciente interés, son escasos los estudios que analizan de manera directa cómo el uso formativo de la IA generativa contribuye al desarrollo de competencias digitales. Una revisión sistemática conducida por Zhao et al. (2021), que analizó 33 artículos publicados entre 2015 y 2021, revela que apenas un 15 % de los trabajos publicados examina factores que pueden influir en su adquisición. Por el contrario, las investigaciones recientes se han enfocado en la relación inversa: cómo un mayor nivel de competencia digital favorece la adopción de estas tecnologías. En concreto, Moravec et al. (2024) muestran que un mayor nivel de competencias digitales se asocia con un uso más frecuente de ChatGPT para realizar pruebas, entretenimiento y adquisición de nuevos conocimientos. De forma complementaria, estudios recientes han encontrado que una mayor alfabetización digital se asocia con actitudes más favorables hacia el uso de tecnologías impulsadas por IA en contextos educativos (Saklaki y Gardikiotis, 2024), lo que refuerza la idea de una interacción dinámica entre competencias y disposición tecnológica. Si bien algunos de los trabajos recientes ofrecen perspectivas valiosas sobre las habilidades necesarias en entornos mediados por inteligencia artificial, persiste una escasez de investigaciones que evalúen de forma sistemática y estructurada la adquisición de competencias digitales específicas entre el alumnado universitario, utilizando marcos estandarizados como el DigComp.

Este estudio realiza tres aportes principales. En primer lugar, ofrece, según nuestro conocimiento, el primer análisis empírico del impacto de la integración didáctica de la IA generativa en el desarrollo de competencias digitales específicas dentro del marco DigComp, a partir de un diseño cuasiexperimental. En segundo lugar, cuantifica la magnitud de este impacto, identificando avances significativos en aspectos fundamentales como la autorregulación digital y la interacción con herramientas tecnológicas. En tercer lugar, al analizar efectos heterogéneos, revela un potencial efecto nivelador: los mayores avances se observan entre estudiantes con niveles iniciales más bajos, lo que sugiere que, aplicada con intencionalidad didáctica, la IA generativa puede contribuir a una mayor equidad en el desarrollo de competencias digitales.

Los resultados evidencian que una implementación didáctica intencionada de la IA generativa puede no solo potenciar el aprendizaje, sino también reducir desigualdades preexistentes. Su uso estratégico contribuye a construir entornos más equitativos e inclusivos, reforzando competencias clave para desenvolverse de forma crítica y autónoma en escenarios digitales cada vez más complejos.

El resto del artículo se estructura en cinco apartados. Primero, se revisa la literatura relevante sobre competencias digitales y el uso de IA generativa en contextos educativos. A continuación, se describe la metodología empleada, que

comprende el diseño experimental y la estrategia de estimación, basada en un modelo de diferencias en diferencias con efectos fijos. Luego, se describen los datos utilizados y se ofrece un análisis descriptivo. En el apartado siguiente, se exponen y discuten los principales resultados, considerando tanto los efectos promedio como los efectos heterogéneos. Finalmente, se concluye con una reflexión sobre las implicaciones derivadas de los hallazgos.

REVISIÓN DE LA LITERATURA

El desarrollo de competencias digitales se ha convertido en un requisito indispensable para afrontar los retos y beneficiarse de las posibilidades que ofrece un entorno cada vez más digitalizado. En el ámbito educativo, resulta esencial que los estudiantes adquieran estas competencias, ya que la capacidad de utilizar eficazmente las tecnologías y acceder a la información constituye un factor clave para su éxito académico, así como para su desempeño profesional y personal (OECD, 2023).

Esta necesidad adquiere especial relevancia en la educación superior, donde los estudiantes son comúnmente denominados “nativos digitales” (Prensky, 2001), bajo la premisa de que, al haber crecido en entornos digitales, poseen habilidades innatas en el uso de herramientas tecnológicas. Sin embargo, pertenecer a esta generación no garantiza un uso competente, crítico ni orientado al aprendizaje, como han señalado diversos estudios (Gallardo-Echenique et al., 2015; Li y Ranieri, 2010; Ng, 2012; Selwyn, 2009). Esto pone de relieve la importancia de formar a los estudiantes en competencias digitales que les permitan adaptarse activamente a un entorno tecnológico en constante evolución y desarrollar una actitud abierta frente a la innovación (Barak, 2018). Por ello, el profesorado debe fomentar activamente el fortalecimiento de estas competencias con el fin de favorecer el aprendizaje de los estudiantes y aprovechar el potencial educativo que ofrecen las tecnologías digitales.

En este contexto, el concepto de competencia digital ha sido ampliamente abordado en la literatura especializada, dando lugar a una diversidad de enfoques y definiciones. Entre las definiciones de competencias digitales más aceptadas en la actualidad, se encuentra aquella que las concibe como un conjunto interrelacionado de habilidades esenciales para desenvolverse con éxito en entornos digitales (List, 2019). Dicha concepción está presente en el Marco Europeo de Competencia Digital para la Ciudadanía (DigComp), desarrollado inicialmente por Ferrari y Punie (2013). Este marco define las competencias digitales como un conjunto de conocimientos, habilidades y actitudes que permiten a los ciudadanos utilizar de forma eficaz las tecnologías digitales para el trabajo, el aprendizaje y la participación social. El modelo identifica cinco áreas clave: alfabetización en información y datos, comunicación y colaboración, creación de contenido digital, seguridad y resolución de problemas, siendo la versión más reciente el DigComp 2.2 (Vuorikari et al., 2022).

Aunque numerosos estudios han adoptado el modelo DigComp como punto de referencia para la evaluación de competencias digitales en contextos universitarios, la mayoría se ha centrado en medir el nivel de dominio de los estudiantes. En contraste, son relativamente escasas las investigaciones que analizan el impacto de intervenciones didácticas en el desarrollo efectivo de estas competencias. Entre las contribuciones más relevantes en esta línea destaca el estudio de Gutiérrez Porlán y Serrano Sánchez (2016), que demuestra cómo una intervención didáctica mediada por tecnologías digitales puede potenciar significativamente las competencias en las cinco áreas del DigComp. Estos hallazgos refuerzan la importancia de integrar estrategias intencionadas y sostenidas para fomentar una alfabetización digital crítica y funcional.

A pesar del creciente interés y la rápida adopción de herramientas de IA generativa, la literatura empírica que examina su impacto en la mejora de las competencias digitales del alumnado universitario sigue siendo limitada. Kasneci et al. (2023) destacan la importancia de acompañar el uso de herramientas como ChatGPT con estrategias didácticas que promuevan la verificación crítica de la información, el uso de fuentes complementarias fiables y la integración de actividades orientadas al desarrollo cognitivo del alumnado. Además, el diseño de prompts eficaces, la interpretación contextualizada de las respuestas de la IA y la reflexión sobre su uso son elementos clave para lograr una interacción significativa con estas tecnologías (Lee y Palmer, 2025). Fomentar estas prácticas en el aula puede favorecer el fortalecimiento de las competencias digitales. Benvenuti et al. (2023) sostienen que la inteligencia artificial representa un recurso de gran valor para los docentes en la promoción de habilidades como la creatividad, el pensamiento crítico y la resolución de problemas. Por su parte, Arseven y Bal (2025) destacan que este potencial radica en su capacidad para facilitar procesos metacognitivos fundamentales para el desarrollo de dichas competencias; sin embargo, ambos advierten que las experiencias educativas que integran la IA con estos propósitos siguen siendo escasas en la literatura. En consecuencia, resulta pertinente analizar en qué medida una integración didáctica y estructurada de estas tecnologías contribuye al fortalecimiento de dichas competencias en los estudiantes.

Una primera aportación desde esta perspectiva es el trabajo de Dalgıç et al. (2024), que demuestra que la integración de ChatGPT en la educación tiene un impacto positivo tanto en los resultados de aprendizaje como en el desarrollo de las competencias digitales de los estudiantes. El estudio señala que la alfabetización digital es una variable mediadora esencial y que, para aprovechar plenamente las ventajas de la inteligencia artificial, los estudiantes deben poseer y fortalecer continuamente sus competencias digitales. Además, concluye que el uso de ChatGPT fomenta directamente habilidades como la búsqueda y análisis de información, la resolución de problemas y el aprendizaje en entornos virtuales, actuando así no solo como una herramienta que exige competencias digitales, sino también como un recurso que contribuye activamente a su desarrollo. Por ello, los autores enfatizan

la importancia de que las instituciones educativas prioricen la formación en competencias digitales para maximizar el potencial de tecnologías avanzadas como ChatGPT. Una segunda aportación relevante es el estudio de Naamati-Schneider y Alt (2024), quienes afirman haber sido los primeros en evaluar empíricamente la hipótesis de que el uso de ChatGPT puede volver obsoletas determinadas competencias digitales al asumir tareas previamente realizadas por los estudiantes, en particular aquellas vinculadas con el acceso y el análisis de la información. No obstante, sus hallazgos también revelan que la incorporación de esta tecnología en el proceso educativo puede potenciar la competencia de evaluación crítica, al exigir que los estudiantes verifiquen la validez y fiabilidad de los contenidos generados por la IA.

El análisis de estudios previos evidencia una creciente necesidad de promover competencias digitales en la educación superior, al tiempo que revela una base empírica aún limitada sobre el impacto de las herramientas de inteligencia artificial generativa en su desarrollo, especialmente por la falta de un marco de referencia unificado que facilite la comparación entre estudios. Si bien estas tecnologías muestran un potencial considerable para fortalecer dichas competencias, su efectividad depende en gran medida de cómo se integran en el proceso de enseñanza y aprendizaje. En este contexto, la presente investigación se propone cuantificar el efecto de una intervención didáctica basada en IA generativa, con el objetivo de aportar evidencia sobre su contribución al desarrollo de las competencias digitales del alumnado universitario, dentro del marco DigComp, así como analizar sus posibles efectos heterogéneos.

METODOLOGÍA

Para evaluar el impacto del uso formativo de la inteligencia artificial generativa (IA) en el aprendizaje, llevamos a cabo un ensayo controlado aleatorizado (RCT, por sus siglas en inglés) con estudiantes matriculados en la asignatura de Microeconomía. Los estudiantes pertenecían a las titulaciones de Administración y Dirección de Empresas y de Marketing y Dirección Comercial de la Universidad Católica de Murcia. Ambos programas ofrecían la asignatura en dos modalidades: presencial y en línea. Además, en el programa de Administración de Empresas, la modalidad presencial estaba disponible tanto en español como en inglés. Así, se conformaron cinco grupos (aulas) en total: tres presenciales y dos en línea.

La asignación del tratamiento se realizó a nivel de grupo (aula), de manera que cada uno de los cinco grupos fue asignado en su totalidad al tratamiento o al grupo de control. La asignación aleatoria se estratificó por modalidad (presencial o en línea) para asegurar una distribución equitativa del tratamiento entre las modalidades.¹

Para mantener un entorno de aprendizaje consistente en todas las modalidades, el profesorado estandarizó los materiales y métodos de enseñanza, y estableció un protocolo de evaluación con el fin de garantizar que todos los grupos recibieran

un contenido idéntico del curso (materiales, recursos y actividades), impartido de manera uniforme y evaluado con los mismos criterios, independientemente del programa. Todo el material del curso estuvo disponible a través de CANVAS, la plataforma institucional de la universidad.

La intervención constó de dos componentes: (1) una formación orientada a la aplicación estratégica de modelos de IA generativa, ofrecida únicamente al grupo de tratamiento, y (2) una serie de autoevaluaciones en línea, aplicadas a ambos grupos, compuestas por preguntas de opción múltiple que cubrían tanto los contenidos teóricos como prácticos de cada unidad del curso. Las autoevaluaciones incluían retroalimentación inmediata, la cual fue diferenciada según el grupo asignado.

La formación se centró en técnicas de elaboración de prompts eficaces y en el uso de modelos de lenguaje de gran escala (LLMs, por sus siglas en inglés) potenciados por IA, con el objetivo general de mejorar el aprendizaje y las competencias digitales de los estudiantes. Para ello, se plantearon los siguientes objetivos específicos, orientados a promover una interacción efectiva con los LLMs:

- Contextualizar la pregunta - Fomentar que los estudiantes describan el problema completo antes de formular preguntas específicas.
- Definir términos y variables clave - Enseñar a los estudiantes a expresar de manera explícita los principios económicos, fórmulas o supuestos relevantes a sus consultas.
- Descomponer problemas complejos - Aconsejar a estructurar las consultas en segmentos paso a paso para obtener respuestas más organizadas.
- Usar preguntas iterativas - resaltar la importancia de reformular, refinar y ampliar las respuestas generadas por la IA para profundizar la comprensión.
- Verificar sus propias soluciones - Fomentar que los estudiantes intenten resolver los problemas por sí mismos antes de recurrir a la IA.

Estos objetivos se abordaron en un taller formativo impartido exclusivamente al grupo de tratamiento durante el primer día de clase. La sesión se centró, principalmente, en el uso de ChatGPT, aunque también se presentaron brevemente otras herramientas, como Copilot y Gemini. El profesorado ofreció una explicación introductoria, acompañada de ejemplos prácticos, tras lo cual los estudiantes interactuaron por sí mismos con las herramientas y pudieron plantear sus dudas.

En las semanas posteriores, los estudiantes debían completar las autoevaluaciones en línea al finalizar cada unidad, a través de la plataforma Canvas del curso de Microeconomía. Cada prueba tenía un límite de tiempo de una hora y, al finalizar, los estudiantes recibían una calificación numérica. Se permitía un máximo de cinco intentos por evaluación, con un intervalo mínimo de un día entre ellos. En cada intento, las preguntas se generaban de forma aleatoria a partir del banco de ítems. La retroalimentación en las autoevaluaciones fue diseñada intencionalmente para diferenciarse entre los grupos de tratamiento. En el grupo tratado, la retroalimentación

proporcionaba una guía paso a paso, con pistas y prompts personalizados alineados con los objetivos específicos enseñados para interactuar eficazmente con los LLMs. Esta retroalimentación formativa estaba diseñada para orientar al estudiantado hacia la respuesta correcta sin proporcionarla explícitamente, fomentando una comprensión más profunda y mejorando su capacidad de resolver problemas de forma autónoma. En lugar de indicar únicamente si la respuesta era correcta o incorrecta, la retroalimentación era adaptativa: los estudiantes podían interactuar con distintos componentes de la retroalimentación y personalizar su experiencia de aprendizaje según sus necesidades. Esto les permitía explorar explicaciones alternativas y refinar su comprensión según su rendimiento y brechas de aprendizaje. En cambio, los estudiantes del grupo de control recibieron únicamente una retroalimentación básica que indicaba si sus respuestas eran correctas o incorrectas, sin explicaciones, pistas ni recursos instruccionales adicionales. Los materiales y datos para replicar la intervención están disponibles en un repositorio abierto.²

Para evaluar el impacto de la intervención se administraron dos cuestionarios: uno el primer día de clase, antes del taller formativo en IA (pre), y otro tras la última autoevaluación (postintervención). Cada cuestionario incluía dos partes: (i) datos demográficos, familiares y de alfabetización económica³, y (ii) una evaluación de las competencias digitales autoinformadas por el estudiantado, basada en el marco desarrollado por Vuorikari et al. (2022). La primera versión de este marco, DigComp 1.0, fue elaborada por Ferrari y Punie (2013), con el objetivo de identificar y definir las competencias digitales relevantes para todos los ciudadanos que viven y trabajan en Europa. Actualmente, el modelo DigComp se ha consolidado como uno de los marcos de referencia más utilizados para el diseño de instrumentos de medición de competencias digitales en contextos universitarios (Mattar et al., 2022).

En la Tabla 1 se presenta una descripción detallada de los ítems respondidos por los estudiantes en relación con las competencias digitales evaluadas. Estas se alinearon con el modelo europeo DigComp 2.2 e incluyeron cuatro áreas clave: alfabetización en información y datos, comunicación y colaboración, seguridad, y resolución de problemas. En esta última, se evaluaron dos dimensiones: por un lado, la interacción con herramientas tecnológicas o dimensión funcional (Comp3), y por otro, la autorregulación digital, entendida como la capacidad de reflexionar metacognitivamente sobre el propio aprendizaje en entornos digitales (Comp4).

Tabla 1
Competencias digitales evaluadas

DigComp			Ítem
Área 1	Alfabetización en información y datos	Comp1	Identificar mis necesidades de información y encontrar datos y contenido a través de una búsqueda sencilla en entornos digitales, como por ejemplo Google.

	DigComp		Ítem
Área 2	Comunicación y colaboración	Comp2	Lograr una comunicación efectiva utilizando herramientas digitales, como, por ejemplo, el correo electrónico, foros de discusión, etc.
Área 5	Resolución de problemas	Comp3	Dar instrucciones simples para que un sistema informático me ayude a resolver un problema o tarea básica, como, por ejemplo, bases de datos, software de hojas de cálculo, software de presentaciones, herramientas de IA, etc.
	Resolución de problemas	Comp4	Reconocer qué necesito mejorar o actualizar en mis competencias digitales.
Área 4	Seguridad	Comp5	Evaluar los beneficios y riesgos antes de permitir que terceros procesen datos personales cuando utilizo entornos digitales o herramientas de IA.

Los ítems, redactados en formato de autoinforme y con opciones cerradas, comenzaban con “soy capaz de...”, para evaluar el nivel de habilidad percibida. Se formularon de forma neutral, de modo que la respuesta indicara el grado de dominio sin inducir dificultad (Clifford et al., 2020). Para su elaboración, se tomaron como base las definiciones y ejemplos del marco DigComp 2.2, adaptados a los objetivos del estudio. Además, se optó por incluir solo cinco ítems para medir las competencias digitales, con el fin de evitar que el cuestionario resultara excesivamente extenso, reduciendo así el riesgo de abandono y favoreciendo la calidad de las respuestas, una estrategia recomendada en este ámbito (Vuorikari et al., 2025).

Los estudiantes valoraron su grado de acuerdo con cada afirmación mediante una escala Likert de 1 a 5, donde 1 correspondía al nivel más bajo (básico) y 5 al más alto (avanzado). Este formato, basado en ítems autoinformados, es uno de los métodos más utilizados para medir la competencia digital (Laanpere, 2019).

La fiabilidad del instrumento se estimó mediante los coeficientes alfa de Cronbach y omega de McDonald, que arrojaron los mismos valores tanto en el pretest (0.75) como en el postest (0.80). Estos coeficientes indican una consistencia interna aceptable en la medición inicial y buena tras la intervención (Hair et al., 2010).

Método de estimación

La estimación se basa en un experimento aleatorizado combinado con un modelo de diferencias en diferencias (DiD) con efectos fijos. En primer lugar, la aleatorización permite asignar a los participantes al grupo de tratamiento y al grupo de control de forma exógena, lo que garantiza la comparabilidad y minimiza los sesgos de selección. A partir de esta asignación, se implementó una estrategia de

diferencias en diferencias que compara la evolución del nivel de competencia entre ambos grupos antes y después de la intervención.

Dado que la asignación del tratamiento fue a nivel de grupo (aula), los errores estándar se agruparon en esa unidad. Para corregir posibles sesgos derivados del reducido número de clusters (cinco), la significatividad estadística se evaluó mediante el método wild cluster bootstrap-t, como se detalla en la sección de robustez.

La especificación econométrica es la siguiente:

$$Y_{it} = \delta(Treat_i \times Post_t) + \beta Post_t + \gamma_i + \epsilon_{it}$$

Donde:

- Y_{it} : variable ficticia que toma el valor 1 si el estudiante reporta el nivel máximo (valor 5) en la competencia analizada.
- $Treat_i \times Post_t$: interacción entre $Treat_i$ y $Post_t$.
- $Treat_i$: variable ficticia que toma valor 1 si el estudiante está en el grupo de tratamiento.
- $Post_t$: dummy que toma valor 1 para las observaciones posteriores a la intervención (cuestionario final).
- γ_i : efectos fijos individuales.

La dicotomización de la variable dependiente facilita una interpretación más directa del efecto del tratamiento, al centrarse en la probabilidad de alcanzar el valor 5, entendido como un umbral claro y exigente de competencia plena. Esta definición atenúa sesgos de deseabilidad social al exigir la respuesta máxima, filtrando así posibles sobreestimaciones moderadas. Aunque supone una pérdida parcial de varianza, la robustez de los resultados se contrasta con modelos ordinales (logit ordinal), presentados en la sección correspondiente.

Tabla 2

Características iniciales

	(1)	(2)	(3)	(4)
Variable	Todos	Tratados	Control	P-valor (2)-(3)
A. Características del Estudiante				
Mujer	0.355 (0.037)	0.356 (0.047)	0.354 (0.060)	0.968
Edad	20.911 (0.328)	20.808 (0.452)	21.077 (0.453)	0.823

	(1)	(2)	(3)	(4)
Variable	Todos	Tratados	Control	P-valor (2)-(3)
(Edad)^2	455.325 (20.568)	454.039 (29.661)	457.385 (24.936)	0.962
Trabaja	0.266 (0.034)	0.269 (0.044)	0.262 (0.055)	0.903
Estudiante de intercambio	0.053 (0.017)	0.058 (0.023)	0.046 (0.026)	0.691
Estudiante repetidor	0.077 (0.021)	0.096 (0.029)	0.046 (0.026)	0.360
Estudiante <i>online</i>	0.089 (0.022)	0.067 (0.025)	0.123 (0.041)	0.760
Nacional de un país occidental	0.716 (0.035)	0.625 (0.048)	0.862 (0.043)	0.512
Alfabetización económica pre-curso	3.840 (0.14)	3.875 (0.176)	3.785 (0.232)	0.800
Competencia digital pre- curso ⁴	0.000 (0.123)	-0.297 (0.155)	0.475 (0.189)	0.141
B. Características del Hogar				
Renta <20000€	0.095 (0.023)	0.096 (0.029)	0.092 (0.036)	0.853
Renta 20000-40000€	0.243 (0.033)	0.269 (0.044)	0.2 (0.055)	0.124
Renta 40000-60000€	0.266 (0.034)	0.269 (0.044)	0.262 (0.055)	0.846
Renta >60000€	0.32 (0.036)	0.288 (0.045)	0.369 (0.06)	0.548
Madre - Educación Primaria	0.065 (0.019)	0.087 (0.028)	0.031 (0.022)	0.301
Madre - Educación Secundaria	0.420 (0.038)	0.337 (0.047)	0.554 (0.062)	0.118
Madre - Educación Terciaria	0.515 (0.039)	0.577 (0.049)	0.415 (0.062)	0.157
Padre - Educación Primaria	0.065 (0.019)	0.077 (0.026)	0.046 (0.026)	0.467

	(1)	(2)	(3)	(4)
Variable	Todos	Tratados	Control	P-valor (2)-(3)
Padre - Educación Secundaria	0.432 (0.038)	0.394 (0.048)	0.492 (0.062)	0.534
Padre - Educación Terciaria	0.503 (0.039)	0.529 (0.049)	0.462 (0.062)	0.652
Observaciones	169	104	65	

Nota: errores estándar robustos agrupados por aula entre paréntesis. El valor *p* en la columna (4) corresponde a la diferencia de medias y se ha calculado ajustando por el bajo número de clusters mediante el método wild cluster bootstrap-t.

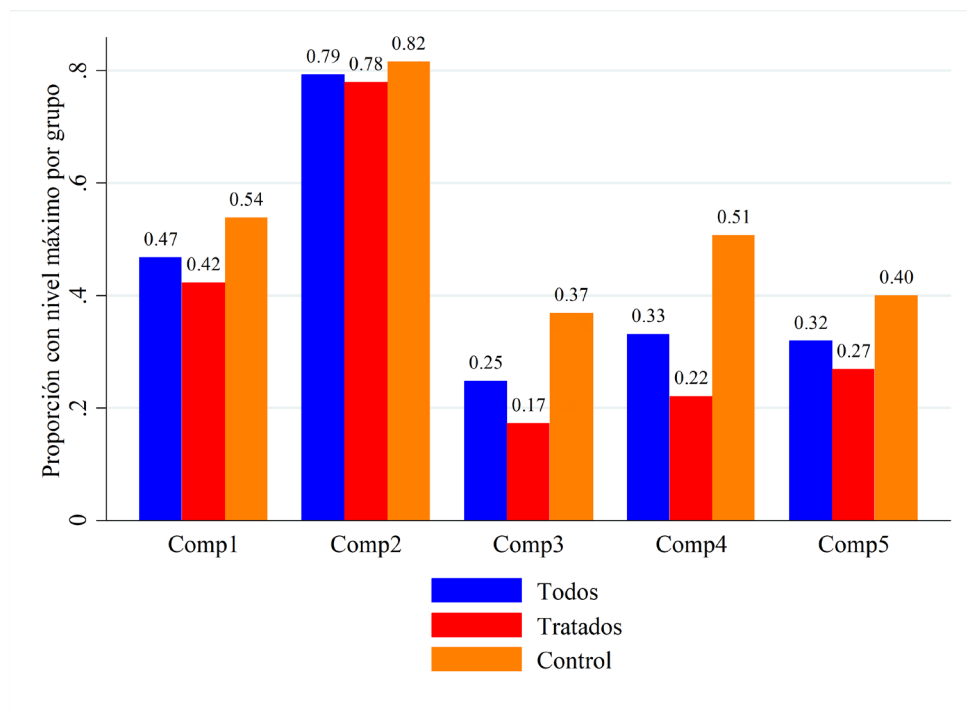
DATOS Y ESTADÍSTICOS DESCRIPTIVOS

Tal y como se ha descrito en la sección anterior, los datos utilizados provienen de dos encuestas llevadas a cabo antes y después de la intervención⁵. La Tabla 2 presenta las características iniciales de los estudiantes designados de forma aleatoria a los grupos de tratamiento y control. El total de la muestra lo conforman 169 estudiantes de los cuales 104 fueron asignados aleatoriamente al grupo de tratamiento y 65 al grupo de control según el grupo (aula) al que pertenecían. No se detectan diferencias estadísticamente significativas entre los grupos en ninguna de las variables demográficas, académicas o socioeconómicas consideradas, lo cual sugiere que el proceso de aleatorización fue apropiado y hay un balance adecuado en las condiciones iniciales.

64 García, C. G., & Pallarés, N. (2026). Impact of generative AI on university students' digital competences: experimental evidence based on the DigComp framework [Impacto de la IA generativa en competencias digitales universitarias: evidencia experimental basada en el marco DigComp]. *RIED-Revista Iberoamericana de Educación a Distancia*, 29(1), pp. 53-77. <https://doi.org/10.5944/ried.45533>

Figura 1

Proporción de estudiantes con nivel máximo (valor 5) en competencias digitales antes de la intervención



De forma complementaria, el análisis por dimensiones presentado en la Figura 1 revela que los niveles iniciales de competencia eran heterogéneos entre áreas. Las proporciones de estudiantes que reportaron haber alcanzado el nivel máximo (valor 5) oscilaron entre 0.25 (Comp3) y 0.79 (Comp2) en el total de la muestra, y entre el 0.17 y el 0.78, respectivamente, dentro del grupo de tratamiento. Esto indica un margen considerable para la mejora, especialmente en dimensiones funcionales como la interacción con herramientas tecnológicas (Comp3), así como en áreas metacognitivas relacionadas con las habilidades de autorregulación digital (Comp4) o con la toma de decisiones informadas en seguridad digital (Comp5). Tal como señalan estudios previos, niveles bajos de competencia digital general suelen asociarse con menor confianza y práctica en seguridad digital (Tomczyk, 2020), lo que podría contribuir a explicar los valores relativamente más bajos en estas dimensiones.

RESULTADOS Y DISCUSIÓN

Los modelos estimados emplean una variable dependiente dicotómica que indica si el estudiante alcanzó el nivel más alto de competencia digital, lo que permite interpretar los coeficientes (multiplicados por 100) como cambios en la probabilidad, en puntos porcentuales (p.p.), de alcanzar dicho nivel.

La Figura 2 ilustra de forma visual los efectos promedio estimados del tratamiento sobre la probabilidad de alcanzar el nivel más alto de competencia digital. Se observa con claridad que las mayores ganancias se concentran en las dimensiones de autorregulación digital (Comp4) e interacción con herramientas tecnológicas (Comp3), seguidas por la competencia en identificación y búsqueda de información (Comp1). En cambio, las competencias asociadas a la comunicación digital (Comp2) y la seguridad digital (Comp5) muestran estimaciones positivas, aunque con menor magnitud y mayor incertidumbre estadística. Esta representación gráfica ofrece una visualización integrada de la magnitud y la precisión estadística de los efectos estimados para cada competencia.

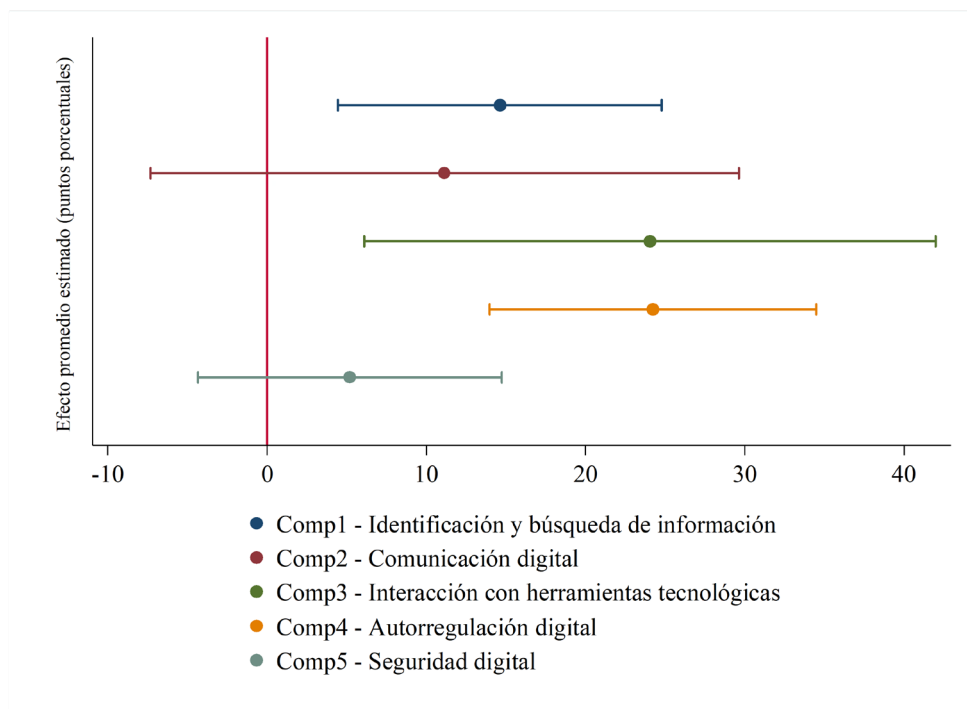
Estos resultados se corresponden con los coeficientes presentados en la Tabla 3, donde se confirma un impacto positivo y significativo del tratamiento en tres de las cinco dimensiones analizadas: Comp1, Comp3 y Comp4. El impacto más fuerte se observa en Comp3 y Comp4, con aumentos de 24 p.p., seguidos por Comp1 con un incremento de 14.6 p.p. En contraste, las competencias ligadas a la comunicación digital (Comp2) y la seguridad digital (Comp5) presentan efectos positivos, pero no significativos a nivel promedio, aunque sí emergen patrones diferenciados cuando se considera el nivel competencial inicial.

Para explorar la existencia de efectos distributivos, se estimaron modelos con interacciones basadas en una variable indicadora que identifica al alumnado cuya competencia digital inicial estaba por encima de la mediana (Q2). Los resultados revelan heterogeneidad en los impactos⁶, con los beneficios más pronunciados concentrados en el subgrupo con niveles iniciales bajos (es decir, por debajo de la mediana), tal y como muestra el Panel B de la Tabla 3. En estudiantes con menor nivel de competencia digital inicial, se observaron mejoras significativas en todas las competencias, con incrementos que oscilan entre los 18 y 38 p.p. En cambio, entre quienes tenían mayor competencia digital inicial, los efectos estadísticamente significativos fueron positivos en el caso de la autorregulación digital (Comp4) con un efecto de 14 p.p., pero negativos en el caso de la seguridad digital (Comp5). Este resultado negativo podría reflejar un ajuste en la autoevaluación de los estudiantes con mayor nivel inicial, quienes tras la intervención podrían haber desarrollado una visión más crítica de sus propias competencias. Aunque el estudio no recoge evidencia empírica directa al respecto, este patrón es coherente con las implicaciones teóricas del efecto Dunning-Kruger, según el cual el aumento del conocimiento puede ir acompañado de una mayor conciencia sobre las propias limitaciones (Dunning, 2011; Kruger y Dunning, 1999).

Todo esto apunta a un efecto compensatorio de la intervención en el contexto estudiado, coherente con investigaciones que muestran cómo las tecnologías, implementadas con orientación didáctica, pueden reducir brechas y fomentar aprendizajes más equitativos.

Figura 2

Efectos promedio del tratamiento sobre la probabilidad de alcanzar el nivel más alto de competencia digital



Nota: coeficientes estimados de Treat*Post, en puntos porcentuales (por 100), con intervalos de confianza al 95 %. Corresponden con el Panel A de la Tabla 3.

Investigaciones recientes sobre IA generativa destacan su potencial para fortalecer competencias digitales. Naamati-Schneider y Alt (2024) evidencian que, en entornos estructurados como el aprendizaje basado en problemas, ChatGPT promueve habilidades como formular prompts eficaces y evaluar críticamente los resultados, subrayando que estas destrezas requieren una implementación didáctica deliberada. De lo contrario, las competencias digitales fundamentales podrían verse desatendidas. En la misma línea, Güner y Er (2025) observan que, sin orientación, los

estudiantes utilizan ChatGPT de manera superficial, mientras que el entrenamiento en prompting y uso estratégico fomenta interacciones más reflexivas y autónomas. Finalmente, estudios experimentales y metaanálisis (Wang et al., 2025; Wang y Fan, 2025) confirman que el *feedback* personalizado de ChatGPT mejora motivación, reflexión y desempeño en tareas complejas, apoyando el desarrollo de habilidades cognitivas y digitales avanzadas.

Los hallazgos son especialmente relevantes por la mejora en competencias metacognitivas como la autorregulación digital (Comp4). Lin et al. (2025) señalan que el uso guiado de herramientas como ChatGPT facilita la planificación autónoma, la organización de ideas y la formulación de preguntas, al mitigar limitaciones de memoria de trabajo y estructuración de la información. Sin embargo, advierten que este apoyo puede generar dependencia del *feedback* automatizado, limitando la autonomía crítica. De ahí la necesidad de acompañar la IA con estrategias didácticas que promuevan una apropiación activa.

Aunque estudios como Hillmayr et al. (2020) se centran en niveles educativos inferiores, sus hallazgos sobre el efecto positivo de las tecnologías en autorregulación y resolución de problemas podrían extrapolarse al ámbito universitario, especialmente entre estudiantes con menor nivel de competencia.

El patrón observado en nuestros resultados, con mayores beneficios entre estudiantes con menor nivel inicial, sugiere que estas tecnologías, implementadas con intencionalidad didáctica, pueden desempeñar un papel compensatorio. Este efecto nivelador destaca en dimensiones clave como la autorregulación digital (Comp4) y la interacción con herramientas tecnológicas (Comp3), esenciales en contextos universitarios que demandan creciente autonomía y entornos orientados a la adquisición de competencias digitales vinculadas al empleo crítico de la IA. Ello coincide con la literatura que sostiene que la IA generativa, como apoyo personalizado, puede desempeñar funciones equivalentes a sistemas de tutoría inteligente sin reemplazar al docente (Hwang et al., 2020).

En el panorama educativo actual destaca la necesidad de reforzar competencias que permitan a los estudiantes adaptarse a entornos dinámicos e inciertos, especialmente las relacionadas con la resolución de problemas, consideradas clave para la formación continua (Csapó y Funke, 2017; OECD, 2013; Rahman, 2019).

Tabla 3

Resultados de los modelos lineales de diferencias en diferencias con efectos fijos (Coeficientes estimados sobre la probabilidad de alcanzar el nivel más alto de competencia digital)

	(1)	(2)	(3)	(4)	(5)
	Comp1	Comp2	Comp3	Comp4	Comp5
A. Promedio					
Treat*Post	0.146** (0.037)	0.112 (0.067)	0.240** (0.065)	0.242*** (0.037)	0.052 (0.034)
R ²	0.017	0.012	0.055	0.038	0.004
B. Competencias digitales pre-curso por encima de la mediana Q2					
Treat*Post	0.374*** (0.041)	0.209** (0.05)	0.294*** (0.045)	0.318*** (0.057)	0.178*** (0.02)
Treat*Post*Q2	-0.539*** (0.105)	-0.230* (0.108)	-0.126 (0.082)	-0.179 (0.095)	-0.298*** (0.009)
Treat*Post+Treat*Post*Q2	-0.165 (0.092)	-0.021 (0.133)	0.168 (0.097)	0.139* (0.065)	-0.12** (0.027)
R ²	0.151	0.039	0.061	0.05	0.043
N	338	338	338	338	338

Nota: errores estándar robustos agrupados por aula entre paréntesis. * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$

Robustez de los resultados

Para evaluar la robustez de los resultados de los modelos lineales DiD (Tabla 3), se estimaron modelos logit ordinales bajo la misma estrategia. Estos modelos, que consideran la naturaleza ordinal de la variable dependiente, permiten calcular efectos marginales sobre la probabilidad de alcanzar el nivel más alto de competencia digital (valor 5). Los resultados (Tabla 4) confirman la validez de las estimaciones al mostrar efectos consistentes.

En promedio (Panel A), el tratamiento muestra efectos positivos y significativos en todas las dimensiones (13-28 p.p.), especialmente en Comp4 y Comp3. Este patrón se acentúa entre estudiantes con menor nivel inicial (Panel B), que muestran mejoras más amplias (21-36 p.p.), mientras que los de nivel alto también se benefician, aunque en menor medida (17-27 p.p.). Las interacciones negativas y significativas refuerzan esta diferencia.

Las diferencias entre modelos responden al tratamiento de la variable dependiente: el modelo lineal opera sobre una versión dicotómica, mientras que el modelo ordinal considera toda la escala y capta efectos no lineales. Ambos coinciden en los patrones clave, con efectos positivos especialmente entre estudiantes con menor nivel inicial. El modelo lineal se prefiere por su claridad interpretativa y control de heterogeneidad no observada mediante efectos fijos.⁷ La coherencia entre enfoques refuerza la robustez de las conclusiones.

Tabla 4
Resultados de los modelos logit ordinales de diferencias en diferencias (Efectos marginales sobre la probabilidad de alcanzar el nivel más alto de competencia digital)

	(1)	(2)	(3)	(4)	(5)
	Comp1	Comp2	Comp3	Comp4	Comp5
A. Promedio					
Treat*Post	0.144*** (0.000)	0.126*** (0.002)	0.175*** (0.008)	0.281*** (0.003)	0.160*** (0.008)
R ²	0.02	0.02	0.021	0.028	0.02
B. Competencias digitales pre-curso por encima de la mediana Q2					
Treat*Post	0.211*** (0.001)	0.209*** (0.000)	0.227*** (0.004)	0.364*** (0.003)	0.302*** (0.003)
Treat*Post*Q2	-0.247*** (0.001)	-0.253*** (0.002)	-0.151*** (0.002)	-0.233*** (0.095)	-0.330*** (0.005)
Treat*Post+Treat*Post*Q2	0.188*** (0.011)	0.170*** (0.012)	0.185*** (0.009)	0.273*** (0.008)	0.183*** (0.017)
R ²	0.111	0.081	0.098	0.042	0.102
N	338	338	338	338	338

Nota: errores estándar robustos agrupados por aula entre paréntesis. * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$

Además de la consistencia observada en los modelos alternativos (logit ordinal), se ha verificado la robustez de los resultados desde el punto de vista de la inferencia estadística. Como se detalla en la sección metodológica, dado que la asignación del tratamiento se realizó a nivel de grupo (aula), con un total de cinco clusters, la significatividad estadística de los coeficientes se evaluó mediante el método wild cluster bootstrap-t propuesto por Cameron et al. (2008), implementado con pesos de Webb (2014) y 1.000 repeticiones.⁸ Los resultados principales se mantienen

70 García, C. G., & Pallarés, N. (2026). Impact of generative AI on university students' digital competences: experimental evidence based on the DigComp framework [Impacto de la IA generativa en competencias digitales universitarias: evidencia experimental basada en el marco DigComp]. *RIED-Revista Iberoamericana de Educación a Distancia*, 29(1), pp. 53-77. <https://doi.org/10.5944/ried.45533>

estables con este procedimiento y las conclusiones generales no se ven afectadas, lo que respalda la validez de las estimaciones.

CONCLUSIONES Y RECOMENDACIONES

Este estudio aporta evidencia empírica sobre el impacto positivo de una intervención didáctica, centrada en el uso de inteligencia artificial generativa, orientada específicamente al desarrollo de competencias digitales en estudiantes universitarios. El análisis revela efectos promedio significativos en tres de las cinco dimensiones evaluadas del marco DigComp: identificación y búsqueda de información (Comp1), interacción con herramientas tecnológicas (Comp3) y autorregulación digital (Comp4), con incrementos que oscilan entre 14.6 y 24 p.p.

Estos avances no solo son relevantes por su magnitud, sino también por el tipo de competencias involucradas. En particular, el incremento en la “alfabetización informacional y de datos” (Comp1) destaca en un contexto donde la abundancia de contenidos generados por sistemas automatizados exige habilidades críticas para filtrar, verificar y contrastar información. Este hallazgo refuerza el papel central de esta competencia en los marcos de ciudadanía digital y aprendizaje autónomo propuestos por Vuorikari et al. (2025).

En lo que respecta a las competencias Comp3 y Comp4, los resultados reflejan mejoras significativas promedio tanto en el uso funcional de la tecnología como en la dimensión metacognitiva del aprendizaje digital. Estas competencias son especialmente críticas en contextos universitarios que demandan una creciente autonomía del estudiantado. Asimismo, el análisis de efectos heterogéneos revela un patrón consistente: los mayores beneficios se concentran entre quienes partían de niveles iniciales más bajos, lo que sugiere que la intervención tuvo un efecto nivelador. En estos casos, se observaron mejoras significativas en todas las competencias, lo que refuerza el potencial compensatorio de este tipo de estrategias.

Los hallazgos sugieren la necesidad de revisar la idea generalizada de que los estudiantes universitarios, por su familiaridad con la tecnología, poseen habilidades digitales desarrolladas. Tal como señalan investigaciones previas, ser “nativo digital” no garantiza un uso competente ni crítico de las herramientas tecnológicas. Por el contrario, los resultados apuntan a la necesidad de intervenciones didácticas explícitas, bien diseñadas y guiadas, que aborden de forma sistemática la adquisición de estas habilidades (Gallardo-Echenique et al., 2015; Ng, 2012).

En conjunto, el estudio contribuye al campo emergente sobre IA generativa en educación, al ofrecer evidencia cuantitativa sobre cómo su integración didáctica puede fortalecer tanto competencias funcionales como metacognitivas. Estos resultados respaldan el uso intencionado de tecnologías emergentes no solo como recurso didáctico, sino también como herramienta para reducir desigualdades y promover entornos de aprendizaje más equitativos e inclusivos en contextos

similares, alineados con los principios del aprendizaje a lo largo de la vida y con los estándares europeos de competencias digitales.

Finalmente, los resultados subrayan la importancia de que el profesorado participe activamente en la incorporación didáctica de estas tecnologías. Esto exige docentes con competencias digitales sólidas y una actitud abierta al cambio, capaces de adaptar sus prácticas a contextos educativos en transformación (Inamorato dos Santos et al., 2023). Además, dado que su impacto varía en función del nivel inicial del estudiantado, su implementación debe guiarse por criterios didácticos que aseguren un uso eficaz, ético y orientado al cierre de brechas. Por lo tanto, aprovechar las posibilidades que ofrece la IA generativa en el ámbito educativo no solo implica introducir nuevas herramientas, sino repensar las condiciones en las que estas pueden contribuir a una formación más justa y adaptada a los desafíos del presente.

No obstante, estos resultados deben interpretarse considerando ciertas limitaciones. En primer lugar, la muestra se compone exclusivamente de estudiantes de una universidad, lo que limita la generalización de los hallazgos a otros contextos educativos. Futuras investigaciones podrían ampliar la muestra incluyendo diversas instituciones y asignaturas, lo que permitiría evaluar la transferibilidad y escalabilidad de la intervención. En segundo lugar, aunque la asignación al tratamiento fue aleatoria, se realizó a nivel de aula por razones didácticas y para evitar *spillovers*, lo que redujo el número de clústeres y podría implicar cierto riesgo de autoselección *ex ante*. Para afrontarlo, los análisis incorporaron inferencia basada en remuestreo, así como pruebas de balance previas (Tabla 2), que muestran una distribución equilibrada de características observables entre condiciones. Asimismo, se incluyeron efectos fijos individuales en los modelos de diferencias en diferencias para controlar posibles sesgos asociados a factores no observados invariables. Este enfoque refuerza la interpretación causal, aunque no sustituye la evidencia que podrían aportar estudios adicionales. En tercer lugar, las variables dependientes provienen de cuestionarios autoinformados, lo que puede introducir sesgos de percepción o deseabilidad social. Para reducir estos posibles sesgos, definimos la variable de interés de forma estricta (solo quienes reportaron el nivel máximo de competencia), con el fin de evitar sobreestimaciones moderadas. Futuros estudios podrían complementarse con medidas más objetivas, como rúbricas de desempeño o el análisis de artefactos digitales.

NOTAS

- ¹ La aleatorización condujo al siguiente reparto de los grupos (aulas): en Administración de Empresas, el grupo presencial en inglés y el grupo en línea fueron asignados al tratamiento; en Marketing y Dirección Comercial, también fue tratado el grupo presencial. El resto de los grupos actuaron como grupo de control: el grupo presencial en Administración de Empresas y el grupo en línea en Marketing y Dirección Comercial.

2. https://figshare.com/projects/Impacto_de_la_IA_Generativa_en_Competencias_Digitales_Universitarias_Evidencia_Experimental_Basada_en_el_Marco_DigComp/262561
3. La alfabetización económica se evaluó en el cuestionario mediante 10 preguntas de opción múltiple, adaptadas del Test of Economic Literacy (TEL-4) y de la sección de microeconomía del Test of Understanding in College Economics (TUCE-4) (Walstad et al., 2007, 2013).
4. La competencia digital de cada estudiante antes del curso se estima mediante un análisis de componentes principales basado en las cinco competencias evaluadas. El índice resultante varía entre -5.662 y 2.341.
5. Además, se empleó información administrativa de matrícula (género, edad, estudiante de intercambio, estudiante repetidor y estudiante online).
6. Adicionalmente, al nivel inicial de competencia digital, se analizó la heterogeneidad del efecto del tratamiento según género, edad y conocimientos previos en economía. Se observaron efectos más intensos en hombres y en estudiantes con menor nivel de alfabetización económica al inicio del curso. En mujeres, la autorregulación digital (Comp4) mostró mejoras significativas, dimensión que presenta efectos positivos en todos los subgrupos. La edad, en cambio, no muestra un efecto diferencial claro.
7. Aunque en términos formales es posible incluir *dummies* individuales en modelos no lineales como el *logit* ordinal, este enfoque no reproduce las propiedades de estimación de los efectos fijos lineales y puede generar sesgos o problemas de identificación, especialmente cuando el número de observaciones por unidad es limitado (Greene, 2004).
8. Este enfoque, adecuado para menos de diez agrupaciones, mejora la precisión de los valores *p* y se mantiene robusto incluso con cinco clústeres.

REFERENCIAS

- Arseven, T. y Bal, M. (2025). Critical literacy in artificial intelligence-assisted writing instruction: A systematic review. *Thinking Skills and Creativity*, 57, 101850. <https://doi.org/10.1016/j.tsc.2025.101850>
- Barak, M. (2018). Are digital natives open to change? Examining flexible thinking and resistance to change. *Computers & Education*, 121, 115-123. <https://doi.org/10.1016/j.compedu.2018.01.016>
- Benvenuti, M., Cangelosi, A., Weinberger, A., Mazzoni, E., Benassi, M., Barbaresi, M. y Orsoni, M. (2023). Artificial intelligence and human behavioral development: A perspective on new skills and competences acquisition for the educational context. *Computers in Human Behavior*, 148, 107903. <https://doi.org/10.1016/j.chb.2023.107903>
- Bastian, A., Liza, L. O. y Efastri, S. M. (2023). Revolutionizing education: How digital literacy is transforming inclusive classrooms in post-COVID-19. *Journal of Public Health*, 45(3), e609-e610. <https://doi.org/10.1093/pubmed/fdado58>
- Blau, I., Shamir-Inbal, T. y Avdiel, O. (2020). How does the pedagogical design of a technology-enhanced collaborative academic course promote digital literacies, self-regulation, and perceived learning of students? *The Internet and Higher Education*, 45, 100722. <https://doi.org/10.1016/j.iheduc.2019.100722>
- Buchholz, B. A., DeHart, J. y Moorman, G. (2020). Digital citizenship during a global

- pandemic: Moving beyond digital literacy. *Journal of Adolescent & Adult Literacy*, 64(1), 11-17. <https://doi.org/10.1002/jaal.1076>
- Cameron, A. C., Gelbach, J. B. y Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics*, 90(3), 414-427. <https://doi.org/10.1162/rest.90.3.414>
- Dalgıç, A., Yaşar, E. y Demir, M. (2024). ChatGPT and learning outcomes in tourism education: The role of digital literacy and individualized learning. *Journal of Hospitality, Leisure, Sport & Tourism Education*, 34, 100481. <https://doi.org/10.1016/j.jhlste.2024.100481>
- Dimitriadou, E. y Lanitis, A. (2023). A critical evaluation, challenges, and future perspectives of using artificial intelligence and emerging technologies in smart classrooms. *Smart Learning Environments*, 10(1), Article 12. <https://doi.org/10.1186/s40561-023-00231-3>
- Dunning, D. (2011). The Dunning-Kruger effect: On being ignorant of one's own ignorance. En J. M. Olson y M. P. Zanna (Eds.), *Advances in Experimental Social Psychology* (Vol. 44, pp. 247-296). <https://doi.org/10.1016/B978-0-12-385522-0.00005-6>
- Clifford, I., Kluzer, S., Troia, S., Jakobsone, M., Zandbergs, U., Vuorikari, R., Punie, Y., Castaño-Muñoz, J., Mediavilla, C., O'Keeffe, W. y Cabrera, M. (2020). *DigCompSAT: A self-reflection tool for the European Digital Competence Framework for citizens* (JRC Technical Report). Publications Office of the European Union. <https://doi.org/10.2760/77437>
- Csapó, B. y Funke, J. (Eds.). (2017). *The nature of problem solving: Using research to inspire 21st century learning*. OECD Publishing. <https://doi.org/10.1787/9789264273955-en>
- Comisión Europea, Dirección General de Educación, Juventud, Deporte y Cultura. (2019). *Key competences for lifelong learning*. Publications Office of the European Union. <https://doi.org/10.2766/569540>
- Ferrari, A. y Punie, Y. (2013). *DIGCOMP: A framework for developing and understanding digital competence in Europe*. Publications Office of the European Union. <https://doi.org/10.2788/52966>
- García Peñalvo, F. J., Llorens-Largo, F. y Vidal, J. (2024). La nueva realidad de la educación ante los avances de la inteligencia artificial generativa. *RIED-Revista Iberoamericana de Educación a Distancia*, 27(1), 9-39. <https://doi.org/10.5944/ried.27.1.37716>
- Gallardo-Echenique, E. E., de Oliveira, J. M., Marqués-Molias, L., Esteve-Mon, F., Wang, Y. y Baker, R. (2015). Digital competence in the knowledge society. *MERLOT Journal of Online Learning and Teaching*, 11(1). https://jolt.merlot.org/vol11no1/Gallardo-Echenique_0315.pdf
- Greene, W. (2004). The behaviour of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects. *The Econometrics Journal*, 7(1), 98-119. <https://doi.org/10.1111/j.1368-423X.2004.00123.x>
- Güner, H. y Er, E. (2025). AI in the classroom: Exploring students' interaction with ChatGPT in programming learning. *Education and Information Technologies*, 30, 12681-12707. <https://doi.org/10.1007/s10639-025-13337-7>
- Gutiérrez Porlán, I. y Serrano Sánchez, J. L. (2016). Evaluation and development of digital competence in future primary school teachers at the University of Murcia. *Journal of New Approaches in Educational Research*, 5(1), 51-56. <https://doi.org/10.7821/naer.2016.1.152>

- Hair, J. F., Jr., Black, W. C., Babin, B. J. y Anderson, R. E. (2010). *Multivariate data analysis* (7ª ed.). Prentice Hall.
- Hillmayr, D., Ziernwald, L., Reinhold, F., Hofer, S. I. y Reiss, K. M. (2020). The potential of digital tools to enhance mathematics and science learning in secondary schools: A context-specific meta-analysis. *Computers & Education*, 153, 103897. <https://doi.org/10.1016/j.compedu.2020.103897>
- Hwang, G. J., Xie, H., Wah, B. W. y Gašević, D. (2020). Vision, challenges, roles and research issues of artificial intelligence in education. *Computers and Education: Artificial Intelligence*, 1, 100001. <https://doi.org/10.1016/j.caeai.2020.100001>
- Inamorato dos Santos, A., Chinkes, E., Carvalho, M. A. G., Solórzano, C. M. V. y Marroni, L. S. (2023). The digital competence of academics in higher education: Is the glass half empty or half full? *International Journal of Educational Technology in Higher Education*, 20, Article 9. <https://doi.org/10.1186/s41239-022-00376-0>
- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Kruger, J. y Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6), 1121-1134. <https://doi.org/10.1037/0022-3514.77.6.1121>
- Laanpere, M. (2019). *Recommendations on assessment tools for monitoring digital literacy within UNESCO's Digital Literacy Global Framework* (Information Paper No. 56; UIS/2019/LO/IP/56). UNESCO Institute for Statistics. <https://unesdoc.unesco.org/ark:/48223/pf0000366740>
- Laupichler, M. C., Aster, A., Schirch, J. y Raupach, T. (2022). Artificial intelligence literacy in higher and adult education: A scoping literature review. *Computers and Education: Artificial Intelligence*, 3, 100101. <https://doi.org/10.1016/j.caeai.2022.100101>
- Lee, D. y Palmer, E. (2025). Prompt engineering in higher education: A systematic review to help inform curricula. *International Journal of Educational Technology in Higher Education*, 22, Article 7. <https://doi.org/10.1186/s41239-025-00503-7>
- Lee, D., Arnold, M., Srivastava, A., Plastow, K., Strelan, P., Ploechl, F., Lekkas, D. y Palmer, E. (2024). The impact of generative AI on higher education learning and teaching: A study of educators' perspectives. *Computers and Education: Artificial Intelligence*, 6, 100221. <https://doi.org/10.1016/j.caeai.2024.100221>
- Li, Y. y Ranieri, M. (2010). Are 'digital natives' really digitally competent?—A study on Chinese teenagers. *British Journal of Educational Technology*, 41(6), 1029-1042. <https://doi.org/10.1111/j.1467-8535.2009.01053.x>
- Lin, C. J., Lee, H. Y., Wang, W. S., Huang, Y. M. y Wu, T. T. (2025). Enhancing reflective thinking in STEM education through experiential learning: The role of generative AI as a learning aid. *Education and Information Technologies*, 30(5), 6315-6337. <https://doi.org/10.1007/s10639-024-13072-5>
- List, A. (2019). Defining digital literacy development: An examination of pre-service teachers' beliefs. *Computers & Education*, 138, 146-158. <https://doi.org/10.1016/j.compedu.2019.03.009>

- Mattar, J., Ramos, D. K. y Lucas, M. R. (2022). DigComp-based digital competence assessment tools: Literature review and instrument analysis. *Education and Information Technologies*, 27(8), 10843-10867. <https://doi.org/10.1007/s10639-022-11034-3>
- Moravec, V., Hynek, N., Gavurová, B. y Rigelský, M. (2024). Who uses it and for what purpose? The role of digital literacy in ChatGPT adoption and utilisation. *Journal of Innovation & Knowledge*, 9(4), 100602. <https://doi.org/10.1016/j.jik.2024.100602>
- Morgan, A., Sibson, R. y Jackson, D. (2022). Digital demand and digital deficit: Conceptualising digital literacy and gauging proficiency among higher education students. *Journal of Higher Education Policy and Management*, 44(3), 258-275. <https://doi.org/10.1080/1360080X.2022.2030275>
- Naamati-Schneider, L. y Alt, D. (2024). Beyond digital literacy: The era of AI-powered assistants and evolving user skills. *Education and Information Technologies*, 29(16), 21263-21293. <https://doi.org/10.1007/s10639-024-12694-z>
- Ng, W. (2012). Can we teach digital natives digital literacy? *Computers & Education*, 59(3), 1065-1078. <https://doi.org/10.1016/j.compedu.2012.04.016>
- OECD. (2013). *PISA 2012 assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy*. OECD Publishing. <https://doi.org/10.1787/9789264190511-en>
- (2023). *OECD Digital Education Outlook 2023: Towards an effective digital education ecosystem*. OECD Publishing. <https://doi.org/10.1787/c74f03de-en>
- Polizzi, G. (2020). Digital literacy and the national curriculum for England: Learning from how the experts engage with and evaluate online content. *Computers & Education*, 152, 103859. <https://doi.org/10.1016/j.compedu.2020.103859>
- Popenici, S. A. D. y Kerr, S. (2017). Exploring the impact of artificial intelligence on teaching and learning in higher education. *Research and Practice in Technology Enhanced Learning*, 12(1), 22. <https://doi.org/10.1186/s41039-017-0062-8>
- Prensky, M. (2001). Digital natives, digital immigrants. *On the Horizon*, 9(5), 1-6. <https://doi.org/10.1108/10748120110424816>
- Prior, D. D., Mazanov, J., Meacheam, D., Heaslip, G. y Hanson, J. (2016). Attitude, digital literacy and self-efficacy: Flow-on effects for online learning behavior. *The Internet and Higher Education*, 29, 91-97. <https://doi.org/10.1016/j.iheduc.2016.01.001>
- Rahman, M. M. (2019). 21st century skill “problem solving”: Defining the concept. *Asian Journal of Interdisciplinary Research*, 2(1), 64-74. <https://doi.org/10.34256/ajir1917>
- Saklaki, A. y Gardikiotis, A. (2024). Exploring Greek students’ attitudes toward artificial intelligence: Relationships with AI ethics, media, and digital literacy. *Societies*, 14(12), 248. <https://doi.org/10.3390/soc14120248>
- Selwyn, N. (2009). The digital native—myth and reality. *Aslib Proceedings: New Information Perspectives*, 61(4), 364-379. <https://doi.org/10.1108/00012530910973776>
- Ting, Y.-L. (2015). Tapping into students’ digital literacy and designing negotiated learning to promote learner autonomy. *The Internet and Higher Education*, 26, 25-32. <https://doi.org/10.1016/j.iheduc.2015.04.004>
- Tomczyk, Ł. (2020). Skills in the area of digital safety as a key component of digital literacy among teachers. *Education and Information Technologies*, 25(1), 471-486. <https://doi.org/10.1007/s10639-019-09980-6>

- Van Laar, E., van Deursen, A. J. A. M., van Dijk, J. A. G. M. y de Haan, J. (2020). Determinants of 21st-century skills and 21st-century digital skills for workers: A systematic literature review. *Sage Open*, 10(1), 1-14. <https://doi.org/10.1177/2158244019900176>
- Vuorikari, R., Kluzer, S. y Punie, Y. (2022). *DigComp 2.2: The digital competence framework for citizens – with new examples of knowledge, skills and attitudes* [EUR 31006 EN; JRC128415]. Publications Office of the European Union. <https://doi.org/10.2760/115376>
- Vuorikari, R., Pokropek, A. y Castaño Muñoz, J. (2025). Enhancing digital skills assessment: Introducing compact tools for measuring digital competence. *Technology, Knowledge and Learning*. Advance online publication. <https://doi.org/10.1007/s10758-025-09825-x>
- Walstad, W. B., Rebeck, K. y Butters, R. B. (2013). *Test of Economic Literacy: Examiner's manual* (4^a ed.). Council for Economic Education. <https://www.econedlink.org/wp-content/uploads/2018/09/TEL-Manual.pdf>
- Walstad, W. B., Watts, M. y Rebeck, K. (2007). *Test of Understanding of College Economics: Examiner's manual* (4^a ed.). Council for Economic Education. <https://www.econedlink.org/wp-content/uploads/2018/09/TUCE-4th.pdf>
- Wang, J. y Fan, W. (2025). The effect of ChatGPT on students' learning performance, learning perception, and higher-order thinking: Insights from a meta-analysis. *Humanities and Social Sciences Communications*, 12, 621. <https://doi.org/10.1057/s41599-025-04787-y>
- Wang, W.-S., Lin, C.-J., Lee, H.-Y., Huang, Y.-M. y Wu, T.-T. (2025). Integrating feedback mechanisms and ChatGPT for VR-based experiential learning: Impacts on reflective thinking and AIoT physical hands-on tasks. *Interactive Learning Environments*, 33(2), 1770-1787. <https://doi.org/10.1080/10494820.2024.2375644>
- Webb, M. D. (2014). *Reworking wild bootstrap based inference for clustered errors* (QED Working Paper No. 1315). Queen's University, Department of Economics. https://qed.econ.queensu.ca/working_papers/papers/qed_wp_1315.pdf
- Zhai, X., Chu, X., Chai, C. S., Jong, M. S. Y., Istenič, A., Spector, J. M., Liu, J., Yuan, J. y Li, Y. (2021). A review of artificial intelligence (AI) in education from 2010 to 2020. *Complexity*, 2021, 8812542. <https://doi.org/10.1155/2021/8812542>
- Zhao, Y., Pinto Llorente, A. M. y Sánchez Gómez, M. C. (2021). Digital competence in higher education research: A systematic literature review. *Computers & Education*, 168, 104212. <https://doi.org/10.1016/j.compedu.2021.104212>

Fecha de recepción del artículo: 1 de junio de 2025

Fecha de aceptación del artículo: 9 de septiembre de 2025

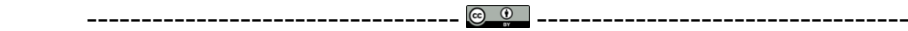
Fecha de aprobación para maquetación: 6 de octubre de 2025

Fecha de publicación en OnlineFirst: 24 de octubre de 2025

Fecha de publicación: 1 de enero de 2026

AI-ED-SAT: diseño y validación de un cuestionario para autoevaluar competencias docentes en IA educativa

AI-ED-SAT: design and validation of a questionnaire for self-assessment of teaching skills in educational AI



Andresa Sartor-Harada - *Universidad Internacional de La Rioja, UNIR (España)*
 Juliana Azevedo Gomes - *Universidad Internacional de La Rioja, UNIR (España)*

RESUMEN

La integración acelerada de la Inteligencia Artificial (IA) en el ámbito educativo plantea nuevos retos para la formación y evaluación de competencias docentes. Este estudio presenta el diseño, validación y análisis psicométrico del cuestionario AI-ED-SAT, una herramienta de autoevaluación dirigida a docentes para diagnosticar su grado de preparación en el uso pedagógico, ético y curricular de la IA. Desarrollado bajo un enfoque de investigación basado en diseño (Design-Based Research), el instrumento se estructuró en torno a cuatro dimensiones alineadas con los actuales marcos de competencias en IA de la UNESCO (2024a, 2024b): comprensión conceptual de la IA, uso pedagógico de herramientas inteligentes, reflexión ética y crítica e integración curricular. Su construcción se apoyó en una revisión teórica exhaustiva y un proceso iterativo de validación mediante el método Delphi con 12 expertos en tecnología educativa, IA y formación docente. Posteriormente, en una prueba piloto a una muestra de 128 docentes de distintos niveles educativos, se analizó su fiabilidad (α de Cronbach = 0.93), validez de contenido y validez de constructo. Los resultados del análisis factorial exploratorio (AFE) confirmaron la agrupación de los ítems en cuatro factores teóricos, mientras que el análisis factorial confirmatorio (AFC) mostró índices de ajuste excelentes (CFI = 0.96; RMSEA = 0.045). El AI-ED-SAT se configura como una herramienta robusta y actualizada, útil tanto en la investigación como en programas de formación docente. Su enfoque autorreflexivo contribuye a fortalecer la alfabetización crítica y la agencia profesional del profesorado frente a los desafíos de la IA en la educación.

Palabras clave: inteligencia artificial; educación; competencias docentes; cuestionario; validación de instrumento.

ABSTRACT

The accelerated integration of Artificial Intelligence (AI) in education poses new challenges for teacher training and assessment. This study presents the design, validation, and psychometric analysis of the AI-ED-SAT questionnaire, a self-assessment tool designed for teachers to diagnose their level of preparedness in the pedagogical, ethical, and curricular use of AI. The AI-ED-SAT questionnaire was validated within the Design-Based Research approach. It comprises four dimensions aligned with UNESCO's current AI competency frameworks (2024a, 2024b): conceptual understanding of AI, pedagogical use of intelligent tools, ethical and critical reflection, and curricular integration. Its construction was based on an exhaustive theoretical review and an iterative validation process using the Delphi method with 12 experts in educational technology, AI, and teacher training. Subsequently, in a pilot test involving a sample of 128 teachers from various educational levels, the reliability (Cronbach's α = 0.93), content validity, and construct validity were analyzed. The results of the exploratory factor analysis (EFA) confirmed the grouping of items into four theoretical factors, while the confirmatory factor analysis (CFA) showed excellent fit indices (CFI = 0.96; RMSEA = 0.045). The AI-ED-SAT is a robust and up-to-date tool that is useful in both research and teacher training programs. Its self-reflective approach helps strengthen teachers' critical literacy and professional agency in the face of the challenges posed by AI in education.

Keywords: artificial intelligence; education; teacher competencies; questionnaire; instrument validation.

INTRODUCCIÓN

La irrupción de la Inteligencia Artificial (IA) en los ecosistemas educativos está transformando no solo las metodologías de enseñanza-aprendizaje, sino también los marcos de referencia éticos, pedagógicos y organizativos que rigen la práctica docente. Si bien los avances tecnológicos han potenciado el desarrollo de entornos virtuales adaptativos, sistemas de tutoría inteligente y procesos de evaluación automatizada, su implementación en contextos escolares exige una preparación docente específica, situada y reflexiva (Holmes et al., 2019; UNESCO, 2021; UNESCO, 2024b).

Más allá del acceso a dispositivos o plataformas digitales, se hace necesario que los profesionales de la educación desarrollen competencias críticas en IA, que les permitan comprender su funcionamiento, valorar sus implicaciones sociales y éticas, y diseñar prácticas pedagógicas coherentes con los principios de equidad, inclusión y sentido formativo. En este marco, las competencias docentes en IA deben ser entendidas como un conjunto multidimensional de conocimientos, habilidades y actitudes que van más allá de la alfabetización digital tradicional (Celik, 2023; Chen et al., 2020).

Sin embargo, la literatura evidencia una escasez de instrumentos específicos y validados que permitan evaluar estas competencias en relación con la IA de forma rigurosa, contextualizada y formativa. La mayoría de los modelos existentes se centran en indicadores genéricos de competencia digital, lo que limita su capacidad para abordar los desafíos particulares que plantea la integración de la IA en la educación (Zawacki-Richter et al., 2019). Este vacío dificulta tanto la investigación empírica como el diseño de programas de formación docente orientados a las nuevas demandas tecnológicas.

En respuesta a esta necesidad, el presente estudio presenta el diseño, validación y análisis psicométrico del AI-ED-SAT (Artificial Intelligence in Education – Self-Assessment Tool), un cuestionario de autoevaluación dirigido a docentes de distintos niveles educativos, con el propósito de diagnosticar su nivel de competencia en torno al uso de la IA en el ámbito escolar. El instrumento se estructura en torno a cuatro dimensiones clave –comprensión conceptual de la IA, uso pedagógico de herramientas, ética y consideraciones en el aula, e integración curricular– y se ha desarrollado mediante un enfoque metodológico mixto, combinando revisión teórica, validación por juicio de expertos (método Delphi) y análisis factorial (exploratorio y confirmatorio). Su objetivo es doble: ofrecer una herramienta rigurosa para la investigación educativa y, al mismo tiempo, servir como recurso formativo que promueva la reflexión profesional y la mejora continua del profesorado frente a los desafíos de la IA en la educación.

Aunque el cuestionario ha sido desarrollado íntegramente en castellano, se optó por mantener un acrónimo en inglés con el objetivo de facilitar su visibilidad y adopción en futuras investigaciones internacionales. Esta decisión responde a la creciente presencia del inglés como lengua vehicular en el campo de la IA

educativa, lo que puede favorecer su inclusión en bases de datos académicas, redes de colaboración global y procesos de traducción y adaptación transcultural.

Educación y tecnologías emergentes: el nuevo paradigma docente

El avance de las tecnologías emergentes, y en particular de la IA, está reconfigurando profundamente el escenario educativo contemporáneo. Diversas investigaciones recientes subrayan la necesidad urgente de redefinir las competencias docentes ante esta transformación (Celik, 2023; Holmes et al., 2019; UNESCO, 2023). Lejos de limitarse al dominio técnico, estas competencias comprenden una comprensión crítica del impacto de la tecnología en los procesos de enseñanza-aprendizaje, el uso ético de los datos, la garantía de inclusión y el diseño de experiencias personalizadas y significativas.

La transformación digital ha generado cambios estructurales en la educación, afectando tanto a las herramientas disponibles como a las dinámicas pedagógicas, los marcos normativos y los roles profesionales del cuerpo docente. En este contexto, tecnologías como la realidad aumentada, el big data, el blockchain o la propia IA no solo abren nuevas posibilidades didácticas, sino que imponen la necesidad de una reconfiguración profunda de las competencias profesionales del profesorado (Bo, 2024; UNESCO, 2024b).

Este nuevo paradigma exige al docente desempeñar funciones más complejas: diseñador de experiencias de aprendizaje, mediador crítico entre tecnología y currículo, y gestor ético del conocimiento digital. En consecuencia, la integración tecnológica en el aula no puede reducirse a un uso instrumental o técnico, sino que requiere un enfoque pedagógico contextualizado, capaz de valorar sus implicaciones sociales, éticas y educativas, y adaptarlas a las necesidades y trayectorias del estudiantado (Cabero-Almenara et al., 2020).

Diversas iniciativas internacionales, como el marco DigCompEdu (Ghomí y Redecker, 2019) o las directrices de la UNESCO para la competencia docente en TIC y actualmente, las específicas para el uso de IA, han establecido referentes para el desarrollo de competencias digitales en el profesorado. No obstante, dichos marcos suelen adoptar una perspectiva generalista, centrada en habilidades transversales y tecnologías ya consolidadas, sin abordar con suficiente profundidad los desafíos específicos que plantean las tecnologías emergentes, y en particular, la IA. Esta limitación evidencia la necesidad de avanzar hacia modelos que integren dimensiones específicas de apropiación crítica de estas tecnologías, orientados a fomentar nuevas alfabetizaciones digitales que sitúen al docente como agente activo de transformación dentro de los ecosistemas educativos contemporáneos.

Aunque DigCompEdu (Ghomí y Redecker, 2019) constituye un marco ampliamente reconocido para evaluar la competencia digital docente, su orientación generalista lo limita para abordar en profundidad las particularidades de la Inteligencia Artificial.

Este instrumento describe áreas competenciales transversales aplicables a diversas tecnologías digitales, sin diferenciar entre herramientas consolidadas y emergentes, ni integrar de forma explícita dimensiones éticas y curriculares específicas para la IA. En contraste, el AI-ED-SAT se focaliza exclusivamente en el uso educativo de la Inteligencia Artificial, incorporando indicadores vinculados a la comprensión conceptual de algoritmos, la aplicación pedagógica contextualizada, la reflexión ética y la integración curricular estratégica. De este modo, el AI-ED-SAT no pretende reemplazar a marcos como DigCompEdu, sino complementarlos, ofreciendo un nivel de detalle adaptado a los retos y oportunidades singulares que plantea la IA en los entornos educativos.

Inteligencia Artificial en la educación: potencial, riesgos y controversias

La incorporación de la Inteligencia Artificial en el ámbito educativo representa una de las transformaciones más significativas de los últimos años, con implicaciones directas en las prácticas docentes, los modelos pedagógicos y los procesos de enseñanza-aprendizaje. Su presencia se manifiesta en múltiples formas: sistemas de tutoría inteligente, plataformas adaptativas, análisis de aprendizaje, chatbots, motores de recomendación o herramientas de retroalimentación automatizada (Holmes et al., 2019; Zawacki-Richter et al., 2019).

Esta expansión tecnológica ofrece oportunidades relevantes para personalizar el aprendizaje, optimizar los procesos evaluativos y aliviar las tareas administrativas del profesorado. Sin embargo, también plantea desafíos sustanciales que no pueden ser ignorados. Entre ellos se encuentran los sesgos algorítmicos, la opacidad de los sistemas automatizados, la vulneración de la privacidad y el riesgo de delegar decisiones pedagógicas fundamentales a sistemas que operan con lógicas ajenas al juicio humano (Chesterman, 2021; Ferrante, 2021; Huang, 2023; Selwyn, 2019).

En este sentido, la IA no puede ser concebida únicamente como una herramienta funcional, sino como un fenómeno sociotécnico complejo, que implica decisiones políticas, pedagógicas y éticas (Flores Vivar y García Peñalvo, 2023). Su implementación educativa exige que el profesorado no solo adquiera habilidades operativas, sino que desarrolle también una alfabetización crítica que le permita comprender el funcionamiento de los algoritmos, identificar sus posibles sesgos y valorar sus implicaciones sociales y educativas (García Peñalvo et al., 2024; Lane et al., 2023; Luckin, 2017).

Un referente ético reciente y relevante es el *Safe AI Education Manifesto* (Alier Forment et al., 2024), que establece recomendaciones claras para el uso responsable de la IA en contextos educativos. Entre sus principios fundamentales se destacan la transparencia en los sistemas, la rendición de cuentas, la protección de la privacidad estudiantil, la promoción de la equidad y el fomento de una alfabetización crítica frente a las tecnologías algorítmicas. Estos principios refuerzan la dimensión ética del AI-ED-SAT, especialmente en los ítems relacionados con la identificación de

sesgos, la toma de decisiones automatizada y los impactos educativos, contribuyendo a consolidar el instrumento como una herramienta integradora de principios éticos actualizados y relevantes para la práctica docente contemporánea.

A pesar de este panorama, la formación docente en IA continúa siendo incipiente. Las propuestas existentes suelen centrarse en aspectos instrumentales, sin incorporar una mirada integral que contemple tanto su dimensión pedagógica como ética (Delgado et al., 2024; Holmes y Tuomi, 2022). Ante este vacío, resulta fundamental diseñar instrumentos válidos y fiables que permitan evaluar no solo el grado de familiarización del profesorado con la IA, sino también su capacidad para integrarla de forma crítica, contextualizada y pedagógicamente significativa. Para ello, es necesario articular procesos metodológicos rigurosos, que combinen juicio experto con análisis estadísticos robustos, como proponen Sampieri (2018) y Muñiz y Fonseca-Pedrero (2019).

Competencias docentes en IA: más allá de la alfabetización digital

La irrupción de la Inteligencia Artificial (IA) en los entornos educativos ha consolidado su posición como tecnología disruptiva, con aplicaciones que abarcan desde la personalización del aprendizaje hasta la automatización de tareas evaluativas y administrativas (Holmes et al., 2019; Owan et al., 2023). Su implementación plantea retos importantes que exigen al profesorado una comprensión profunda de sus fundamentos técnicos y una reflexión crítica sobre los riesgos asociados, como los sesgos algorítmicos, la opacidad en la toma de decisiones, la dependencia tecnológica o las desigualdades en el acceso (Celik, 2023; Sharma et al., 2024).

En este contexto, el desarrollo de competencias específicas en IA se presenta como una prioridad estratégica para los sistemas educativos. A diferencia de las competencias digitales genéricas —centradas en la gestión de recursos digitales o entornos virtuales—, las competencias docentes en IA requieren una comprensión multidimensional que integre aspectos técnicos, éticos, pedagógicos y curriculares. Diversos documentos conceptuales han comenzado a elaborar este nuevo perfil competencial, entre ellos, los marcos de competencias en IA para docentes y estudiantes de la UNESCO (2024a, 2024b), el documento *AI and the Future of Teaching* de la UNESCO (2021), la actualización del marco DigCompEdu al contexto de IA (Caena y Redecker, 2019), y los estudios de Chen et al. (2020), Chiu et al. (2024) y Ng et al. (2023).

Khreisat et al. (2024) proponen una estructuración en cuatro ámbitos interrelacionados:

1. Comprensión de los principios y funcionamiento de la IA.
2. Capacidad de utilizar herramientas de IA con fines pedagógicos.
3. Conocimiento crítico sobre sus implicaciones éticas y sociales.
4. Integración curricular desde un enfoque interdisciplinario.

Esta propuesta refleja una concepción ampliada del rol docente, que trasciende el uso instrumental de la tecnología. En lugar de limitarse a operar herramientas, el docente debe posicionarse como un agente reflexivo y crítico, capaz de evaluar el sentido pedagógico, las consecuencias sociales y la pertinencia curricular de la IA en sus prácticas educativas. Tal como advierten Holmes et al. (2019), Zawacki-Richter et al. (2019) y Ayuso del Puerto y Gutiérrez Esteban (2022), muchos marcos existentes aún se mantienen en un plano descriptivo o eminentemente técnico, lo que limita su capacidad para orientar un desarrollo profesional auténticamente transformador.

Desde esta perspectiva, se hace necesaria la creación de instrumentos de evaluación que respondan a esta complejidad. Dichos instrumentos deben ir más allá de la medición del conocimiento declarativo, e incluir dimensiones como la competencia pedagógica, la disposición crítica y la integración curricular. El instrumento AI-ED-SAT, desarrollado en el marco de este estudio, se inscribe en esta lógica, al estructurarse en torno a cuatro dimensiones clave: comprensión conceptual, aplicación pedagógica, dimensión ética e integración curricular. Esta aproximación busca captar la complejidad del desarrollo profesional docente en la era de la IA, reconociendo la necesidad de una alfabetización crítica que permita a los educadores posicionarse activamente frente a los desafíos de las tecnologías emergentes.

Evaluación de competencias docentes: construcción y validación de instrumentos

En un contexto de transformación educativa impulsado por tecnologías emergentes, contar con instrumentos válidos y fiables para diagnosticar el nivel de competencia docente frente a la Inteligencia Artificial se vuelve crucial. La autoevaluación constituye, en este sentido, una estrategia formativa y reflexiva que permite identificar fortalezas y áreas de mejora desde una perspectiva situada (Panadero et al., 2019; Rachbauer et al., 2025).

Si bien existen numerosos instrumentos para medir la competencia digital docente (Cabero-Almenara et al., 2020; Ghomi y Redecker, 2019), la mayoría se enfoca en habilidades técnicas genéricas, sin abordar de manera específica las complejidades pedagógicas, éticas y curriculares que plantea la IA. Esta limitación ha generado una necesidad urgente de modelos evaluativos que capten no solo la familiaridad tecnológica del profesorado, sino también su capacidad de integrar la IA de forma crítica, contextualizada y pedagógicamente significativa (Celik, 2023; Delgado et al., 2024).

La construcción de instrumentos con estas características requiere un enfoque metodológico riguroso que combine criterios teóricos, empíricos y prácticos. En los últimos años, se ha consolidado una metodología mixta que incluye la revisión sistemática de literatura para la elaboración de ítems, la validación mediante juicio de expertos (como el método Delphi) y el análisis psicométrico (validez de contenido

y constructo, fiabilidad interna) (Diefes-Dux et al., 2010; Muñiz y Fonseca-Pedrero, 2019; Sampieri, 2018). En este proceso, los enfoques participativos e interactivos, como la Design-Based Research (DBR), han demostrado ser especialmente útiles para garantizar la pertinencia semántica y pedagógica del instrumento.

La autoevaluación, además, se ha consolidado como una herramienta eficaz para fomentar la reflexión profesional, ofrecer datos para el diseño de planes formativos personalizados y promover la agencia docente (Borge et al., 2005). A diferencia de los modelos externos de evaluación, sitúa al profesorado como protagonista del diagnóstico y la mejora.

En este marco se inscribe el instrumento AI-ED-SAT, concebido para evaluar de manera integral las competencias docentes relacionadas con la IA en contextos educativos. Su desarrollo parte de una revisión crítica de la literatura especializada y de marcos de referencia recientes que abordan las implicaciones pedagógicas, éticas y curriculares de la IA (Celik, 2023; Holmes et al., 2019; Zawacki-Richter et al., 2019).

A partir de este análisis, se identificaron cuatro dimensiones clave que configuran un perfil competencial orientado a una apropiación crítica y formativa de la IA:

1. **Comprensión conceptual de la IA**
Incluye conocimientos básicos sobre algoritmos, aprendizaje automático y sistemas adaptativos, con foco en sus aplicaciones educativas. Esta dimensión busca situar al docente como un actor informado frente a los discursos tecnológicos dominantes (Holmes et al., 2019).
2. **Uso pedagógico de herramientas de IA**
Evalúa la capacidad de aplicar herramientas de IA con propósitos didácticos claros, considerando el nivel de los estudiantes y el contexto educativo. Incluye diseño de actividades, personalización del aprendizaje y evaluación automatizada, superando un uso meramente técnico (Chen et al., 2020).
3. **Ética y consideraciones críticas sobre IA**
Abarca la identificación de riesgos como sesgos algorítmicos, problemas de privacidad o automatización de decisiones. Esta dimensión promueve una alfabetización crítica, posicionando al docente como mediador ético (Celik, 2023).
4. **Integración curricular de la IA**
Mide la capacidad de incorporar la IA en la planificación didáctica y curricular, mediante proyectos interdisciplinarios, metodologías activas y adaptación a nuevos escenarios tecnológicos (Rachbauer et al., 2025).

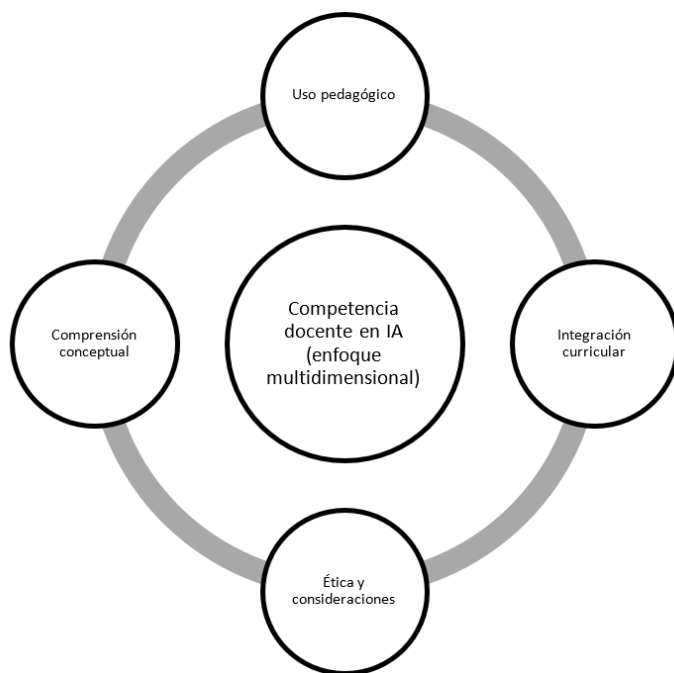
Cabe señalar que, si bien algunas competencias pueden parecer solaparse —en particular entre el uso pedagógico de herramientas de IA y su integración curricular—, ambas se distinguen por su nivel de aplicación: mientras la segunda dimensión se refiere al uso didáctico en el aula (microplanificación), la cuarta

se enfoca en la incorporación estratégica de la IA en la planificación curricular general (macroplanificación), incluyendo enfoques interdisciplinarios y alineación institucional. Esta diferenciación permite evitar ambigüedades y captar distintos niveles de apropiación docente.

Estas dimensiones se articulan de manera complementaria, permitiendo una evaluación holística del rol docente en el ecosistema digital. A diferencia de modelos fragmentarios o excesivamente técnicos, el AI-ED-SAT propone un marco integrado donde la acción pedagógica, el conocimiento conceptual, la visión curricular y el compromiso ético se entrelazan en un mismo dispositivo evaluativo.

Figura 1

Esquema del instrumento



La Figura 1 presenta una representación esquemática del modelo AI-ED-SAT, destacando la interrelación entre sus dimensiones y su fundamento teórico. El orden gráfico de las dimensiones: Uso Pedagógico, Comprensión conceptual, Integración curricular y Ética y consideraciones no implica jerarquía, sino que refleja una disposición circular y no lineal que enfatiza su interdependencia. Las dimensiones del modelo no siguen un orden de adquisición predeterminado, sino que configuran un marco multidimensional que puede desarrollarse de forma

paralela, complementaria o combinada, según el perfil del docente, su experiencia previa y el contexto educativo en el que se desempeñe.

En suma, la incorporación de la IA en el ámbito educativo exige una comprensión integral por parte del profesorado, que combine el dominio conceptual con una apropiación crítica, ética y pedagógica. Pese a los avances en la definición teórica de estas competencias, sigue existiendo una carencia de instrumentos específicos, válidos y fiables que permitan diagnosticar dichas competencias de manera rigurosa y contextualizada.

El AI-ED-SAT se concibe como una respuesta a esta necesidad. Su diseño metodológico, basado en la combinación de enfoques cualitativos y cuantitativos, incluyendo Design-Based Research, juicio de expertos y análisis factorial, busca asegurar la solidez teórica y psicométrica del instrumento, así como su utilidad formativa en la mejora de la práctica docente.

METODOLOGÍA

Diseño metodológico del instrumento

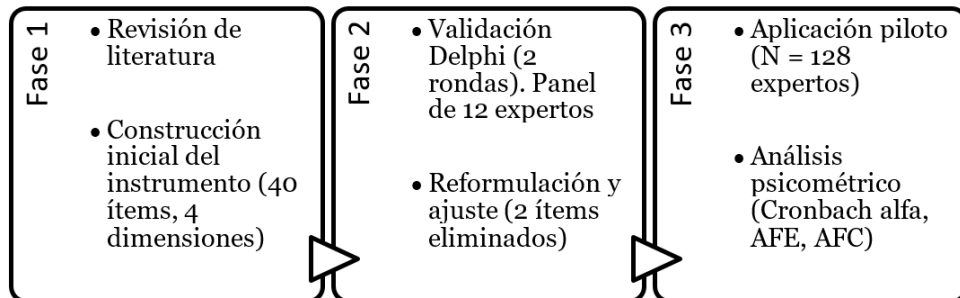
Este estudio se enmarca en un enfoque cuantitativo de alcance descriptivo y correlacional, orientado al diseño, validación y análisis psicométrico de un instrumento de autoevaluación docente sobre el uso de la Inteligencia Artificial (IA) en educación. Metodológicamente, se integraron los principios del Design-Based Research (DBR) (Brown, 1992) con las orientaciones actuales para la validación de instrumentos educativos (Muñiz y Fonseca-Pedrero, 2019; Sampieri, 2018).

El proceso se estructuró en tres fases principales:

1. Construcción del cuestionario, a partir de una revisión sistemática de literatura reciente.
2. Validación de contenido mediante juicio de expertos, utilizando el método Delphi.
3. Aplicación piloto y análisis psicométrico del instrumento.

Figura 2

Etapas del proceso de elaboración y validación



Se llevó a cabo una prueba piloto a modo de intervención, en la que se aplicó el instrumento a individuos con características similares a la muestra objetivo. La finalidad de esta prueba fue ajustar aquellos ítems que pudieran requerir modificaciones en términos de operatividad, redacción u otros aspectos, así como optimizar la apariencia del instrumento (Hernández-Sampieri et al., 2006).

La revisión bibliográfica realizada sobre las competencias docentes en el uso de la Inteligencia Artificial (IA) en educación condujo a la configuración de la primera estructura del cuestionario de autoevaluación en torno a dos dimensiones fundamentales que cubren el espectro de análisis de la práctica docente: (1) la integración pedagógica de la IA en el aula y (2) la competencia docente en IA y su impacto en la enseñanza.

El primer borrador se construyó siguiendo un proceso fundamentado en la metodología basada en el diseño (DBR, por sus siglas en inglés) descrita por Brown (1992). El proceso se estructuró en tres fases: en la primera, se llevó a cabo una revisión de la literatura sobre IA en educación y competencias docentes en entornos digitales; en la segunda, se definieron las dimensiones y se elaboraron los ítems que compondrían el cuestionario; y en la tercera, se validó el instrumento mediante el método Delphi con un panel de expertos y el aprendizaje (Chen et al., 2020; Holmes et al., 2019; Luckin, 2017; UNESCO, 2024b).

El instrumento, denominado AI-ED-SAT (Artificial Intelligence in Education – Self-Assessment Tool), fue diseñado con el propósito de evaluar el nivel de competencia docente en cuatro dimensiones clave:

1. Conocimientos generales sobre IA y su aplicación educativa
2. Uso pedagógico de herramientas de IA
3. Ética y consideraciones sobre IA en el aula
4. Integración curricular de la IA

Se elaboraron inicialmente 40 ítems, redactados como afirmaciones, y distribuidos equitativamente en las cuatro dimensiones (10 ítems por dimensión). Las respuestas se estructuraron en una escala de Likert de 5 puntos:

1 = Muy en desacuerdo | 2 = En desacuerdo | 3 = Ni de acuerdo ni en desacuerdo
| 4 = De acuerdo | 5 = Muy de acuerdo

La decisión de incluir un número equilibrado de 10 ítems por dimensión respondió a criterios de simetría estructural y comparabilidad entre dimensiones, buscando asegurar un análisis psicométrico equilibrado. Si bien la naturaleza de las dimensiones podría haber permitido una extensión o reducción de ítems, se optó por una distribución equitativa inicial como punto de partida para la validación, en línea con prácticas comunes en instrumentos similares (Muñiz y Fonseca-Pedrero, 2019). Posteriormente, el análisis de validez permitió refinar el instrumento, reduciendo el total a 38 ítems en la versión final.

Esta escala permite captar niveles diferenciados de autoevaluación sin sobrecargar la toma de decisiones, y es adecuada para análisis factoriales. La validación del instrumento buscaba garantizar que los indicadores reflejaran adecuadamente la integración de la IA en la educación y las competencias docentes necesarias para su implementación efectiva.

La versión final del cuestionario, compuesta por 38 ítems distribuidos en cuatro dimensiones, puede consultarse íntegramente en el Anexo 1. La inclusión del instrumento completo permite ofrecer una visión detallada de los indicadores empleados para evaluar la competencia docente en IA y favorece su posible replicabilidad en estudios posteriores o programas de formación.

Validación de contenido del instrumento: el método Delphi

La segunda fase del estudio tuvo como objetivo validar el contenido del borrador preliminar del cuestionario previamente elaborado. Para ello, se empleó el método Delphi, ampliamente reconocido en el ámbito de las Ciencias de la Educación por su capacidad para clarificar problemáticas complejas mediante un proceso estructurado de consulta entre expertos organizados en un panel (Reguant-Álvarez y Torrado-Fonseca, 2016). Esta técnica, de naturaleza reiterada, promueve el consenso informado a través de rondas sucesivas de evaluación, permitiendo una convergencia progresiva de opiniones fundamentadas (Linstone y Turoff, 2002; Reguant-Álvarez y Torrado-Fonseca, 2016).

Los instrumentos utilizados para la recolección de datos consistieron en cuestionarios diseñados para valorar dimensiones e ítems específicos, incluyendo apartados para observaciones cualitativas. El proceso de recopilación se organizó con

una temporalidad flexible, aunque delimitada, con el fin de facilitar la participación efectiva de un panel de expertos distribuidos geográficamente y con disponibilidad horaria diversa.

En esta etapa, se procedió a la conformación del panel de expertos, priorizando la representatividad y diversidad profesional de los participantes por encima del número absoluto. Para ello, se definieron criterios de selección orientados a garantizar un equilibrio entre el conocimiento actualizado sobre la integración de la Inteligencia Artificial (IA) en contextos educativos y las competencias docentes requeridas para su aplicación efectiva.

Si bien la composición del panel incluyó una mayor proporción de formadores universitarios, esta elección respondió a la necesidad de contar con expertos con una visión sistemática del desarrollo docente y experiencia en formación continua, clave para evaluar competencias transversales como las vinculadas a la IA. No obstante, se aseguró la presencia de profesionales de primaria y secundaria con experiencia directa en la aplicación de tecnologías de IA en el aula, así como de especialistas acreditados en inteligencia artificial con producción científica y participación en proyectos de innovación educativa.

A partir de la estructura preliminar del instrumento, se estableció como criterio central la inclusión de profesionales provenientes de tres perfiles clave: (1) docentes de distintos niveles educativos con experiencia en el uso pedagógico de tecnologías emergentes; (2) especialistas en IA aplicada a la educación; y (3) formadores de docentes en el ámbito de la tecnología educativa. Esta estrategia permitió integrar una perspectiva multidisciplinar coherente con los objetivos del estudio.

Si bien no existe un consenso normativo sobre el tamaño ideal de los paneles Delphi, diversos autores sugieren orientaciones prácticas. En general, se considera que un mínimo de diez participantes asegura la estabilidad del consenso, dado que muestras inferiores a siete pueden comprometer la representatividad (Reguant-Álvarez y Torrado-Fonseca, 2016). En consonancia con esta premisa, Skulmoski et al. (2007) argumentan que, para paneles relativamente homogéneos, una muestra entre diez y quince expertos puede resultar metodológicamente sólida.

Se invitó a participar a doce expertos que cumplieran los criterios establecidos. La invitación incluía una caracterización detallada de su perfil profesional, información sobre sus funciones docentes e investigadoras, así como una autoevaluación de su grado de adecuación en relación con el propósito del estudio.

Todos los expertos aceptaron la invitación, conformando un panel con un alto nivel de especialización. El 75 % del grupo contaba con título de doctorado, e incluía siete formadores universitarios en programas de formación docente, tres docentes de educación básica (primaria y secundaria) con experiencia directa en el uso de IA en el aula, y dos especialistas con trayectoria combinada en investigación académica y capacitación docente en IA educativa. Ambos expertos cuentan con publicaciones indexadas sobre IA en educación, han participado como asesores en

proyectos de integración tecnológica en instituciones escolares y colaboran en redes internacionales sobre innovación educativa y ética digital.

Además, se procuró una distribución equilibrada por género en la composición del panel, que incluyó 7 mujeres (58 %) y 5 varones (42 %), en línea con los criterios de representatividad y diversidad establecidos en el diseño metodológico.

Entre los criterios comunes de inclusión se encuentran:

- Experiencia mínima de cinco años en tecnología educativa, formación docente o investigación aplicada sobre IA.
- Participación en iniciativas de innovación o proyectos de formación en competencias digitales.
- Representación de distintos niveles del sistema educativo (educación básica y superior)

Tabla 1

Composición del panel de expertos Delphi

Categoría	Cantidad
Total de expertos	12
Con grado de doctorado	9
Formadores universitarios	7
Docentes primaria y secundaria	3
Especialistas en IA y formación docente	2

Las investigadoras responsables supervisaron la recogida de datos durante la fase de validación, llevada a cabo mediante un proceso iterativo de consulta a expertos utilizando el correo electrónico como canal de comunicación principal. El procedimiento se estructuró en dos rondas sucesivas, número considerado adecuado para facilitar la convergencia de opiniones y alcanzar un consenso metodológicamente sólido (Linstone y Turoff, 2002). Desde el inicio, se informó a los participantes sobre esta estructura y el grado de implicación requerido, con el objetivo de garantizar transparencia y compromiso por parte del panel (Reguant-Álvarez y Torrado-Fonseca, 2016).

La primera ronda tuvo como propósito principal la evaluación cuantitativa y cualitativa de los ítems del instrumento, centrando el análisis en dos dimensiones fundamentales: relevancia conceptual y claridad lingüística. A nivel cuantitativo, se solicitó a los expertos que valoraran cada ítem mediante una escala tipo Likert de cinco puntos, donde 1 representaba “nada relevante” o “poco claro”, y 5 indicaba “muy relevante” o “muy claro”. Paralelamente, se habilitaron espacios para comentarios abiertos, lo que permitió recoger sugerencias cualitativas orientadas a mejorar la

redacción, precisión o adecuación de los ítems. Este enfoque mixto, sustentado en trabajos previos (Gallant y Luthy, 2020; Grand-Guillaume-Perrenoud et al., 2023; Paulin et al., 2024; Ramírez, 2019; Ramírez-Montoya y Lugo-Ocando, 2020), favorece el enriquecimiento del instrumento desde una perspectiva interpretativa.

A partir del análisis de esta primera fase, se procedió a reformular seis ítems con base en las observaciones cualitativas y se eliminaron dos por considerarse redundantes o conceptualmente débiles. Estos ajustes condujeron al diseño del formulario para la segunda ronda, centrada en la reevaluación de los ítems modificados. Nuevamente, se empleó una escala tipo Likert de cinco puntos para valorar su pertinencia y claridad tras las modificaciones. El objetivo de esta segunda iteración fue confirmar la validez de los cambios introducidos y consolidar una versión final que reflejara el consenso del panel. Los resultados obtenidos permitieron cerrar el proceso con una estructura definitiva compuesta por 38 ítems, distribuidos en cuatro dimensiones teóricas.

El análisis de los resultados de ambas rondas se realizó mediante una combinación de técnicas cuantitativas:

- Estadística descriptiva: cálculo de media, desviación típica y percentiles para cada ítem.
- Coeficiente de concordancia de Kendall (W): utilizado para medir el grado de acuerdo entre los expertos.
- Criterio de aceptación basado en el percentil 80: se estableció como umbral mínimo para considerar aceptados los ítems, de acuerdo con Mauri et al. (2007).

Este enfoque analítico permitió identificar con precisión los ítems que alcanzaron altos niveles de consenso, así como aquellos que requerían ajustes o eliminación.

Aplicación piloto y análisis psicométrico

Una vez finalizado el proceso de validación por juicio de expertos, la versión revisada del instrumento fue sometida a una aplicación piloto con una muestra de 128 docentes procedentes de diversos contextos educativos en España y América Latina. La selección fue intencional, con criterios de accesibilidad y diversidad institucional.

Perfil de la muestra:

- Educación primaria: 34 %
- Educación secundaria: 31 %
- Educación superior: 35 %
- Promedio de experiencia docente: 9,6 años
- Distribución por género: 69 % mujeres, 31 % varones

La aplicación se llevó a cabo en modalidad *online*. La muestra estuvo compuesta por docentes procedentes de siete países de habla hispana: España ($n = 49$), México ($n = 27$), Argentina ($n = 18$), Colombia ($n = 12$), Chile ($n = 9$), Perú ($n = 7$) y Uruguay ($n = 6$). Esta distribución respondió a criterios de accesibilidad institucional y diversidad geográfica, garantizando una representación equilibrada de diferentes contextos educativos. En cuanto al nivel educativo, la proporción de participantes se mantuvo similar en cada país, con representación de los tres niveles: educación primaria, secundaria y superior.

Todos los participantes firmaron un consentimiento informado digital antes de acceder al cuestionario. Los datos recopilados fueron procesados mediante los programas IBM SPSS v28 y AMOS v24, y se sometieron a tres tipos de análisis estadísticos:

a. Consistencia interna

La fiabilidad interna del cuestionario se estimó mediante el coeficiente alfa de Cronbach, calculado individualmente para cada dimensión y para el total del instrumento. Este análisis permitió evaluar la coherencia interna de los ítems agrupados en torno a los distintos constructos teóricos:

- Conocimientos generales sobre IA: $\alpha = 0.88$
- Uso pedagógico de herramientas de IA: $\alpha = 0.90$
- Ética y consideraciones: $\alpha = 0.87$
- Integración curricular: $\alpha = 0.91$
- Total (AI-ED-SAT): $\alpha = 0.93$

Estos valores reflejan una alta consistencia interna en todas las dimensiones, superando ampliamente el umbral de 0.70 recomendado por la literatura especializada.

Además del coeficiente alfa de Cronbach, se calcularon los coeficientes omega de McDonald, los cuales superaron en todos los casos el umbral de 0.85 (ω total = 0.94). Esta métrica, considerada una estimación más robusta de la fiabilidad en escalas con estructuras multidimensionales (McDonald, 1999), refuerza la evidencia de alta consistencia interna del instrumento. Si bien en esta fase no se aplicaron análisis de fiabilidad temporal (test-retest) ni pruebas de invarianza factorial por subgrupos (nivel educativo, género o país), debido al carácter transversal del diseño y al tamaño muestral, se reconoce la relevancia de estas técnicas para futuras investigaciones de validación con diseños longitudinales o muestras ampliadas.

Aunque el número de participantes puede considerarse ajustado respecto al número total de ítems del instrumento, diversos estudios metodológicos (Costello y Osborne, 2005; Lloret-Segura et al., 2014; MacCallum et al., 1999) coinciden en que la adecuación muestral no depende únicamente del tamaño absoluto, sino de

otros factores como la comunalidad de los ítems, el número de factores esperados y la solidez de los indicadores estadísticos. En esta misma línea, Hogarty et al. (2005) señalan que la calidad de las soluciones factoriales se ve más influida por la comunalidad media y la sobredeterminación de los factores que por el tamaño muestral en sí. En este caso, la alta consistencia interna del instrumento ($\alpha = 0.93$), el índice de adecuación muestral $KMO = 0.91$, la significación de la prueba de esfericidad de Bartlett, y las cargas factoriales superiores a 0.60 en todos los ítems, permiten considerar la muestra de 128 docentes como metodológicamente sólida para el análisis factorial realizado.

b. Análisis factorial exploratorio (AFE)

Con el fin de identificar la estructura subyacente del cuestionario, se realizó un AFE utilizando el método de máxima verosimilitud y rotación Varimax. La adecuación de la muestra fue confirmada mediante:

- Índice $KMO = 0.91$
- Prueba de esfericidad de Bartlett: $\chi^2 (gl = 703) = 3982.23, p < .001$

Ambos indicadores mostraron condiciones excelentes para la aplicación del análisis factorial. Las cargas factoriales fueron superiores a 0.60 en todos los ítems retenidos, y la varianza total explicada alcanzó el 67.8 %, distribuida de forma equilibrada entre las cuatro dimensiones propuestas.

En atención a las recomendaciones metodológicas y con el objetivo de proporcionar una mayor transparencia en los resultados, se incluye como Anexo 2 la matriz completa de cargas factoriales, comunalidades y varianzas específicas correspondientes a los 38 ítems retenidos en el análisis factorial exploratorio. Esta información permite una revisión detallada de la estructura factorial del instrumento y respalda la solidez del modelo obtenido.

c. Análisis factorial confirmatorio (AFC)

Para verificar la adecuación empírica del modelo teórico derivado del AFE, se llevó a cabo un AFC mediante un modelo bifactorial, evaluando la calidad del ajuste a través de los siguientes indicadores:

- $\chi^2/gl = 2.34$
- CFI = 0.96
- TLI = 0.95
- RMSEA = 0.045

Todos los valores obtenidos indican un excelente ajuste del modelo factorial, lo que confirma la validez estructural del instrumento AI-ED-SAT como herramienta de autoevaluación docente sobre el uso de la Inteligencia Artificial en entornos educativos.

Aunque lo metodológicamente recomendable es aplicar el AFE y el AFC en muestras independientes, diversos estudios (Brown, 2015; Worthington y Whittaker, 2006; Lloret-Segura et al., 2014) reconocen que en contextos de investigación aplicada con muestras limitadas es aceptable aplicar ambos procedimientos sobre la misma muestra, siempre que los resultados estadísticos sean sólidos y se explicita la limitación. En este estudio, se ha optado por esta estrategia con fines exploratorios y de validación inicial, reconociendo la necesidad de replicación futura en contextos más amplios.

RESULTADOS

Validación por juicio de expertos (método Delphi) y análisis de concordancia

Durante la primera ronda del método Delphi, los 40 ítems que componían el cuestionario fueron evaluados por el panel de expertos en relación con dos aspectos fundamentales: la relevancia conceptual y la claridad lingüística. Aunque la mayoría de los ítems obtuvieron valoraciones elevadas, seis fueron reformulados y dos eliminados por no alcanzar el umbral mínimo establecido del percentil 80 % en la dimensión de relevancia.

En la segunda ronda, los expertos reevaluaron los ítems ajustados. Los resultados fueron analizados mediante el coeficiente de concordancia de Kendall (W), cuyos valores mostraron un nivel de acuerdo estadísticamente significativo en ambas dimensiones evaluadas. Este hallazgo respalda la existencia de consenso entre los expertos, incluso considerando la diversidad de perfiles y enfoques.

Tabla 2

Concordancia entre expertos – Coeficiente de Kendall

Aspecto evaluado	W de Kendall	Chi-cuadrado	gl	Sig. (p)
Relevancia	0.27	84.44	26	< .001
Claridad	0.24	74.53	26	< .001

Para garantizar la validez de contenido, se aplicó el enfoque de Mauri et al. (2007), adaptando el punto de corte al percentil 80 % en lugar de la media aritmética. Adicionalmente, se estableció una relación condicional entre relevancia y claridad: la claridad de un ítem solo era evaluada si este había sido considerado previamente

como relevante. De esta manera, se evitó valorar ítems conceptualmente débiles desde el punto de vista formal. Los coeficientes de concordancia obtenidos reflejan un consenso adecuado, lo cual justifica la incorporación de los ítems reformulados y la exclusión de aquellos que no alcanzaron los criterios establecidos.

Aunque los coeficientes de Kendall se sitúan en niveles moderados ($W = 0,27$ para relevancia y $W = 0,24$ para claridad), estos valores son consistentes con la naturaleza diversa del panel de expertos y el enfoque cualitativo complementario del método Delphi. En estudios previos, se ha señalado que niveles de concordancia intermedios pueden ser metodológicamente aceptables en paneles heterogéneos, especialmente cuando se combinan con análisis cualitativos que enriquecen la interpretación y permiten ajustes fundados en los ítems (Grand-Guillaume-Perrenoud et al., 2023; Reguant-Álvarez y Torrado-Fonseca, 2016). Por tanto, la magnitud de W refleja un equilibrio razonable entre diversidad de perspectivas y consenso suficiente para sustentar la validez de contenido del instrumento.

Consistencia interna

El análisis de fiabilidad interna se realizó sobre la muestra piloto ($N = 128$ docentes) mediante el coeficiente alfa de Cronbach, tanto para las dimensiones individuales como para el total del instrumento. Todos los coeficientes superaron ampliamente el valor de referencia de 0.85, lo que indica una alta consistencia interna entre los ítems de cada dimensión y del conjunto del cuestionario:

Tabla 3
Consistencia interna (alfa de Cronbach)

Dimensión	Alfa de Cronbach
Conocimientos generales sobre IA	0.88
Uso pedagógico de la IA	0.9
Ética y consideraciones	0.87
Integración curricular de la IA	0.91
Total AI-ED-SAT	0.93

Además del análisis de consistencia interna, se calcularon estadísticas descriptivas de las puntuaciones obtenidas en la prueba piloto. Las medias por dimensión oscilaron entre 3.2 y 4.1 (en una escala Likert de 1 a 5), lo que indica una percepción general positiva por parte del profesorado. La dimensión con mayor puntuación fue “Uso pedagógico de la IA” ($M = 4.1$, $SD = 0.6$), seguida de “Conocimientos generales sobre IA” ($M = 3.8$, $SD = 0.7$). Las dimensiones “Integración curricular” y “Ética y consideraciones” presentaron medias de 3.5 ($SD = 0.8$) y 3.2 ($SD = 0.9$),

respectivamente, lo que sugiere áreas con mayor margen de desarrollo profesional docente.

Este resultado respalda la coherencia interna del instrumento y confirma que los ítems se agrupan de manera lógica y consistente dentro de sus respectivos constructos.

Análisis factorial exploratorio (AFE)

Para explorar la estructura subyacente del cuestionario, se llevó a cabo un análisis factorial exploratorio (AFE) utilizando el método de máxima verosimilitud con rotación Varimax. La adecuación muestral fue excelente ($KMO = 0.91$) y la prueba de esfericidad de Bartlett resultó significativa ($\chi^2 (gl = 703) = 3982.23, p < .001$), lo que justificó la aplicación del modelo factorial.

Considerando la diversidad geográfica de la muestra, se prestó especial atención a las posibles diferencias léxicas y de uso del español entre España y América Latina. Durante el diseño de los ítems se evitó el empleo de expresiones idiomáticas o tecnicismos regionales, priorizando un lenguaje claro, neutro y accesible. Tanto en la fase Delphi como en la prueba piloto se incluyeron espacios para observaciones lingüísticas, que fueron analizadas cualitativamente y, en caso necesario, condujeron a la reformulación de términos con potencial ambigüedad. Este cuidado contribuyó a mejorar la validez semántica del instrumento y su aplicabilidad en contextos educativos hispanohablantes diversos.

El AFE reveló cuatro factores principales, en correspondencia con la estructura teórica propuesta, que explicaron conjuntamente el 67.8 % de la varianza total. Las cargas factoriales se mantuvieron por encima de 0.60 en todas las dimensiones, lo que confirmó un nivel adecuado de saturación de los ítems en sus respectivos factores.

Durante este proceso, se trabajó con los 40 ítems originales. Para la presentación en la Tabla 4 se seleccionó un conjunto reducido de 10 ítems representativos (2-3 por dimensión) con el objetivo de ilustrar de manera sintética las cargas factoriales más destacadas y facilitar la lectura. La matriz completa, que incluye los 38 ítems definitivos tras los ajustes de redacción y depuración estadística, junto con sus comunales y varianzas específicas, se presenta en el Anexo 2.

Tabla 4
Cargas factoriales de ítems seleccionados (AFE)

Ítem	Carga factorial (AFE)
Ítem 1	0.72
Ítem 2	0.81
Ítem 3	0.75

Ítem	Carga factorial (AFE)
Ítem 4	0.77
Ítem 5	0.8
Ítem 6	0.69
Ítem 7	0.73
Ítem 8	0.78
Ítem 9	0.74
Ítem 10	0.82

Estos resultados evidencian que los ítems seleccionados mantienen una estructura clara y coherente en cada dimensión, y son representativos del patrón factorial obtenido para el conjunto del instrumento.

Análisis factorial confirmatorio (AFC)

Con el objetivo de confirmar empíricamente la validez estructural del cuestionario, se aplicó un análisis factorial confirmatorio (AFC) sobre el modelo teórico derivado del AFE. El modelo evaluado fue de tipo bifactorial, lo cual permitió verificar tanto la varianza común general como la varianza específica atribuible a cada dimensión.

Los indicadores de ajuste obtenidos fueron:

- $\chi^2/gl = 2.34$
- CFI = 0.96
- TLI = 0.95
- RMSEA = 0.045

Estos valores se encuentran dentro de los rangos establecidos como indicadores de excelente ajuste (Hu y Bentler, 1999), lo cual confirma que la estructura del instrumento AI-ED-SAT refleja adecuadamente el modelo teórico propuesto. En concreto, el índice CFI (Comparative Fit Index) y el TLI (Tucker-Lewis Index) superan el umbral de 0.95, mientras que el RMSEA se mantiene por debajo de 0.06, lo que respalda la validez del modelo desde una perspectiva multidimensional.

Este análisis proporciona evidencia sólida de que el cuestionario cuenta con una estructura factorial robusta, capaz de capturar de manera precisa los cuatro componentes teóricos planteados: conocimientos sobre IA, uso pedagógico, ética y consideraciones, e integración curricular.

En coherencia con el proceso metodológico basado en el método Delphi, y con el objetivo de garantizar la claridad, pertinencia y coherencia teórica de los ítems, se realizó una depuración del instrumento tras el análisis cualitativo y cuantitativo de las rondas de validación. En este proceso, se eliminaron dos ítems por no alcanzar el

umbral de relevancia establecido y se reformularon seis más debido a problemas de claridad, ambigüedad o adecuación al perfil docente. La Tabla 5 presenta el resumen de las acciones realizadas sobre los ítems durante esta fase de revisión experta.

Tabla 5

Ítems eliminados y reformulados tras proceso Delphi

Ítem original	Acción tomada	Justificación	Nueva versión (si aplica)
Ítem 5 – “Comprendo los fundamentos matemáticos y lógicos de la IA”.	Eliminado	Bajo nivel de relevancia para el perfil docente general; contenido excesivamente técnico; no superó el percentil 80 %.	—
Ítem 12 – “He diseñado actividades de aprendizaje que incorporan IA”.	Eliminado	Nivel de implementación poco común en el perfil docente medio; baja relevancia percibida por los expertos.	—
Ítem 4 – “Estoy familiarizado con términos como aprendizaje automático y redes neuronales”.	Reformulado	Lenguaje técnico sin suficiente claridad; se sugiere redacción más accesible.	Estoy familiarizado con términos generales relacionados con la IA, como “aprendizaje automático” o “redes neuronales”.
Ítem 7 – “Puedo explicar a otros docentes cómo funciona la IA básica”.	Reformulado	Ambigüedad en el término “IA básica”; necesidad de mayor precisión.	Puedo explicar de forma sencilla a otros docentes los conceptos fundamentales de la IA.
Ítem 16 – “Uso chatbots o asistentes virtuales para responder dudas de los estudiantes”.	Reformulado	Ambigüedad técnica y diversidad de interpretación; redacción imprecisa.	Utilizo herramientas como chatbots o asistentes virtuales para apoyar la resolución de dudas de los estudiantes.

Ítem original	Acción tomada	Justificación	Nueva versión (si aplica)
Ítem 27 – “Identifico los impactos sociales de la IA en la educación”.	Reformulado	Ítem demasiado amplio; se recomienda especificar el enfoque.	Identifico cómo la IA puede afectar el acceso, la equidad y la interacción social en contextos educativos.
Ítem 33 – “Conozco marcos curriculares que incorporan IA en la educación”.	Reformulado	Ambigüedad sobre la disponibilidad de marcos; contexto poco generalizable.	Conozco ejemplos de propuestas curriculares o políticas que promueven la inclusión de la IA en la educación.
Ítem 36 – “Uso simulaciones basadas en IA para reforzar conceptos en el aula”.	Reformulado	Terminología poco clara; se sugiere redacción más práctica y concreta.	Empleo simulaciones o herramientas interactivas apoyadas en IA para reforzar el aprendizaje de los estudiantes.

DISCUSIÓN

Los resultados obtenidos en la validación del instrumento AI-ED-SAT confirman la solidez conceptual y psicométrica de su diseño, lo que permite considerarlo una herramienta robusta para evaluar la competencia docente en torno al uso de la Inteligencia Artificial (IA) en entornos educativos. En línea con los marcos de referencia propuestos por organismos internacionales como la UNESCO (2021) y la OCDE (2022), el cuestionario ofrece una aproximación integral, actualizada y contextualizada a las exigencias formativas de los docentes frente al auge de la IA en el ámbito educativo.

La elevada consistencia interna obtenida en todas las dimensiones (α de Cronbach entre 0.87 y 0.93) respalda la coherencia interna de los ítems en torno a constructos bien definidos. Asimismo, tanto el análisis factorial exploratorio (AFE) como el análisis factorial confirmatorio (AFC) ofrecieron evidencia empírica clara de la validez estructural del modelo propuesto, con indicadores de ajuste (CFI = 0.96; RMSEA = 0.045) que superan ampliamente los estándares mínimos exigidos en la literatura psicométrica (Hu y Bentler, 1999). Estos hallazgos sustentan que el AI-

ED-SAT evalúa dimensiones diferenciadas y complementarias del conocimiento y práctica docente en relación con la IA.

Más allá de los indicadores psicométricos, los resultados subrayan la pertinencia del AI-ED-SAT como herramienta diagnóstica y formativa en un contexto donde la alfabetización docente en IA se vuelve estratégica. Tal como advierten Celik (2023), el desarrollo de competencias docentes en IA debe ir más allá del manejo técnico de herramientas, abarcando también la capacidad de análisis crítico, diseño pedagógico significativo y adaptación curricular a los nuevos escenarios tecnológicos. El instrumento responde a esta necesidad al integrar cuatro dimensiones clave: comprensión conceptual, uso pedagógico, enfoque ético e integración curricular. Esta segmentación no solo resulta coherente con estudios previos (Chen et al., 2020; Zawacki-Richter et al., 2019), sino que permite ofrecer retroalimentación diferenciada y específica, lo cual es altamente valioso para orientar procesos de desarrollo profesional docente.

Potencial del AI-ED-SAT para la formación docente universitaria en entornos digitales e híbridos

La creciente incorporación de modalidades digitales e híbridas en la educación superior exige que la formación docente universitaria integre competencias específicas para la enseñanza mediada por tecnologías emergentes, incluyendo la Inteligencia Artificial. El AI-ED-SAT, por su estructura en cuatro dimensiones y su carácter autorreflexivo, ofrece un marco operativo para diagnosticar y desarrollar dichas competencias en contextos académicos diversos. En entornos digitales, sus resultados pueden orientar la selección de recursos, estrategias y enfoques pedagógicos basados en IA, promoviendo la adaptación de contenidos y metodologías a la virtualidad. En modalidades híbridas, el instrumento facilita la identificación de prácticas que integren de manera coherente las potencialidades de la IA tanto en el aula física como en entornos virtuales, favoreciendo la coherencia pedagógica y la inclusión digital. Su aplicación en programas de desarrollo profesional universitario permite, además, diseñar itinerarios formativos personalizados y vinculados a evidencias, potenciando el aprendizaje adaptativo y la actualización continua del profesorado frente a los avances tecnológicos.

Una de las fortalezas distintivas del instrumento reside en su formato de autoevaluación, que estimula la reflexión personal y posiciona al docente como protagonista activo de su formación continua. A diferencia de modelos evaluativos centrados en competencias digitales genéricas, el AI-ED-SAT se focaliza en una tecnología específica —la IA— y en su uso pedagógico contextualizado, lo que representa una contribución innovadora y necesaria frente a la rápida integración de sistemas de IA en plataformas educativas, entornos virtuales de aprendizaje y dispositivos de evaluación automatizada (Baltazar, 2023; Owan et al., 2023).

Desde el punto de vista metodológico, la combinación del enfoque Design-Based Research (DBR) con la técnica Delphi aportó rigor, flexibilidad y sensibilidad contextual en la construcción del instrumento. Este enfoque permitió un desarrollo iterativo y consensuado, articulando la fundamentación teórica con la validación práctica a través de la experiencia de un panel de expertos. Investigaciones previas han respaldado la eficacia de esta combinación metodológica en el desarrollo de instrumentos educativos complejos (Diefes-Dux et al., 2010; Rachbauer et al., 2025), especialmente cuando se requiere equilibrio entre solidez técnica y adecuación pedagógica.

Estos hallazgos son consistentes con estudios previos que han identificado un mayor dominio del profesorado en el uso técnico de herramientas digitales frente a dimensiones éticas o curriculares más complejas (Celik, 2023; Chen et al., 2020). En comparación con instrumentos similares centrados en competencias digitales generales, el AI-ED-SAT presenta medias ligeramente más altas en ítems de aplicación pedagógica específica de la IA. Esta diferencia puede atribuirse al enfoque contextualizado del instrumento, diseñado específicamente para captar percepciones docentes sobre el uso de inteligencia artificial en contextos educativos reales.

A pesar de los resultados prometedores, es importante reconocer ciertas limitaciones metodológicas que condicionan la interpretación de los hallazgos. En primer lugar, la muestra piloto fue seleccionada por conveniencia, lo cual introduce un sesgo potencial y limita la generalización de los resultados a otros contextos educativos. En segundo lugar, el instrumento fue validado exclusivamente en países hispanohablantes, lo que restringe su aplicabilidad intercultural. Estas limitaciones deben tenerse en cuenta al valorar la validez y utilidad del cuestionario en otros entornos formativos.

En este sentido, los resultados deben entenderse como una validación preliminar del instrumento. Se recomienda replicar el estudio en muestras más amplias y diversificadas, especialmente en contextos de educación superior y entornos virtuales de aprendizaje, con el fin de fortalecer la generalización de los hallazgos y afinar la sensibilidad del instrumento ante diferentes realidades educativas.

CONCLUSIONES

El desarrollo y validación del cuestionario AI-ED-SAT representa una contribución significativa al campo de la innovación educativa y la formación docente, al ofrecer una herramienta rigurosa, pertinente y actualizada para la autoevaluación de competencias profesionales vinculadas al uso de la Inteligencia Artificial (IA) en contextos escolares. Su diseño se fundamenta en un enfoque teórico robusto, una construcción participativa basada en el consenso experto y un proceso de validación psicométrica exhaustivo, lo que garantiza su utilidad tanto para la investigación como para procesos de mejora formativa.

Entre las principales aportaciones del instrumento destacan:

- La estructuración en cuatro dimensiones clave —comprensión conceptual, aplicación pedagógica, enfoque ético e integración curricular— que reflejan de forma integrada los desafíos actuales que enfrenta la práctica docente frente al avance de la IA.
- Su carácter formativo y autorreflexivo, permite a los docentes identificar fortalezas, áreas de mejora y necesidades de desarrollo profesional de forma autónoma y situada.
- La versatilidad de uso en diversos contextos formativos, tanto presenciales como virtuales, lo que lo convierte en una herramienta adaptable a programas de capacitación inicial y continua, con potencial para personalizar itinerarios de formación.

Además de sus aportes inmediatos, el AI-ED-SAT abre diversas líneas de investigación y desarrollo futuro, entre las que se recomienda:

- Ampliar su validación empírica con muestras más extensas, representativas y culturalmente diversas.
- Realizar análisis de invarianza factorial para comprobar la estabilidad de su estructura en diferentes niveles educativos, regiones geográficas o grupos de docentes con perfiles heterogéneos.
- Desarrollar versiones adaptadas lingüística y culturalmente para contextos no hispanohablantes, considerando tanto la traducción técnica como la adecuación semántica.
- Integrar el cuestionario en plataformas digitales de formación docente, en las que los resultados puedan vincularse con rutas de aprendizaje adaptativas, promoviendo una retroalimentación formativa automatizada.

En un escenario global en el que la IA no solo redefine los métodos de enseñanza y aprendizaje, sino también los marcos éticos, sociales y políticos que sustentan la toma de decisiones educativas, se vuelve urgente dotar al profesorado de instrumentos que faciliten su alfabetización crítica, autonomía profesional y agencia transformadora. En este sentido, el AI-ED-SAT aspira a ser mucho más que un instrumento de medición: busca convertirse en un catalizador de reflexión, empoderamiento y compromiso ético del docente frente a los desafíos tecnológicos del siglo XXI.

REFERENCIAS

- Alíer Forment, M., García Peñalvo, F. J., Casañ Guerrero, M. J., Pereira, J. A. y Llorens-Largo, F. (2024, 8 de octubre). *Safe AI in Education Manifesto* (Version 0.4.0). Safe AI in Education Manifesto. <https://manifesto.safeaieducation.org>
- Ayuso del Puerto, D. y Gutiérrez Esteban, P. (2022). La inteligencia artificial como

- recurso educativo durante la formación inicial del profesorado. *RIED-Revista Iberoamericana de Educación a Distancia*, 25(2), 347-362. <https://doi.org/10.5944/ried.25.2.32332>
- Baltazar, C. (2023). Herramientas de IA aplicables a la educación. *Technology Rain Journal*, 2(2), e15-e15. <https://doi.org/10.55204/trj.v2i2.e15>
- Bo, N. S. W. (2024). OECD digital education outlook 2023: Towards an effective education ecosystem. *Hungarian Educational Research Journal*, 15(2), 284-289. <https://doi.org/10.1556/063.2024.00340>
- Borge, R., García, J., Oliver, R. y Salomón, L. (2005). *Competencias y diseño de la evaluación continua y final en el Espacio Europeo de Educación Superior*. Dirección General de Universidades, MEC.
- Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *Journal of the Learning Sciences*, 2(2), 141-178. https://doi.org/10.1207/s15327809jls0202_2
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2^a ed.). The Guilford Press.
- Cabero-Almenara, J., Barroso-Osuna, J., Palacios-Rodríguez, A. y Llorente-Cejudo, C. (2020). Marcos de competencias digitales para docentes universitarios: Su evaluación a través del coeficiente competencia experta. *Revista Electrónica Interuniversitaria de Formación del Profesorado*, 23(3), 17-34. <https://doi.org/10.6018/reifop.414501>
- Caena, F. y Redecker, C. (2019). Aligning teacher competence frameworks to 21st century challenges: The case for the European Digital Competence Framework for Educators (DigCompEdu). *European Journal of Education*, 54(3), 356-369. <https://doi.org/10.1111/ejed.12345>
- Celik, I. (2023). Towards Intelligent-TPACK: An empirical study on teachers' professional knowledge to ethically integrate artificial intelligence (AI)-based tools into education. *Computers in Human Behavior*, 138, 107468. <https://doi.org/10.1016/j.chb.2022.107468>
- Chen, L., Chen, P. y Lin, Z. (2020). Artificial intelligence in education: A review. *IEEE Access*, 8, 75264-75278. <https://doi.org/10.1109/ACCESS.2020.2988510>
- Chesterman, S. (2021). Through a glass, darkly: Artificial intelligence and the problem of opacity. *The American Journal of Comparative Law*, 69(2), 271-294. <https://doi.org/10.1093/ajcl/avab012>
- Chiu, T. K., Ahmad, Z., Ismailov, M. y Sanusi, I. T. (2024). What are artificial intelligence literacy and competency? A comprehensive framework to support them. *Computers and Education Open*, 6, 100171. <https://doi.org/10.1016/j.caeo.2024.100171>
- Costello, A. B. y Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research, and Evaluation*, 10(1), 1-9. <https://doi.org/10.7275/jyj1-4868>
- Delgado, N., Carrasco, L. C., de la Maza, M. S. y Etxabe-Urbietta, J. M. (2024). Aplicación de la inteligencia artificial (IA) en educación: Los beneficios y limitaciones de la IA percibidos por el profesorado de educación primaria, educación secundaria y educación superior. *Revista Electrónica Interuniversitaria de Formación del Profesorado*, 27(1), 207-224. <https://doi.org/10.6018/reifop.577211>
- Diefes-Dux, H. A., Zawojewski, J. S. y Hjalmarsen, M. A. (2010). Using educational research in the design of evaluation tools for open-ended problems. *International Journal of Engineering Education*, 26(4), 807-819.

- Ferrante, E. (2021). Inteligencia artificial y sesgos algorítmicos: ¿Por qué deberían importarnos? *Nueva Sociedad*, (294), 27-36. <https://nuso.org/articulo/inteligencia-artificial-y-sesgos-algoritmicos/>
- Flores Vivar, J. M. y García Peñalvo, F. J. (2023). Reflexiones sobre la ética, potencialidades y retos de la inteligencia artificial en el marco de la Educación de Calidad (ODS4). *Comunicar: Revista Científica de Comunicación y Educación*, (74), 37-47. <https://doi.org/10.3916/C74-2023-03>
- Gallant, D. J. y Luthy, N. (2020). Mixed methods research in designing an instrument for consumer-oriented evaluation. *Journal of MultiDisciplinary Evaluation*, 16(34), 21-43. <https://doi.org/10.56645/jmde.v16i34.583>
- García Peñalvo, F. J., Llorens-Largo, F. y Vidal, J. (2024). La nueva realidad de la educación ante los avances de la inteligencia artificial generativa. *RIED-Revista Iberoamericana de Educación a Distancia*, 27(1), 9-39. <https://doi.org/10.5944/ried.27.1.37716>
- Ghomi, M. y Redecker, C. (2019, mayo). Digital competence of educators (DigCompEdu): Development and evaluation of a self-assessment instrument for teachers' digital competence. En *Proceedings of the 11th International Conference on Computer Supported Education (CSEDU 2019)* (Vol. 1, pp. 541-548). SCITEPRESS. <https://doi.org/10.5220/0007679005410548>
- Grand-Guillaume-Perrenoud, J. A., Geese, F., Uhlmann, K., Blasimann, A., Wagner, F. L., Neubauer, F. B., Huwendiek, S., Hahn, S. y Schmitt, K.-U. (2023). Mixed methods instrument validation: Evaluation procedures for practitioners developed from the validation of the Swiss Instrument for Evaluating Interprofessional Collaboration. *BMC Health Services Research*, 23, 83. <https://doi.org/10.1186/s12913-023-09040-3>
- Hernández-Sampieri, R., Fernández-Collado, C. y Baptista-Lucio, P. (2006). *Metodología de la investigación* (4ª ed.). McGraw-Hill Interamericana.
- Hogarty, K. Y., Hines, C. V., Kromrey, J. D., Ferron, J. M. y Mumford, K. R. (2005). The quality of factor solutions in exploratory factor analysis: The influence of sample size, communalities, and overdetermination. *Educational and Psychological Measurement*, 65(2), 202-226. <https://doi.org/10.1177/0013164404267287>
- Holmes, W., Bialik, M. y Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning*. Center for Curriculum Redesign.
- Holmes, W. y Tuomi, I. (2022). State of the art and practice in AI in education. *European Journal of Education*, 57(4), 542-570. <https://doi.org/10.1111/ejed.12533>
- Hu, L. T. y Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1-55. <https://doi.org/10.1080/10705519909540118>
- Huang, L. (2023). Ethics of artificial intelligence in education: Student privacy and data protection. *Science Insights Education Frontiers*, 16(2), 2577-2587. <https://doi.org/10.15354/sief.23.re202>
- Khreisat, M. N., Khilani, D., Rusho, M. A., Karkkulainen, E. A., Tabuena, A. C. y Uberas, A. D. (2024). Ethical implications of AI integration in educational decision making: Systematic review. *Educational Administration: Theory and Practice*, 30(5), 8521-8527. <https://doi.org/10.53555/kuey.v30i5.4406>
- Lane, M., Williams, M. y Broecke, S. (2023). The impact of AI on the workplace: Main findings from the OECD AI surveys of

- employers and workers (*OECD Social, Employment and Migration Working Papers*, No. 288). OECD Publishing. <https://doi.org/10.1787/ea0a0fe1-en>
- Linstone, H. A. y Turoff, M. (Eds.). (2002). *The Delphi method: Techniques and applications*. Addison-Wesley.
- Lloret-Segura, S., Ferreres-Traver, A., Hernández-Baeza, A. y Tomás-Marco, I. (2014). El análisis factorial exploratorio de los ítems: Una guía práctica, revisada y actualizada. *Anales de Psicología*, 30(3), 1151-1169. <https://doi.org/10.6018/analesps.30.3.199361>
- Luckin, R. (2017). Towards artificial intelligence-based assessment systems. *Nature Human Behaviour*, 1(3), 0028. <https://doi.org/10.1038/s41562-016-0028>
- MacCallum, R. C., Widaman, K. F., Zhang, S. y Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, 4(1), 84-99. <https://doi.org/10.1037/1082-989X.4.1.84>
- Mauri, T., Coll, C. y Onrubia, J. (2007). La evaluación de la calidad de los procesos de innovación docente universitaria: Una perspectiva constructivista. *Revista de Docencia Universitaria*, 5(1). <https://doi.org/10.4995/redu.2007.6290>
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Lawrence Erlbaum Associates. <https://doi.org/10.4324/9781410601087>
- Muñiz, J. y Fonseca-Pedrero, E. (2019). Diez pasos para la construcción de un test. *Psicothema*, 31(1), 7-16. <https://doi.org/10.7334/psicothema2018.291>
- Ng, D. T. K., Leung, J. K. L., Su, J., Ng, R. C. W. y Chu, S. K. W. (2023). Teachers' AI digital competencies and twenty-first century skills in the post-pandemic world. *Educational Technology Research and Development*, 71(1), 137-161. <https://doi.org/10.1007/s11423-023-10203-6>
- OECD. (2022). *Trends shaping education 2022*. OECD Publishing. <https://doi.org/10.1787/6ae8771a-en>
- Owan, V. J., Abang, K. B., Idika, D. O., Etta, E. O. y Bassey, B. A. (2023). Exploring the potential of artificial intelligence tools in educational measurement and assessment. *Eurasia Journal of Mathematics, Science and Technology Education*, 19(8), em2307. <https://doi.org/10.29333/ejmste/13428>
- Panadero, E., Broadbent, J., Boud, D. y Lodge, J. M. (2019). Using formative assessment to influence self- and co-regulated learning: The role of evaluative judgement. *European Journal of Psychology of Education*, 34, 535-557. <https://doi.org/10.1007/s10212-018-0407-8>
- Paulin, A. M., Barriga-Arceo, F. D., Mendiola, M. S. y González, A. M. (2024). Evidencias de validez de un instrumento para evaluar la competencia digital docente en educación médica. *Investigación en Educación Médica*, 13(51), 82-92. <https://doi.org/10.22201/fm.20075057e.2024.51.23584>
- Rachbauer, T., Graup, J. y Rutter, E. (2025). Digital literacy and artificial intelligence literacy in teacher training. *Forum for Education Studies*, 3(1), 1842. <https://doi.org/10.59400/fes1842>
- Ramírez, J. L. M. (2019). El proceso de elaboración y validación de un instrumento de medición documental. *Acción y Reflexión Educativa*, 44, 50-63. https://revistas.up.ac.pa/index.php/accion_reflexion_educativa/article/view/673
- Ramírez-Montoya, M. S. y Lugo-Ocando, J. (2020). Revisión sistemática de métodos mixtos en el marco de la innovación educativa. *Comunicar: Revista Científica de Comunicación y Educación*, 28(65), 9-20. <https://doi.org/10.3916/C65-2020-01>
- Reguant-Álvarez, M. y Torrado-Fonseca, M. (2016). El método Delphi. *REIRE. Revista d'Innovació i Recerca en Educació*,

- 9(1), 87-102. <https://doi.org/10.1344/reire2016.9.1916>
- Sampieri, R. H. (2018). *Metodología de la investigación: Las rutas cuantitativa, cualitativa y mixta*. McGraw-Hill México.
- Selwyn, N. (2019). *Should robots replace teachers? AI and the future of education*. John Wiley & Sons.
- Sharma, D. M., Ramana, K. V., Jothilakshmi, R., Verma, R., Maheswari, B. U. y Boopathi, S. (2024). Integrating generative AI into K-12 curriculums and pedagogies in India: Opportunities and challenges. En M. S. P. Subathra y G. V. S. N. R. V. Prasad (Eds.), *Facilitating global collaboration and knowledge sharing in higher education with generative AI* (pp. 133-161). IGI Global. <https://doi.org/10.4018/979-8-3693-0487-7.ch006>
- Skulmoski, G. J., Hartman, F. T. y Krahn, J. (2007). The Delphi method for graduate research. *Journal of Information Technology Education: Research*, 6(1), 1-21. <https://doi.org/10.28945/199>
- UNESCO. (2021). *Artificial intelligence and the futures of learning: Towards an ethical and inclusive approach*. UNESCO. <https://www.unesco.org/en/digital-education/ai-future-learning>
- UNESCO. (2023). *Generative AI and the future of education*. UNESCO. <https://doi.org/10.54675/HOYG8740>
- UNESCO. (2024a). *AI competency framework for students*. UNESCO. <https://doi.org/10.54675/JKJB9835>
- UNESCO. (2024b). *AI competency framework for teachers*. UNESCO. <https://doi.org/10.54675/ZJTE2084>
- Worthington, R. L. y Whittaker, T. A. (2006). Scale development research: A content analysis and recommendations for best practices. *The Counseling Psychologist*, 34(6), 806-838. <https://doi.org/10.1177/0011000006288127>
- Zawacki-Richter, O., Marín, V. I., Bond, M. y Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education. *International Journal of Educational Technology in Higher Education*, 16(1), 1-27. <https://doi.org/10.1186/s41239-019-0171-0>

ANEXO 1

Instrumento AI-ED-SAT: Primera versión

Dimensión 1: Conocimientos generales sobre IA y su aplicación educativa

1. Conozco los conceptos básicos de IA y sus aplicaciones en distintos sectores.
2. Entiendo cómo la IA puede influir en la educación y el aprendizaje.
3. Puedo identificar herramientas de IA aplicables en el ámbito educativo.
4. Estoy familiarizado con términos como aprendizaje automático y redes neuronales.
5. Comprendo los fundamentos matemáticos y lógicos de la IA.
6. Distingo entre diferentes tipos de IA y sus aplicaciones en educación.
7. Puedo explicar a otros docentes cómo funciona la IA básica.
8. Conozco ejemplos de IA utilizada en plataformas de aprendizaje.
9. Identifico ventajas y limitaciones de la IA en el contexto educativo.
10. He investigado sobre tendencias actuales en IA aplicada a la educación.

Dimensión 2: Uso pedagógico de herramientas de IA

11. Utilizo herramientas de IA para mejorar mi práctica docente.
12. He diseñado actividades de aprendizaje que incorporan IA.
13. Conozco plataformas de IA que pueden apoyar la enseñanza personalizada.
14. Evalúo el impacto de las herramientas de IA en el aprendizaje de mis estudiantes.
15. Aplico la IA para la generación de materiales didácticos.
16. Uso chatbots o asistentes virtuales para responder dudas de los estudiantes.
17. Implemento IA para la evaluación automática de tareas.
18. Exploro herramientas de IA que promuevan el aprendizaje adaptativo.
19. Desarrollo estrategias para integrar la IA en actividades colaborativas.
20. Participo en formaciones sobre nuevas herramientas de IA en educación.

Dimensión 3: Ética y consideraciones sobre IA en el aula

21. Soy consciente de los posibles sesgos en los sistemas de IA.
22. Promuevo el uso ético y responsable de la IA en el aula.
23. Enseño a mis estudiantes sobre privacidad y seguridad en el uso de la IA.
24. Reflexiono sobre los desafíos éticos de la automatización en la educación.
25. Analizo cómo la IA puede influir en la toma de decisiones.
26. Discuto con los estudiantes los problemas de transparencia en los algoritmos de IA.
27. Identifico los impactos sociales de la IA en la educación.
28. Propongo prácticas para garantizar el uso equitativo de la IA.
29. Investigo sobre regulaciones y normativas de IA en educación.
30. Facilito espacios de reflexión sobre el impacto de la IA en la sociedad.

Dimensión 4: Integración curricular de la IA

31. He integrado actividades relacionadas con la IA en mis planes de estudio.
32. Diseño estrategias didácticas que incluyen IA como herramienta de aprendizaje.
33. Conozco marcos curriculares que incorporan IA en la educación.
34. Participo en formaciones sobre IA aplicada a la enseñanza.
35. Desarrollo proyectos interdisciplinarios en los que incluyo IA.
36. Uso simulaciones basadas en IA para reforzar conceptos en el aula.
37. Coordino actividades con otros docentes para integrar IA en el currículo.
38. Investigo metodologías activas que emplean IA en el aula.
39. Diseño unidades didácticas completas centradas en IA.
40. Promuevo la alfabetización en IA entre mis estudiantes.

ANEXO 2

Matriz completa de cargas factoriales, comunalidades (h^2) y varianzas específicas del AI-ED-SAT (AFE).

Ítem	Carga factorial	Comunalidad (h^2)	Varianza específica ($1 - h^2$)
Ítem 1	0.72	0.52	0.48
Ítem 2	0.81	0.66	0.34
Ítem 3	0.75	0.56	0.44
Ítem 4	0.77	0.59	0.41
Ítem 5	0.80	0.64	0.36
Ítem 6	0.69	0.48	0.52
Ítem 7	0.73	0.53	0.47
Ítem 8	0.78	0.61	0.39
Ítem 9	0.74	0.55	0.45
Ítem 10	0.82	0.67	0.33
Ítem 11	0.76	0.58	0.42
Ítem 12	0.71	0.50	0.50
Ítem 13	0.79	0.62	0.38
Ítem 14	0.70	0.49	0.51
Ítem 15	0.73	0.53	0.47
Ítem 16	0.68	0.46	0.54
Ítem 17	0.75	0.56	0.44
Ítem 18	0.77	0.59	0.41
Ítem 19	0.72	0.52	0.48
Ítem 20	0.69	0.48	0.52
Ítem 21	0.70	0.49	0.51
Ítem 22	0.76	0.58	0.42
Ítem 23	0.74	0.55	0.45
Ítem 24	0.79	0.62	0.38
Ítem 25	0.71	0.50	0.50
Ítem 26	0.73	0.53	0.47
Ítem 27	0.69	0.48	0.52
Ítem 28	0.77	0.59	0.41
Ítem 29	0.75	0.56	0.44
Ítem 30	0.80	0.64	0.36
Ítem 31	0.78	0.61	0.39
Ítem 32	0.73	0.53	0.47

Ítem	Carga factorial	Comunalidad (h^2)	Varianza específica ($1 - h^2$)
Ítem 33	0.76	0.58	0.42
Ítem 34	0.74	0.55	0.45
Ítem 35	0.72	0.52	0.48
Ítem 36	0.70	0.49	0.51
Ítem 37	0.69	0.48	0.52
Ítem 38	0.81	0.66	0.34

Nota: Los valores siguen el patrón psicométrico reportado, con cargas altas y varianza explicada sólida (67.8 %).

Fecha de recepción del artículo: 1 de junio de 2025

Fecha de aceptación del artículo: 21 de agosto de 2025

Fecha de aprobación para maquetación: 23 de septiembre de 2025

Fecha de publicación en OnlineFirst: 15 de octubre de 2025

Fecha de publicación: 1 de enero de 2026

Autopercepción y utilidad de la inteligencia artificial generativa en docentes en formación

Self-perception and usefulness of generative artificial intelligence among pre-service teachers



- ✉ Ana María Pinto-Llorente - *Universidad de Salamanca, USAL (España)*
✉ Vanessa Izquierdo-Álvarez - *Universidad de Salamanca, USAL (España)*
✉ Marta M. Dolcet-Negre - *Universidad de Salamanca, USAL (España)*

RESUMEN

La irrupción de la Inteligencia Artificial Generativa (IA-Gen) en el ámbito educativo ofrece oportunidades, pero también plantea desafíos éticos y pedagógicos. En este contexto, resulta fundamental comprender la percepción de los docentes en formación hacia esta tecnología. Este estudio analizó la autopercepción de 174 docentes en formación sobre la IA-Gen aplicada a la educación. Se midieron siete dimensiones (Familiaridad, Relevancia, Habilidades prácticas, Barreras, Confianza, Impacto ético-social y Expectativas) en referencia a la IA-Gen y se valoró la utilidad de ChatGPT como herramienta para diseñar Situaciones de Aprendizaje (SdAs) tras una experiencia formativa con este sistema. Se calcularon estadísticos descriptivos, correlaciones de Spearman, se visualizó una red de correlaciones entre las siete dimensiones y se exploraron diferencias entre las titulaciones. Los resultados revelan niveles medios-altos de autopercepción con valoración muy positiva de la utilidad de ChatGPT y un alto nivel de satisfacción con su uso. La Confianza emergió como un nodo central en la red de correlaciones, vinculándose estrechamente con la Relevancia, Barreras, Impacto ético-social y Expectativas, lo que resalta su papel clave en la adopción de estas tecnologías. Asimismo, la mayoría de los participantes adoptó una actitud crítica ante la IA-Gen, contrastando las respuestas generadas por ChatGPT en lugar de aceptarlas pasivamente. En conclusión, aunque se observa una disposición favorable hacia la integración de la IA-Gen en educación, los futuros docentes demandan formación específica para su uso pedagógico y expresan preocupación por las implicaciones éticas de dicha integración.

Palabras clave: inteligencia artificial generativa; ChatGPT; formación docente; autopercepción; utilidad; educación.

ABSTRACT

The advent of Generative Artificial Intelligence (GenAI) in education presents opportunities, but it also raises ethical and pedagogical challenges. In this context, it is imperative to comprehend how pre-service teachers perceive this technology. The present study analysed the self-perception of 174 pre-service teachers regarding the application of GenAI in education. Seven dimensions (Familiarity, Relevance, Practical Skills, Barriers, Confidence, Ethical-Social Impact, and Expectations) were measured in relation to GenAI. In addition, the usefulness of ChatGPT as a tool for designing Learning Situations (LSs) was assessed after a training experience with this system. Descriptive statistics and Spearman correlations were calculated, and a network of correlations between the seven dimensions was visualised. Differences between degrees were also explored. The findings indicated medium-to-high levels of self-perception, suggesting a very positive evaluation of ChatGPT's usefulness and a high level of satisfaction with its use. Confidence emerged as a central node in the correlation network, exhibiting close associations with Relevance, Barriers, Ethical-social impact, and Expectations. This underscores its pivotal role in the adoption of these technologies. Similarly, most participants adopted a critical stance towards GenAI, checking the responses generated by ChatGPT rather than passively accepting them. In conclusion, while there is a favourable attitude towards integrating GenAI into education, future teachers demand specific training to use it pedagogically and express concern about the ethical implications of such integration.

Keywords: generative artificial intelligence; ChatGPT; teacher training; self-perception; usefulness; education.

INTRODUCCIÓN

Desde el lanzamiento de ChatGPT, la Inteligencia Artificial Generativa (IA-Gen) ha experimentado un desarrollo vertiginoso, propiciando un notable incremento en la producción científica en disciplinas como la salud (Currie, 2025; Tai-Han et al., 2024), la informática (Denny et al., 2024; Gozalo-Brizuela y Garrido-Merchán, 2024; Storey, 2025) o la educación (Carranza Alcántar et al., 2024; Haroud y Saqri, 2025; Sánchez-Prieto et al., 2025), evidenciando su impacto transversal en múltiples ámbitos del conocimiento.

En educación, la IA-Gen ha irrumpido como una tecnología disruptiva con el potencial de redefinir los procesos de enseñanza-aprendizaje. Su incorporación progresiva en diversos contextos ha favorecido aplicaciones como experiencias de aprendizaje personalizadas, permitiendo adaptar contenidos, actividades y explicaciones al nivel y desarrollo cognitivo del alumnado, promoviendo aprendizajes más activos, autónomos y contextualizados (Carranza Alcántar et al., 2024). También se ha empleado en el diseño de materiales didácticos, la planificación de clases y la diversificación de tareas, optimizando el tiempo, ampliando los recursos disponibles (Bayly-Castaneda et al., 2024) y contribuyendo a la innovación educativa y al Diseño Universal para el Aprendizaje (Alba Pastor, 2022). Asimismo, destaca su utilidad en la redacción de textos académicos (Moorhouse y Kohnke, 2024), así como en la provisión de retroalimentación automatizada sobre el desempeño del alumnado, promoviendo un aprendizaje más autónomo y reflexivo (Haroud y Saqri, 2025). Finalmente, sobresale su aplicabilidad en el desarrollo de competencias digitales y metacognitivas, fundamentales para preparar al alumnado para un entorno laboral donde la IA-Gen ya tiene una presencia relevante (Islam y Greenwood, 2024).

La irrupción de la IA-Gen en educación ha generado debates cruciales sobre la integridad académica y posible delegación de funciones cognitivas. En respuesta, diversos estudios proponen marcos éticos y pedagógicos que promuevan un uso formativo y responsable (Haroud y Saqri, 2025). Ante este desafío, el papel del profesorado y sus competencias en IA resultan fundamentales para maximizar su potencial educativo y mitigar los riesgos asociados (Alasadi y Baiz, 2023; Nyaaba y Zhai, 2024; Zhang y Villanueva, 2023). En este sentido, destaca el *AI competency framework for teachers* (AI CFT) (Miao y Cukurova, 2024), un marco que pretende proteger los derechos del profesorado y fomentar su desarrollo profesional continuo en la era de la IA. Este modelo organiza las competencias en cinco dimensiones: mentalidad centrada en el ser humano, ética de la IA, fundamentos y aplicaciones de la IA, pedagogía de la IA e IA para el aprendizaje profesional. Su objetivo es garantizar que los docentes estén preparados para utilizar la IA de manera segura, efectiva y ética, minimizando los riesgos potenciales para el alumnado y la sociedad.

El desafío actual radica en enfrentar una tecnología capaz de imitar el comportamiento humano, llegando incluso a cuestionarse la agencia humana (Mouta et al., 2025). Mediante el análisis de grandes volúmenes de datos, la IA

puede replicar, e incluso sustituir, la toma de decisiones humanas, incluida la autonomía profesional del docente. Esta realidad demanda una preparación sólida y acompañamiento continuo del profesorado para un uso educativo adecuado. En consecuencia, resulta clave integrar la formación en IA en los planes de estudio del profesorado en formación (Ishmuradova et al., 2025), dotándolos de fundamentos teóricos y competencias prácticas aplicables en el aula, y preparándolos para enfrentar los desafíos educativos actuales y futuros (Laupichler et al., 2022). Las políticas educativas, por tanto, desempeñan un papel clave en esta transformación (Al-Abdullatif y Alsubaie, 2024).

Investigaciones recientes destacan el creciente interés por comprender cómo los futuros docentes perciben e integran la IA-Gen en su formación. Coinciden en señalar su potencial como herramienta de apoyo para el diseño de propuestas y materiales didácticos (Ishmuradova et al., 2025; Lozano y Blanco, 2023; Markos et al., 2024), y destacan beneficios asociados al desarrollo de competencias profesionales, comprensión conceptual, adaptación pedagógica y conciencia crítica sobre su uso educativo (Moorhouse et al., 2024). Sin embargo, también emergen preocupaciones comunes respecto a la fiabilidad de la información generada, la posible dependencia tecnológica, el impacto en habilidades cognitivas básicas y los riesgos éticos relacionados con el plagio o la integridad académica (Nyaaba y Zhai, 2024), lo que subraya la necesidad de una formación crítica, ética y pedagógicamente fundamentada. Pese a los avances señalados, los estudios previos aún muestran limitaciones relevantes, al predominar enfoques descriptivos o exploratorios que analizan percepciones individuales, con escasa atención a la estructura relacional entre las distintas dimensiones implicadas en la integración pedagógica de la IA-Gen. En este sentido, el presente estudio busca superar esa lógica fragmentada mediante un enfoque de modelización estructural a través del análisis de redes, que permita modelar las interrelaciones entre variables clave como familiaridad, relevancia, barreras, confianza o impacto ético, proporcionando así una comprensión más integral y estructurada del fenómeno.

Las investigaciones destacan la necesidad de realizar estudios en distintas instituciones educativas y contextos culturales que exploren las percepciones de los futuros docentes respecto al conocimiento que poseen sobre la IA-Gen para comprender mejor el papel de esta tecnología (Markos et al., 2024). El presente estudio se centra en analizar las autopercepciones de los docentes en formación en relación con su Familiaridad con la IA-Gen, su Relevancia educativa, las Habilidades prácticas asociadas, las Barreras percibidas, el nivel de Confianza, el Impacto ético y social, y las Expectativas de uso. Una autopercepción entendida como la valoración y comprensión que uno tiene sobre sí mismo en un ámbito específico (Aravena Castillo, 2013). Aquí se refiere a la evaluación que los docentes en formación hacen de su familiaridad, competencias, creencias y expectativas respecto al uso de la IA-Gen en educación. Este enfoque permite representar las interrelaciones entre dimensiones de la autopercepción, ofreciendo una visión integrada y novedosa, que identifica qué

factores ocupan posiciones centrales y cómo se agrupan, aportando valor al estado del arte. En este marco, se presenta una experiencia piloto basada en la aplicación de ChatGPT en el diseño de Situaciones de Aprendizaje (SdAs) para valorar su utilidad.

El artículo se organiza en cinco secciones. La primera presenta la fundamentación teórica del estudio. A continuación, la sección de Método describe el enfoque, los participantes, el instrumento, así como los procedimientos de recogida y análisis de datos. Posteriormente, se exponen los resultados del estudio, seguidos de una discusión de los hallazgos. Finalmente, se ofrecen las conclusiones junto con recomendaciones para docentes e instituciones, implicaciones teóricas y prácticas, y posibles líneas futuras de investigación y transferencia.

MÉTODO

Este estudio adopta un enfoque cuantitativo, adecuado para la recolección y análisis de datos medibles (Hernández-Sampieri y Mendoza-Torres, 2023). Se enmarca en un diseño de investigación evaluativa, pertinente cuando se busca valorar procesos o intervenciones en contextos educativos (Arias, 2012), en este caso, la autopercepción de docentes en formación sobre la IA-Gen y su satisfacción con la experiencia de uso de ChatGPT en el diseño de SdAs.

Se plantean los siguientes objetivos (O):

- O1. Analizar la autopercepción de los docentes en formación sobre la IA-Gen en relación con su Familiaridad, Relevancia, Habilidades prácticas, Barreras, Confianza, Impacto ético y social, y Expectativas.
- O2. Conocer la valoración que realizan los docentes en formación sobre la utilidad de ChatGPT como herramienta de apoyo en el diseño de SdAs.

El primero responde a la necesidad, identificada en estudios previos, de comprender de forma integral cómo los futuros docentes conciben la IA-Gen en diversos aspectos, más allá de casos aislados (Markos et al., 2024). El segundo surge ante la creciente incorporación de herramientas de IA-Gen en educación y la falta de evidencias sobre cómo los docentes en formación perciben su utilidad pedagógica (Nyaaba y Zhai, 2024).

A partir de estos objetivos, se formularon las siguientes preguntas de investigación (PI):

- PI1. ¿Cuál es la autopercepción que tienen los docentes en formación sobre la IA-Gen en relación con su Familiaridad, Relevancia, Habilidades prácticas, Barreras, Confianza, Impacto ético y social, y Expectativas?

PI2. ¿Cómo valoran los docentes en formación la utilidad de ChatGPT como herramienta de apoyo en el diseño de SdAs?

La experiencia desarrollada tuvo como objetivo explorar el potencial pedagógico de ChatGPT en el diseño de SdAs dirigidas a las etapas de Educación Infantil y Primaria. Los participantes utilizaron principalmente ChatGPT-3.5, con acceso limitado a ChatGPT-4, según la disponibilidad del servicio. La experiencia se organizó en tres sesiones prácticas de dos horas durante tres semanas. La primera tuvo un carácter introductorio orientada a familiarizar al alumnado con el uso educativo de la IA-Gen y la formulación de prompts efectivos. Se proporcionaron prompts organizados en secuencias lógicas que facilitaron la construcción progresiva de las propuestas y prompts de revisión y análisis para analizar, ajustar y mejorar las producciones generadas, promoviendo la coherencia pedagógica, la toma de decisiones fundamentada y la mejora continua del diseño. La experiencia contó con acompañamiento docente para resolver aspectos técnicos y fomentar el pensamiento crítico y la toma de decisiones informadas. Se establecieron criterios de calidad para evaluar las respuestas generadas, centrados en la coherencia discursiva, la veracidad de la información y la adecuación curricular. Desde un enfoque activo, los estudiantes exploraron el uso educativo de la IA-Gen y fortalecieron sus competencias digitales en un entorno innovador.

Participantes

Para la selección de la muestra, se empleó un muestreo no probabilístico por conveniencia, justificado por el acceso directo a los participantes y la pertinencia de los contenidos. Se excluyeron aquellos estudiantes que no completaron la experiencia. Este tipo de muestreo puede conllevar ciertos sesgos, especialmente relacionados con la motivación o la familiaridad tecnológica. El tamaño muestral ($n = 185$) vino definido por el número de estudiantes matriculados. Participaron 174 estudiantes de la Universidad de Salamanca de titulaciones del ámbito educativo. Así, 37 pertenecían al Grado en Maestro en Educación Infantil (21,26 %), 58 al Grado en Maestro en Educación Primaria (33,33 %), 19 a la Doble titulación de Grado en Maestro en Educación Primaria y Educación Infantil (10,91 %), 27 al M. U. en Innovación en Didácticas Específicas para Educación Infantil y Primaria (15,51 %), y 33 al M.U. en Las TIC en Educación: Análisis y Diseño de Procesos, Recursos y Prácticas Formativas (18,96 %).

La participación estudiantil se realizó desde diferentes asignaturas relacionadas con tecnología e innovación educativa. Los estudiantes de los Grados participaron en la asignatura Las TIC en Educación, los del Máster en Innovación en la asignatura Avances en Tecnologías Digitales para la Innovación Educativa, y los del Máster TIC en la asignatura Líneas de Investigación en Tecnología Educativa.

En cuanto a la caracterización de los participantes, el grupo estuvo compuesto por 38 hombres (21,84 %) y 136 mujeres (78,16 %), con edades comprendidas entre los 17 y los 47 años y una media de edad de 21,02 años, con una desviación típica de 4,13.

Instrumento

El instrumento utilizado para recoger los datos fue un cuestionario validado en el trabajo de Espinoza-San Juan et al. (2024), elaborado mediante Google Forms y distribuido a través del campus virtual (Tabla 1). Este constaba de tres secciones. La primera recogía datos sociodemográficos: titulación, edad y sexo. La segunda incluía una escala compuesta por 22 ítems y siete dimensiones vinculadas a la autopercepción sobre la IA-Gen en educación: Familiaridad, Relevancia, Habilidades prácticas, Barreras, Confianza, Impacto ético y social, y Expectativas. Cada dimensión estaba integrada por entre dos y cuatro ítems tipo Likert con escala de respuesta de 1 (totalmente en desacuerdo) a 5 (totalmente de acuerdo). La tercera recogía la valoración sobre la utilidad de ChatGPT en el diseño de SdAs mediante cinco ítems: dos preguntas tipo Likert sobre Satisfacción y Utilidad percibida, y tres ítems de opción múltiple sobre la Calidad de las respuestas generadas, la Actitud adoptada por el participante y la Formación.

Tabla 1
Ítems del cuestionario

Segunda sección		
Dimensión	Ítem	Texto
Nivel de familiaridad con la IA-Gen	F01	Estoy familiarizado con conceptos básicos de la IA-Gen (como aprendizaje automático, redes neuronales, algoritmos, etc.).
	F02	Considero que poseo una comprensión amplia y detallada sobre las aplicaciones que puede tener la IA-Gen en educación.
	F03	Durante mi formación universitaria, he tenido la oportunidad de profundizar en el uso de la IA-Gen en el ámbito educativo.
	F04	He adquirido conocimientos sobre IA-Gen a través de cursos en línea, talleres y métodos de autoaprendizaje.
Percepción sobre la relevancia de la IA-Gen en educación	R01	Incorporar la IA-Gen es importante para actualizar las formas de enseñar y de aprender.
	R02	Las herramientas basadas en IA-Gen tienen el potencial de mejorar la calidad de la educación.
	R03	La integración de la IA-Gen en la educación es una tendencia actual.

Habilidades prácticas con herramientas de IA-Gen	H01	Me siento capaz de utilizar herramientas específicas de IA-Gen en educación (como chatbot educativos, sistemas de recomendación de contenidos, etc.).
	H02	He tenido experiencias prácticas exitosas, utilizando herramientas de IA-Gen para la educación (como chatbots educativos, sistemas de recomendación de contenidos, etc.).
	H03	Considero que puedo enseñar a otros a utilizar herramientas de IA-Gen en educación (como chatbots educativos, sistemas de recomendación de contenidos, etc.).
	H04	He compartido o enseñado mis experiencias prácticas con herramientas de IA-Gen en educación a otras personas. (como chatbots educativos, sistemas de recomendación de contenidos, etc.).
Barreras percibidas para la integración de la IA-Gen	B01	La falta de formación específica y especializada es un obstáculo para la integración de la IA-Gen en educación.
	B02	Las preocupaciones sobre la privacidad y seguridad de la información personal son una barrera para la adopción de la IA-Gen.
	B03	La resistencia al cambio por parte de algunos educadores es un desafío para la adopción de la IA-Gen.
Confianza en la integración de la IA-Gen en la enseñanza	C01	Me siento seguro/a de mi capacidad para integrar la IA-Gen en mis futuras clases.
	C02	Creo que la IA-Gen puede ser una herramienta valiosa en mi práctica docente.
Impacto ético y social de la IA en educación.	I01	Es importante considerar las implicaciones éticas de la IA-Gen en educación como el sesgo, la privacidad y la transparencia.
	I02	El desarrollo de herramientas de IA-Gen deben estar orientados hacia la inclusividad de todos los usuarios.
	I03	La adopción de la IA-Gen en educación exige una evaluación constante de su impacto ético y social.
Expectativas futuras sobre la IA-Gen en el campo educativo	E01	Espero que la IA-Gen tome un papel central y decisivo en la educación en el futuro.
	E02	La IA-Gen puede adaptar el aprendizaje según las necesidades individuales de cada estudiante.
	E03	Es probable que busque más formación en el uso de herramientas y técnicas de IA-Gen, específicamente para la educación.

Tercera sección

Dimensión	Ítem	Texto
Satisfacción	S01	Indica el grado de satisfacción con la experiencia realizada.

Utilidad percibida	U01	¿Consideras que Chat GPT es una buena herramienta para ayudar a los maestros en el diseño de SdAs?
Calidad de las respuestas generadas	Ca01	Consideras que las respuestas obtenidas con ChatGPT en su mayoría han sido: • Incorrectas • Parcialmente incorrectas • Incoherentes • Coherentes • Incompletas • Completas • Parcialmente correctas • Correctas
Actitud	A01	En relación con las respuestas que has obtenido de ChatGPT: • Has confiado plenamente en su veracidad y las has utilizado. • Has corroborado la veracidad de la respuesta contrastándola con otras fuentes y la has modificado parcialmente. • Has corroborado la veracidad de la respuesta contrastándola con otras fuentes y la has modificado completamente.
Formación	Fo01	Si quisiéramos que ChatGPT fuera utilizado por los maestros, consideras que: • Los alumnos y docentes deberían recibir formación sobre la herramienta en específico y sobre la IA-Gen en general. • Podríamos recibir una breve introducción sobre ChatGPT en particular, su contexto y ciertas advertencias sobre su uso. • La herramienta es de uso sencillo y es fácil adaptarse a ella. No es necesario ningún tipo de capacitación.

Fuente: Elaboración propia.

Recogida y análisis de datos

La recogida de datos se llevó a cabo entre octubre de 2024 y enero de 2025, con participación voluntaria, anónima y con consentimiento informado.

Se llevaron a cabo análisis descriptivos y relacionales para caracterizar estadísticamente las dimensiones evaluadas sobre la autopercepción y la utilidad de la IA-Gen en educación.

Cada dimensión del cuestionario se trató como una variable compuesta, resumida mediante su media (μ) y desviación estándar (σ), y se evaluó su normalidad con la prueba de Shapiro-Wilk para orientar la elección entre técnicas paramétricas o

no paramétricas. Aunque se presentan medias y desviaciones estándar, práctica habitual en investigación educativa, para analizar la relación entre la edad de los participantes y las distintas dimensiones se empleó el coeficiente de correlación de Spearman (ρ), dada su robustez al no requerir supuestos de normalidad ni linealidad, y su adecuación al tipo de escala utilizada.

Se exploraron diferencias entre los grupos de titulación en las dimensiones del cuestionario mediante la prueba de Kruskal-Wallis para muestras independientes al no cumplirse los supuestos de normalidad (Shapiro-Wilk) ni de homogeneidad de varianzas (Levene). Los resultados se presentan como mediana [rango intercuartílico], junto con el correspondiente valor p para la comparación entre grupos.

Se realizaron análisis exploratorios y de representación estructural mediante redes para explorar la estructura relacional entre las dimensiones del cuestionario. A partir de un análisis de correlación entre dimensiones compuestas, se generó un dendrograma jerárquico basado en distancias $1 - \rho$ (Spearman), lo que permitió identificar agrupamientos conceptuales entre las dimensiones, identificar la existencia de ejes temáticos interrelacionados en las autopercepciones de los participantes como conocimiento, valoración y obstáculos. Esta estructura facilitó la organización visual de una red de correlaciones, facilitando una representación intuitiva y relacional de los vínculos entre los constructos evaluados. Todos los análisis se realizaron utilizando el software libre R (versión 4.4.3) (R Core Team, 2024). Para favorecer la transparencia y la reproducibilidad, el código empleado para los análisis está disponible en el repositorio OSF: <https://osf.io/c83dh> (Pinto-Llorente et al., 2025). Dado que los datos contienen información sensible de estudiantes, estarán disponibles previa solicitud a los autores.

RESULTADOS

A continuación, se presentan los resultados obtenidos de acuerdo con los análisis realizados y las preguntas de investigación.

PI1. ¿Cuál es la autopercepción que tienen los docentes en formación sobre la IA-Gen en relación con su Familiaridad, Relevancia, Habilidades prácticas, Barreras, Confianza, Impacto ético y social y Expectativas?

Los resultados descriptivos muestran niveles medios-altos de autopercepción sobre la IA-Gen. Se observa un nivel moderado en las dimensiones Confianza ($\mu = 3,91$; $\sigma = 0,68$), Expectativas ($\mu = 3,72$; $\sigma = 0,68$), Familiaridad ($\mu = 3,56$; $\sigma = 0,73$) y Habilidades prácticas ($\mu = 3,56$; $\sigma = 0,76$), y niveles altos en Impacto ético y social ($\mu = 4,34$; $\sigma = 0,65$), Barreras ($\mu = 4,18$; $\sigma = 0,60$) y Relevancia ($\mu = 4,08$; $\sigma = 0,57$).

Se analizaron las asociaciones entre la edad de los participantes y las dimensiones del cuestionario. Los resultados muestran correlaciones de Spearman positivas y estadísticamente significativas entre la edad y las dimensiones Relevancia ($\rho = 0,30$; $p < 0,01$), Expectativas ($\rho = 0,20$; $p < 0,01$), e Impacto ético y social ($\rho = 0,40$;

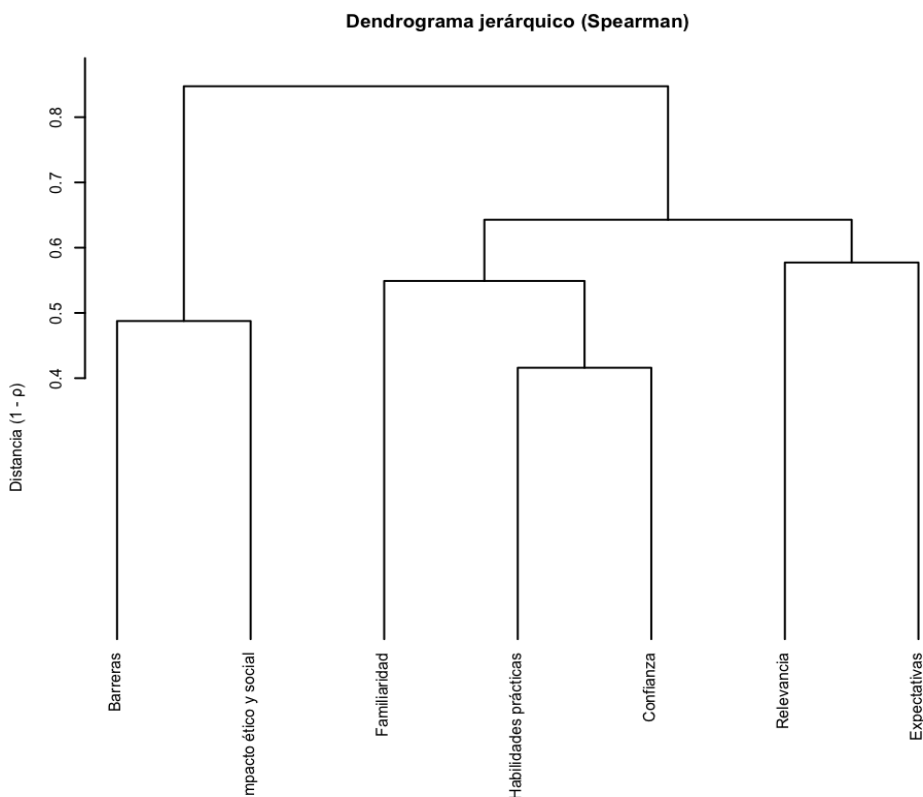
$p < 0,001$). También se analizaron las diferencias en las dimensiones relacionadas con la autopercepción de la IA-Gen en función de la titulación que cursaban los participantes. Se encontraron diferencias significativas ($p < 0,05$) en la dimensión Expectativas. El M. U. en las TIC en Educación obtuvo la mediana más alta (4 [3,67; 4,33]), mientras que las demás titulaciones presentaron valores ligeramente inferiores, reflejando una variabilidad moderada entre grupos. Por otro lado, en la dimensión Familiaridad, se detectó una diferencia cercana al umbral de significación estadística ($p = 0,051$). La familiaridad fue menor en el M. U. en Innovación en Didácticas Específicas (3,25 [3; 3,75]) y mayor en el Grado en Maestro en Educación Infantil y el M. U. en las TIC en Educación (ambos con una mediana de 3,75). En cuanto a la dimensión Barreras, aunque las diferencias no alcanzaron significación estadística ($p > 0,05$), el M. U. en Innovación en Didácticas Específicas mostró una mayor percepción de barreras (4,67 [4,33; 4,67]), en comparación con los Grados y el M. U. en las TIC en Educación (mediana de 4). En referencia a la dimensión Confianza, no se hallaron diferencias significativas ($p > 0,05$). Sin embargo, la Doble titulación de Grado en Maestro presentó la mediana más baja (3,50 [3,50; 4]), mientras que el resto de las titulaciones se situó en 4. En la dimensión Habilidades prácticas, no se observaron diferencias estadísticamente relevantes entre titulaciones ($p > 0,05$). Las medianas se mantuvieron entre 3,25 y 3,75, sin mostrar un patrón diferencial claro. Finalmente, en la dimensión Impacto ético y social, se observaron diferencias estadísticamente significativas ($p < 0,001$) entre titulaciones. El M. U. en Innovación en Didácticas Específicas obtuvo la mediana más elevada (5 [4,67; 5]), seguido del M. U. en las TIC en Educación (4,67 [4; 5]), frente a valores más bajos en las titulaciones de Grado.

Se calcularon los tamaños de efecto para cada comparación inferencial por titulación. La variable Edad ($\eta^2\_H = 0,678$) presentó el tamaño de efecto más elevado, con una diferencia marcada entre titulaciones. Entre las dimensiones del cuestionario, Impacto ético y social ($\eta^2\_H = 0,142$) mostró una diferencia de magnitud moderada. En cambio, dimensiones como Expectativas ($\eta^2\_H = 0,036$), Relevancia ($\eta^2\_H = 0,036$) y Familiaridad ($\eta^2\_H = 0,032$) presentaron tamaños de efecto pequeños, aunque consistentes con valores de p próximos o inferiores a 0,05. Barreras ($\eta^2\_H = 0,022$) y Confianza ($\eta^2\_H = 0,021$) mostraron efectos aún más reducidos, mientras que en Habilidades prácticas no se observaron efectos relevantes ($\eta^2\_H = -0,003$). En cuanto a la variable categórica Sexo, el tamaño de efecto fue moderado (V de Cramer = 0,372), indicando una distribución desigual entre titulaciones.

El dendrograma generado a partir de las correlaciones de Spearman (Figura 1) evidenció una organización jerárquica clara entre las dimensiones evaluadas. Se identificó un primer clúster, denominado Conocimiento, compuesto por Familiaridad, Habilidades prácticas y Confianza. Un segundo clúster, Valoración, agrupó las dimensiones Relevancia y Expectativas. Finalmente, Impacto ético y social, y Barreras conformaron el clúster Obstáculos.

Figura 1

Agrupamiento jerárquico de las dimensiones de autopercepción sobre la IA-Gen



El dendrograma fue utilizado como base para la representación de una red de correlaciones entre dimensiones (Figura 2), construida a partir de la matriz de correlaciones de Spearman. La red correlacional representa cada dimensión como un nodo. La disposición de nodos refleja fielmente los clústeres emergentes del análisis jerárquico (Conocimiento en blanco, Valoración en azul y Obstáculos en naranja), reforzando la validez empírica de la organización conceptual adoptada. El grosor de cada línea representa la intensidad de la correlación.

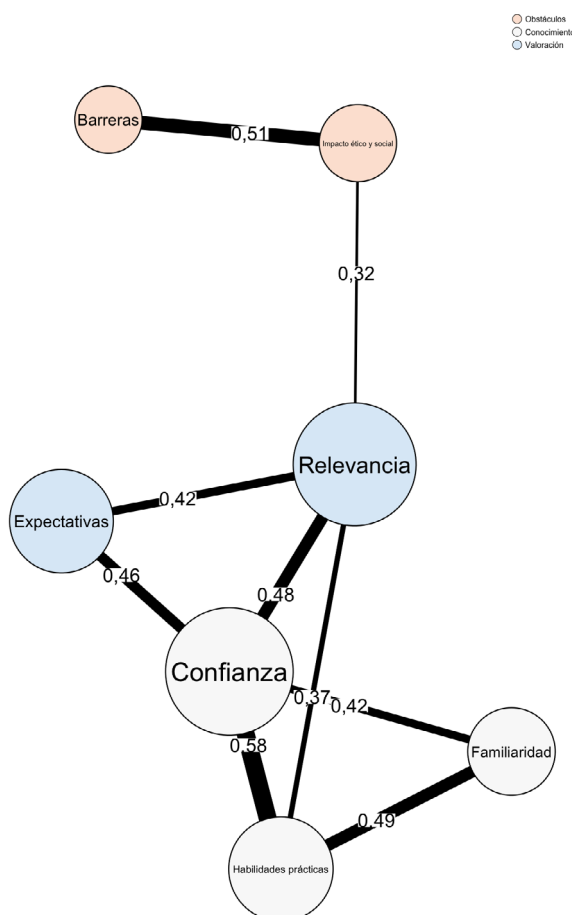
Se calcularon las métricas de centralidad de las dimensiones en la red de correlaciones, considerando un umbral de $|\rho| \geq 0,3$ para definir la presencia de

aristas. Las medidas incluyeron el grado (número de conexiones significativas), la intermediación y la cercanía.

Las dimensiones Relevancia y Confianza destacaron por tener el mayor grado (4 conexiones cada una), así como los valores más altos de intermediación (8,5 y 3,0; respectivamente), lo que indica que actúan como nodos centrales en la red. También Habilidad presentó una conectividad moderada (grado = 3). Por el contrario, dimensiones como Barreras, Familiaridad y Expectativas mostraron menor grado y baja intermediación, lo que sugiere un rol más periférico.

Figura 2

Red de correlaciones de las dimensiones de autopercepción sobre la IA-Gen en educación



PI2. ¿Cómo valoran los docentes en formación la utilidad de ChatGPT como herramienta de apoyo en el diseño de SdAs?

La valoración de ChatGPT como herramienta de apoyo en el diseño de SdAs fue, en general, muy positiva. La variable Satisfacción alcanzó un valor de 4,45 ($\sigma = 0,68$), y la Utilidad percibida de 4,24 ($\sigma = 0,84$). Para evaluar la calidad de las respuestas generadas, se empleó un ítem dicotómico (“Sí”/“No”). Los resultados que se detallan corresponden únicamente al grupo que respondió afirmativamente. La mayoría valoró las respuestas como coherentes ($n = 82$; 66,70 %), un menor porcentaje las consideró incorrectas ($n = 20$; 16,30 %), completas ($n = 19$; 15,40 %), parcialmente correctas ($n = 17$; 13,80 %), parcialmente incorrectas ($n = 14$; 11,40 %), incompletas ($n = 14$; 11,40 %) o incoherentes ($n = 7$; 5,70 %). Ninguna fue evaluada como completamente correcta.

Respecto a la Actitud frente a la respuesta generada, la mayoría ($n = 138$, 79,30 %) señaló que contrastó la respuesta y la modificó parcialmente. Una proporción menor ($n = 27$, 15,50 %) indicó que la utilizó directamente, sin realizar cambios. Solo un grupo muy reducido ($n = 9$, 5,20 %) rechazó por completo la respuesta proporcionada por la IA-Gen.

En coherencia con esta actitud crítica y reflexiva, un amplio porcentaje de la muestra ($n = 117$, 67,20 %) manifestó que considera necesaria la formación sobre el uso de ChatGPT y otras herramientas de IA-Gen para docentes y estudiantes. Un 27,60 % ($n = 48$) indicó que dicha formación debería dirigirse exclusivamente a docentes, y solo un 5,20 % ($n = 9$) consideró que no era necesaria.

Se exploraron correlaciones para identificar factores asociados a una valoración más positiva. Los resultados revelan correlaciones positivas y estadísticamente significativas entre la Utilidad percibida de ChatGPT y las siguientes dimensiones:

- Confianza ($\rho = 0,36$; $p < 0,001$)
- Relevancia ($\rho = 0,23$; $p = 0,0019$)
- Habilidades prácticas ($\rho = 0,20$; $p = 0,0088$)

Estas correlaciones indican que quienes tienen más Confianza, poseen mayores Habilidades prácticas y perciben la IA-Gen como Relevante en educación, tienden a valorar de forma más positiva la Utilidad de ChatGPT para el diseño de SdAs.

Se analizó el coeficiente de fiabilidad interna ω de McDonald para las siete dimensiones de la segunda sección del cuestionario. Todas mostraron niveles adecuados de consistencia interna: Familiaridad ($\omega = 0,80$), Relevancia ($\omega = 0,72$), Habilidades prácticas ($\omega = 0,85$), Barreras ($\omega = 0,70$), Impacto ético-social ($\omega = 0,81$) y Expectativas ($\omega = 0,74$). En el caso de la dimensión Confianza, compuesta por solo dos ítems, no fue posible calcular el coeficiente. Del mismo modo, las cuatro dimensiones de la tercera sección, formadas por un único ítem, tampoco permitieron su estimación. Respecto a la dimensión Calidad de las respuestas generadas, se obtuvo

$\omega = 0,65$. La literatura establece valores aceptables entre 0,70 y 0,90, incluyendo valores superiores a 0,65 (Ventura-León y Caycho-Rodríguez, 2017).

DISCUSIÓN

Esta investigación se centró en conocer la autopercepción de los docentes en formación sobre la IA-Gen y la utilidad de ChatGPT en el diseño de SdAs. Respecto a la primera pregunta de investigación, las dimensiones con niveles más altos fueron Impacto ético y social, Barreras y Relevancia, lo que evidencia una conciencia crítica hacia la IA-Gen. En cuanto al Impacto ético y social, numerosos estudios revelan los mismos desafíos encontrados en cuanto a sesgo, privacidad o transparencia (Ishmuradova et al., 2025; Lozano y Blanco, 2023; Markos et al., 2024), así como la falta de acceso a la IA-Gen que podría contribuir a ampliar las brechas educativas generadas por la falta de acceso equitativo (Kasneci et al., 2023) incluso limitando el acceso al conocimiento (UNESCO, 2023). Por otra parte, Jo (2024) encontró que la preocupación por la privacidad se asocia negativamente con su uso. Asimismo, subrayan la necesidad de integrar la IA-Gen en los contextos educativos como herramienta de apoyo docente y de formar específicamente al profesorado en su uso (Ishmuradova et al., 2025; Lozano y Blanco, 2023; Markos et al., 2024). En cuanto a la dimensión Relevancia, los estudios previos señalan los enormes beneficios que puede ofrecer la IA-Gen para mejorar el proceso de enseñanza-aprendizaje (Ishmuradova et al., 2025; Markos et al., 2024).

Por su parte, las dimensiones Confianza, Expectativas, Familiaridad y Habilidades prácticas obtuvieron niveles medios. Estos resultados están alineados con los de Kelly et al. (2023), quienes evidencian que la confianza y familiaridad con la IA-Gen suele ser escasa, incrementándose con la experiencia. Otros estudios sugieren que demostrar el beneficio del uso de chatbots de IA en educación puede favorecer su adopción y uso eficaz (Jo, 2024). En relación con las Expectativas, los resultados se alinean con otras investigaciones que señalan la capacidad de la IA-Gen para adaptar el aprendizaje a las necesidades del alumnado, lo que refuerza su papel en la educación actual y futura (Ruiz-Rojas et al., 2023; Sánchez Vera, 2024; Sánchez-Prieto et al., 2025). En cuanto a las Habilidades prácticas, los resultados coinciden con los de Ng et al. (2023) quienes señalan que los futuros docentes necesitan más experiencias para conocer, usar y enseñar con herramientas de IA-Gen. Cabe señalar que, dado el carácter correlacional del estudio, estas relaciones no implican causalidad.

La investigación evidenció una organización jerárquica de las dimensiones estudiadas, agrupadas con base en tres clústeres: Conocimiento, Valoración y Obstáculos respecto a la IA-Gen. Esta organización resulta útil para orientar las acciones futuras con docentes en formación, al ofrecer un marco conceptual claro sobre sus percepciones y necesidades.

Los resultados de la red muestran que las dimensiones Relevancia y Confianza ocupan posiciones centrales en la estructura del cuestionario, actuando como nodos de conexión clave, facilitando la articulación entre bloques como Familiaridad, Habilidades prácticas y Expectativas. En términos interpretativos, esto sugiere que los que se sienten seguros integrando la IA-Gen en su práctica docente (alta Confianza) y que la perciben como una herramienta valiosa (alta Relevancia), tienden también a mostrar mayor conocimiento previo, mayores habilidades prácticas y mayores expectativas de uso futuro. Algunos informes señalan que la confianza en la IA-Gen incrementa su intención de uso (Ivanov et al., 2024). Su papel articulador, confirmado por su posición nodal en esta investigación, ya había sido señalado como puente entre actitudes favorables y preocupaciones críticas. En general, los resultados evidencian un equilibrio entre actitudes positivas y conciencia crítica. Las puntuaciones elevadas en Expectativas y Relevancia conviven con valores también altos en Impacto ético-social y Barreras prácticas. Este patrón refleja que los encuestados reconocen el potencial de la IA-Gen para mejorar la enseñanza (Kasneci et al., 2023), sin ignorar riesgos y limitaciones. La percepción de obstáculos se modera con expectativas altas, indicando una actitud equilibrada que oscila entre optimismo fundado y prudencia crítica.

La inclusión de los tamaños de efecto permitió interpretar con mayor precisión la relevancia práctica de las diferencias entre titulaciones. Aunque algunas comparaciones no alcanzaron significación estadística, los tamaños de efecto pequeños en dimensiones como Expectativas, Relevancia o Familiaridad sugieren diferencias sutiles pero potencialmente relevantes en contextos formativos. Se observaron patrones claros según la titulación: los estudiantes de grado mostraron niveles más bajos en Expectativas y Familiaridad con la IA-Gen, así como una leve tendencia a percibir más Barreras que los de máster. Esto sugiere que los programas de posgrado promueven mayor confianza y expectativas, posiblemente por un currículo más avanzado o una experiencia previa más amplia.

Destaca especialmente la dimensión ética y social, en la que el estudiantado de máster superó claramente al de grado. En conjunto, estos hallazgos refuerzan la necesidad de considerar la magnitud real de las diferencias para orientar propuestas educativas ajustadas al perfil del alumnado. En grado, conviene activar el interés y la familiarización; en posgrado, reforzar una mirada crítica y reflexiva sobre la IA-Gen.

Estudios previos indican que la familiaridad y predisposición hacia tecnologías innovadoras tiende a ser mayor en niveles superiores (Saihi et al., 2025). Por tanto, las intervenciones formativas deberían adaptarse al perfil: en grado, reforzar motivación y exposición temprana a la IA-Gen; y en posgrado, profundizar en competencias críticas y colaborativas (Saihi et al., 2025; Wang y Wang, 2022). Sin embargo, esta diferenciación debe tomarse con precaución, pues la muestra fue autoseleccionada y no se aplicó triangulación cualitativa, limitando la generalización especialmente entre niveles formativos. Estos hallazgos son coherentes con investigaciones previas que destacan la importancia de la alfabetización digital crítica (Wang et al., 2024) y

el acompañamiento formativo para facilitar la adopción responsable de tecnologías emergentes en educación (Bearman y Ajjawi, 2023).

Estas reflexiones cobran especial relevancia en contextos de educación en línea o híbrida, donde el uso de herramientas como ChatGPT puede apoyar el diseño y desarrollo de SdAs, favoreciendo la personalización y adaptación didáctica en modalidades no presenciales.

Las implicaciones para la formación docente resultan evidentes, los programas deben integrar de forma equilibrada componentes técnicos, críticos y éticos. Se recomienda desarrollar competencias técnicas mediante instrucción práctica con herramientas como ChatGPT, comprensión básica de sus algoritmos y actividades didácticas específicas (Wang y Wang, 2022). Por otra parte, fomentar el pensamiento crítico y la alfabetización digital, por ejemplo, a través de entrenamiento para evaluar la calidad de los contenidos generados por IA-Gen y promoción de estrategias para verificar información y analizar sesgos (Kasneci et al., 2023). También favorecer la reflexión ética y deontológica mediante debates y sensibilización sobre los desafíos éticos (privacidad, equidad, dependencia tecnológica) en línea con el llamado a un enfoque pedagógico ético en IA-Gen (Kasneci et al., 2023).

Atendiendo a la segunda pregunta de investigación, se alcanzaron valores altos en Satisfacción y Utilidad percibida con ChatGPT. Estos resultados sugieren que esta tecnología puede desempeñar un papel relevante en procesos de planificación didáctica y producción de contenidos. Un hallazgo destacado fue la Actitud crítica del alumnado que optó mayoritariamente por verificar la información y adaptarla, mostrando competencia metacognitiva en el uso de IA-Gen. Esto es relevante dado que IA-Gen puede producir respuestas erróneas o sesgadas, confirmando que el uso efectivo de ChatGPT depende también del juicio pedagógico y ético. Además, la mayoría consideró necesaria formación específica, reforzando la idea de que su integración requiere acompañamiento formativo, coherente con estudios previos que destacan la necesidad de marcos éticos, pedagógicos y operativos para el uso adecuado de IA-Gen (Ivanov et al., 2024; Kasneci et al., 2023).

En resumen, el surgimiento de la IA-Gen ha generado gran interés por conocer las necesidades de los futuros docentes para su incorporación eficaz en las aulas (Ogunleye et al., 2024). Las evidencias revelan una predisposición positiva y conocimiento sobre los beneficios educativos (Buyakova et al., 2024). Whitbread et al. (2025) alertan de la necesidad de establecer políticas que aseguren formación y abordaje de desafíos. Esto incluye programas de alfabetización en IA para que los docentes integren estas tecnologías con confianza y eficacia (Al-Abdullatif, 2024; Lozano y Blanco, 2023; Omar et al., 2024), promoviendo un uso informado y responsable (Nyaaba y Zhai, 2024; Markos et al., 2024; Okunade, 2024).

CONCLUSIONES

Este estudio, de enfoque cuantitativo y diseño evaluativo, tuvo como propósito analizar la autopercepción del profesorado en formación sobre la IA-Gen y valorar su experiencia de uso de ChatGPT en el diseño de SdAs. Con este fin, se aplicó un cuestionario tras una experiencia formativa, lo que permitió recopilar datos objetivos sobre dimensiones clave, así como sobre la percepción de utilidad de la herramienta. Los resultados evidencian una disposición favorable hacia su integración, con niveles elevados de Familiaridad, Habilidades prácticas, Expectativas positivas y valoración del Impacto ético y social. La dimensión Confianza emerge como un eje estructural clave, actuando como nexo entre el Conocimiento, la Actitud y la disposición al uso.

El análisis de la red de correlaciones revela que la Confianza docente no solo refleja autoeficacia, sino también una actitud predispuesta a la adopción crítica de tecnologías emergentes. A su vez, se observa una dualidad entre entusiasmo por el potencial de la IA-Gen y conciencia de sus riesgos éticos y prácticos. Esta coexistencia subraya la necesidad de programas formativos que combinen entrenamiento instrumental con espacios de reflexión crítica. Los hallazgos también sugieren que la percepción de barreras no se asocia principalmente con déficits técnicos, sino con sensibilidades éticas y pedagógicas. La posición periférica de esta dimensión en la red refuerza la importancia de abordarla desde enfoques contextualizados y no exclusivamente desde la capacitación tecnológica.

Con base en la literatura científica y los hallazgos exploratorios mediante análisis de redes, proponemos un modelo conceptual que articula las relaciones esperadas entre las distintas dimensiones de autopercepción sobre la IA-Gen en docentes en formación.

En este modelo, Familiaridad y Habilidades prácticas se agrupan en un clúster denominado Conocimiento previo, que actúan como antecedente y favorece el desarrollo de Confianza en el uso de herramientas de IA-Gen. Esta Confianza se concibe como una dimensión mediadora central, que influye tanto en la percepción de Relevancia como en una menor percepción de Barreras para su implementación. A su vez, ambas inciden en la construcción de Expectativas futuras respecto al uso educativo de esta tecnología.

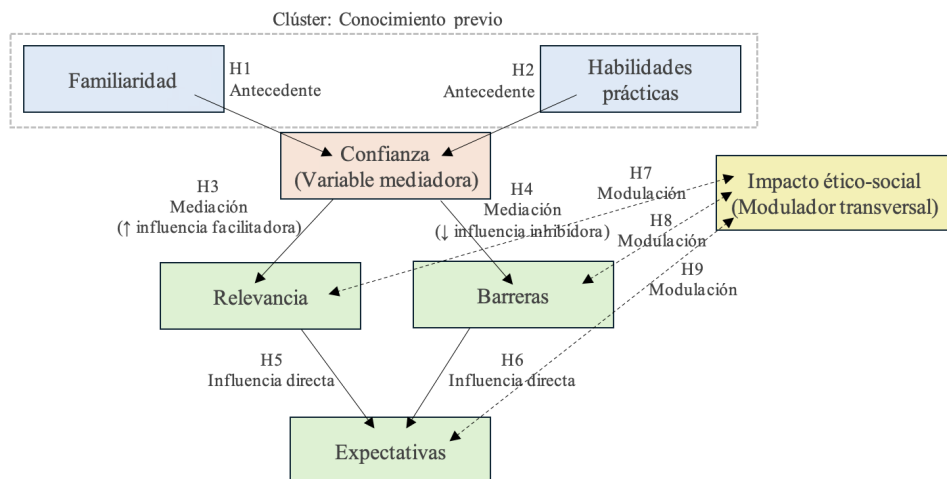
Impacto ético y social se posiciona como una dimensión transversal que modula críticamente las actitudes hacia la adopción de la IA-Gen, especialmente en quienes perciben mayor relevancia y confianza, pero también mayor conciencia sobre los desafíos éticos y sociales. Esta disposición crítica se asocia con una actitud reflexiva observable en el uso de herramientas como ChatGPT.

El modelo anticipa que la Confianza opera como nodo articulador entre el conocimiento previo y la disposición futura hacia la IA-Gen, mientras que el Impacto ético actúa como un regulador crítico de estas actitudes.

Este modelo se alinea con el *Technology Acceptance Model* (TAM), centrado en la utilidad percibida y actitud hacia la tecnología, y con el AI CFT, que integra competencias éticas y pedagógicas para un uso responsable de la IA. En síntesis, el modelo articula las relaciones hipotéticas representadas en la Figura 3.

Figura 3

Modelo conceptual: Relaciones hipotéticas entre dimensiones de autopercepción sobre IA-Gen



A partir de estos resultados, se plantean cinco líneas de actuación para integrar eficazmente la IA-Gen en la formación docente: (1) formación continua que combine uso técnico e implicaciones éticas; (2) aprendizaje práctico con actividades aplicadas; (3) desarrollo del pensamiento crítico; (4) apoyo inicial para fortalecer la confianza docente; y (5) evaluación continua del impacto ético y social. Asimismo, se insta a que las instituciones integren estos principios en sus políticas y currículos, promoviendo entornos híbridos y colaborativos que faciliten el uso equitativo de la IA-Gen y fortalezcan comunidades de práctica en el desarrollo profesional docente.

La IA-Gen ofrece una oportunidad estratégica para transformar la educación, enriqueciendo la enseñanza, fomentando la innovación y democratizando el acceso al conocimiento. Su implementación requiere una reflexión pedagógica profunda, políticas claras y formación continua, enfocándose no solo en capacidades tecnológicas, sino en su integración didáctica adaptada a cada contexto y promoviendo una cultura digital crítica. Desde esta perspectiva, los hallazgos profundizan en la actitud del profesorado en formación hacia una adopción crítica de la IA-Gen y establecen bases para que futuros docentes integren esta tecnología en sus aulas, formando a estudiantes de etapas obligatorias en su uso ético y creativo.

Este estudio presenta algunas limitaciones. Su diseño transversal impide establecer causalidad, y se basa únicamente en datos autoinformados, sin medidas objetivas de uso real, lo que puede afectar la precisión. Además, no incluye comparación con otras tecnologías ni seguimiento longitudinal, limitando el análisis

del valor específico de ChatGPT y la evolución de percepciones a lo largo del tiempo. La muestra no es representativa de todos los niveles formativos, y falta triangulación cualitativa que aporte mayor profundidad.

Por ello, se recomienda que futuras investigaciones sean longitudinales e interinstitucionales, integren métodos mixtos con triangulación cualitativa (entrevistas, grupos focales) y medidas objetivas de uso, además de comparar sistemáticamente la IA-Gen con otras tecnologías educativas para identificar sus ventajas y limitaciones específicas.

Agradecimientos

Esta investigación se realizó con el marco del Proyecto de Innovación docente: Aprendizaje personalizado y adaptativo a través de sistemas de inteligencia artificial generativa en Ciencias de la Educación (ID2024/260), apoyado por el proyecto SA217P23 financiado por la Junta de Castilla y León.

REFERENCIAS

- Al-Abdullatif, A. M. (2024). Modeling teachers' acceptance of generative artificial intelligence use in higher education: The role of AI literacy, intelligent TPACK, and perceived trust. *Education Sciences*, 14(11), 1209. <https://doi.org/10.3390/educsci14111209>
- Al-Abdullatif, A. M. y Alsubaie, M. A. (2024). ChatGPT in learning: Assessing students' use intentions through the lens of perceived value and the influence of AI literacy. *Behavioral Sciences*, 14(9), 845. <https://doi.org/10.3390/bs14090845>
- Alasadi, E. A. y Baiz, C. R. (2023). Generative AI in education and research: Opportunities, concerns, and solutions. *Journal of Chemical Education*, 100(8), 2965-2971. <https://doi.org/10.1021/acs.jchemed.3c00323>
- Alba Pastor, C. (2022). *Enseñar pensando en todos los estudiantes: El modelo de diseño universal para el aprendizaje (DUA)*. Ediciones SM.
- Aravena Castillo, F. (2013). Developing the collaborative model in the initial formation: The autoperception of the professional performance of the beginner teacher in action. *Estudios Pedagógicos*, 39(1), 27-44. <https://doi.org/10.4067/S0718-07052013000100002>
- Arias, F. G. (2012). *El proyecto de investigación*. Episteme.
- Bayly-Castaneda, K., Ramírez-Montoya, M.-S. y Morita-Alexander, A. (2024). Crafting personalized learning paths with AI for lifelong learning: A systematic literature review. *Frontiers in Education*, 9, Article 1424386. <https://doi.org/10.3389/feduc.2024.1424386>
- Bearman, M. y Ajjawi, R. (2023). Learning to work with the black box: Pedagogy for a world with artificial intelligence. *British Journal of Educational Technology*, 54(5), 1160-1173. <https://doi.org/10.1111/bjet.13337>
- Buyakova, K. I., Dmitriev, Ya. A., Ivanova, A. S., Feshchenko, A. V. y Yakovleva, K. I. (2024). Students' and teachers' attitudes towards the use of tools with generative artificial intelligence at the university. *The Education and Science Journal*, 26(7),

- 160-193. <https://doi.org/10.17853/1994-5639-2024-7-160-193>
- Carranza Alcántar, M. del R., Macías González, G. G., Gómez Rodríguez, H., Jiménez Padilla, A. A. y Jacobo Montes, F. M. (2024). Percepciones docentes sobre la integración de aplicaciones de IA generativa en el proceso de enseñanza universitario. *REDU. Revista de Docencia Universitaria*, 22(2), 21-40. <https://doi.org/10.4995/redu.2024.22027>
- Currie, G. M. (2025). Generative artificial intelligence in nuclear medicine education. *Journal of Nuclear Medicine Technology*, 53(1), 72-79. <https://doi.org/10.2967/jnmt.124.268323>
- Denny, P., Prather, J., Becker, B. A., Finnie-Ansley, J., Hellas, A., Leinonen, J., Luxton-Reilly, A., Reeves, B. N., Santos, E. A. y Sarsa, S. (2024). Computing education in the era of generative AI. *Communications of the ACM*, 67(2), 56-67. <https://doi.org/10.1145/3624720>
- Espinoza-San Juan, J., Raby, M. D. y Sagredo-Lillo, E. (2024). Validación de un cuestionario sobre las percepciones y usos de la IA-Gen entre estudiantes de pedagogía. *RISTI: Revista Ibérica de Sistemas e Tecnologías de Informação*, (70), 574-585.
- Gozalo-Brizuela, R. y Garrido-Merchán, E. E. (2024). A survey of generative AI applications. *Journal of Computer Science*, 20(8), 801-818. <https://doi.org/10.3844/jcssp.2024.801.818>
- Haroud, S. y Saqri, N. (2025). Generative AI in higher education: Teachers' and students' perspectives on support, replacement, and digital literacy. *Education Sciences*, 15(4), 396. <https://doi.org/10.3390/educsci15040396>
- Hernández-Sampieri, R. y Mendoza-Torres, C. P. (2023). *Metodología de la investigación*. McGraw-Hill.
- Ishmuradova, I. I., Zhdanov, S. P., Kondrashev, S. V., Erokhova, N. S., Grishnova, E. E. y Volosova, N. Y. (2025). Pre-service science teachers' perception on using generative artificial intelligence in science education. *Contemporary Educational Technology*, 17(3), ep579. <https://doi.org/10.30935/cedtech/16207>
- Islam, G. y Greenwood, M. (2024). Generative artificial intelligence as hypercommons: Ethics of authorship and ownership. *Journal of Business Ethics*, 192(3), 659-663. <https://doi.org/10.1007/s10551-024-05741-9>
- Ivanov, S., Soliman, M., Tuomi, A., Alkathiri, N. A. y Al-Alawi, A. N. (2024). Drivers of generative AI adoption in higher education through the lens of the theory of planned behaviour. *Technology in Society*, 77, 102521. <https://doi.org/10.1016/j.techsoc.2024.102521>
- Jo, H. (2024). From concerns to benefits: A comprehensive study of ChatGPT usage in education. *International Journal of Educational Technology in Higher Education*, 21, 35. <https://doi.org/10.1186/s41239-024-00471-4>
- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeiffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Kelly, A., Sullivan, M. y Strampel, K. (2023). Generative artificial intelligence: University student awareness, experience, and confidence in use across disciplines. *Journal of University Teaching & Learning Practice*, 20(6). <https://doi.org/10.53761/1.20.6.12>
- Laupichler, M. C., Aster, A., Schirch, J. y Raupach, T. (2022). Artificial intelligence literacy in higher and adult education: A scoping literature review. *Computers*

- and Education: Artificial Intelligence, 3, 100101. <https://doi.org/10.1016/j.caeai.2022.100101>
- Lozano, A. y Blanco Fontao, C. (2023). Is the education system prepared for the irruption of artificial intelligence? A study on the perceptions of students of primary education degree from a dual perspective: Current pupils and future teachers. *Education Sciences*, 13(7), 733. <https://doi.org/10.3390/educsci13070733>
- Markos, A., Prentzas, J. y Sidiropoulou, M. (2024). Pre-service teachers' assessment of ChatGPT's utility in higher education: SWOT and content analysis. *Electronics*, 13(10), 1985. <https://doi.org/10.3390/electronics13101985>
- Miao, F. y Cukurova, M. (2024). *AI competency framework for teachers*. UNESCO. <https://doi.org/10.54675/ZJTE2084>
- Moorhouse, B. L. y Kohnke, L. (2024). The effects of generative AI on initial language teacher education: The perceptions of teacher educators. *System*, 122, 103290. <https://doi.org/10.1016/j.system.2024.103290>
- Moorhouse, B. L., Wan, Y., Wu, C., Kohnke, L., Ho, T. Y. y Kwong, T. (2024). Developing language teachers' professional generative AI competence: An intervention study in an initial language teacher education course. *System*, 125, 103399. [https://doi.org/10.1016/j/system.2024.103399](https://doi.org/10.1016/j.system.2024.103399)
- Mouta, A., Torrecilla-Sánchez, E. M. y Pinto-Llorente, A. M. (2025). Comprehensive professional learning for teacher agency in addressing ethical challenges of AIED: Insights from educational design research. *Education and Information Technologies*, 30, 3343-3387. <https://doi.org/10.1007/s10639-024-12946-y>
- Ng, D. T. K., Leung, J. K. L., Su, J., Ng, R. C. W. y Chu, S. K. W. (2023). Teachers' AI digital competencies and twenty-first century skills in the post-pandemic world. *Educational Technology Research and Development*, 71(1), 137-161. <https://doi.org/10.1007/s11423-023-10203-6>
- Nyaaba, M. y Zhai, X. (2024). Generative AI professional development needs for teacher educators. *Journal of AI*, 8(1), 1-13. <https://doi.org/10.61969/jai.1385915>
- Ogunleye, B., Zakariyyah, K. I., Ajao, O., Olayinka, O. y Sharma, H. (2024). A systematic review of generative AI for teaching and learning practice. *Education Sciences*, 14(6), 636. <https://doi.org/10.3390/educsci14060636>
- Okunade, A. I. (2024). The role of artificial intelligence in teaching of science education in secondary schools in Nigeria. *European Journal of Computer Science and Information Technology*, 12(1), 57-67. <https://doi.org/10.37745/ejsit2013/vol12n15767>
- Omar, A., Shaqour, A. Z. y Khlaif, Z. N. (2024). Attitudes of faculty members in Palestinian universities toward employing artificial intelligence applications in higher education: Opportunities and challenges. *Frontiers in Education*, 9, 1414606. <https://doi.org/10.3389/educ.2024.1414606>
- Pinto-Llorente, A. M., Izquierdo-Álvarez, V. y Dolcet-Negre, M. M. (2025). *Autopercepción y utilidad de la inteligencia artificial generativa en docentes en formación* [Open Science Framework]. <https://osf.io/c83dh>
- R Core Team. (2024). *R: A language and environment for statistical computing* (Version 4.4.3) [Computer software]. R Foundation for Statistical Computing. <https://www.r-project.org/>
- Ruiz-Rojas, L. I., Acosta-Vargas, P., De-Moreta-Llovet, J. y Gonzalez-Rodriguez, M. (2023). Empowering education with generative artificial intelligence tools: Approach with an instructional design matrix. *Sustainability*, 15(15), Article 11524. <https://doi.org/10.3390/su151511524>

- Saihi, A., Ben-Daya, M. y Hariga, M. (2025). The moderating role of technology proficiency and academic discipline in AI-chatbot adoption within higher education: Insights from a PLS-SEM analysis. *Education and Information Technologies*, 30, 5843-5881. <https://doi.org/10.1007/s10639-024-13023-0>
- Sánchez Vera, M. del M. (2024). La inteligencia artificial como recurso docente: Usos y posibilidades para el profesorado. *Educación*, 60(1), 33-47. <https://doi.org/10.5565/rev/educar.1810>
- Sánchez-Prieto, J. C., Izquierdo-Álvarez, V., del Moral-Marcos, M. T. y Martínez-Abad, F. (2025). Inteligencia artificial generativa para autoaprendizaje en educación superior: Diseño y validación de una máquina de ejemplos. *RIED-Revista Iberoamericana de Educación a Distancia*, 28(1), 59-81. <https://doi.org/10.5944/ried.28.1.41548>
- Storey, V. C., Yue, W. T., Zhao, J. L. y Lukyanenko, R. (2025). Generative artificial intelligence: Evolving technology, growing societal impact, and opportunities for information systems research. *Information Systems Frontiers*. <https://doi.org/10.1007/s10796-025-10581-7>
- Tai-Han, L., Hsing-Yi, C., Ming-Jr, J., Chih-Kai, C., Chong-Lih, P., Guo-Shiou, L., Jyh-Cherng, Y., Ming-Shen, D., Cheng-Ping, Y. y Hung-Sheng, S. (2024). An advanced machine learning model for a web-based artificial intelligence-based clinical decision support system application: Model development and validation study. *Journal of Medical Internet Research*, 26, e56022. <https://doi.org/10.2196/56022>
- UNESCO. (2023). *Education 2030 agenda*. <https://www.unesco.org/en/digital-education/artificial-intelligence>
- Ventura-León, J. L. y Caycho-Rodríguez, T. (2017). El coeficiente omega: Un método alternativo para la estimación de la confiabilidad. *Revista Latinoamericana de Ciencias Sociales, Niñez y Juventud*, 15(1), 625-627.
- Wang, K., Ruan, Q., Zhang, X., Fu, C. y Duan, B. (2024). Pre-service teachers' GenAI anxiety, technology self-efficacy, and TPACK: Their structural relations with behavioral intention to design GenAI-assisted teaching. *Behavioral Sciences*, 14(5), 373. <https://doi.org/10.3390/bs14050373>
- Wang, Y. Y. y Wang, Y. S. (2022). Development and validation of an artificial intelligence anxiety scale: An initial application in predicting motivated learning behavior. *Interactive Learning Environments*, 30(4), 619-634. <https://doi.org/10.1080/10494820.2019.1674887>
- Whitbread, M., Hayes, C., Prabhakar, S. y Upsher, R. (2025). Exploring university staffs' perceptions of using generative artificial intelligence at university. *Education Sciences*, 15(3), 367. <https://doi.org/10.3390/educsci15030367>
- Zhang, C. y Villanueva, L. E. (2023). Generative artificial intelligence preparedness and technological competence. *International Journal of Education and Humanities*, 11(2), 164-170. <https://doi.org/10.54097/ijeh.v11i2.13753>

Fecha de recepción del artículo: 1 de junio de 2025

Fecha de aceptación del artículo: 18 de agosto de 2025

Fecha de aprobación para maquetación: 22 de septiembre de 2025

Fecha de publicación en OnlineFirst: 14 de octubre de 2025

Fecha de publicación: 1 de enero de 2026

Docentes y divulgación científica en redes sociales: desarrollo y validación inicial de instrumentos de evaluación

Teachers and science communication on social media: development and initial validation of assessment instruments



Dieter Reynaldo Fuentes-Cancell - *Universidad de Valladolid, UVA (España)*

Odiel Estrada-Molina - *Universidad de Valladolid, UVA (España)*

Mónica Gutiérrez-Ortega - *Universidad de Valladolid, UVA (España)*

RESUMEN

En el contexto actual de educación digital y ciencia abierta, el profesorado universitario no solo genera conocimiento mediante su producción científica y profesional, sino que además debe saber divulgarlo a públicos académicos y no especializados. Aunque vinculada a la competencia digital, la competencia mediática aporta un enfoque crítico-comunicativo necesario para este proceso, especialmente en redes sociales digitales (RSD). Este estudio diseña y valida dos constructos para evaluar la competencia mediática del profesorado (CMP): uno centrado en el uso general de RSD y otro específico para LinkedIn. Ambos modelos integran el Marco Común de Competencia Digital Docente 2.2 y el modelo de alfabetización mediática de Ferrés y Piscitelli (2012). La validación se realizó mediante revisión teórica, juicio de expertos ($n = 30$) y análisis factorial exploratorio (AFE). La idoneidad de los datos se confirmó mediante índice KMO (> 0.80) y prueba de esfericidad de Bartlett; ambos instrumentos mostraron alta fiabilidad ($\alpha > 0.85$; $\omega > 0.87$). El AFE identificó seis dimensiones teóricas, explicando el 78.2 % de la varianza para RSD y el 78.8 % para LinkedIn. Estos hallazgos ofrecen una aproximación inicial a la estructura de la CMP y su potencial diagnóstico en la formación docente, destacando el papel estratégico de LinkedIn como entorno profesional para la divulgación académica. Se subraya la necesidad de incorporar estrategias comunicativas y mediáticas en la formación del profesorado. Como línea futura, se recomienda realizar análisis factorial confirmatorio con muestras más amplias para consolidar los resultados obtenidos.

Palabras clave: competencia mediática; profesores universitarios; redes sociales digitales; evaluación; validez de contenido; confiabilidad.

ABSTRACT

In the current context of digital education and open science, university professors are not only knowledge creators through their scientific and professional output but are also expected to disseminate it to both academic and non-specialist audiences. While related to digital competence, media literacy provides a critical-communicative perspective essential for effective science dissemination, particularly via digital social networks (DSNs). This study designed and validated two constructs to assess teachers' media competence (TMC): one focused on general DSN use and the other specifically on LinkedIn. Both models integrate the Common Framework for Digital Teaching Competence 2.2 and the media literacy model by Ferrés and Piscitelli (2012). The validation process combined theoretical review, expert judgment ($n = 30$), and exploratory factor analysis (EFA). Data suitability was confirmed through KMO (> 0.80) and Bartlett's test of sphericity, with both instruments showing high reliability ($\alpha > 0.85$; $\omega > 0.87$). The EFA identified six theoretical dimensions, explaining 78.2% of the variance for DSN and 78.8% for LinkedIn. These findings provide an initial approximation of the TMC structure and its potential for diagnostic use in teacher training, highlighting LinkedIn's strategic role as a professional environment for academic dissemination. The incorporation of communicative and media strategies in teacher education is underscored. Future research should include confirmatory factor analysis with larger samples to consolidate these initial results.

Keywords: media competence; university teachers; digital social networks; assessment; content validity; reliability.

INTRODUCCIÓN

En el marco de la educación digital y la ciencia abierta, comunicar conocimiento científico en redes sociales digitales (RSD) se ha convertido en una competencia clave del profesorado. No obstante, su formación en este ámbito enfrenta retos conceptuales y metodológicos. En este contexto, este estudio se alinea con el Marco de Competencia Digital Docente 2.2 (INTEF, 2022), particularmente en sus áreas 2 (Contenidos digitales) y 6 (Desarrollo de la competencia digital del alumnado), por su enfoque en la creación, difusión y análisis crítico de contenidos digitales. Junto al modelo de alfabetización mediática de Ferrés y Piscitelli (2012), ambos marcos sustentan el diseño y validación de dos constructos para evaluar la competencia mediática docente (CMP). Se analiza la estructura factorial de los indicadores y su aplicabilidad en contextos formativos.

Competencia mediática del profesorado en el contexto de las redes sociales digitales

La competencia digital y la CMP son conceptos complementarios y fundamentales en el contexto educativo actual (Marcelo-Martínez et al., 2025). La primera se enfoca en el uso pedagógico de herramientas tecnológicas para mejorar la enseñanza y el aprendizaje (Cabero-Almenara et al., 2020), mientras que la segunda enfatiza la alfabetización crítica sobre los medios y su impacto en la construcción del conocimiento (Ferrés y Piscitelli, 2012). Ambas son esenciales para la divulgación científica en RSD, donde los docentes actúan como creadores y mediadores del conocimiento académico (Buzón et al., 2019).

La divulgación científica en RSD se ha convertido en un elemento clave para acercar el conocimiento académico a la sociedad, ampliar el impacto de las investigaciones y fomentar la interacción con diferentes públicos (Valencia-Oliveros y Martín-Gutiérrez, 2024). La creciente digitalización y el auge de plataformas como Twitter o X (Vásquez-Rocca, 2024), Instagram (Romero Rodríguez et al., 2019), Facebook (Verstappen y Opgenhaffen, 2024) y YouTube (Marchal et al., 2025) han generado nuevas oportunidades para comunicar la ciencia más allá de los círculos académicos tradicionales. Sin embargo, su uso entre los investigadores sigue siendo limitado, lo que pone de manifiesto la necesidad de fortalecer la formación en comunicación digital (Said-Hung et al., 2024).

En un contexto de desinformación creciente, estas plataformas no solo facilitan la difusión del conocimiento, sino que también permiten la interacción con el público y la adaptación a nuevos formatos digitales (Martin-Neira et al., 2023). A pesar de su potencial en educación y ciencias sociales, el reto radica en equilibrar la accesibilidad del mensaje sin perder rigor científico, especialmente en redes como LinkedIn e Instagram, donde el contenido debe ser breve y atractivo (Gallego et al., 2023).

LinkedIn se ha consolidado como una plataforma clave para la divulgación científica, diferenciándose de otras RSD por su enfoque profesional y su capacidad de conectar a investigadores con comunidades académicas, empresariales y del sector educativo (Knight, 2019). Su algoritmo favorece la visibilidad del contenido científico, permitiendo la difusión de artículos, preprints y hallazgos de investigación, así como la interacción en comunidades especializadas (Baruffaldi et al., 2017). Estudios recientes han identificado que LinkedIn es la red social más utilizada por docentes universitarios para compartir su producción científica, superando a plataformas como ResearchGate y Academia.edu, debido a su capacidad de ampliar el impacto del conocimiento y facilitar la transferencia entre la academia y la industria (Huang et al., 2019; Knight, 2019; Ordóñez Castillo et al., 2024; Said-Hung et al., 2024; Tolbert, 2024). No obstante, persisten barreras para su adopción generalizada, como la falta de formación en comunicación digital y el escaso reconocimiento institucional de estas actividades (Argüello-Gutiérrez y Moreno-López, 2024). A pesar de estas limitaciones, LinkedIn ofrece un entorno propicio para la construcción de una identidad académica digital, la difusión del conocimiento y la generación de redes colaborativas, lo que resalta la necesidad de impulsar estrategias institucionales que promuevan su uso efectivo en la divulgación científica.

Investigaciones previas proponen modelos generales de CMP (Gümüş y Kukul, 2023; Jiang y Yu, 2024; Vergili y Kara, 2024) pero no la han estructurado en la divulgación científica en RSD. Aunque identifican habilidades mediáticas para fines educativos y científicos, no establecen un constructo específico que sirva de base para la formación del profesorado. Por lo cual, responde a esta laguna mediante una propuesta integradora que articula el MRCDD 2.2 (INTEF, 2022) y el modelo de alfabetización mediática de Ferrés y Piscitelli (2012).

La integración permite un análisis integral de la CMP en divulgación científica. En este sentido, diversas revisiones sistemáticas recientes respaldan dicha convergencia. Claro et al. (2024) subrayan la importancia de consolidar la operacionalización de la competencia digital docente dentro de marcos normativos como el MRCDD 2.2, mientras que Olaya Guerrero et al. (2025) enfatizan el papel fundamental de las estrategias institucionales en la formación docente en competencia digital. Desde el ámbito de la alfabetización mediática, Fuentes Cancell et al. (2021) destacan que un uso efectivo de las RSD en educación superior requiere no solo habilidades tecnológicas, sino también pensamiento crítico para evaluar el impacto y la credibilidad de los medios.

En síntesis, la complementariedad entre el MRCDD 2.2 y el modelo de Ferrés y Piscitelli justifica su integración para evaluar la CMP en divulgación científica en RSD, consolidando un marco robusto para la formación del profesorado en contextos digitales. Esta combinación permite consolidar un modelo sólido para la formación docente en entornos digitales, alineado con las exigencias actuales de la comunicación académica.

A partir de estos referentes, este estudio aborda la necesidad de fortalecer la participación del profesorado en la comunicación y difusión del conocimiento a través de RSD, contribuyendo a la consolidación de su impacto académico y profesional (Hidalgo y Hidalgo, 2024; Jiang y Yu, 2024). Para ello, se propone el diseño y validación de dos constructos específicos mediante un enfoque teórico y empírico. Los hallazgos permitirán estructurar un modelo de evaluación aplicable tanto a la formación docente como a investigaciones futuras sobre alfabetización mediática y divulgación científica en entornos digitales.

Problema y objetivos de la investigación

En este contexto, el estudio se plantea el siguiente problema: ¿Cómo se estructura la CMP para la divulgación científica en RSD, con especial atención a LinkedIn, y qué implicaciones tiene para su formación docente? Asimismo, se examina la necesidad de incorporar esta competencia en los programas de formación inicial y continua del profesorado.

Para responder esta pregunta, el estudio se propone alcanzar los siguientes objetivos específicos: (1) diseñar y validar dos constructos que definan la CMP para la divulgación científica en RSD y en LinkedIn; (2) analizar la estructura factorial de dichos constructos; (3) evaluar la fiabilidad y validez de los indicadores que los componen; y (4) identificar las implicaciones pedagógicas derivadas del modelo para la formación docente en contextos digitales.

MÉTODOS

Diseño de investigación

Este estudio sigue un enfoque exploratorio basado en la adecuación conceptual y en la validación estructural de los constructos (Hernández-Sampieri y Mendoza, 2018). El diseño se estructura en tres fases: (1) validación teórica y evaluación experta de los indicadores; (2) análisis de la estructura factorial de los constructos; y (3) evaluación de la fiabilidad y validez de los indicadores definidos.

Instrumentos

Se diseñaron dos instrumentos para evaluar la competencia mediática del profesorado en la divulgación científica: uno enfocado en RSD y otro en LinkedIn. Ambos se fundamentan en el MRCDD 2.2 y el modelo de alfabetización mediática de Ferrés y Piscitelli (2012).

El instrumento para RSD abarca seis dimensiones (lenguaje, tecnología, interacción, producción y difusión, ideología y valores, y estética) con 34 indicadores,

mientras que el de LinkedIn, con la misma estructura, incluye 28 indicadores adaptados a la divulgación en entornos profesionales. Aunque comparten una base conceptual, el instrumento para RSD evalúa la divulgación en plataformas generales, mientras que el de LinkedIn se centra en estrategias comunicativas dentro de una red profesional. Ambos instrumentos fueron validados mediante juicio de expertos y análisis factorial exploratorio. La fiabilidad fue alta en ambos casos, con coeficientes de consistencia interna superiores a $\alpha = 0.85$ y $\omega = 0.87$ ($\alpha = 0.91$ y $\omega = 0.93$ para RSD; $\alpha = 0.89$ y $\omega = 0.90$ para LinkedIn), lo que asegura su pertinencia y estabilidad.

El MRCDD 2.2 se selecciona por su alineación con el DigCompEdu y su contribución al Espacio Europeo de Educación, mientras que el modelo de Ferrés y Piscitelli (2012) es un referente en alfabetización mediática y ampliamente citado en la literatura científica. Si bien existen diversos modelos de competencias y alfabetización mediática como el de Buckingham (2003), UNESCO (Grizzle et al., 2021), Hobbs (2010) y el de Area y Guarro (2012), el planteamiento de Ferrés y Piscitelli destaca por su enfoque integral y su impacto en el ámbito académico.

El MRCDD 2.2 estructura la competencia digital docente en cinco áreas: compromiso profesional, enseñanza y aprendizaje, evaluación y retroalimentación, empoderamiento del alumnado y desarrollo de la competencia digital. Por su parte, el modelo de Ferrés y Piscitelli (2012) define la competencia mediática en seis dimensiones, permitiendo un análisis integral de la CMP en divulgación científica (Tabla 1).

Los instrumentos desarrollados –uno para la divulgación en RSD y otro específico para LinkedIn– responden a fases complementarias dentro del proceso de modelización de la competencia mediática docente. No constituyen un único instrumento consolidado, sino aproximaciones diferenciadas y adaptadas a contextos diversos de divulgación científica, orientadas a validar constructos específicos para futuros desarrollos de un modelo integral.

Tabla 1

Correspondencia entre las Dimensiones de Ferrés y Piscitelli (2012) y el MRCDD 2.2

Dimensiones de la competencia mediática	Áreas del MRCDD 2.2	Relación y puntos en común
1. Lenguajes	Área 2: Contenidos digitales	La competencia mediática enfatiza la alfabetización en múltiples lenguajes (audiovisual, digital, icónico, etc.), mientras que la competencia digital docente fomenta la creación, selección y uso de contenidos digitales en la enseñanza.

Dimensiones de la competencia mediática	Áreas del MRCDD 2.2	Relación y puntos en común
2. Tecnología	Área 1: Compromiso profesional Área 3: Enseñanza y aprendizaje	La competencia mediática considera la tecnología como herramienta para la comunicación, mientras que el MRCDD enfatiza su uso pedagógico, su integración en la enseñanza y su aplicación profesional.
3. Procesos de interacción	Área 5: Empoderamiento del alumnado Área 6: Desarrollo de la competencia digital del alumnado	Ambos modelos destacan la interacción en entornos digitales. El MRCDD resalta la colaboración y la ciudadanía digital, mientras que la competencia mediática aborda la participación activa en medios y redes.
4. Procesos de producción y difusión	Área 2: Contenidos digitales Área 6: Desarrollo de la competencia digital del alumnado	La producción de mensajes propios y la interacción con los ajenos es clave en ambos modelos, vinculando la creación de contenidos digitales con la alfabetización mediática.
5. Ideología y valores	Área 5: Empoderamiento del alumnado Área 6: Desarrollo de la competencia digital del alumnado	El análisis crítico de los medios y la comprensión de la ideología subyacente en los mensajes se relacionan con la formación de ciudadanos digitales responsables y críticos.
6. Dimensión estética	Área 2: Contenidos digitales	La competencia mediática enfatiza el valor estético de los medios, lo que se vincula con la capacidad de los docentes de diseñar contenidos digitales atractivos y efectivos en el MRCDD.

Fuente: Elaboración propia.

Procedimiento

La validación teórica de los constructos se realizó mediante revisión bibliográfica y juicio de expertos (septiembre de 2024 a enero de 2025), asegurando su alineación con el MRCDD 2.2 y el modelo de Ferrés y Piscitelli (2012). Para ello, se diseñó un cuestionario digital en el que 30 especialistas evaluaron claridad, pertinencia y coherencia de los indicadores. El proceso garantizó el anonimato, la confidencialidad y el cumplimiento de criterios éticos. La selección de expertos con amplia trayectoria permitió una validación sólida del contenido, cuyos resultados se detallan en el apartado de análisis de datos.

Análisis de datos

La muestra se seleccionó mediante muestreo intencional, considerando la trayectoria en alfabetización mediática, tecnología educativa o formación docente. Participaron 30 especialistas procedentes de universidades de España, Portugal y América Latina, el 70 % con más de 10 años de experiencia. Los criterios de inclusión fueron: contar con publicaciones o proyectos relevantes y experiencia en validación de instrumentos. La muestra, compuesta por un 60 % de hombres y con una edad media de 45.2 años ($\sigma = 7.1$), aportó diversidad disciplinar y geográfica, garantizando un juicio experto integral. Para la validación, se aplicó un cuestionario digital para evaluar la claridad, pertinencia y coherencia de los indicadores, siguiendo criterios éticos y de confidencialidad.

En términos de trayectoria profesional, el 70 % de los participantes contaba con más de 10 años de experiencia en docencia universitaria y formación del profesorado en entornos digitales. Además, un 20 % estaba especializado en investigación sobre alfabetización mediática, y un 10 % en el desarrollo de tecnologías aplicadas a la educación. La selección de los expertos se realizó en función de su experiencia y trayectoria en los ámbitos mencionados, garantizando la representación de distintos perfiles académicos y profesionales. Este enfoque permitió obtener una evaluación multidimensional del instrumento, asegurando su validez desde una perspectiva integral y transdisciplinar.

La muestra de 30 expertos utilizada en este estudio se considera metodológicamente adecuada para la validación de contenido, en consonancia con las recomendaciones de la literatura especializada en desarrollo de instrumentos. Autores como DiFazio et al. (2018) y Polit y Beck (2006) sostienen que entre 5 y 30 expertos son suficientes para este tipo de análisis, siempre que se garantice la especialización temática y la diversidad profesional de los participantes. En lugar de optar por muestras amplias con participantes no especializados –como ocurre en otros estudios similares–, se priorizó la calidad y pertinencia del juicio experto, seleccionando profesionales con una trayectoria consolidada en alfabetización mediática, tecnología educativa y formación docente. Esta elección responde a la necesidad de validar constructos altamente especializados con un enfoque teórico riguroso. Asimismo, se aplicaron procedimientos estadísticos robustos como el Índice de Validez de Contenido (IVC) y el coeficiente de concordancia de Fleiss Kappa (Fleiss et al., 2003), que aportan evidencia cuantitativa sobre el nivel de acuerdo entre evaluadores, incluso en muestras reducidas pero altamente calificadas.

A partir de las observaciones de los expertos, se realizaron ajustes para mejorar la precisión del instrumento, se eliminaron redundancias fusionando indicadores con superposición conceptual y se incorporaron elementos sobre pensamiento crítico y verificación de fuentes para fortalecer la alfabetización mediática. Estos ajustes garantizan la validez teórica y metodológica de los constructos, asegurando

su alineación con los modelos de referencia y su aplicabilidad en la formación del profesorado.

El Análisis Factorial Exploratorio (AFE) mediante SPSS v. 29.0 permitió evaluar la validez de constructo e identificar la estructura subyacente de los indicadores (López-Aguado y Gutiérrez-Provecho, 2019). Se utilizaron los criterios de adecuación muestral de Kaiser-Meyer-Olkin y la prueba de esfericidad de Bartlett. Mediante el método de Máxima Verosimilitud (ML) con rotación Varimax, se identificaron los factores asegurando su independencia. La confiabilidad del instrumento se evaluó mediante los coeficientes de consistencia interna Alfa de Cronbach (α) y Omega de McDonald (ω) (Viladrich et al., 2017). Dado su carácter exploratorio, el AFE se aplicó a una muestra intencional de 30 expertos. Si bien los resultados aportan indicios sobre la estructura factorial de los constructos, esta muestra no permite una validación concluyente. Por ello, será necesario replicar el análisis con muestras superiores a 200 docentes universitarios en futuros estudios.

Como complemento visual al análisis cuantitativo, se elaboraron mapas de calor que permiten representar gráficamente la intensidad de las cargas factoriales entre los indicadores y los factores extraídos en los AFE. Este recurso facilita la identificación de agrupamientos empíricos y posibles cargas cruzadas, reforzando la interpretación de los resultados. Cabe señalar que, al tratarse de un estudio basado en AFE, no se presentan ecuaciones estructurales ni diagramas de modelos latentes, ya que estos corresponden al enfoque confirmatorio (AFC), propio de fases posteriores de validación empírica.

RESULTADOS

Validación teórica y análisis de concordancia

El constructo CMP para la divulgación científica en RSD se estructura según las dimensiones propuestas por Ferrés y Piscitelli (2012), describiéndose sus indicadores y relación con el MRCDD 2.2 (Tabla 2).

Tabla 2

Dimensión e indicadores del constructo CMP para la divulgación científica en RSD

Id	Ámbito	Indicadores
<i>Dimensión lenguaje</i>		
Id.01	Análisis	Identifica la tipología de información científica adecuada para publicación en diferentes RSD.
Id.02		Evalúa qué información publicada en RSD es relevante y alineada con objetivos profesionales.
Id.03		Selecciona funcionalidades de RSD según objetivos de divulgación científica.
Id.04	Expresión	Sintetiza contenidos académicos en RSD destacando objetivo, alcance y resultados principales.
Id.05		Diseña infografías, posters, imágenes animadas y videos para divulgar resultados científicos.
Id.06		Transforma y adapta recursos digitales existentes en RSD para divulgación científica.
Id.07		Redacta contenido divulgativo en RSD utilizando un lenguaje claro, accesible y adaptado a diferentes audiencias.
<i>Dimensión tecnología</i>		
Id.08	Análisis	Comprende funcionalidades clave de RSD para divulgar actividad investigadora.
Id.09		Distingue RSD profesionales y manejar sus funcionalidades básicas.
Id.10		Selecciona RSD específicas para publicar resultados investigativos según objetivos definidos.
Id.11		Utiliza herramientas digitales para programar publicaciones y analizar métricas de impacto en RSD científicas.
Id.12	Expresión	Usa entornos hipermediales, transmediáticos y multimodales para divulgación científica.
Id.13		Diseña recursos digitales (imágenes, videos, infografías) para resultados científicos.
Id.14		Publica resultados científicos en RSD de manera efectiva.
<i>Dimensión interacción</i>		
Id.15	Análisis	Selecciona contenidos científicos adecuados para divulgar en RSD.
Id.16		Identifica recursos digitales atractivos para divulgación en RSD.
Id.17		Percibe motivaciones y emociones en interacciones en RSD.
Id.18		Aplica normas éticas y de derechos de autor en contenidos de RSD.

Id	Ámbito	Indicadores
Id.19	Expresión	Fomenta colaboración con otros investigadores mediante RSD.
Id.20		Interactúa con diversidad cultural respetando particularidades en RSD.
Id.21		Diseña estrategias para fomentar la participación en publicaciones científicas en RSD.
<i>Dimensión producción y difusión</i>		
Id.22	Análisis	Selecciona fuentes de publicación en RSD (cuentas personales, grupales o institucionales).
Id.23		Selecciona estrategias adecuadas para la difusión de contenidos científicos en múltiples plataformas digitales.
Id.24	Expresión	Colabora en la divulgación de producción científica usando funcionalidades de RSD.
Id.25		Comparte resultados científicos en RSD de manera comprensible.
Id.26		Produce contenidos científicos adaptados al lenguaje de RSD.
<i>Ideología y valores</i>		
Id.27	Análisis	Evalúa fiabilidad y relevancia de información científica en RSD.
Id.28		Contrasta información científica publicada en RSD con estándares académicos.
Id.29		Identifica información en RSD con fines contradictorios a normativas científicas.
Id.30	Expresión	Redactar contenidos en RSD respetando ética y principios científicos.
Id.31		Interactúa desde una perspectiva inclusiva y respetuosa con divergencias científicas.
<i>Dimensión Estética</i>		
Id.32	Análisis	Adapta contenido científico a particularidades de cada RSD sin duplicar información
Id.33	Expresión	Diseña contenidos atractivos que generen interés en audiencias diversas.
Id.34		Produce mensajes creativos y originales que promuevan debate científico en RSD.

Fuente: Elaboración propia.

Una vez definidos los indicadores, se establecen sus niveles de logro según el MRCDD 2.2 presentándose ejemplos de desempeño (Tabla 3).

Tabla 3
Niveles de progresión para la CMP en RSD

Etapa	Nivel	Indicadores de logro	Ejemplos de desempeño
Inicial	A1 - Básico	Identifica RSD apropiadas para divulgación científica, publica contenido básico, reconoce información relevante.	Publica un hallazgo científico sencillo en una red social académica; sigue cuentas de divulgación científica.
	A2 - Explorador	Interactúa con publicaciones científicas en diferentes redes, explora opciones de formato y presentación de contenido.	Comenta o comparte publicaciones científicas en Twitter o LinkedIn; identifica tendencias científicas en redes.
Intermedia	B1 - Intermedio	Diseña publicaciones científicas con imágenes, infografías o videos, usa métricas básicas para evaluar impacto.	Crea una publicación con gráficos explicativos; revisa analíticas de una publicación en redes.
	B2 - Intermedio superior	Participa en debates científicos, adapta su contenido a diferentes plataformas, sigue normas de ética digital.	Comenta activamente en un hilo de discusión académica; adapta una publicación de Twitter para ResearchGate.
Avanzada	C1 - Experto	Planifica estrategias de publicación, usa herramientas de automatización, fomenta la interacción y colaboración científica en redes.	Programa publicaciones usando herramientas como Hootsuite o Buffer; crea hilos de Twitter sobre un estudio científico.
	C2 - Experto superior	Lidera debates científicos en redes, evalúa impacto de su divulgación y diseña estrategias avanzadas de engagement.	Organiza un Twitter Space sobre un tema de investigación; usa LinkedIn Analytics para mejorar la difusión científica.

Fuente: Elaboración propia.

El constructo: CMP para la divulgación científica en LinkedIn se adaptó a la especificidad de LinkedIn, definiendo 28 indicadores y niveles de progresión. En la Tabla 4 se muestran sus 28 indicadores.

Tabla 4

Dimensiones e indicadores del constructo CMP para la divulgación científica en LinkedIn

Id	Ámbito	Indicadores
<i>Dimensión lenguaje</i>		
Id.35	Expresión	Crea publicaciones profesionales que destaquen logros científicos y su relevancia.
Id.36		Diseña presentaciones breves y atractivas para publicaciones en formato de LinkedIn Stories.
Id.37		Redacta artículos de blog científico en LinkedIn manteniendo un tono accesible y profesional.
Id.38	Análisis	Identifica grupos y comunidades en LinkedIn relevantes para compartir investigaciones científicas.
Id.39		Reconoce publicaciones científicas destacadas dentro de la red profesional.
Id.40		Sintetiza investigaciones científicas en publicaciones que combinen texto y gráficos profesionales.
<i>Dimensión tecnología</i>		
Id.41	Expresión	Usa herramientas multimedia para enriquecer publicaciones (videos, presentaciones, infografías).
Id.42		Participa en eventos virtuales o seminarios webs organizados a través de LinkedIn Live.
Id.43		Automatiza el seguimiento de interacciones mediante integraciones con herramientas de gestión de contactos.
Id.44		Utiliza herramientas digitales para programar publicaciones y analizar métricas de impacto en LinkedIn.
Id.45	Análisis	Comprende la funcionalidad de LinkedIn Analytics para evaluar el impacto de publicaciones científicas.
Id.46		Identifica oportunidades de conexión profesional basadas en intereses de investigación compartidos.
<i>Dimensión Proceso de interacción</i>		
Id.47	Expresión	Interactúa con publicaciones científicas mediante comentarios que aporten valor profesional.
Id.48		Participa en debates científicos dentro de grupos profesionales en LinkedIn.
Id.49		Colabora en proyectos compartidos mediante funciones de LinkedIn como los mensajes grupales.
Id.50		Diseña estrategias para fomentar la participación en publicaciones científicas en LinkedIn.

Id	Ámbito	Indicadores
Id.51	Análisis	Identifica perfiles clave de expertos e instituciones en el área de investigación.
Id.52		Reconoce publicaciones de alto impacto para generar discusiones científicas en LinkedIn.
<i>Dimensión Proceso de producción y difusión</i>		
Id.53	Expresión	Diseña publicaciones específicas para destacar resultados científicos en el feed de LinkedIn.
Id.54		Crea contenidos para LinkedIn que cumplan estándares de derechos de autor y confidencialidad.
Id.55		Selecciona estrategias adecuadas para la difusión de contenidos científicos en LinkedIn y otras plataformas profesionales.
Id.56	Análisis	Selecciona formatos de publicación adecuados para compartir resultados científicos (artículos, videos, etc.).
<i>Dimensión Ideología y valores</i>		
Id.57	Expresión	Redacta contenidos científicos inclusivos y éticos que respeten la diversidad profesional.
Id.58		Actúa con responsabilidad al compartir investigaciones, respetando la propiedad intelectual.
Id.59	Análisis	Evalúa la credibilidad de perfiles y fuentes de información compartidas en LinkedIn.
Id.60		Identifica contenido relevante que promueva valores éticos y científicos.
<i>Dimensión Estética</i>		
Id.61	Expresión	Crea publicaciones originales que motiven interacción y participación profesional en LinkedIn.
Id.62	Análisis	Diseña contenidos que proyecten profesionalismo y atractivo visual (imágenes de alta calidad y gráficos).

Fuente: Elaboración propia.

Definidos los indicadores, se establecieron sus niveles de logro y ejemplos de desempeño (Tabla 5).

Tabla 5
Niveles de progresión del segundo constructo

Etapas de progresión	Niveles de progresión	Indicadores de logro	Ejemplos de desempeño
Inicial	A1 - Básico	Crea publicaciones profesionales que destaquen logros científicos y su relevancia; Reconoce publicaciones científicas destacadas; Identifica perfiles clave de expertos e instituciones; Comprende la funcionalidad básica de LinkedIn.	Publica un logro científico personal en LinkedIn; identifica una publicación científica relevante y compartirla.
	A2 - Explorador	Identifica grupos y comunidades en LinkedIn relevantes; Interactúa con publicaciones científicas mediante comentarios; Selecciona formatos adecuados; Crea publicaciones originales que motiven interacción. Explora LinkedIn Analytics para identificar tendencias en publicaciones científicas.	Unirse a un grupo de LinkedIn y comentar publicaciones relevantes; elige entre usar texto o video para compartir un resultado.
Intermedia	B1 - Intermedio	Redacta artículos de blog científico; Sintetiza investigaciones en publicaciones con texto y gráficos; Usa herramientas multimedia; Evalúa credibilidad de perfiles y fuentes; Diseña contenidos visualmente atractivos.	Escribe un artículo en LinkedIn Blog; combina infografía y texto para sintetizar una investigación; evalúa un perfil de investigador antes de conectarse.
	B2- Intermedio superior	Diseña presentaciones breves para LinkedIn Stories; Participa en debates científicos en grupos; Diseña publicaciones específicas para destacar resultados; Crea contenidos éticos y que respeten derechos de autor. Planifica una estrategia de publicaciones utilizando herramientas de programación.	Diseña un LinkedIn Story sobre un avance científico; comenta en un debate grupal sobre una publicación académica; crea una publicación que resuma un resultado científico con gráficos de alta calidad.

Etapas de progresión	Niveles de progresión	Indicadores de logro	Ejemplos de desempeño
Avanzada	C1 - Experto	Participa en eventos virtuales a través de LinkedIn Live; Identifica conexiones basadas en intereses compartidos; Colabora en proyectos usando funciones de LinkedIn; Crea publicaciones inclusivas y éticas.	Realiza un seminario en LinkedIn Live; crea una publicación inclusiva que celebre la diversidad en la investigación científica; envía mensajes grupales a colaboradores sobre un proyecto conjunto.
	C2 - Experto superior	Automatiza seguimiento de interacciones; Comprende LinkedIn Analytics para evaluar impacto; Diseña estrategias avanzadas de divulgación; Lidera debates científicos y genera impacto. Implementa estrategias de engagement mediante preguntas, encuestas y contenido interactivo en LinkedIn.	Configura una automatización con herramientas externas para gestionar interacciones; usa LinkedIn Analytics para medir el impacto de una publicación; lidera un debate en un grupo profesional sobre un tema de alto impacto científico.

Fuente: Elaboración propia.

Análisis Factorial Exploratorio

Se confirmó la idoneidad de los datos para el análisis factorial mediante KMO (0.85 y 0.83) y prueba de esfericidad de Bartlett. Todas las cargas factoriales superaron 0.40, asegurando una contribución significativa de los ítems a sus respectivos factores. Los factores identificados explicaron el 78.2 % para el constructo CMP en RSD (Tabla 6) y el 78.8 % de la varianza total en RSD y LinkedIn (Tabla 7), respectivamente.

Tabla 6

Análisis Factorial Exploratorio del primer constructo

Indicador	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
ID1	0.59	0.88	0.77	0.70	0.48
ID2	0.48	0.43	0.83	0.70	0.75
ID3	0.41	0.88	0.82	0.51	0.49
ID4	0.49	0.55	0.66	0.62	0.55

Indicador	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
ID5	0.71	0.47	0.55	0.58	0.63
ID6	0.79	0.50	0.66	0.70	0.42
ID7	0.70	0.49	0.43	0.87	0.88
ID8	0.80	0.55	0.45	0.74	0.62
ID9	0.46	0.65	0.42	0.85	0.53
ID10	0.73	0.56	0.66	0.67	0.49
ID11	0.88	0.79	0.87	0.85	0.70
ID12	0.86	0.44	0.50	0.42	0.56
ID13	0.59	0.54	0.81	0.58	0.54
ID14	0.67	0.47	0.80	0.44	0.89
ID15	0.79	0.50	0.40	0.81	0.75
ID16	0.76	0.79	0.44	0.58	0.46
ID17	0.83	0.71	0.57	0.43	0.56
ID18	0.56	0.76	0.72	0.84	0.64
ID19	0.46	0.76	0.78	0.68	0.79
ID20	0.65	0.66	0.61	0.41	0.45
ID21	0.42	0.72	0.56	0.65	0.85
ID22	0.52	0.61	0.78	0.51	0.44
ID23	0.54	0.48	0.86	0.80	0.72
ID24	0.84	0.80	0.49	0.85	0.67
ID25	0.80	0.85	0.56	0.46	0.51
ID26	0.61	0.81	0.83	0.40	0.66
ID27	0.61	0.51	0.46	0.57	0.87
ID28	0.56	0.66	0.75	0.58	0.89
ID29	0.88	0.53	0.65	0.55	0.54
ID30	0.42	0.70	0.65	0.43	0.54
ID31	0.85	0.52	0.47	0.64	0.89
ID32	0.52	0.74	0.78	0.52	0.76
ID33	0.58	0.72	0.72	0.67	0.45
ID34	0.82	0.56	0.49	0.42	0.70

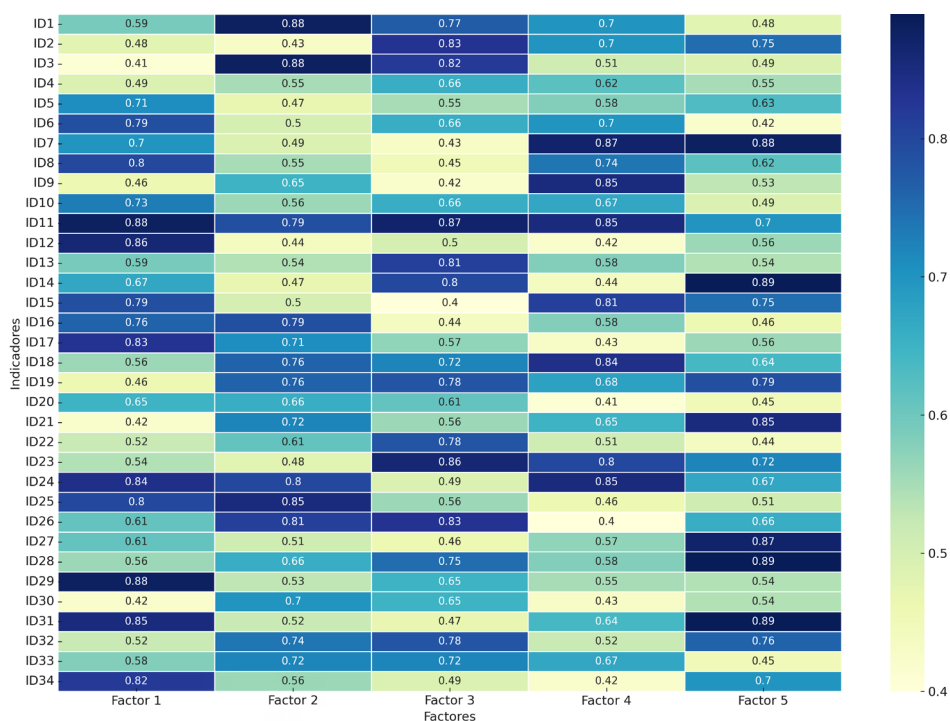
Fuente: Elaboración propia.

Con el fin de facilitar la interpretación visual de la estructura factorial resultante del AFE, se elaboró un mapa de calor que representa la intensidad de las cargas factoriales de los 34 indicadores en los cinco factores identificados empíricamente.

Esta representación permite observar patrones de agrupamiento, identificar ítems con cargas cruzadas y verificar la consistencia interna del modelo propuesto. La Figura 1 muestra el mapa de calor de las cargas factoriales correspondientes al constructo CMP para RSD. Esta representación permite visualizar la intensidad de las saturaciones y la coherencia interna de los factores, facilitando la identificación de agrupamientos y posibles cargas cruzadas. Su interpretación complementa los resultados del AFE y aporta evidencia gráfica de la consistencia del modelo.

Figura 1

Mapa de calor de cargas factoriales del primer constructo



Como puede observarse en el gráfico, la mayoría de los indicadores presentan cargas elevadas en un único factor, lo que refuerza la validez de la estructura empírica propuesta. No obstante, algunos ítems muestran cargas relevantes en más de una dimensión, lo que sugiere una posible interdependencia entre los componentes del modelo, especialmente entre las dimensiones de interacción, tecnología y producción y difusión. A continuación, se presenta la Tabla 7, correspondiente al AFE del segundo constructo, centrado en la competencia mediática del profesorado para la divulgación científica en LinkedIn.

Tabla 7*Análisis Factorial Exploratorio del segundo constructo*

Indicador	Factor 1	Factor 2	Factor 3	Factor 4
ID1	0.74	0.41	0.66	0.51
ID2	0.72	0.49	0.75	0.59
ID3	0.87	0.47	0.57	0.46
ID4	0.86	0.84	0.53	0.73
ID5	0.81	0.68	0.66	0.52
ID6	0.45	0.85	0.85	0.72
ID7	0.57	0.57	0.76	0.85
ID8	0.84	0.79	0.72	0.44
ID9	0.48	0.85	0.70	0.40
ID10	0.45	0.73	0.40	0.48
ID11	0.67	0.75	0.73	0.51
ID12	0.76	0.52	0.56	0.77
ID13	0.72	0.82	0.73	0.68
ID14	0.45	0.58	0.53	0.52
ID15	0.89	0.60	0.85	0.72
ID16	0.80	0.65	0.69	0.65
ID17	0.50	0.76	0.54	0.41
ID18	0.72	0.49	0.87	0.88
ID19	0.86	0.59	0.41	0.86
ID20	0.61	0.88	0.88	0.83
ID21	0.55	0.59	0.83	0.56
ID22	0.48	0.68	0.87	0.75
ID23	0.69	0.45	0.71	0.90
ID24	0.47	0.66	0.84	0.77
ID25	0.75	0.75	0.58	0.55
ID26	0.80	0.81	0.83	0.86
ID27	0.66	0.65	0.80	0.72
ID28	0.75	0.80	0.85	0.57

Fuente: Elaboración propia.

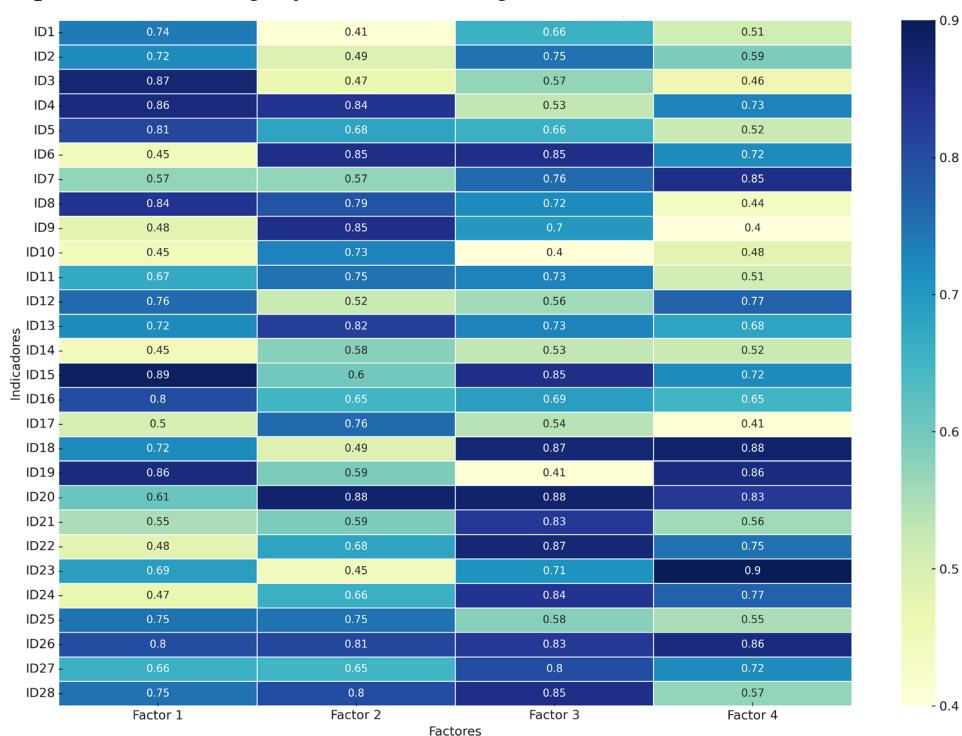
Aunque el modelo teórico de CMP para la divulgación científica en RSD fue diseñado con seis dimensiones (Lenguaje, Tecnología, Interacción, Producción y Difusión, Ideología y Valores, Estética), el AFE reveló una estructura empírica diferente. Esto sugiere que algunas dimensiones teóricas pudieron agruparse en un

mismo factor en la práctica, reflejando la interdependencia entre ellas. En particular, dimensiones como Interacción, Tecnología y Producción y Difusión se agruparon, lo que indica que el uso tecnológico influye directamente en la manera en que se interactúa y se difunde el conocimiento.

Para complementar la matriz de cargas factoriales obtenida en el AFE del segundo constructo, se elaboró un mapa de calor (Figura 2) que muestra la concentración de cargas elevadas en factores específicos. Este análisis visual refuerza la validez preliminar del modelo y permite observar la coherencia empírica de los indicadores.

Figura 2

Mapa de calor de cargas factoriales del segundo constructo



El mapa de calor correspondiente al segundo constructo revela, al igual que en el caso anterior, una concentración clara de cargas elevadas en torno a factores específicos, lo que indica una buena definición empírica de las dimensiones propuestas. La consistencia de los resultados obtenidos en ambos análisis factoriales exploratorios respalda la validez estructural preliminar de los instrumentos desarrollados. Estas evidencias empíricas fortalecen la aplicabilidad del modelo

en contextos formativos y ofrecen una base sólida para futuras investigaciones que profundicen en la evaluación de la competencia mediática docente en RSD.

Evaluación de la fiabilidad

El análisis de fiabilidad confirmó la coherencia interna del instrumento, con coeficientes de $\alpha \geq 0.70$ y $\omega \geq 0.70$ en todas las dimensiones.

En el primer constructo (CMP para la divulgación científica en RSD), los valores de α oscilaron entre 0.84 y 0.89, con un coeficiente global de 0.91, mientras que ω varió entre 0.86 y 0.91, alcanzando un valor global de 0.93. En el segundo constructo (CMP para la divulgación científica en LinkedIn), los coeficientes por dimensión mostraron una tendencia similar, con un α global de 0.89 y un ω global de 0.90. Estos valores indican una alta coherencia interna en las dimensiones, lo que refuerza la estabilidad del modelo (Tabla 8).

El análisis de correlación ítem-total mostró coeficientes superiores a 0.30 en todos los indicadores, lo que garantiza su adecuada discriminación dentro de cada dimensión. No se identificaron ítems con valores críticos que sugirieran su eliminación o modificación. Esto confirma que cada ítem contribuye significativamente a la medición del constructo al que pertenece.

Tabla 8

Resultados del análisis de Alfa de Cronbach y Omega de McDonald

Constructo	Dimensión	Alfa de Cronbach (α)	Omega de McDonald (ω)
CMP para la divulgación científica en RSD	Lenguaje	0.86	0.88
	Tecnología	0.89	0.91
	Interacción	0.85	0.87
	Producción y Difusión	0.88	0.90
	Ideología y Valores	0.87	0.89
	Estética	0.84	0.86
	Global	0.91	0.93

CMP para la divulgación científica en LinkedIn	Lenguaje	0.85	0.87
	Tecnología	0.88	0.89
	Interacción	0.84	0.85
	Producción y Difusión	0.87	0.88
	Ideología y Valores	0.86	0.87
	Estética	0.83	0.84
	Global	0.89	0.90

Fuente: Elaboración propia.

Los resultados del AFE mostraron una reducción de las seis dimensiones teóricas iniciales a cinco factores en RSD y cuatro en LinkedIn. Esta agrupación sugiere integraciones conceptuales, particularmente entre las dimensiones de interacción, tecnología y producción/difusión, las cuales tienden a confluir empíricamente. Este patrón podría explicarse por la convergencia funcional que caracteriza a la práctica docente en entornos digitales, donde la interacción en redes, el uso estratégico de la tecnología y la creación de contenidos se desarrollan de forma interrelacionada. Además, la naturaleza profesional y académica de plataformas como LinkedIn parece potenciar esta integración, diluyendo las fronteras entre dimensiones previstas teóricamente. Estos hallazgos permiten avanzar en la comprensión empírica de la competencia mediática docente, ofreciendo una base para su interpretación crítica y su posible aplicación en la formación docente y la divulgación científica. A continuación, se discuten sus implicaciones teóricas y prácticas.

DISCUSIÓN Y CONCLUSIONES

Los resultados obtenidos ofrecen una aproximación empírica a la estructura de la CMP para la divulgación científica en RSD, diferenciando entre el uso generalista de plataformas digitales y la comunicación académica en entornos profesionales como LinkedIn. La reducción de las dimensiones teóricas iniciales y las agrupaciones observadas sugieren la necesidad de reconsiderar ciertos solapamientos conceptuales, en línea con la práctica docente digital y la integración funcional de la interacción, la tecnología y la producción de contenidos. Los altos índices de fiabilidad y la consistencia factorial preliminar respaldan la pertinencia teórica del instrumento, en consonancia con los marcos del MRCDD 2.2 y la alfabetización mediática (Ferrés y Piscitelli, 2012). Estos hallazgos coinciden con estudios previos que subrayan la importancia de fortalecer la CMP en contextos académicos (Cabero-Almenara et al., 2020; Suelves et al., 2022).

Estos resultados se alinean con investigaciones previas que han abordado la evaluación y desarrollo de la competencia digital docente desde perspectivas

psicométricas y formativas. Por ejemplo, Cabero-Almenara et al. (2020) validaron marcos de competencia mediante juicio experto, destacando la importancia del análisis de consistencia interna y pertinencia de los indicadores. Palacios-Rodríguez et al. (2025) realizaron una macroevaluación comparativa entre España y Portugal basada en el marco DigCompEdu, subrayando la necesidad de enfoques diagnósticos más específicos. En el ámbito conceptual, estudios como el de Marcelo-Martínez et al. (2025) han puesto de relieve la brecha entre el uso instrumental de las RSD y su aprovechamiento pedagógico en la formación docente. Este trabajo amplía dichas contribuciones al integrar la alfabetización mediática en la evaluación de la competencia docente, con una especificidad contextual (LinkedIn) que no había sido abordada previamente.

Los hallazgos también refuerzan la idea de que la alfabetización mediática y digital es esencial en la formación docente, en línea con investigaciones que han evaluado la CMP y sus aplicaciones pedagógicas (Jiang y Yu, 2024; Simons et al., 2017). En particular, este estudio aporta una diferenciación entre la CMP en RSD y su especificidad en LinkedIn, lo que no ha sido suficientemente abordado en estudios previos.

No obstante, existen discrepancias con estudios previos que han abordado la CMP y la digital como un constructo global sin diferenciar sus aplicaciones específicas en la divulgación académica (Ho et al., 2024; Kara et al., 2017; Luan et al., 2023; Vergili y Kara, 2024). Investigaciones anteriores (Palacios-Rodríguez et al., 2025) han analizado la CMP en términos generales, centradas en la competencia digital del profesorado en RSD generalistas y profesionales. Esta distinción aporta un enfoque novedoso al resaltar que la comunicación científica requiere estrategias específicas que no están completamente cubiertas en los modelos generales de alfabetización mediática docente.

Este estudio contribuye de forma original al campo de la formación docente y la alfabetización mediática al proponer dos constructos diferenciados y validados empíricamente para evaluar la competencia mediática del profesorado en la divulgación científica. A diferencia de modelos anteriores, la propuesta aquí presentada permite una evaluación específica para contextos profesionales como LinkedIn, respondiendo a las exigencias actuales de la comunicación académica en entornos digitales. Esta especificidad amplía el campo de aplicación de la CMP y proporciona un marco evaluativo útil tanto para diseñar programas formativos como para futuras investigaciones en el área.

Además de sus aplicaciones formativas, el modelo validado ofrece implicaciones relevantes para las políticas institucionales en educación superior. La inclusión de indicadores específicos sobre divulgación científica en RSD permite a las universidades diseñar planes de acción más alineados con los principios de la ciencia abierta y la transferencia de conocimiento. Esto resulta especialmente pertinente en el contexto de los Objetivos de Desarrollo Sostenible (ODS), que promueven la comunicación científica inclusiva y accesible. Institucionalizar el reconocimiento

de la competencia mediática como dimensión evaluable en la carrera docente puede favorecer no solo la visibilidad de la investigación, sino también una cultura académica más conectada con la sociedad. En esta línea, la propuesta se articula como una herramienta útil para diagnosticar necesidades, definir itinerarios formativos y establecer indicadores de calidad en la divulgación académica digital, aspectos cada vez más valorados por agencias de aseguramiento de la calidad y convocatorias competitivas de financiación.

Los resultados del AFE muestran una estructura factorial que, aunque no coincide exactamente con la propuesta teórica de seis dimensiones, presenta una organización de los factores que se alinea con hallazgos previos en estudios sobre competencias mediáticas y digitales. Diversas investigaciones (García et al., 2022; Romero-Rodríguez et al., 2019) han encontrado o valorado estructuras similares, donde las dimensiones de tecnología, interacción y producción tienden a agruparse debido a su interdependencia en contextos digitales. Este patrón sugiere que, en la práctica, las competencias mediáticas pueden no manifestarse como constructos estrictamente independientes, sino como dimensiones interrelacionadas que responden a la dinámica del entorno digital. Futuros estudios podrían aplicar un AFC para evaluar la estabilidad del modelo y compararlo con estructuras factoriales reportadas en la literatura.

La integración de la CMP en la formación docente y en las estrategias institucionales resulta clave para fortalecer la divulgación científica en RSD. Los resultados confirman la solidez de la validación de contenido, sustentada en criterios teóricos claros y el juicio de expertos, y aportan valor teórico para futuras investigaciones y programas formativos. LinkedIn destaca como herramienta esencial para la visibilización de la producción científica y la transferencia de conocimiento. A nivel institucional, resulta necesario reconocer la divulgación en RSD como actividad académica legítima y alinear la formación docente con el MRCDD 2.2 para favorecer la alfabetización mediática.

Desde una perspectiva aplicada, los instrumentos desarrollados ofrecen un punto de partida para el diagnóstico y la orientación de planes de formación y políticas institucionales orientadas a una divulgación científica ética y eficaz. No obstante, la evidencia factorial obtenida es preliminar y requiere ser contrastada con muestras amplias antes de consolidar un modelo definitivo. Las políticas educativas deben promover el uso crítico de estas plataformas, facilitando recursos y formación acordes con los principios de la ciencia abierta y los Objetivos de Desarrollo Sostenible.

Limitaciones y líneas de trabajo futuro

Entre las principales limitaciones del estudio destaca el tamaño muestral reducido ($n = 30$), que obliga a considerar el análisis factorial exploratorio (AFE) como una aproximación preliminar. Aunque se trata de una muestra intencional de expertos con especialización en alfabetización mediática, tecnología educativa y

formación docente, la validez factorial del instrumento requiere ser contrastada con muestras amplias y heterogéneas.

Como línea futura, se plantea realizar un AFC para comprobar la estructura propuesta de la CMP, bajo un modelo jerárquico de segundo orden con seis dimensiones agrupadas en un factor general. El AFC se aplicará a una muestra superior a 300 docentes universitarios, empleando máxima verosimilitud robusta (MLR) con AMOS y evaluando la bondad de ajuste mediante χ^2/df , CFI, TLI, RMSEA y SRMR. Asimismo, se propone la aplicación empírica de los instrumentos en contextos formativos reales para contrastar su utilidad diagnóstica, realizar validación cruzada y desarrollar estudios de fiabilidad test-retest. También se plantea ampliar su aplicación a muestras internacionales y comparar resultados entre diferentes plataformas sociales, lo que permitirá reforzar su validez predictiva y su utilidad comparativa.

Agradecimientos

Dieter Reynaldo Fuentes Cancell ha sido financiado con cargo a la convocatoria de contratos predoctorales UVa 2024, cofinanciada por el Banco Santander.

REFERENCIAS

- Area, M. y Guarro, A. (2012). La alfabetización informacional y digital: fundamentos pedagógicos para la enseñanza y el aprendizaje competente. *Revista Española de Documentación Científica*, 35(Monográfico), 46-74. <https://doi.org/10.3989/redc.2012.mono.977>
- Argüello-Gutiérrez, C. y Moreno-López, R. (2024). Attitudes and practices of educational researchers towards the use of social media to disseminate science. *Journal of Information Science*. Advance online publication. <https://doi.org/10.1177/01655515241245958>
- Baruffaldi, S. H., Di Maio, G. y Landoni, P. (2017). Determinants of PhD holders' use of social networking sites: An analysis based on LinkedIn. *Research Policy*, 46(4), 740-750. <https://doi.org/10.1016/j.respol.2017.01.014>
- Buckingham, D. (2003). Media education and the end of the critical consumer. *Harvard Educational Review*, 73(3), 309-327. <https://doi.org/10.17763/haer.73.3.c149w3g81t381p67>
- Buzón, V. M., Pérez, R. G. y Catalán, Á. R. (2019). Explorando factores predictores de la competencia digital en las redes sociales virtuales. *Pixel-Bit. Revista de Medios y Educación*, 56(56), 51-69. <https://doi.org/10.12795/pixelbit.2019.i56.03>
- Cabero-Almenara, J., Romero-Tena, R. y Palacios-Rodríguez, A. (2020). Evaluation of teacher digital competence frameworks through expert judgement: The use of the expert competence coefficient. *Journal of New Approaches in Educational Research*, 9(2), 275-283. <https://doi.org/10.7821/naer.2020.7.578>
- Claro, M., Castro-Grau, C., Ochoa, J. M., Hinostroza, J. E. y Cabello, P. (2024). Systematic review of quantitative research on digital competences of in-service school teachers. *Computers & Education*,

- 215, 105030. <https://doi.org/10.1016/j.compedu.2024.105030>
- DiFazio, R. L., Strout, T. D., Vessey, J. A. y Lulloff, A. (2018). Item generation and content validity of the child-adolescent bullying scale. *Nursing Research*, 67(4), 294-304. <https://doi.org/10.1097/NNR.0000000000000283>
- Ferrés, J. y Piscitelli, A. (2012). La competencia mediática: propuesta articulada de dimensiones e indicadores. *Comunicar*, 19(38), 75-82. <https://doi.org/10.3916/C38-2012-02-08>
- Fleiss, J. L., Levin, B. y Cho Paik, M. (2003). *Statistical Methods for Rates and Proportions* (3.ª ed.). John Wiley and Sons. <https://doi.org/10.1002/0471445428>
- Fuentes Cancell, D. R., Estrada Molina, O. y Delgado Yanes, N. (2021). Las redes sociales digitales: una valoración socioeducativa. Revisión sistemática. *Revista Fuentes*, 23(1), 41-52. <https://doi.org/10.12795/revistafuentes.2021.v23.i1.11947>
- Gallego, A. M., Lacerda, J. D. S. y Araujo, A. C. C. (2023). La divulgación científica en Instagram: el reto del discurso audiovisual científico ante los contenidos efímeros. *Revista de Comunicación de la SEECI*, 56, 148-175. <https://doi.org/10.15198/seeci.2023.56.e823>
- García, M. Á. G., Muñoz-Repiso, A. G.-V. y Duarte, M. A. A. (2022). Competencias digitales de los docentes en formación: dimensiones y componentes que promueven su desarrollo. *Civilizar*, 22(42), e20220105. <https://doi.org/10.22518/jour.ccsch/20220105>
- Grizzle, A., Wilson, C., Tuazon, R., Cheung, C. K., Lau, J., Fischer, R., Gordon, D., Akyempong, K., Singh, J., Carr, P. R., Stewart, K., Tayie, S., Suraj, O., Jaakkola, M., Thésée, G., Gulston, C., Andzongo Menyeng, B. P. y Zibi Fama, P. A. (2021). *Media and information literate citizens: think critically, click wisely!* UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000377068>
- Gümüş, M. M. y Kukul, V. (2023). Developing a digital competence scale for teachers: validity and reliability study. *Education and Information Technologies*, 28(3), 2747-2765. <https://doi.org/10.1007/s10639-022-11213-2>
- Hernández-Sampieri, R. y Mendoza, C. (2018). *Metodología de la investigación: Las rutas cuantitativa, cualitativa y mixta* (6.ª ed.). McGraw-Hill.
- Hidalgo, M. y Hidalgo, M. (2024). Análisis del concepto de Competencia Digital Docente: una revisión sistemática de la literatura. *Revista Latinoamericana de Tecnología Educativa - RELATEC*, 23(1), 25-41. <https://doi.org/10.17398/1695-288X.23.1.25>
- Ho, H. T. N., Dinh, V. H. y Phan, Q. A. (2024). The important role of New Media Literacy (NML) in Vietnam: Adaptation and Validation of the NML Scale. *Communication Reports*, 37(1), 28-44. <https://doi.org/10.1080/08934215.2023.2272270>
- Hobbs, R. (2010). *Digital and Media Literacy: A Plan of Action*. The Aspen Institute. <https://www.aspeninstitute.org/publications/digital-media-literacy-plan-action-2>
- Huang, L., Clarke, A., Heldsinger, N. y Tian, W. (2019). The communication role of social media in social marketing: a study of the community sustainability knowledge dissemination on LinkedIn and Twitter. *Journal of Marketing Analytics*, 7(2), 64-75. <https://doi.org/10.1057/s41270-019-00053-8>
- Instituto Nacional de Tecnologías Educativas y de Formación del Profesorado (INTEF). (2022). *Marco de Referencia de la Competencia Digital Docente 2.2*. Ministerio de Educación y Formación Profesional. https://intef.es/wp-content/uploads/2022/03/MRCDD_Vo6B_GTTA.pdf

- Jiang, L. y Yu, N. (2024). Developing and validating a Teachers' Digital Competence Model and Self-Assessment Instrument for secondary school teachers in China. *Education and Information Technologies*, 29(7), 8817-8842. <https://doi.org/10.1007/s10639-023-12182-w>
- Kara, M., Caner, S., Günay Gökben, A., Cengiz, C., İlgör Şimşek, E. y Yıldırım, S. (2017). Validation of an instrument for preservice teachers and an investigation of their new media literacy. *Journal of Educational Computing Research*, 56(7), 1005-1029. <https://doi.org/10.1177/0735633117731380>
- Knight, M. (2019). Teaching responsible social media practices in business and professional communication: the importance of LinkedIn. *Business and Professional Communication Quarterly*, 82(4), 399-400. <https://doi.org/10.1177/2329490619884740>
- López-Aguado, M. y Gutiérrez-Provecho, L. (2019). Cómo realizar e interpretar un análisis factorial exploratorio utilizando SPSS. *REIRE Revista d'Innovació i Recerca en Educació*, 12(2), 1-14. <https://doi.org/10.1344/reire2019.12.227057>
- Luan, L., Liang, J. C., Chai, C. S., Lin, T. B. y Dong, Y. (2023). Development of the new media literacy scale for EFL learners in China: a validation study. *Interactive Learning Environments*, 31(1), 244-257. <https://doi.org/10.1080/10494820.2020.1774396>
- Marcelo-Martínez, P., Yot-Domínguez, C. y Yanes Cabrera, C. (2025). Conectados fuera, desconectados dentro. Las redes sociales en la formación inicial docente. *RIED-Revista Iberoamericana de Educación a Distancia*, 28(1), 83-106. <https://doi.org/10.5944/ried.28.1.41343>
- Marchal, N., Hoes, E., Klüser, K. J., Hamborg, F., Alizadeh, M., Kubli, M. y Katzenbach, C. (2025). How Negative Media Coverage Impacts Platform Governance: Evidence from Facebook, Twitter, and YouTube. *Political Communication*, 42(2), 215-233. <https://doi.org/10.1080/10584609.2024.2377992>
- Martin-Neira, J. I., Trillo-Domínguez, M. y Olvera-Lobo, M. D. (2023). El periodismo científico ante la desinformación: decálogo de buenas prácticas en el entorno digital y transmedia. *Revista ICONO 14. Revista científica de Comunicación y Tecnologías emergentes*, 21(1), 1-28. <https://doi.org/10.7195/ri14.v21i1.1949>
- Olaya Guerrero, J. C., Contreras Contreras, F. y Salinas Ponce, Á. F. B. (2025). Competencias digitales en los docentes universitarios: una revisión sistemática. *Revista InveCom*, 5(1). <https://doi.org/10.5281/zenodo.12659838>
- Ordóñez Castillo, G. A., Montesdeoca Estrada, S. J., Henríquez Mendoza, E. F., Santín Picoita, F. G. y Granda Cruz, C. A. (2024). Divulgación científica y plataformas digitales: Caso de los profesores de comunicación en tres universidades ecuatorianas. *Estudios y Perspectivas Revista Científica y Académica*, 4(3), 3147-3166. <https://doi.org/10.61384/r.c.a.v4i3.602>
- Palacios-Rodríguez, A., Llorente-Cejudo, C., Lucas, M. y Bem-haja, P. (2025). Macroevaluación de la competencia digital docente. Estudio DigCompEdu en España y Portugal. *RIED-Revista Iberoamericana de Educación a Distancia*, 28(1), 177-196. <https://doi.org/10.5944/ried.28.1.41379>
- Polit, D. F. y Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? Critique and recommendations. *Research in Nursing & Health*, 29(5), 489-497. <https://doi.org/10.1002/nur.20147>
- Romero Rodríguez, J. M., Campos Soto, M. N. y Gómez García, G. (2019). Follow me y dame like: Hábitos de uso de Instagram de los futuros maestros. *Revista Interuniversitaria de Formación del*

- Profesorado*, 33(1), 83-96. <https://doi.org/10.47553/rifop.v33i1.72046>
- Romero-Rodríguez, L. M., Contreras-Pulido, P. y Pérez-Rodríguez, M. A. (2019). Media competencies of university professors and students. Comparison of levels in Spain, Portugal, Brazil and Venezuela. *Cultura y Educación*, 31(2), 326-368. <https://doi.org/10.1080/11356405.2019.1597564>
- Said-Hung, E., Martín-Gutiérrez, Á. y Marcano, B. (2024). A study of social media use for scientific communication and dissemination among Spanish education researchers. *Knowledge Management & E-Learning: An International Journal*, 16(2), 237-258. <https://doi.org/10.34105/j.kmel.2024.16.012>
- Simons, M., Meeus, W. y T'Sas, J. (2017). Measuring media literacy for media education: development of a questionnaire for teachers' competencies. *Journal of Media Literacy Education*, 9(1), 99-115. <https://doi.org/10.23860/JMLE-2017-9-1-7>
- Suelves, D. M., Méndez, V. G. y Mas, J. A. R. L. (2022). Análisis de la competencia digital en el futuro profesorado a través de un diseño mixto. *Revista de Educación a Distancia (RED)*, 22(70), 1-17. <https://doi.org/10.6018/red.523071>
- Tolbert, C. L. (2024). The leadership identity presented on LinkedIn. *The Journal of Values-Based Leadership*, 18(1), 11. <https://doi.org/10.22543/1948-0733.1539>
- Valencia-Oliveros, N. Y. y Martín-Gutiérrez, Á. (2024). Divulgación en RRSS de las profesoras universitarias del campo educativo en España. *Profesorado, Revista de Currículum y Formación del Profesorado*, 28(3), 311-331. <https://doi.org/10.30827/profesorado.v28i3.29601>
- Vásquez-Rocca, L. (2024). Comunicación científica en Twitter. Autoría, función y estructuras prototípicas de académicos en Chile. *Galaxia (São Paulo)*, 49, e66109. <https://doi.org/10.1590/1982-2553202466109>
- Vergili, M. y Kara, M. (2024). An investigation of students and teachers' new media literacy: the contributing characteristics with the moderator role of gender. *Research and Practice in Technology Enhanced Learning*, 19, 029. <https://doi.org/10.58459/rptel.2024.19029>
- Verstappen, M. y Opgenhaffen, M. (2024). Making it fit: how science news gets remediated for Facebook and Instagram. *Journalism Studies*, 25(9), 1010-1028. <https://doi.org/10.1080/1461670X.2023.2263799>
- Viladrich, C., Angulo-Brunet, A. y Doval, E. (2017). Un viaje alrededor de alfa y omega para estimar la fiabilidad de consistencia interna. *Anales de Psicología / Annals of Psychology*, 33(3), 755-782. <https://doi.org/10.6018/analesps.33.3.268401>

Fecha de recepción del artículo: 1 de junio de 2025

Fecha de aceptación del artículo: 31 de julio de 2025

Fecha de aprobación para maquetación: 4 de septiembre de 2025

Fecha de publicación en OnlineFirst: 11 de octubre de 2025

Fecha de publicación: 1 de enero de 2026

Tecnologías inmersivas en la universidad: perfiles del profesorado y obstáculos para su integración

Immersive technologies in higher education: faculty profiles and barriers to integration



- Julio Cabero-Almenara - *Universidad de Sevilla, US (España)*
 Antonio Palacios-Rodríguez - *Universidad de Sevilla, US (España)*
 Julio Barroso-Osuna - *Universidad de Sevilla, US (España)*
 Carmen Siles-Rojas - *Universidad de Sevilla, US (España)*

RESUMEN

Introducción: El estudio analiza las barreras percibidas por el profesorado universitario para integrar tecnologías inmersivas (RA, RV y RM) en la educación superior. A pesar de su potencial pedagógico, estas tecnologías enfrentan obstáculos técnicos, pedagógicos, económicos, institucionales y éticos/sociales. **Metodología:** Se empleó un enfoque cuantitativo, no experimental, mediante encuesta online a 775 docentes de universidades españolas e iberoamericanas con experiencia en tecnologías XR. El instrumento incluyó variables sociodemográficas y 23 ítems sobre dificultades percibidas, analizados mediante Análisis de Correspondencias Múltiples (ACM). **Resultados:** El ACM permitió identificar cuatro perfiles docentes: (1) crítico y experimentado, (2) promedio o de transición, (3) tecnopositivo o innovador, y (4) crítico selectivo. Las dificultades económicas fueron las más destacadas, seguidas de las pedagógicas y técnicas. Las percepciones variaron según edad, género, área disciplinar y contexto institucional. **Discusión:** Las barreras no son homogéneas y responden a factores estructurales y culturales. Las dificultades técnicas y económicas afectan más a docentes mayores o de instituciones con menos recursos. Las pedagógicas y éticas preocupan especialmente a docentes de Humanidades y Ciencias Sociales. La resistencia institucional al cambio también emerge como un obstáculo clave. **Conclusiones:** El estudio evidencia la necesidad de estrategias diferenciadas de formación y apoyo institucional. Se recomienda avanzar hacia investigaciones longitudinales y cualitativas que profundicen en la evolución de estas percepciones y en el impacto de las políticas de innovación educativa.

Palabras clave: tecnologías inmersivas; educación superior; realidad virtual y aumentada; dificultades; análisis de correspondencias múltiples.

ABSTRACT

Introduction: The study analyses the barriers perceived by university faculties to integrate immersive technologies (AR, VR and MRI) in higher education. Despite their pedagogical potential, these technologies face technical, pedagogical, economic, institutional and ethical/social barriers. **Methodology:** A quantitative, non-experimental approach was used by means of an online survey of 775 teachers from Spanish and Latin American universities with experience in XR technologies. The instrument included socio-demographic variables and 23 items on perceived difficulties, analysed by Multiple Correspondence Analysis (MCA). **Results:** The MCA identified four teacher profiles: (1) critical and experienced, (2) average or transitional, (3) technopositive or innovative, and (4) selective critical. Economic difficulties were the most prominent, followed by pedagogical and technical difficulties. Perceptions varied according to age, gender, subject area and institutional context. **Discussion:** Barriers are not homogeneous and respond to structural and cultural factors. Technical and economic difficulties affect older teachers or those from institutions with fewer resources. Pedagogical and ethical barriers are of particular concern to teachers in the humanities and social sciences. Institutional resistance to change also emerges as a key obstacle. **Conclusions:** The study evidences the need for differentiated training strategies and institutional support. It is recommended to move towards longitudinal and qualitative research that delves into the evolution of these perceptions and the impact of educational innovation policies.

Keywords: immersive technologies; higher education; virtual and augmented reality; difficulties; multiple correspondence analysis.

INTRODUCCIÓN

La significación que la Realidad Aumentada (RA), la Realidad Virtual (RV) y la Realidad Mixta (RM) están adquiriendo en el terreno educativo universitario se observa directamente por el volumen de investigaciones que en los últimos tiempos se están llevando a cabo sobre ellas. Fruto de estas investigaciones es el aumento de metaanálisis y revisiones sistemáticas sobre diferentes aspectos de ellas, como: la adquisición de aprendizajes por los estudiantes (Yu y Xu, 2022; Victoria Maldonado et al., 2024), las teorías y enfoques pedagógicos utilizados para su diseño (Marougkas et al., 2024; Mohsen y Alangari, 2024), los aspectos éticos y psicológicos que se producen en su utilización (Mogrovejo-Zambrano et al., 2024), las posibilidades que tiene su utilización en las versiones de escritorio e inmersiva (Hamilton et al., 2021; Liu et al., 2024) o la carga cognitiva que implica su utilización (Luna et al., 2023).

Con el objeto de aclarar los términos que se van a utilizar, se puede señalar que existe una fuerte diferencia entre la RA y RV, “ya que, en la segunda, los datos virtuales sustituyen a los físicos, creándose una nueva realidad. Por el contrario, en la realidad aumentada, las dos realidades se superponen en distintas capas de información en formatos diversos (imágenes generadas por ordenador, secuencias de vídeo, animaciones, etc.) para configurar una nueva realidad que es con la que interacciona la persona” (Cabero-Almenara et al., 2022, p. 12). Por su parte, al analizar la RM, es importante señalar que es una tecnología que combina elementos de la RA y RV. Gracias a esta integración, es posible generar objetos virtuales que permiten a los usuarios interactuar con entornos tridimensionales, tanto mediante la inmersión total en espacios virtuales propios de la RV, como a través de la superposición de contenidos digitales sobre el mundo real, característica de la RA.

Estos metaanálisis también se refieren al objeto central de la presente investigación: ¿cuáles son las dificultades que perciben los docentes y gestores de centros para su incorporación a la formación? (Obeidallah et al., 2023; Sales et al., 2023).

Tanto las investigaciones como las revisiones sistemáticas apuntan a que tanto los profesores, como los estudiantes o los directores de centros perciben una serie de dificultades para la incorporación de estas tecnologías emergentes en la formación universitaria. Entre estas dificultades destacan en primer lugar las de índole económico (Mulders et al., 2020; Toala-Palma et al., 2020; Menjivar et al., 2021; Angulo et al., 2023; Sales et al., 2023; Familoni y Onyebuchi, 2024). Esta dimensión no se limita exclusivamente a los costes iniciales asociados a la implementación de estas tecnologías, como pueden ser la adquisición de dispositivos, los costos de ubicar los objetos producidos en plataformas digitales específicas, la inversión en software específico para su producción, o los costos derivados de las licencias de uso. Sino que también engloba aquellos gastos que se derivan del mantenimiento continuo de estos recursos y de su inevitable actualización periódica.

A éstos se incorpora la velocidad con la que avanza la tecnología utilizada en estos dispositivos, lo que obliga a inversiones constantes para su actualización y su adecuación a los nuevos estándares que vayan apareciendo. Por tanto, no se trata únicamente de una inversión puntual, sino de una inversión económica constante, lo que supone una dificultad clara para aquellas instituciones con recursos económicos limitados.

Por otra parte, no se debe olvidar que, en estas dificultades de tipo económico, se encuentra la dificultad de conectividad existente en ciertas instituciones educativas, ya que, para una correcta utilización y conseguir una aceptable sensación de inmersión y desplazamiento por el objeto, se requieren una conexión a internet estable y un ancho de banda suficiente (Sales et al., 2023; Subirats-Blanco y Conde del Rio, 2024).

Una diversidad de autores (Obeidallah et al., 2023; Mogrovejo-Zambrano et al., 2024) señalan como otra de las dificultades, y que en cierta medida se deriva de la anterior, la falta de equidad de utilización de estas tecnologías, tanto para las instituciones como para los estudiantes. Por tanto, se podría hablar de la existencia de una brecha digital que afecta tanto a los estudiantes como a las propias instituciones educativas (ordenadores, gafas de realidad virtual, tabletas o smartphones de última generación). Esta brecha no solo refleja diferencias en el acceso a la tecnología, sino que también evidencia profundas desigualdades en las posibilidades reales de participación en entornos de aprendizaje producidos por estas tecnologías y, en consecuencia, no poder utilizar las posibilidades que ofrecen (Cabero-Almenara et al., 2025).

En cierta medida, los aspectos apuntados crean otra dificultad, que se refiere a la equidad de utilización de estas tecnologías, tanto para las instituciones como para los estudiantes (Mogrovejo-Zambrano et al., 2024). Por ello se podría hablar de la existencia de una brecha digital entre estudiantes e instituciones, ya que ni todos los estudiantes, ni todas las instituciones tienen acceso a dispositivos para participar plenamente en actividades basadas en RA, RV o RM, y ello podrá repercutir en quedar excluidas de estas innovaciones, ampliando la brecha digital entre instituciones.

Otro gran grupo de dificultades apuntadas por diferentes autores se centran en la falta de formación del profesorado para la utilización de la RA, RV y RM en los contextos formativos y la resistencia al cambio que muestran no solo para incorporar las tecnologías digitales en la enseñanza, sino sobre todo las tecnologías emergentes que van apareciendo (AlGerafi et al., 2023; Marín-Díaz et al., 2022; Perifanou et al., 2023; Sales et al., 2023). Aspecto al que se le une, por una parte, la dificultad de su integración en los escenarios de aprendizaje (Mulders et al., 2020), y por otra, la falta de relación con las teorías del aprendizaje (Caballero-Garriazo et al., 2023). Todo ello como consecuencia de la poca investigación existente sobre la forma de utilización y diseño de estos objetos de aprendizaje (Radianti et al., 2020; Sales et al., 2023). Y la escasez de estos objetos en formato abierto (Upadhyay et al., 2024).

Por otra parte, no se debe olvidar que en las investigaciones que han encontrado resultados significativos, en estas los tamaños de efecto no son muy amplios (Coban et al., 2022). Aunque algunas investigaciones sugieren que, sin una adecuada integración de estrategias de aprendizaje en los escenarios de RA, la eficacia de los resultados de aprendizaje puede ser limitada, o incluso menos efectiva que los métodos tradicionales de aprendizaje con apoyo tecnológico (Jingru et al., 2025).

Otro tipo de dificultades se refiere a la existencia de diferentes impactos psicológicos en su utilización, como, por ejemplo, la ansiedad, fatiga y sensación de desorientación que producen determinados tipos de utilización como la inmersiva (Mulders et al., 2020; Cevikbas et al., 2023; Obeidallah et al., 2023; Mogrovejo-Zambrano et al., 2024) y la carga cognitiva que requiere que el estudiante invierta si los objetos no son correctamente diseñados (Angulo et al., 2023; Bautista et al., 2025; Bermejo et al., 2023; Cevikbas et al., 2023).

Los resultados apuntados por los metaanálisis, permiten identificar unos grandes bloques de dificultades que pudieran ser determinantes en la aplicación de las tecnologías de la RA, RV y RM: técnicas, pedagógicas, económicas, éticas, sociales y psicológicas, así como la resistencia al cambio por parte de las instituciones y del profesorado.

De todas formas, de cara a la integración de las tecnologías que se comentan, es importante también contemplar modelos de análisis de las tecnologías como los modelos de aceptación de las tecnologías TAM (Technology Acceptance Model) y UTAUT (Unified Theory of Technology Adoption and Use of Technology), y los modelos de integración de las tecnologías TPACK (Technological Pedagogical Content Knowledge) y el modelo SAMR (Sustitución y Aumento, Modificación y Redefinición). Modelos formulados en su momento por Davis (1989), Venkatesh (2000), Mishra y Koehler (2006) y Puentedura (2006), respectivamente.

De todos estos modelos han sido el TAM y el UTAUT, los que han despertado más investigaciones respecto a las tecnologías de la RA, RV y RM (Cabero-Almenara y Pérez Díez, 2018; Cabero-Almenara et al., 2021); apuntando una serie de conclusiones que fundamentalmente van dirigidas a que son tecnologías que despiertan un alto grado de aceptación que los lleva a mostrar a los posibles usuarios una alta intención de su utilización. Por lo que se refiere a los otros modelos incorporan dos ideas a considerar, por una parte, que una correcta incorporación de estas tecnologías requiere pensar en que la tecnología no se posiciona sola, sino que depende del contenido que transmita y de las metodologías que con ella se aplique, y por otra, que la incorporación de cualquier tecnología no es directa en la enseñanza, sino que implica un proceso lento de incorporación.

MÉTODO

La investigación que se presenta es no experimental y del tipo de encuestas (Tourrón, 2023) y con ella se persigue alcanzar los siguientes objetivos:

1. Identificar las principales dificultades percibidas en la implementación de la RA, RV y RM en el contexto de la educación superior, considerando aspectos técnicos (DT), pedagógicos (DP), económicos (DE), éticos, sociales y psicológicos (DES), así como la resistencia al cambio por parte de las instituciones y del profesorado (DRC).
2. Explorar la asociación entre las dificultades percibidas y las características sociodemográficas del profesorado, a fin de determinar tipologías diferenciadas de docentes en la percepción de barreras frente al uso de RV.
3. Determinar agrupaciones o perfiles de dificultad en función de las dificultades percibidas y las características sociodemográficas del profesorado, con el fin de ofrecer una caracterización comprensiva de los obstáculos a la integración de la RV en la educación superior.

Instrumento

Para la presente investigación se diseñó un instrumento de recogida de información específicamente adaptado a los objetivos del estudio. Este cuestionario consta de dos secciones. La primera está orientada a recabar datos sociodemográficos del participante, tales como género, edad, entre otros aspectos personales. La segunda sección, estructurada mediante una escala tipo Likert, incluye un total de 23 ítems destinados a evaluar, en una escala del 1 al 10, el grado de percepción respecto a la influencia de diversas dificultades.

Dichos ítems se distribuyen en función de las categorías de dificultad presentadas anteriormente en los objetivos a alcanzar: dificultades técnicas (4 ítems), pedagógicas (9 ítems), económicas (3 ítems), relacionadas con la resistencia al cambio por parte de instituciones y profesorado (2 ítems), y aquellas de índole ética, social y psicológica (5 ítems).

Para la validación del cuestionario, se aplicó una prueba piloto a algunos miembros del “Grupo de Investigación Didáctica” de la Universidad de Sevilla y del “Grupo de Investigación de Tecnología Educativa” de la Universidad de Málaga, y a los investigadores del proyecto MEREVIA que habían publicado artículos y comunicaciones de la aplicación de la RA, RV y RM. Posteriormente, se aplicaron los análisis estadísticos de la alfa de Cronbach y la omega de McDonald, para la obtención del índice de fiabilidad.

Esta investigación cuenta con todas las garantías éticas necesarias para asegurar la protección de los derechos y la privacidad de los participantes. En primer lugar, todos los datos recogidos han sido tratados de forma estrictamente anónima, garantizando que no se pueda identificar a ninguna persona a partir de la información obtenida. Además, los participantes fueron debidamente informados sobre los objetivos y procedimientos del estudio, y dieron su consentimiento informado de manera libre y voluntaria, aceptando también una cláusula específica de cesión de datos para fines exclusivamente científicos. El protocolo seguido en esta investigación cumple con

los principios éticos recogidos en la Declaración de Helsinki, asegurando el respeto, la dignidad y la protección de los individuos implicados. Asimismo, se ha seguido rigurosamente el Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos (Reglamento General de Protección de Datos - RGPD), garantizando así un uso responsable, seguro y transparente de la información recabada.

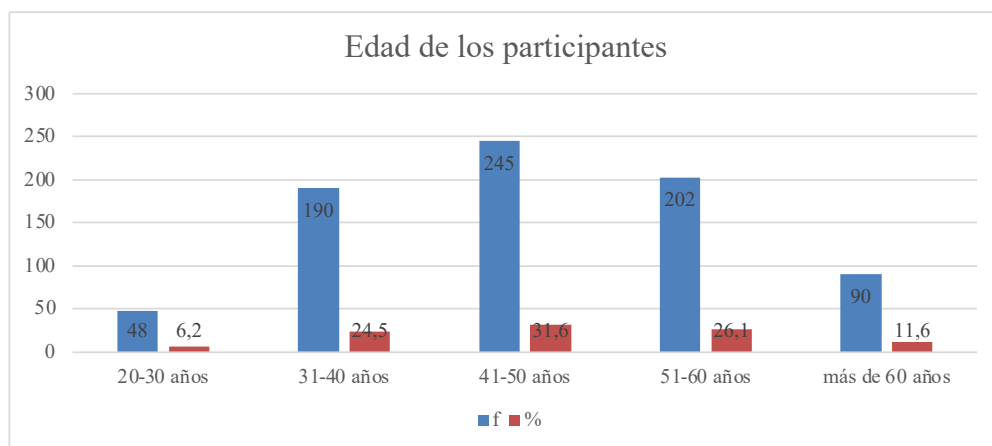
Muestra

La muestra del estudio estuvo compuesta por un total de 775 participantes. De ellos, 389 (50,2 %) se identificaron como hombres, 378 (48,8 %) como mujeres, mientras que 8 personas (1,0 %) optaron por no responder a la pregunta sobre el género.

En la Figura 1 se presentan las frecuencias y porcentajes de los participantes en función de su edad.

Figura 1

Frecuencia y porcentaje de la edad de los participantes



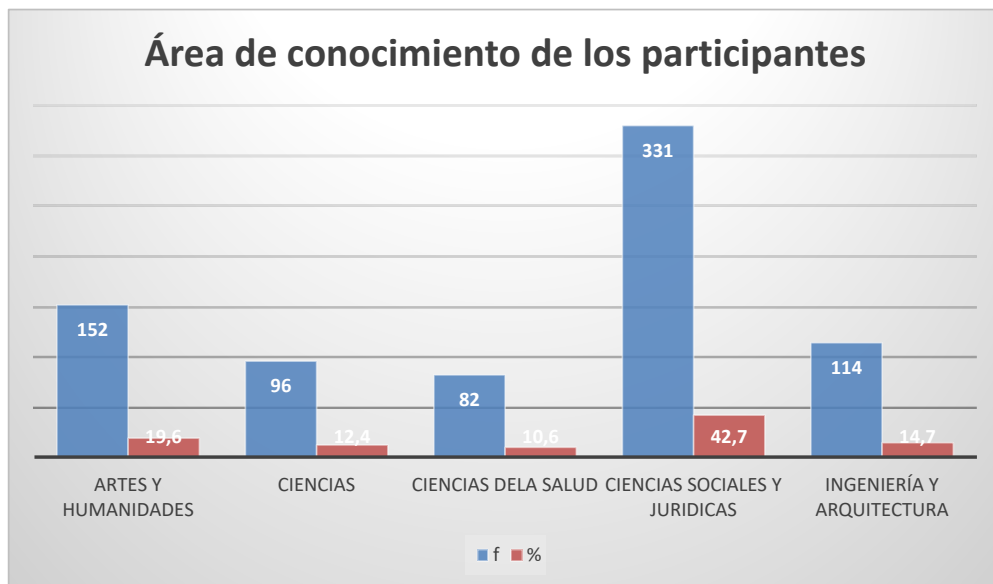
Como se puede observar el 82 % de la distribución se repartía entre los 31-60 años. Destacando los comprendidos entre 41-50 años.

De los encuestados, 395 (51,0 %) trabajaban en universidades españolas y 380 (49,0 %) en iberoamericanas.

Respecto al área de conocimiento al que corresponden los participantes en la Figura 2, se presentan las frecuencias y los porcentajes a las que pertenecía.

Figura 2

Frecuencia y porcentaje de la edad de los participantes



Como se observa, el mayor porcentaje de participantes pertenecían al área de conocimiento de “Ciencias Sociales y Jurídicas”, que podría explicarse por el procedimiento seguido para la selección de los participantes, ya que suelen ser uno de los colectivos para interesados por la incorporación de estas tecnologías a la práctica educativa.

La selección de los participantes se llevó a cabo mediante diversos procedimientos estratégicos. En primer lugar, se incluyeron personas que hubieran publicado, entre el año 2020 y febrero de 2025, algún artículo relacionado con la RA, RV y RM en revistas especializadas en lengua española. En segundo lugar, se consideraron aquellas personas que presentaron comunicaciones sobre dichas tecnologías en el Congreso Edutec 2024. Finalmente, también se incorporó profesorado perteneciente a distintos grupos de investigación de universidades españolas y latinoamericanas, de quienes se disponía de información que evidenciaba su trabajo en la incorporación de estas tecnologías en contextos educativos.

Procedimiento

Previamente al análisis principal, se llevó a cabo un estudio de fiabilidad de las escalas utilizadas para medir las distintas dificultades percibidas en el uso de la RA,

RV y RM en educación superior. Se calcularon los índices de fiabilidad interna Alfa de Cronbach y Omega de McDonald para cada una de las dimensiones incluidas en el cuestionario: dificultades técnicas (DT), pedagógicas (DP), económicas (DE), resistencia al cambio institucional y del profesorado (DRC), y dificultades éticas, sociales y psicológicas (DES). Estos coeficientes permiten estimar la consistencia interna de los ítems que componen cada dimensión, y son considerados indicadores robustos para evaluar la fiabilidad de escalas psicométricas (Dunn et al., 2014). Los valores obtenidos se situaron dentro de los rangos aceptables en investigación educativa, indicando una adecuada fiabilidad para cada subescala.

Además, se calcularon medias y desviaciones típicas para cada dimensión de dificultad percibida, con el objetivo de ofrecer una descripción inicial del comportamiento de los datos y explorar el grado de dificultad que los participantes atribuyen a cada una de las categorías evaluadas. Esta descripción estadística preliminar proporciona una base sólida para la posterior interpretación de los resultados del ACM, al contextualizar el nivel de dificultad percibido por el profesorado en relación con aspectos técnicos, pedagógicos, económicos, institucionales y éticos.

Para el análisis de perfiles y tipologías de los docentes, aplicamos la prueba estadística del ACM, estadístico que se emplea cuando se desea analizar simultáneamente más de dos variables categóricas y que permite representar gráficamente tanto a las categorías de las variables como a los individuos (casos o sujetos) en un espacio reducido de dos o más dimensiones, con el objetivo de visualizar asociaciones, similitudes y estructuras latentes en los datos (Bond y Michailides, 1997). Se tienen muchas variables cualitativas (nominales u ordinales tratadas como nominales), y se persigue con ella identificar perfiles característicos de individuos o grupos. Como se señala en Cornejo (1988, p. 97) con ella se persigue “profundizar en las relaciones de dependencia que se establece entre dos variables cualitativas observadas en una misma población”.

Esta aproximación permite identificar patrones latentes de asociación entre variables categóricas y construir un espacio factorial que represente gráficamente la configuración de dichas percepciones, facilitando así la segmentación e interpretación de perfiles docentes (Greenacre, 2007; Le Roux y Rouanet, 2010).

El empleo del ACM en este contexto se justifica por la naturaleza categórica de los datos, derivados de escalas de valoración ordinal que recogen los niveles percibidos de dificultad en cada dimensión. La técnica permite reducir la complejidad de la información y representar en un espacio bidimensional las relaciones más significativas entre categorías, conservando una proporción considerable de la varianza original (Abdi y Valentin, 2007; Escofier y Pagès, 2008). El análisis no solo permite observar cómo se agrupan las respuestas, sino también cómo interactúan con posibles factores sociodemográficos o institucionales, ofreciendo una visión más integral de las barreras para la adopción de tecnologías inmersivas (Le Roux y Rouanet, 2010).

En la investigación social, se ha utilizado para el análisis de diferentes temáticas, como el locus control (Algañaraz, 2016), la imagen del docente en la prensa (Cabero-Almenara y Loscertales-Abril, 1998), las características de los premios Nobel o la resolución de problemas matemáticos como metodología.

RESULTADOS

El índice de fiabilidad global del instrumento analizado a través de la alfa de Cronbach y la omega de McDonald, fueron de 0,915 y 0,909, respectivamente. Y por lo que se refiere a las diferentes dimensiones identificadas, en la Tabla 1, se ofrecen los valores alcanzados.

Tabla 1
Índice de fiabilidad del instrumento

Variable	Índice de fiabilidad	
	α de Cronbach	Ω de McDonald
Dificultades técnicas	,783	,778
Dificultades pedagógicas	,847	,830
Dificultades económicas	,789	,783
Dificultades por resistencia al cambio de las instituciones	,736	,734
Dificultades éticas, sociales y psicológicas	,799	,792

Valores que de acuerdo con Roco-Videla et al. (2024), señalan un nivel alto de fiabilidad del instrumento.

En línea con los objetivos planteados, y con el propósito de obtener una visión descriptiva inicial sobre el grado de dificultad percibido en el uso de la RA, RV y RM en educación superior, se procedió al análisis de las puntuaciones medias y desviaciones estándar correspondientes a cada una de las dimensiones evaluadas. En la Tabla 2, se presentan las puntuaciones medias y desviaciones típicas alcanzadas en cada de las dificultades apuntadas y en el total del instrumento.

Tabla 2

Media y desviaciones estándar de las diferentes dimensiones que conforman el instrumento

	Media	Desv. estándar
Dificultades técnicas	7,14	1,86
Dificultades pedagógicas	7,40	1,47
Dificultades económicas	8,04	1,71
Dificultades por resistencia al cambio de las instituciones	7,20	2,11
Dificultades éticas, sociales y psicológicas	7,13	1,71
Total del instrumento	7,36	1,32

El análisis de las dimensiones teniendo en cuenta las variables sociodemográficas planteadas también revela información a destacar. En cuanto al género, las mujeres reportan mayores niveles de dificultades en todas las dimensiones evaluadas, destacando especialmente en las dificultades económicas (8,14) y éticas, sociales y psicológicas (7,37). Por ramas de conocimiento, el profesorado de Ingeniería y Arquitectura presenta las puntuaciones medias más elevadas en el total del instrumento (7,62), con especial énfasis en las dificultades económicas (8,55), mientras que Ciencias Naturales reporta las más bajas (7,3). En relación con el tipo de universidad, el profesorado iberoamericano manifiesta más dificultades en todas las dimensiones, con una media general de 7,67 frente a 7,13 en el caso español. Por edad, se observa una tendencia creciente: los mayores de 61 años reportan las puntuaciones más altas en casi todas las dimensiones, alcanzando una media global de 7,75, especialmente en la resistencia institucional al cambio (8,06).

Tras la descripción inicial de las puntuaciones medias en cada dimensión de dificultad, se procedió a realizar un ACM con el fin de explorar de manera más profunda la estructura relacional entre las distintas dificultades percibidas y las variables sociodemográficas del profesorado, como el género y la edad. Esta técnica resulta especialmente adecuada para abordar uno de los objetivos centrales del estudio: identificar perfiles de percepción diferencial de las barreras para la integración de las tecnologías de la RA, RV y RM en función de características personales o institucionales. Dado que los datos utilizados provienen de escalas categorizadas (tratadas como cualitativas), el ACM permite visualizar gráficamente las asociaciones latentes entre las categorías, revelando patrones de agrupamiento y proximidad que podrían no ser evidentes a través de análisis estadísticos tradicionales. En este sentido, el ACM proporciona un marco analítico potente para interpretar de forma integrada las distintas dimensiones del fenómeno estudiado y generar una segmentación significativa del profesorado según su percepción de las dificultades para adoptar tecnologías inmersivas en entornos universitarios (Cardona et al., 2021).

Con el objetivo de facilitar la interpretación de los datos en el análisis de correspondencias múltiples, las puntuaciones obtenidas en las escalas de dificultades (originalmente medidas en una escala de 1 a 10) fueron transformadas en categorías cualitativas. Esta recodificación permitió agrupar los valores en tres niveles: Bajo (puntuaciones de 1 a 4), Medio (puntuaciones de 5 a 8) y Alto (puntuaciones de 9 a 10). Esta estrategia de categorización responde a la necesidad de adaptar los datos a los requisitos del análisis factorial aplicado, dado que el ACM trabaja con variables cualitativas, y contribuye además a una lectura más intuitiva de los perfiles y asociaciones entre niveles de dificultad y variables sociodemográficas, sin perder la riqueza informativa del instrumento original.

Como parte del procedimiento de análisis, el ACM fue sometido a un proceso iterativo de optimización hasta alcanzar un criterio de convergencia aceptable. Este proceso asegura la estabilidad y la validez de la solución factorial obtenida, permitiendo representar adecuadamente las relaciones entre las categorías incluidas en el análisis. En la Tabla 3 se muestra el historial de iteraciones realizado durante el ajuste del modelo.

Tabla 3
Historial de iteraciones

Número de iteración	Varianza explicada		
	Total	Incremento	Pérdida
29	38,699544	,000009	24,300456

El algoritmo de estimación alcanzó una convergencia estable en la iteración 29, con una varianza total explicada de 38,6 % y un incremento marginal final de solo 0,000009, lo que indica que se llegó a una solución óptima (Greenacre, 2007). Este valor es común en el Análisis de Correspondencias Múltiples (ACM) al trabajar con variables categóricas relacionadas con percepciones y experiencias subjetivas, como las dificultades enfrentadas por el profesorado al integrar tecnologías emergentes (Abdi y Valentin, 2007; Le Roux y Rouanet, 2010). La pérdida residual fue de 24,3, reflejando variabilidad no explicada, lo cual es esperable en ACM y no compromete la validez del modelo (Escofier y Pagès, 2008). En conjunto, los resultados confirman la estabilidad y fiabilidad de la solución, validando el uso interpretativo del espacio bidimensional para analizar las relaciones entre tipos de dificultades percibidas y características del profesorado (Greenacre, 2007). Esta evidencia apoya el diseño de estrategias más precisas para la formación y acompañamiento institucional en la adopción de RV/RA/RM en la docencia universitaria (Le Roux y Rouanet, 2010).

Para evaluar la consistencia interna del instrumento y la calidad de la solución factorial obtenida a través del ACM, se presenta a continuación un resumen del modelo (Tabla 4).

Tabla 4
Resumen del modelo

Dimensión	Alfa de Cronbach	Total (autovalor)	Inercia	% de varianza
1	,930	10,192	,309	40,884
2	,888	7,207	,218	21,840

El análisis revela una estructura factorial robusta y explicativa, con dos dimensiones principales que capturan una parte significativa de la variabilidad en las percepciones del profesorado sobre las dificultades para integrar tecnologías emergentes. La primera dimensión, con un alfa de Cronbach de 0,930 y un autovalor de 10,192, explica el 40,884 % de la varianza, reflejando una alta coherencia interna entre las categorías agrupadas (Greenacre, 2007; Le Roux y Rouanet, 2010).

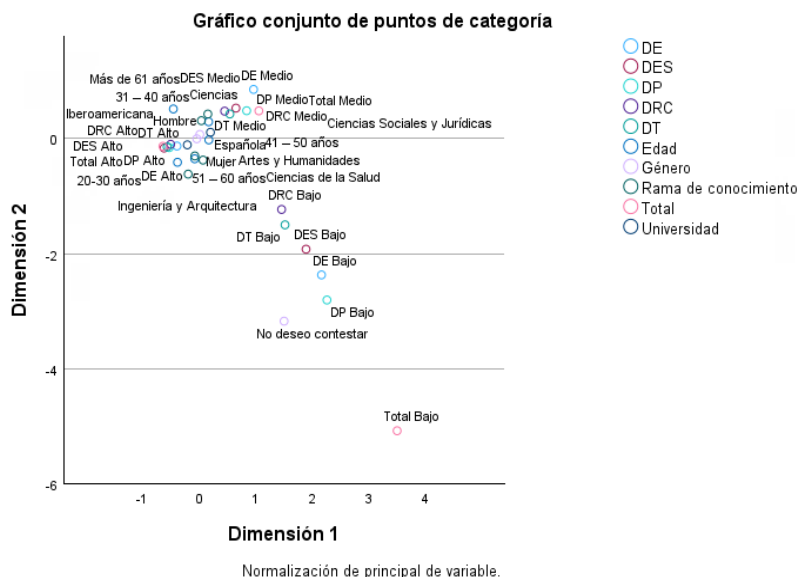
La inercia total explicada alcanza el 76,4 %, un valor inusualmente alto en el contexto del ACM, dada la complejidad de los datos categóricos (Abdi y Valentin, 2007; Escofier y Pagès, 2008). Asimismo, el promedio del alfa de Cronbach (0,913) y del autovalor (35,058) refuerza la fiabilidad psicométrica del modelo, indicando que las dimensiones son sólidas metodológica y conceptualmente (Greenacre, 2007).

Esta configuración permite identificar agrupamientos lógicos de barreras – técnicas, pedagógicas, económicas, institucionales o éticas/sociales– y perfilar distintos tipos de docentes según su vivencia de dichas dificultades. La solidez del modelo proporciona una base interpretativa fiable para analizar cómo variables como el tipo de universidad, la edad, el género o el área disciplinar influyen en la percepción de barreras, ofreciendo insumos valiosos para diseñar políticas de formación, inversión tecnológica y estrategias de acompañamiento más contextualizadas y eficaces.

A continuación, se presenta el gráfico conjunto de puntos de categoría (Figura 3), resultado del ACM, que permite observar visualmente las asociaciones entre las distintas categorías de dificultad percibida y las variables sociodemográficas y académicas analizadas (edad, género, rama de conocimiento, y ubicación de la universidad en la cual trabaja).

Figura 3

Gráfico conjunto de puntos de categoría (nota: DE = Dificultades Económicas, DES; DES = Dificultades Éticas, Sociales y Psicológicas; DP = Dificultades Pedagógicas; DRC = Dificultades de Resistencia al Cambio; DT = Dificultades Técnicas)



En este espacio bidimensional, las proximidades entre puntos indican relaciones o similitudes en los perfiles de respuesta, mientras que las distancias sugieren diferenciación entre categorías. En el siguiente apartado, se analizan los resultados más destacados del análisis, contemplando los diferentes tipos de dificultades que se han considerado en el estudio: técnicas, pedagógicas, económicas, por resistencia al cambio de las instituciones y del tipo éticas, sociales y psicológicas.

Dificultades

El análisis de correspondencias múltiples revela una clara polarización en la percepción del profesorado universitario respecto a las distintas barreras para la integración de tecnologías inmersivas (RA, RV, RM). Las dificultades técnicas (DT) y económicas (DE) se asocian con docentes mayores, hombres y de universidades iberoamericanas, mientras que las dificultades pedagógicas (DP), éticas/sociales/psicológicas (DES) y de resistencia institucional al cambio (DRC) presentan una segmentación más matizada, aunque coherente. En general, quienes perciben altos niveles de dificultad lo hacen de forma acumulativa en varias dimensiones, mientras que los perfiles con baja percepción suelen coincidir en minimizar múltiples

barreras, lo que indica patrones consistentes de respuesta. Estas diferencias sugieren la necesidad de estrategias diferenciadas de formación, apoyo tecnológico e institucional, atendiendo tanto a perfiles críticos como tecnopositivos.

Dificultades totales

La variable Total actúa como un eje integrador que resume la percepción global de dificultades. “Total Alto” se agrupa con puntuaciones altas en todas las dimensiones, representando a un profesorado crítico que percibe obstáculos estructurales importantes para la adopción de tecnologías XR. “Total Medio”, ubicado cerca del centro del gráfico, refleja una posición más equilibrada y representativa de la mayoría del profesorado. Por su parte, “Total Bajo” se posiciona en un cuadrante separado, junto a todas las demás categorías bajas, indicando un subgrupo con visión tecnopositiva, probablemente con mayor experiencia o apoyo institucional. Esta consistencia entre niveles revela una estructura latente en las percepciones docentes y permite identificar perfiles extremos y transicionales.

Variables sociodemográficas

Las variables sociodemográficas muestran efectos diferenciados sobre las percepciones. La edad no parece ejercer una influencia determinante, con una distribución equilibrada entre las distintas franjas etarias. En cuanto al género, hombres y mujeres presentan percepciones similares, pero destaca la categoría “No deseo contestar”, que se asocia con bajos niveles de dificultad y patrones de respuesta diferenciados. La rama de conocimiento introduce mayor diversidad: disciplinas como Ingeniería y Arquitectura tienden a minimizar barreras, mientras que Ciencias Sociales y Humanidades presentan mayores preocupaciones éticas y pedagógicas. Finalmente, la ubicación institucional (España vs. Iberoamérica) no genera una diferenciación fuerte, aunque se detectan matices que podrían ser explorados en futuros estudios. Estos hallazgos permiten delinear perfiles docentes útiles para diseñar intervenciones formativas contextualizadas.

A continuación, se describen los perfiles identificados, desarrollando en profundidad las características que los definen, su posicionamiento frente a las tecnologías emergentes y las implicaciones que estos hallazgos tienen para la innovación educativa en la universidad.

Perfiles

El Perfil 1: Docente Crítico y Experimentado se compone de docentes mayores de 61 años, principalmente en Iberoamérica y en disciplinas como Ciencias Sociales y Jurídicas. Presentan una percepción alta de barreras técnicas, pedagógicas,

económicas, éticas y de resistencia institucional, adoptando una postura crítica ante la incorporación apresurada de tecnologías XR sin reflexión pedagógica o ética (Aquino Negrin et al., 2021; Tapia Cortes, 2020).

El Perfil 2: Docente Promedio o de Transición incluye mayoritariamente mujeres de entre 41 y 50 años del sistema universitario español. Su percepción de dificultades es media en todas las dimensiones, con actitud abierta al cambio si se les brinda formación y apoyo institucional adecuado (Cajas Bravo et al., 2023). Este grupo es clave para facilitar la transición tecnológica desde una postura moderada.

El Perfil 3: Docente Tecnopositivo o Innovador está compuesto por docentes jóvenes (31-40 años), algunos de los cuales no declaran género, y provienen de áreas como ingeniería, ciencias o educación. Su percepción de barreras es baja, y adoptan una actitud activa y entusiasta frente a la innovación, aunque podrían subestimar riesgos éticos o sociales si no media una reflexión crítica (Arancibia Muñoz et al., 2017; UNESCO, 2023).

Por último, el Perfil 4: Docente Crítico Selectivo agrupa mayoritariamente a mujeres entre 51 y 60 años, del ámbito español y de disciplinas humanistas. Aunque no rechazan las tecnologías XR, son especialmente sensibles a las dimensiones éticas y sociales, y reconocen ciertos obstáculos institucionales y económicos. Su enfoque humanista y de justicia educativa les permite liderar desde una mirada crítica, ética y transformadora (UNESCO, 2023; Angulo Salazar et al., 2025).

DISCUSIÓN Y CONCLUSIONES

Los resultados obtenidos en esta investigación permiten avanzar en la comprensión de las barreras que enfrenta el profesorado universitario en la integración de tecnologías inmersivas como la RA, RV y RM en sus prácticas docentes. A través del uso del ACM, se han identificado perfiles claramente diferenciados de docentes, cuyas percepciones sobre las dificultades para la implementación de estas tecnologías emergentes están influenciadas por variables sociodemográficas como la edad, el género, el área de conocimiento o la ubicación de la universidad. En lo que respecta a las dificultades técnicas (DT), el estudio revela que una parte del profesorado – especialmente el más veterano o vinculado a áreas menos tecnológicas – percibe barreras importantes relacionadas con el uso, configuración y mantenimiento de las tecnologías señaladas, como también han apuntado una diversidad de investigaciones (Sales et al., 2023; Subirats-Blanco y Conde del Rio, 2024). Esta dimensión está particularmente influida por la edad, el área disciplinar y la familiaridad previa con entornos digitales, siendo más acentuada en contextos institucionales con menor inversión tecnológica. A pesar de los avances en equipamiento, las brechas técnicas siguen siendo un factor limitante en ciertos perfiles, lo que subraya la necesidad de reforzar el soporte técnico y la capacitación práctica.

En cuanto a las dificultades pedagógicas (DP), los resultados indican que muchos docentes se enfrentan a desafíos a la hora de integrar pedagógicamente

estas tecnologías dentro de marcos didácticos coherentes, como se ha sugerido desde diferentes investigaciones con las cuales coinciden los resultados aquí obtenidos (Mulders et al., 2020; Caballero-Garriazo et al., 2023). Estas dificultades no dependen únicamente del dominio técnico, sino de la comprensión metodológica de cómo aplicar las herramientas de la RA, RV y RM para enriquecer procesos de enseñanza-aprendizaje. Las docentes en transición y con experiencia en innovación educativa muestran actitudes más receptivas, pero reclaman apoyo para rediseñar sus prácticas. Esta dimensión refleja la necesidad urgente de articular propuestas formativas centradas en la transformación pedagógica y no solo en la instrucción operativa.

Las dificultades económicas (DE) emergen como una barrera transversal, aunque especialmente acentuada entre docentes jóvenes o de instituciones con recursos limitados. La percepción de falta de financiación, escasez de equipos o carencia de acceso equitativo a tecnologías inmersivas genera una sensación de imposibilidad práctica para incorporar estas herramientas. Lo encontrado coincide con lo que han apuntado una diversidad de investigaciones (Mulders et al., 2020; Angulo et al., 2023; Familoni y Onyebuchi, 2024; Obeidallah et al., 2023, Sales et al., 2023) respecto a la significación de este tipo de dificultad para incorporar estas tecnologías en la formación universitaria. Esta dimensión se vincula estrechamente con condiciones estructurales, por lo que su superación depende más de decisiones institucionales y políticas públicas que de la acción individual del profesorado. Invertir en infraestructura tecnológica y garantizar su mantenimiento se configura, por tanto, como una condición básica para avanzar hacia una digitalización inclusiva.

Respecto a las dificultades éticas, sociales y psicológicas (DES), se identificó un perfil docente especialmente sensible a las implicaciones que el uso de tecnologías RA, RV y RM pueden tener sobre la subjetividad del alumnado, la equidad educativa, la privacidad o la deshumanización de las relaciones pedagógicas, en definitiva, un problema de equidad (Mogrovejo-Zambrano et al., 2024). Estas preocupaciones son especialmente frecuentes en áreas como las Humanidades y las Ciencias Sociales, y entre docentes con una orientación humanista o crítica. Esta dimensión, muchas veces olvidada en los discursos tecnocéntricos, debe ser considerada seriamente en el diseño de políticas de innovación, mediante la inclusión de marcos éticos y protocolos de uso que garanticen una implementación responsable y centrada en la persona.

Finalmente, las dificultades por resistencia al cambio institucional (DRC) constituyen una de las barreras más relevantes, señaladas incluso por docentes con alta disposición tecnológica. La percepción de estructuras rígidas, falta de liderazgo en innovación, escaso reconocimiento a la docencia y burocracia excesiva frena la motivación y la posibilidad de implementar las tecnologías consideradas en el presente estudio. Esta dimensión pone de manifiesto que no basta con formar al profesorado, sino que es imprescindible transformar la cultura organizacional de las universidades, generando entornos que valoren la experimentación pedagógica,

promuevan redes de colaboración y flexibilicen sus estructuras para adaptarse al cambio.

En conjunto, las cinco dimensiones analizadas permiten comprender la complejidad del fenómeno y refuerzan la necesidad de adoptar un enfoque sistémico e integrado en la promoción de la innovación tecnológica en educación superior. Si bien los perfiles docentes identificados permiten diferenciar actitudes y necesidades específicas, las barreras no pueden abordarse de manera aislada ni individualizada, ya que muchas de ellas responden a condiciones estructurales y culturales del sistema universitario, por una parte, español y, por otra, iberoamericano.

Otro de los principales hallazgos del estudio es la evidencia de una estructura factorial sólida y explicativa, que permitió segmentar al profesorado según sus niveles de percepción de dificultad –técnica, pedagógica, económica, institucional y ética/social–. Esta segmentación reveló no solo la existencia de perfiles extremos, como el del docente tecnopositivo o el crítico y experimentado, sino también tipologías intermedias o híbridas que reflejan posturas de transición o crítica selectiva. La identificación de estos perfiles representa un avance relevante, ya que aporta una base empírica para diseñar políticas de formación, acompañamiento y provisión de recursos más ajustadas a las realidades y necesidades concretas del profesorado.

Asimismo, el estudio constata que, si bien existen diferencias entre áreas de conocimiento –por ejemplo, en la percepción de dificultades técnicas en Ingeniería frente a las éticas en Humanidades–, algunas barreras son transversales a todo el sistema universitario, como las relacionadas con la falta de infraestructura, la escasa formación específica del profesorado o la desconexión con marcos pedagógicos consolidados. Estas barreras, lejos de ser meramente individuales, reflejan limitaciones estructurales del entorno universitario que deben ser abordadas desde una perspectiva institucional integral. El análisis también evidenció que, aunque la edad y la experiencia profesional son variables que influyen en la percepción de ciertas dificultades, no actúan como predictores únicos o deterministas, sino que interactúan con factores culturales, organizativos y actitudinales.

En este sentido, el uso del ACM ha demostrado ser una herramienta metodológica valiosa para revelar la complejidad del fenómeno estudiado. A diferencia de otros enfoques más unidimensionales, este método permitió visualizar gráficamente relaciones latentes entre múltiples variables, favoreciendo la identificación de patrones de percepción que no serían fácilmente detectables mediante análisis tradicionales. Esta aproximación resulta especialmente pertinente en el ámbito educativo, donde la diversidad de perfiles docentes y de contextos institucionales demanda enfoques interpretativos ricos y multifactoriales.

Las implicaciones prácticas derivadas de este estudio son de gran relevancia para el diseño y la implementación de estrategias que faciliten la integración efectiva de tecnologías inmersivas –RA, RV y RM– en la docencia universitaria. En primer lugar, la identificación clara de perfiles docentes diferenciados, que reflejan percepciones variadas sobre las barreras técnicas, pedagógicas, económicas,

éticas y organizativas, subraya la necesidad de desarrollar formaciones específicas y adaptadas a las características y necesidades particulares de cada grupo. Esto implica diseñar programas de capacitación que no solo aborden aspectos técnicos operativos, sino que también fortalezcan las competencias metodológicas para una integración pedagógica coherente, así como sensibilicen respecto a las dimensiones éticas y sociales de estas tecnologías. Asimismo, el reconocimiento de que las dificultades técnicas predominan en docentes con menor familiaridad digital o en áreas menos tecnológicas exige un refuerzo del soporte técnico continuo, así como la simplificación y estandarización de herramientas para facilitar su uso. Por otro lado, las limitaciones económicas y estructurales, estrechamente ligadas a las condiciones institucionales, ponen en evidencia la necesidad de impulsar políticas de inversión sostenida en infraestructura tecnológica, que garanticen el acceso equitativo y el mantenimiento adecuado de los recursos, especialmente en universidades con menor capacidad financiera. Además, las barreras institucionales relacionadas con la resistencia al cambio y la rigidez organizativa exigen una transformación cultural profunda que promueva el liderazgo en innovación, el reconocimiento formal de la docencia y la creación de espacios colaborativos que incentiven la experimentación y el intercambio de buenas prácticas. Finalmente, las preocupaciones éticas y sociales detectadas demandan la inclusión de marcos normativos y protocolos claros que orienten un uso responsable y centrado en el bienestar del alumnado, asegurando que la incorporación de tecnologías inmersivas contribuya a una educación más inclusiva, equitativa y humana.

Pese a los aportes del presente trabajo, es necesario reconocer algunas limitaciones que deben ser consideradas al interpretar los resultados. En primer lugar, el estudio se basa en una muestra intencionada de docentes con experiencia o interés en el uso de tecnologías inmersivas, lo que puede implicar un sesgo de selección positivo hacia este tipo de herramientas. Esto significa que los perfiles identificados podrían diferir si se ampliara la muestra a docentes sin contacto previo con estas tecnologías. De todas formas, ello supondría también otra limitación, el fuerte desconocimiento que los docentes tienen de estas tecnologías.

En segundo lugar, la naturaleza autoinformada del cuestionario y su aplicación en línea implican ciertas limitaciones metodológicas inherentes, como la posibilidad de respuestas socialmente deseables o la dificultad para controlar la comprensión homogénea de los ítems. Si bien el cuestionario mostró una alta fiabilidad interna, sería conveniente contrastar los resultados con estudios observacionales o entrevistas en profundidad.

Además, la categorización de las variables ordinales en niveles (Bajo, Medio, Alto) responde a criterios estadísticos y de interpretación gráfica, pero esta recodificación puede haber suavizado matices importantes en las respuestas individuales. Finalmente, el ACM, aunque robusto, implica una reducción de la información original a un espacio bidimensional, por lo que una parte de la variabilidad inevitablemente queda sin representar en los planos interpretativos.

Las líneas futuras de investigación deberían avanzar hacia un enfoque más longitudinal y mixto. Sería relevante desarrollar estudios que no solo capten la percepción de barreras en un momento determinado, sino que analicen la evolución de estas percepciones tras la participación del profesorado en acciones formativas o en experiencias prácticas con tecnologías XR. Ello permitiría evaluar el impacto real de la capacitación y la infraestructura en la transformación de actitudes, creencias y prácticas docentes.

Por otra parte, sería interesante replicar la investigación pasado un par de años, y por tanto la realización de dos congresos de Edutec y la publicación de nuevos artículos en revistas científicas, para analizar qué dificultades han aumentado y cuáles se ha reducido y mantenido. Y realizar un estudio comparativo entre ambos momentos.

Asimismo, se propone complementar los análisis cuantitativos con estudios cualitativos que profundicen en la vivencia subjetiva de las dificultades percibidas, especialmente en los perfiles extremos detectados (por ejemplo, el docente tecnopositivo o el crítico selectivo). La integración de entrevistas, grupos focales o análisis de casos permitiría enriquecer la comprensión de los significados que los docentes atribuyen a estas tecnologías, así como de los factores emocionales, éticos o ideológicos que median su aceptación o rechazo.

Es importante asumir, para la interpretación correcta de los resultados, el procedimiento seguido para la obtención de los participantes, que si bien es usual en los estudios de este tipo realizado y el criterio utilizado justificaba su selección, sin lugar a duda incorporan una limitación. Al mismo tiempo la descompensación por áreas de conocimiento puede implicar una limitación por ello sería conveniente, replicar la investigación con procedimientos de selección de la muestra estratificados, con la aplicación del “coeficiente de competencia experta” (Cabero-Almenara y Barroso-Osuna, 2013) y considerar en la muestra el porcentaje existente de profesorado por cada área de conocimiento; sin embargo esto último plantea una limitación, ya que no se ha trabajado con profesorado de universidad, sino con quienes publican sobre la incorporación de la RA, RV y RM.

Finalmente, sería pertinente ampliar la muestra a otros contextos universitarios internacionales, así como explorar las relaciones entre las barreras identificadas y variables como la cultura institucional, los modelos pedagógicos predominantes o el tipo de liderazgo académico. Esta mirada más amplia y comparativa permitiría generar marcos interpretativos transferibles, que orienten el diseño de políticas globales y locales más coherentes con las necesidades reales del profesorado universitario en la era de la transformación digital. En este sentido, resulta fundamental vertebrar programas de formación específicos que respondan a los distintos niveles de dominio tecnológico y a los perfiles docentes identificados, de modo que la capacitación sea verdaderamente contextualizada y eficaz, atendiendo tanto a quienes se encuentran en fases iniciales de familiarización con las tecnologías inmersivas como a aquellos con mayor experiencia o una orientación crítica. Como línea futura, también se

propone incorporar análisis estadísticos complementarios que permitan contrastar diferencias significativas entre perfiles docentes y variables sociodemográficas, para afinar aún más las estrategias formativas y de acompañamiento.

Financiación

La investigación tiene financiación a través del Programa Estatal para Promover la Investigación Científica y Tecnológica y su Transferencia, dentro del marco del Plan Estatal de Investigación Científica, Técnica y de Innovación 2021-2023. Ministerio de Ciencia e Innovación (MEREVIA (PID2022-136430OB7-IOO)).

REFERENCIAS

- Abdi, H. y Valentin, D. (2007). Multiple correspondence analysis. En N. J. Salkind (Ed.), *Encyclopedia of measurement and statistics* (pp. 651-657). SAGE Publications.
- Algañaraz, V. H. (2016). El “Análisis de Correspondencias Múltiples” como herramienta metodológica de síntesis teórica y empírica. Su aporte al estudio del locus universitario privado argentino (1955-1983). *Revista Latinoamericana de Metodología de las Ciencias Sociales*, 6(1), e003.
- AlGerafi, M. A. M., Zhou, Y., Oubibi, M. y Wijaya, T. (2023). Unlocking the potential: A comprehensive evaluation of augmented reality and virtual reality in education. *Electronics*, 12(18), 3953. <https://doi.org/10.3390/electronics12183953>
- Angulo, G., Lewis, F., Plante, P. y Brassard, C. (2023). Estado del arte sobre el uso de la realidad virtual, la realidad aumentada y el video 360° en educación superior. *EduTec. Revista Electrónica de Tecnología Educativa*, 84, 35-51. <https://doi.org/10.21556/edutec.2023.84.2769>
- Angulo Salazar, D. B., Andy Shiguango, N. P., Tipan Lugmania, N. D. y Vera Tigrero, M. del P. (2025). Las TIC en el perfil docente: Análisis de los conocimientos y uso de la tecnología. *Estudios y Perspectivas Revista Científica y Académica*, 5(1), 1495-1514.
- Aquino Negrin, O. y Hernández Romero, G. (2021). Tecnologías de la Información y Comunicación como herramientas de trabajo en docentes mayores de 60 años: Uso de las TIC en docentes mayores de 60 años. *Revista Científica Sinapsis*, 1(19). <https://doi.org/10.37117/s.v19i1.429>
- Arancibia Muñoz, M. L., Valdivia Zamorano, I., Araneda Riveros, S. M. y Cabero-Almenara, J. (2017). Tipologías para la innovación tecnológica en docentes de educación superior a partir de un análisis de conglomerados: Un estudio exploratorio. *Revista de Educación a Distancia (RED)*, 17(55). <https://doi.org/10.6018/red/55/5>
- Bautista, L. E., Maradei, F. y Pedraza, G. (2025). Análisis de la disposición espacial de contenido en entornos de realidad aumentada y su efecto en la carga cognitiva de los usuarios. *Pixel-Bit. Revista de Medios y Educación*, 72, 39-69.
- Bermejo, B., Cortes, D., Oskam, J., Moilanen, T., Loijas, J., Govender, P., Hussey, J., Schmidt, A. L., Burbach, R., King, D., O'Connor, C. y Dunlea, D. (2023). AR/VR teaching-learning experiences in higher education institutions (HEI): A systematic literature review. *Informatics*,

- 10(2), 45. <https://doi.org/10.3390/informatics10020045>
- Bond, J. y Michailides, G. (1997). Interactive correspondence analysis in a dynamic object-oriented environment. *Journal of Statistical Software*, 2(8), 1-30. <https://doi.org/10.18637/jss.v002.i08>
- Caballero-Garriazo, J., Rojas-Huacanca, J., Sánchez-Castro, A. y Lázaro-Aguirre, A. (2023). Revisión sistemática sobre la aplicación de la realidad virtual en la educación universitaria. *Revista Electrónica Educare*, 27(3), 1-18. <https://doi.org/10.15359/ree.27-3.17271>
- Cabero-Almenara, J. y Barroso-Osuna, J. (2013). La utilización del juicio de experto para la evaluación de TIC: el coeficiente de competencia experta. *Bordón*, 65(2), 25-38. <https://doi.org/10.13042/brp.2013.65202>
- Cabero-Almenara, J., Barroso-Osuna, J. y Martínez-Roig, R. (2021). Mixed, augmented and virtual reality applied to the teaching of mathematics for architects. *Applied Sciences*, 11(15), 7125. <https://doi.org/10.3390/app11157125>
- Cabero-Almenara, J. y Loscertales-Abril, F. (Eds.). (1998). *¿Cómo nos ven los demás? La imagen del profesor y la enseñanza en los medios de comunicación social*. Secretariado de Publicaciones de la Universidad de Sevilla.
- Cabero-Almenara, J. y Pérez Díez, J. L. (2018). Validación del modelo TAM de adopción de la realidad aumentada mediante ecuaciones estructurales. *Estudios sobre Educación*, 34, 129-153. <https://doi.org/10.15581/004.34.129-153>
- Cabero-Almenara, J., Rodríguez Gallego, M. y Llorente-Cejudo, C. (2025). Realidad mixta, virtual y aumentada: Tecnologías para el aprendizaje. *Texto libre. Linguagem e Tecnologia*, 18, 1-19.
- Cabero-Almenara, J., Valencia, R. y Llorente, C. (2022). Ecosystem of emerging technologies: Augmented, virtual and mixed reality. *Tecnología, Ciencia y Educación*, 23, 7-22. <https://doi.org/10.51302/tce.2022.1148>
- Cajas Bravo, T. V., Silva Infantes, M. y Dávila Morán, R. C. (2023). Percepción docente sobre el uso de la tecnología en la educación superior. *Revista Conrado*, 19(90), 326-335.
- Cardona, C., Padilla, L. y Gutierrez, J. (2021). *Análisis de correspondencia múltiple (ACM)*. <https://rpubs.com/ocamilocardona/813536>
- Cevikbas, M., Bulut, N. y Kaiser, G. (2023). Exploring the benefits and drawbacks of AR and VR technologies for learners of mathematics: Recent developments. *Systems*, 11(5), 244. <https://doi.org/10.3390/systems11050244>
- Coban, M., Bolat, Y. y Goksu, I. (2022). The potential of immersive virtual reality to enhance learning: A meta-analysis. *Educational Research Review*, 36, 100452. <https://doi.org/10.1016/j.edurev.2022.100452>
- Cornejo, J. M. (1988). *Técnicas de investigación social. El análisis de correspondencia*. PPU.
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319-340. <https://doi.org/10.2307/249008>
- Dunn, T. J., Baguley, T. y Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, 105(3), 399-412. <https://doi.org/10.1111/bjop.12046>
- Escofier, B. y Pagès, J. (2008). *Analyses factorielles simples et multiples: Objectifs, méthodes et interprétation* (4^a ed.). Dunod.
- Familoni, B. T. y Onyebuchi, N. C. (2024). Augmented and virtual reality in US education: A review: Analyzing the impact, effectiveness, and future prospects of AR/VR tools in enhancing learning experiences. *International Journal of*

- Applied Research in Social Sciences*, 6(4), 642-663. <https://doi.org/10.51594/ijarss.v6i4.1043>
- Greenacre, M. (2007). *Correspondence analysis in practice* (2ª ed.). Chapman and Hall/CRC.
- Hamilton, D., McKechnie, J., Edgerton, E. y Wilson, C. (2021). Immersive virtual reality as a pedagogical tool in education: A systematic literature review of quantitative learning outcomes and experimental design. *Journal of Computer Assisted Learning*, 37(5), 1-32. <https://doi.org/10.1007/s40692-020-00169-2>
- Jingru, Z., Yahaya, W. y Sanmugam, M. (2025). An educational mobile augmented reality system to enhance undergraduate. *Interactive Learning Environments*.
- Le Roux, B. y Rouanet, H. (2010). *Multiple correspondence analysis*. SAGE Publications. <https://doi.org/10.4135/9781412993906>
- Liu, J., Liu, Q., Yu, S., Ma, J., Liu, M. y Wu, L. (2024). How do autonomy and learner characteristics combine to influence learners' learning outcomes and cognitive load in virtual reality learning environments? A fuzzy-set qualitative comparative analysis approach. *Education and Information Technologies*, 29(1), 77-101. <https://doi.org/10.1007/s10639-023-12262-x>
- Luna, A., Ortiz Colón, A. M. y Rodríguez Moreno, J. (2023). Tecnologías inmersivas en el aprendizaje autorregulado: Revisión sistemática de literatura científica. *Digital Education Review*, 44, 105-113. <https://doi.org/10.1344/der.2023.44.105-113>
- Marín-Díaz, V., Sampedro-Requena, B. E., Vega-Gea, E. M. y Ruiz-Palmero, J. (2022). Conocimientos tecnopedagógicos del profesorado de secundaria sobre la realidad mixta. *Revista Tecnología, Ciencia y Educación*, (23), 23-48. <https://doi.org/10.51302/tce.2022.1984>
- Maroungkas, A., Troussas, C., Krouska, A. y Sgouropoulou, C. (2024). How personalized and effective is immersive virtual reality in education? A systematic literature review for the last decade. *Multimedia Tools and Applications*, 83, 18185-18233. <https://doi.org/10.1007/s11042-023-15986-7>
- Menjivar, E., Sanchezs, E., Ruiz Palmero, J. y Valenzuela, T. (2021). Revisión de la producción científica sobre la realidad virtual entre 2016 y 2020 a través de Scopus y WoS. *EDMETIC, Revista de Educación Mediática y TIC*, 10(2), 26-55. <https://doi.org/10.21071/edmetic.v10i2.13422>
- Mishra, P. y Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record*, 108(6), 1017-1054. <https://doi.org/10.1111/j.1467-9620.2006.00684.x>
- Mogrovejo-Zambrano, J. N., Montalván-Vélez, C. L., Barragan-Espinoza, G. M., & Cabrera-Davila, M. A. (2024). Fenomenología de la realidad virtual: Explorando la experiencia humana en entornos digitales inmersivos. *Journal of Economic and Social Science Research*, 4(1), 149-159. <https://doi.org/10.55813/gaea/jessr/v4/n1/88>
- Mohsen, M. A. y Alangari, T. S. (2024). Analyzing two decades of immersive technology research in education: Trends, clusters, and future directions. *Education and Information Technologies*, 29(3), 3571-3587. <https://doi.org/10.1007/s10639-023-11968-2>
- Mulders, M., Buchner, J. y Kerres, M. (2020). A framework for the use of immersive virtual reality in learning environments. *International Journal of Emerging Technologies in Learning (iJET)*, 15(24), 208-224. <https://doi.org/10.3991/ijet.v15i24.16615>
- Obeidallah, R., Al Ahmad, A. y Qutishat, D. (2023). Challenges of extended reality

- technology in higher education: A review. *Journal of Emerging Technologies in Learning (iJET)*, 18(14), 39-50. <https://doi.org/10.3991/ijet.v18i14.39871>
- Perifanou, M., Economides, A. A. y Nikou, S. A. (2023). Teachers' views on integrating augmented reality in education: Needs, opportunities, challenges and recommendations. *Future Internet*, 15(1), 20. <https://doi.org/10.3390/fi15010020>
- Puentedura, R. (2006). *Transformation, technology, and education*. Hippasus. <http://hippasus.com/resources/tte/>
- Radianti, J., Majchrzak, T. A., Fromm, J. y Wohlgenannt, I. (2020). A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & Education*, 147, 103778. <https://doi.org/10.1016/j.compedu.2019.103778>
- Roco-Videla, Á., Aguilera-Eguía, R. y Olguin-Barraza, M. (2024). Ventajas del uso del coeficiente de omega de McDonald frente al alfa de Cronbach. *Nutrición Hospitalaria*, 41(1), 262-263. <https://doi.org/10.20960/nh.04879>
- Sales, S., Macédo, R., De Oliveira, V., Souza, W., Gama, L., De Brito, S. y Winkler, I. (2023). Augmented reality in education for industry 4.0: What are the barriers to adoption? *Cuadernos de Educación y Desarrollo*, 15(1), 481-505. <https://doi.org/10.55905/cuadvi15n1-025>
- Subirats-Blanco, R. y Conde del Rio, M. A. (2024). Integración de la realidad aumentada en la enseñanza del inglés en educación primaria: Mejorando el aprendizaje lingüístico y la adquisición de vocabulario. *EDMETIC*, 13(2), art.7. <https://doi.org/10.21071/edmetic.v13i2.17072>
- Tapia Cortes, C. (2020). Tipologías de uso educativo de las Tecnologías de la Información y Comunicación: Una revisión sistemática de la literatura. *Edutec. Revista Electrónica De Tecnología Educativa*, (71), 16-34. <https://doi.org/10.21556/edutec.2020.71.1489>
- Toala-Palma, J., Arteaga-Mera, J., Quintana, J. y Santana, I. (2020). Virtual reality as a tool for educational innovation. *Episteme Koinonia*, 3(5). <https://fundacionkoinonia.com.ve/ojs/index.php/epistemekoinonia/article/view/835>
- Tourrón, J. (2023). *Análisis de datos y medidas en educación* (Vol. II). Unir.
- UNESCO. (2023). *Informe GEM 2023: Tecnología en la educación. Una herramienta en los términos de los docentes*. UNESCO.
- Upadhyay, B., Chalil Madathil, K., Hegde, S., Anderson, D., Wooldridge, E., Presley, D., Perez, L. y Reid, B. (2024). Barriers toward the implementation of extended reality (XR) technologies to support education and training in workforce development programs. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 68(1), 265-269. <https://doi.org/10.1177/10711813241275080>
- Venkatesh, V. (2000). Determinants of perceived ease of use: Integrating control, intrinsic motivation, and emotion into the technology acceptance model. *Information Systems Research*, 11(4), 342-365. <https://doi.org/10.1287/isre.11.4.342.11872>
- Victoria Maldonado, J. J., Fuentes-Cabrera, A., Fernández-Cerero, J. y Sadio-Ramos, F. J. (2024). Influencia de la realidad virtual en el rendimiento académico en educación secundaria a través de un meta-análisis. *Pixel-Bit. Revista de Medios y Educación*, 71, 107-121. <https://doi.org/10.12795/pixelbit.104279>
- Yu, Z. y Xu, W. (2022). A meta-analysis and systematic review of the effect of virtual reality technology on users' learning outcomes. *Computer Applications in Engineering Education*, 30(6), 1470-1484. <https://doi.org/10.1002/cae.22532>

Fecha de recepción del artículo: 1 de junio de 2025

Fecha de aceptación del artículo: 31 de julio de 2025

Fecha de aprobación para maquetación: 5 de septiembre de 2025

Fecha de publicación en OnlineFirst: 12 de octubre de 2025

Fecha de publicación: 1 de enero de 2026

Desarrollo y validación de la escala de Aprendizaje Autorregulado con IA (AI-SRL)

Development and validation of the AI-Enhanced Self-Regulated Learning (AI-SRL) scale



- ✉ Mehmet Ali Ayaz - *Giresun University, GRÜ (Turquía)*
- ✉ Fatih Karataş - *Neuşehir Hacı Bektaş Veli University, NEVÜ (Turquía)*
- ✉ Erkan Yüce - *Aksaray University, ASÜ (Turquía)*

RESUMEN

La integración de la Inteligencia Artificial (IA) en entornos educativos brinda nuevas oportunidades para mejorar el aprendizaje autorregulado (SRL) en los estudiantes. Sin embargo, las herramientas actuales carecen de precisión para evaluar cómo la IA influye en las capacidades autorregulatorias de los estudiantes. Este estudio presenta la Escala de Aprendizaje Autorregulado Mejorado por IA (AI-SRL), diseñada para evaluar el aprendizaje autorregulado en entornos educativos asistidos por IA. A través de un riguroso proceso de desarrollo de escala, que incluyó revisión de literatura, consulta con expertos y validación sistemática, se construyó un instrumento de medición integral. En el desarrollo se empleó un enfoque de validación en dos fases con muestras separadas para análisis factoriales exploratorios y confirmatorios. La escala final validada consta de 22 ítems organizados en cinco dimensiones: Competencia en IA, Conciencia de Aprendizaje, Estrategias de Aprendizaje, Implicación y Eficiencia, y Colaboración Ética. Estos factores capturan la naturaleza multifacética de cómo los estudiantes regulan su aprendizaje al usar herramientas de IA, desde competencia técnica hasta consideraciones éticas. La escala demostró sólidas propiedades psicométricas, con excelente consistencia interna y robusta validez de constructo. Este instrumento validado tiene aplicaciones prácticas para docentes que buscan optimizar la integración de IA en sus aulas, investigadores que investigan la intersección de IA y aprendizaje autorregulado, e instituciones desarrollando currículos mejorados con IA. La Escala AI-SRL proporciona un marco confiable para evaluar en qué medida los estudiantes aprovechan las herramientas de IA mientras mantienen sus capacidades autorregulatorias, contribuyendo así a una implementación de IA más efectiva y responsable.

Palabras clave: IA; AIED; aprendizaje autorregulado; AI-SRL; escala AI-SRL.

ABSTRACT

The integration of Artificial Intelligence (AI) into educational settings offers new opportunities for enhancing self-regulated learning (SRL) among students. However, current tools lack precision in measuring how AI influences learners' self-regulatory capabilities. This study introduces the AI-Enhanced Self-Regulated Learning (AI-SRL) Scale, designed to assess SRL in AI-supported educational environments. Through a rigorous scale development process involving a literature review, expert consultation, and systematic validation, we developed a comprehensive measurement instrument. The development employed a two-phase validation approach with separate samples for exploratory and confirmatory factor analyses. The final validated scale comprises 22 items organized into five dimensions: AI Competence, Learning Awareness, Learning Strategies, Engagement and Efficiency, and Ethical Collaboration. These factors capture the multifaceted nature of how students regulate their learning when using AI tools, from technical proficiency to ethical considerations. The scale demonstrated strong psychometric properties, with excellent internal consistency and robust construct validity. This validated instrument has practical applications for educators aiming to optimize AI integration in their classrooms, researchers investigating the intersection of AI and SRL, and institutions developing AI-enhanced curricula. The AI-SRL Scale provides a reliable framework for assessing how effectively students leverage AI tools while maintaining their self-regulatory capabilities, thereby contributing to more effective and responsible AI implementation.

Keywords: AI; AIED; self-regulated learning; AI-SRL; AI-SRL scale.

INTRODUCCIÓN

El aprendizaje autorregulado (AAR), componente crítico para el éxito en el aprendizaje digital y permanente, implica establecer metas, supervisar el progreso y ajustar conductas para alcanzar el éxito académico (Hilpert et al., 2023; Lai, 2024). En la era digital, el AAR es cada vez más crucial para la formación continua, pues permite a las personas adaptarse a necesidades educativas cambiantes y promueve la resolución de problemas (Adams et al., 2024; Hilpert et al., 2023; Kong y Yang, 2024). La evolución de las prácticas educativas y la integración de tecnologías avanzadas como la IA ofrecen oportunidades para potenciar el AAR y afrontar sus desafíos.

Las bases teóricas del AAR se fundamentan en dos marcos integrales que han moldeado la psicología educativa durante más de tres décadas. El modelo de fases cíclicas de Zimmerman, basado en la teoría social cognitiva, concibe el AAR como un proceso estructurado en tres fases recurrentes: planificación, ejecución y autorreflexión (Panadero, 2017; Schunk y Zimmerman, 2012; Zimmerman, 1989). La fase de planificación abarca el análisis de la tarea —incluyendo fijar metas y la planificación estratégica— junto con creencias motivacionales como la autoeficacia, las expectativas de resultado y el interés intrínseco. La fase de ejecución implica estrategias de autocontrol como la atención focalizada y la autoinstrucción, combinadas con la autoobservación mediante la monitorización metacognitiva del progreso. La fase de autorreflexión incluye el autojuicio, donde se evalúa el desempeño contra estándares, y procesos de autorreacción con respuestas de satisfacción y reacciones adaptativas que influyen en futuros aprendizajes (Panadero, 2017; Tinajero et al., 2024).

Como complemento, el marco de Pintrich (2000) presenta una taxonomía de cuatro fases —planificación/activación, monitorización, control y reacción/reflexión— aplicadas a cuatro áreas de regulación: cognición, motivación/afecto, conducta y contexto (Tinajero et al., 2024; Yot-Domínguez y Marcelo, 2017). Este marco es valioso por su tratamiento explícito de la regulación motivacional y los factores contextuales, y proporciona una matriz de 16 celdas de procesos regulatorios que abarca de forma exhaustiva los componentes del AAR. El modelo de Pintrich enfatiza que el AAR es un proceso activo y constructivo donde los aprendices establecen metas, y luego intentan monitorizar, regular y controlar su cognición, motivación y conducta, guiados por sus objetivos y el entorno (Panadero, 2017). Estos modelos fundamentales, aunque desarrollados para formatos convencionales, constituyen la arquitectura teórica esencial para investigar cómo la IA transforma los procesos establecidos del AAR.

La integración de la IA en entornos educativos transforma la estructura del aprendizaje, creando oportunidades y desafíos para fomentar el AAR (Chang et al., 2023; Grüneke et al., 2024). Las tecnologías de IA ofrecen formas innovadoras para potenciar las estrategias de AAR, al proveer retroalimentación personalizada, apoyo

flexible y observación en tiempo real que ayudan al aprendiz a gestionar eficazmente su aprendizaje (Huang et al., 2023; Molenaar et al., 2023; Ng et al., 2024). La investigación demuestra que la IA apoya la regulación metacognitiva, cognitiva y conductual en las cuatro fases de Pintrich; por ejemplo, los paneles de analíticas de aprendizaje mejoran la conciencia metacognitiva a través del seguimiento del progreso y la visualización del rendimiento (Gkintoni et al., 2025; Seo et al., 2021; Wang y Lin, 2023). Además, las herramientas de IA apoyan estrategias metacognitivas con soporte en la fijación de metas y planificación, alineándose con la fase de planificación de Zimmerman (Anggoro y Pratiwi, 2023; Liu et al., 2024). Estos sistemas rastrean conductas y ofrecen información sobre los procesos de AAR, y ofrecen soluciones para su desarrollo (Chang et al., 2023; Jin et al., 2023; Molenaar et al., 2023).

Aunque las herramientas de IA pueden apoyar habilidades metacognitivas y la monitorización en tiempo real (Dahri et al., 2024; Jin et al., 2023), los investigadores temen que la dependencia excesiva de la IA obstaculice el desarrollo de habilidades esenciales (Boguslawski et al., 2025). La literatura destaca una tensión entre la capacidad de la IA para dar apoyo personalizado y el riesgo de mermar la autonomía del aprendiz por la descarga cognitiva, donde el uso frecuente de la IA se asocia negativamente con el pensamiento crítico (Chiu, 2024; Gerlich, 2025; Kong y Yang, 2024). Los estudios revelan que, si bien los aprendices perciben la IA como útil para la regulación metacognitiva, cognitiva y conductual, prefieren el apoyo humano para sus necesidades motivacionales. Esto sugiere que la IA tiene dificultades para apoyar la regulación de la motivación, un componente crítico del marco de Pintrich (Lan y Zhou, 2025; Seo et al., 2021). Los estudios también enfatizan la importancia de equilibrar el apoyo de la IA y la guía humana en entornos educativos (Kong y Yang, 2024; Lai, 2024). Si bien la IA ofrece supervisión y retroalimentación detalladas, los educadores son cruciales para diseñar actividades y dar apoyo emocional (Chiu et al., 2023). La investigación sugiere que una implementación eficaz de la IA en el AAR requiere considerar cómo estas herramientas pueden complementar, y no reemplazar, la interacción humana, sobre todo en el desarrollo metacognitivo y la motivación del estudiante (Wang y Lin, 2023; Xia et al., 2023).

No obstante, el potencial de la IA para mejorar plenamente el AAR aún no se ha materializado debido a varias barreras. Mientras algunos estudios destacan que la IA puede aliviar la ansiedad ante el aprendizaje y mejorar la motivación (Biju et al., 2024), otros señalan la necesidad de una integración equilibrada para mantener la participación activa del estudiante (Grüneke et al., 2024). Además, han surgido preocupaciones éticas de gran relevancia, entre ellas la privacidad de los datos y la vigilancia, el sesgo algorítmico que puede perpetuar desigualdades educativas, y desafíos a la integridad académica, dado que el contenido generado por IA complica la evaluación del aprendizaje auténtico (Gkintoni et al., 2025; Walter, 2024). Para afrontar estos retos, los investigadores enfatizan la necesidad crucial de desarrollar

instrumentos de medición fiables para evaluar el impacto de la IA en los procesos de AAR (Baker, 2023; Grüneke et al., 2024; Jin et al., 2023; Molenaar et al., 2023).

Basado en fundamentos teóricos, este estudio conceptualiza el AAR potenciado por IA en cinco dimensiones que amplían y operacionalizan los marcos de Zimmerman y Pintrich para el contexto tecnológico actual. El factor de Competencia en IA se alinea con la fase de planificación, que incluye la autoeficacia en el uso de herramientas de IA y su planificación estratégica. La Conciencia del Aprendizaje corresponde a la monitorización metacognitiva de ambos modelos y capta la percepción del estudiante sobre su comprensión y progreso con la IA. Los factores de Estrategias de Aprendizaje, Implicación y Eficiencia se corresponden con las fases de desempeño y control de ambos marcos, y refleja la selección de estrategias cognitivas y la adaptación, así como la regulación conductual y motivacional necesaria para perseverar en las tareas de aprendizaje. De manera crítica, el factor de Colaboración Ética representa una extensión necesaria a los modelos clásicos, operacionaliza la regulación del contexto moderno (la cuarta área regulatoria de Pintrich) que aborda el sesgo algorítmico, la privacidad de los datos y las preocupaciones sobre la integridad académica: demandas regulatorias únicas de la era de la IA no anticipadas en los marcos tradicionales de AAR (Gkintoni et al., 2025; Walter, 2024).

El desarrollo de la Escala AI-SRL es clave para evaluar diversos entornos de tecnología educativa. En plataformas con evaluación formativa en tiempo real como Socrative, sus dimensiones de Conciencia del Aprendizaje y Estrategias de Aprendizaje miden si el *feedback* instantáneo promueve la monitorización metacognitiva y el ajuste estratégico (Bauer et al., 2025). En Sistemas de Gestión del Aprendizaje (SGA) como Moodle con asistentes de IA, las subescalas de Competencia en IA e Implicación y Eficiencia evalúan si los tutores IA 24/7 mejoran la autorregulación del estudiante sin supervisión directa. Con respecto a los Sistemas de Tutoría Inteligente que proporcionan andamiaje adaptativo y rutas de aprendizaje personalizadas, la escala completa ofrece una evaluación integral de todo el ciclo regulatorio, desde la planificación y el uso de estrategias hasta las consideraciones éticas, y proporciona retroalimentación valiosa para el diseño de sistemas que sean tanto inteligentes como pedagógicamente sólidos (Wu y Chiu, 2025). Con el uso creciente de IA generativa en tareas académicas, la escala ofrece métodos estructurados para investigar la regulación de la indagación, la evaluación crítica y el uso ético, cerrando brechas en la comprensión de los ecosistemas digitales de aprendizaje (Li et al., 2025; Mohamed et al., 2025). Por ello, este estudio aborda las siguientes preguntas:

PI1: ¿Cómo puede medirse de manera integral el aprendizaje autorregulado potenciado por IA con una escala que operacionalice y extienda los marcos teóricos establecidos por Zimmerman y Pintrich para contextos tecnológicos contemporáneos?

PI2: ¿Cuál es la estructura factorial y cuáles son las propiedades psicométricas de la Escala AI-SRL cuando se valida mediante procedimientos de análisis factorial exploratorio y confirmatorio?

PI3: ¿En qué medida la Escala AI-SRL funciona como un instrumento fiable y válido para capturar las dimensiones específicas del AAR en entornos educativos apoyados por IA?

Revisión de la literatura

Los avances recientes en IA han transformado el panorama del aprendizaje autorregulado (AAR) en entornos educativos, creando tanto oportunidades sin precedentes como desafíos complejos que exigen un examen teórico y empírico. La literatura revela un panorama matizado donde las tecnologías de IA muestran un gran potencial para apoyar las dimensiones cognitivas y metacognitivas del AAR (Seo et al., 2021; Wang y Lin, 2023), pero también presentan desafíos para la regulación motivacional y plantean preocupaciones éticas sobre la agencia del aprendiz y su bienestar digital (Lan y Zhou, 2025).

Las bases teóricas del AAR, en particular el modelo cíclico de Zimmerman y el marco integral de Pintrich, han demostrado ser notablemente robustas en contextos de aprendizaje potenciados por IA. La estructura cíclica de tres fases de Zimmerman —planificación, ejecución y autorreflexión— se corresponde eficazmente con las aplicaciones de IA (Chiu, 2024), las cuales ofrecen un apoyo específico en cada fase. La investigación de Lan y Zhou (2025) encontró que, aunque el 50 % de las aplicaciones de IA apoyan principalmente la fase de ejecución, el apoyo integral en las tres fases sigue siendo escaso (solo el 21 % de las aplicaciones), siendo la fase de autorreflexión la que menos apoyo recibe. Esto sugiere una necesidad continua de guía humana en los procesos reflexivos. Del mismo modo, el marco de Pintrich y sus cuatro áreas de regulación (cognitiva, motivacional/afectiva, conductual y contextual) encuentra claros paralelismos en los entornos de aprendizaje mediados por IA (Yot-Domínguez y Marcelo, 2017), aunque la regulación motivacional sigue siendo particularmente problemática (Jin et al., 2023; Seo et al., 2021).

Las tecnologías de IA han demostrado un apoyo significativo a los componentes centrales del AAR en diversas disciplinas, sobre todo, en los dominios cognitivo y metacognitivo. Estudios empíricos recientes revelan relaciones significativas entre la autoeficacia en el uso de IA y el rendimiento académico, siendo la aplicación práctica de estas herramientas el predictor más fuerte de la autoeficacia (Bećirović et al., 2025). La investigación de Liang et al. (2023) halló que la interacción estudiante-IA mejoró significativamente el rendimiento a través de una mayor implicación cognitiva ($\beta = 0.046$), mientras que los sistemas de retroalimentación basados en IA mejoran la conciencia metacognitiva del estudiante a través del seguimiento del progreso y la visualización del rendimiento en tiempo real (Gkintoni et al., 2025; Molenaar et al., 2023). La implementación de la IA en el AAR revela complejas dinámicas de apoyo y obstaculización. Aunque herramientas como ChatGPT han mejorado las habilidades de investigación y han apoyado el aprendizaje autodirigido a través de

la autonomía (Li et al., 2025), otros estudios revelan preocupantes correlaciones negativas ($r = -0.39$) entre la dependencia excesiva de la IA y el pensamiento crítico, especialmente en aprendices de 17-25 años (Gerlich, 2025). Este efecto de “descarga cognitiva” sugiere que el uso frecuente de la IA puede impedir el desarrollo de habilidades metacognitivas independientes, lo que hace que los aprendices dependan de los sistemas en lugar de desarrollar capacidades de autorregulación autónomas (Gerlich, 2025). La investigación de Glick et al. (2024) demostró que, si bien los módulos de formación con IA aumentaron el uso de herramientas de planificación y la fijación de metas, el equilibrio entre el apoyo de la IA y la agencia del aprendiz sigue siendo una preocupación crítica.

Quizá el hallazgo más significativo en la literatura es que las aplicaciones de IA tienen dificultades para apoyar la regulación motivacional, un componente crítico del AAR (Seo et al., 2021). Múltiples revisiones sistemáticas confirman que, si bien los aprendices perciben la IA como útil para la regulación metacognitiva, cognitiva y conductual, prefieren recibir apoyo humano para satisfacer sus necesidades motivacionales (Jin et al., 2023; Lan y Zhou, 2025). Desde la Teoría de la Autodeterminación, aunque la IA puede apoyar la competencia con retroalimentación personalizada y la autonomía mediante el control de las rutas de aprendizaje (Chiu, 2024; Li et al., 2025), la vinculación sigue siendo su aspecto más débil.

La evidencia empírica reciente (2020-2025) apunta cada vez más hacia los sistemas híbridos de regulación humano-IA como el enfoque óptimo para apoyar el AAR. La investigación distingue entre la autorregulación centrada en el humano, donde la IA es un facilitador que provee herramientas y datos para el control del aprendiz, y la autorregulación centrada en la IA, donde esta genera ciclos de AAR a partir de datos (Lan y Zhou, 2025). Las implementaciones más efectivas parecen implicar una interacción en la que la IA se integra en los ciclos de AAR centrados en el humano a través de las dimensiones conductual, emocional y cognitiva, manteniendo la agencia humana (Wu y Chiu, 2025).

En general, aunque la IA tiene un potencial transformador para el AAR, la evidencia sugiere que una implementación eficaz requiere un enfoque equilibrado que aproveche sus capacidades tecnológicas preservando los elementos humanos fundamentales del AAR. El desarrollo de nuevas herramientas de medición como la Escala AI-SRL es crucial para obtener una comprensión más profunda de la eficacia e impacto de las herramientas de IA, abordar la compleja interacción entre la mejora cognitiva y los posibles déficits motivacionales, así como guiar futuras integraciones que mantengan la agencia del aprendiz mientras se aprovechan las fortalezas de la IA en personalización y la gestión de la carga cognitiva.

MÉTODO

En este estudio, el objetivo fue desarrollar una escala para medir los niveles de AAR de los participantes en el contexto del uso de la IA. A pesar de las directrices

establecidas, los investigadores a menudo enfrentan limitaciones en el desarrollo de escalas, que incluyen las características de la muestra, cuestiones metodológicas y desafíos psicométricos. Para optimizar futuras prácticas, es crucial considerar estas posibles limitaciones y tomar decisiones metodológicas informadas durante todo el proceso (Morgado et al., 2017). Por tanto, y reconociendo estas limitaciones, este estudio emplea una metodología robusta para el desarrollo de una escala fiable y válida.

Participantes

Se utilizaron dos muestras distintas para el análisis factorial exploratorio (AFE) y el análisis factorial confirmatorio (AFC). La muestra del AFE consistió en 504 participantes, mientras que la del AFC incluyó 424. Los participantes fueron reclutados en línea mediante un muestreo por conveniencia e intencional (Creswell, 2014).

El muestreo por conveniencia se usó para facilitar el acceso a los participantes, procedentes de universidades con las que teníamos algún tipo de conexión. Nuestro muestreo fue intencional para seleccionar participantes capaces de comprender fácilmente los ítems de la escala. La mayoría era de departamentos de enseñanza de lenguas extranjeras o de literatura. En el caso de participantes de otros departamentos, se verificó que pudieran entender los ítems sin dificultad. Se priorizó la recolección de datos principalmente de estos departamentos para garantizar la claridad en la comprensión de los ítems, los cuales estaban en inglés. Esto podría haber afectado en cierta medida la generalizabilidad de las respuestas y la interpretación de los factores en algunos constructos. Futuros estudios para validar la escala podrían recolectar datos de poblaciones académicas más diversas para mejorar la generalizabilidad y los estadísticos. Los datos demográficos de nuestros participantes se resumen en la siguiente tabla.

Tabla 1
Datos demográficos

Género	n	Edad	n	Año/Curso	n	Departamento/ Carrera	n
Hombre	334	18 o menos	195	Curso preparatorio de inglés	212	Enseñanza de la Lengua Inglesa	288
Mujer	594	19	202	1.º año	184	Lengua y Literatura Inglesa	262
Total	928	20	228	2.º año	260	Grado en Farmacia	149

Género	n	Edad	n	Año/Curso	n	Departamento/ Carrera	n
		21	193	3. ^{er} año	180	Relaciones Internacionales	31
		22	195	4. ^o año o superior	92	Psicología	28
		23 +	202	Total	928	Derecho	25
		Total	928			Turismo y Orientación	23
						Ingeniería	24
						Otros	98
						Total	928

Como indica la Tabla 1, se contó con un total de 928 participantes para los dos primeros pasos (AFE y AFC) del estudio. De ellos, 594 (64.0 %) eran mujeres y 334 (36.0 %) hombres. Según el departamento académico, 288 participantes (31.0 %) eran de Enseñanza de la Lengua Inglesa, 262 (28.2 %) de Lengua y Literatura Inglesa, y 149 (16.1 %) de Grado en Farmacia. La distribución por curso incluyó a 212 participantes (22.8 %) de cursos preparatorios de inglés, 184 (19.8 %) de primer año, 260 (28.0 %) de segundo, 180 (19.4 %) de tercero y 92 (9.9 %) de cuarto o superior. También se recolectó información adicional sobre el uso de la tecnología y la IA en el aprendizaje por parte de los participantes. La Tabla 2 resume esta información.

Tabla 2

Uso de tecnología e IA en el aprendizaje por parte de los participantes

Nivel de uso de tecnología en el aprendizaje	n	Experiencia con IA	n	Frecuencia de uso de la IA para aprender	n
Muy bajo	8	1 año	543	Cada día	104
Bajo	46	2 años	239	Varias veces por semana	212
Medio	395	3 años	81	Una vez por semana	141
Alto	338	4 años	29	Varias veces al mes	180
Muy alto	141	5 años	13	Una vez al mes	189
Total	928	6+ años	23	Rara vez / si es necesario	102
		Total	928	Total	928

Como se observa en la Tabla 2, los participantes indicaron su nivel de uso de la tecnología en el aprendizaje; la gran mayoría reportó niveles medios (395, 42.5 %), altos (338, 36.4 %) y muy altos (141, 14 %). Aquellos con un nivel bajo o muy bajo representaron el 5.8 % de los participantes. También proporcionaron información sobre su experiencia con la IA: 543 de ellos (58.5 %) tenían un año de experiencia, 239 tenían 2 años (25.7 %), 81 tenían 3 años (8.7 %), y el resto tenía 4, 5, 6 o más años de experiencia. Finalmente, recabamos datos sobre la frecuencia de uso de la IA para aprender. El número de participantes que la usaba cada día fue de 104 (11.2 %), 212 (aproximadamente el 22.8 %) la usaba varias veces por semana, 141 (15.2 %) una vez por semana, 180 (19.3 %) varias veces al mes, y 189 (20.3 %) una vez al mes. Además, 102 participantes (aproximadamente el 11 %) reportaron usar la IA para aprender cuando lo necesitaban.

Generación de ítems

El desarrollo de escalas constituye un proceso fundamental en la investigación social y conductual, el cual suele comprender tres fases principales: la generación de ítems, la construcción de la escala y la evaluación de la misma (Boateng et al., 2018). La fase de generación de ítems se inicia con la definición precisa del constructo que se desea medir y la elaboración de un banco preliminar de ítems (Boateng et al., 2018).

En este estudio, el banco inicial de ítems se elaboró a través de una revisión exhaustiva de la literatura sobre AAR e IA. La validez de contenido suele evaluarse mediante una revisión realizada por expertos (Teeluckdharry et al., 2021). Los ítems de nuestra escala se diseñaron para capturar las dimensiones clave del AAR en el contexto de la IA, asegurando su coherencia con los marcos teóricos existentes. El banco inicial de ítems se sometió a un riguroso proceso de validación de contenido por un panel de siete expertos. La selección de los expertos se realizó con base en criterios específicos: (1) poseer un doctorado en Tecnología Educativa, Psicología Educativa o un campo afín; (2) un mínimo de cinco años de experiencia en investigación; y (3) al menos dos publicaciones revisadas por pares sobre AAR o la aplicación de la IA en la educación. Esta estrategia de muestreo intencional aseguró un alto nivel de experiencia tanto teórica como práctica. El panel proporcionó retroalimentación sobre la claridad, la relevancia y la representatividad de los ítems. La validez de contenido se evaluó con el Índice de Validez de Contenido (IVC). Se conservaron aquellos ítems que alcanzaron un $IVC \geq .80$ por parte de al menos 6 de los 7 expertos. El promedio del IVC a nivel de escala ($IVC-E/Prom$) fue de .92, superando el umbral recomendado de .90.

Los ítems fueron revisados o eliminados basándose en esta retroalimentación, lo que resultó en una escala preliminar para la prueba piloto. La literatura señala varios tamaños de muestra para estudios piloto. Johanson y Brooks (2009) destacan que Isaac y Michael (1995) y Hill (1998) sugirieron entre 10 y 30 participantes, mientras

que Mooney y Duval (1993) propusieron entre 30 y 50, y Hertzog (2008) entre 25 y 40 para el desarrollo de escalas. En nuestro caso, la prueba piloto se llevó a cabo con una muestra de 33 participantes que representaban características similares a las de los grupos objetivo del estudio. El objetivo de la prueba piloto fue identificar problemas relacionados con la redacción de los ítems, las opciones de respuesta o la extensión global de la escala. A partir de los resultados, se realizaron revisiones menores para mejorar la claridad y la fiabilidad de los ítems.

Análisis Factorial Exploratorio (AFE)

La fase de construcción de la escala incluye la prueba preliminar, la administración del cuestionario, la reducción de ítems y la extracción de factores (Boateng et al., 2018). Tras la aplicación de la escala al primer grupo de participantes, se procedió al siguiente paso. La fase de evaluación de la escala implica la prueba de la dimensionalidad, la fiabilidad y la validez (Teeluckdharry et al., 2021). Se evaluó la adecuación de los datos de la primera muestra ($n = 504$) para el análisis factorial. La medida de adecuación muestral de Kaiser-Meyer-Olkin (KMO) fue de .96, lo que indica una adecuación excelente (Kaiser y Rice, 1974). Además, la Prueba de Esfericidad de Bartlett fue significativa, $\chi^2 = 9846.80, p < .001$, lo que confirmó que la matriz de correlaciones no era una matriz de identidad y que, por tanto, resultaba apropiada para el análisis. Tras esto, se llevó a cabo un Análisis Factorial Exploratorio (AFE) con SPSS 22 para examinar y confirmar la estructura factorial de la Escala AI-SRL. El AFE se utilizó para revelar la estructura factorial subyacente e identificar grupos de ítems relacionados, estableciendo así la validez de constructo de la escala al reducir los datos a un número menor de factores significativos (Field, 2018).

Se empleó el Análisis de Componentes Principales (ACP) como método de extracción y la rotación promax —un método de rotación oblicua— que permite correlaciones entre factores (Kline, 2014; Muthén y Muthén, 2017). Aunque se probaron la Factorización de Ejes Principales (FEP) y la Máxima Verosimilitud (MV), no generaron soluciones factoriales estables o interpretables. Dado nuestro objetivo de explorar las dimensiones subyacentes y reducir los ítems a componentes coherentes, se seleccionó el ACP por su robustez e interpretabilidad en este contexto, en concordancia con recomendaciones previas para el desarrollo inicial de escalas (Fabrigar y Wegener, 2012; Jolliffe, 2014). Para garantizar aún más la validez, se utilizó una muestra distinta para realizar un Análisis Factorial Confirmatorio (AFC) mediante AMOS, que confirmó la estructura derivada del ACP con índices de ajuste del modelo satisfactorios.

El AFE sugirió en la mayoría de los casos una solución de cuatro factores, y en algunos casos cinco, basándose en autovalores superiores a uno. Tras evaluar un gran número de modelos con AFE y AFC, se adoptó la solución de cinco factores por ser la que ofreció el mejor ajuste del modelo. En la etapa de AFE, se mantuvieron los ítems con cargas factoriales de 0.4 o superiores, umbral considerado significativo

(Floyd y Widaman, 1995; Hair et al., 2010). Durante el AFE, se eliminaron tres ítems debido a cargas cruzadas que no superaban el umbral de diferencia de 0.10 (Büyüköztürk, 2018): “No comparto mi información personal al usar herramientas de IA”, “Verifico la originalidad del trabajo generado por IA antes de usarlo” y “Me aseguro de que mis compañeros también usen la IA de forma ética”. En la etapa de AFC, se eliminaron nueve ítems adicionales debido a cargas estandarizadas bajas o marginales, redundancia con ítems de mayor carga o índices de modificación problemáticos entre factores. Estos incluían ítems correspondientes a Competencia en IA (ej., “Puedo aprender rápidamente a usar nuevas herramientas de IA”), de Conciencia del Aprendizaje reflexiva (ej., “Evalúo si las herramientas de IA me dan información precisa”), y de Colaboración Ética (ej., “Discuto cuestiones éticas sobre IA con compañeros”). Las decisiones de eliminación se basaron en criterios estadísticos y conceptuales para asegurar el rigor psicométrico y la coherencia teórica de la escala final. Los ítems eliminados y su justificación se detallan en el Apéndice B. El AFE reveló una estructura de cinco factores que explicó el 62.42 % de la varianza total. Los ítems mostraron cargas elevadas en sus respectivos factores, con valores entre 0.44 y 0.96.

Tabla 3
Autovalores del análisis paralelo

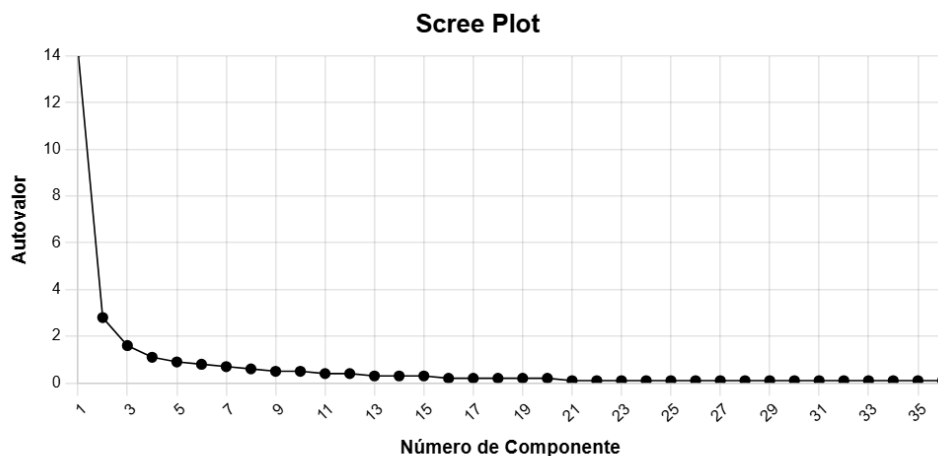
Factor	Autovalores iniciales	% de varianza	% acumulados
Factor 1	14.183	43.99	43.99
Factor 2	2.220	6.95	50.95
Factor 3	1.652	5.25	56.20
Factor 4	1.006	3.13	59.34
Factor 5	.996	3.080	62.42

Como muestra la Tabla 3, los autovalores de los primeros cuatro factores son superiores a uno, y el del quinto es 0.99, justo por debajo de 1.00. Decidimos mantenerlo como factor en la escala final porque estaba respaldado por la literatura previa y los análisis posteriores. Se realizó un análisis paralelo empleando 1000 matrices de datos generadas aleatoriamente con el mismo número de variables (31) y participantes (504). Los primeros tres factores superaron claramente el percentil 95 de los autovalores aleatorios, lo que justificó firmemente su retención. Los factores 4 y 5 tuvieron autovalores ligeramente por debajo del umbral del percentil 95, pero se mantuvieron basándose en: (a) su alineación teórica con los marcos de AAR, (b) cargas factoriales adecuadas (todas > .40), y (c) una mejora en el ajuste del modelo en el AFC posterior. Esta decisión se encuentra en consonancia con las recomendaciones de considerar múltiples criterios más allá de los autovalores para

la retención de factores (Fabrigar y Wegener, 2012). Asimismo, la solución de cinco factores fue respaldada por un gráfico de sedimentación, que ilustra los factores potenciales dentro de la escala (Eisinga et al., 2013).

Figura 1

Gráfico de sedimentación del análisis factorial exploratorio inicial



El gráfico de sedimentación también respaldó una solución de cinco factores, ya que la pendiente del gráfico se aplanó considerablemente después del quinto factor. El Apéndice C presenta las cargas factoriales de cada ítem del modelo inicial, lo que describe la naturaleza subdimensional de la escala. Tras este análisis, se eliminaron los ítems 21, 24 y 29. Los 31 ítems restantes formaron una estructura de cinco factores, que luego fue sometida a un AFC utilizando los datos del segundo grupo de estudio. Este análisis posterior dio como resultado la escala final de 22 ítems.

Análisis Factorial Confirmatorio (AFC)

Una vez confirmado el modelo inicial del AFE, recolectamos nuevos datos basados en este último modelo, los cuales se usaron para el AFC. Sun (2005) indica tres propósitos principales para realizar un AFC: evaluación de la validez de constructo, comparación de patrones de respuesta y comparación de modelos competidores. En nuestro caso, lo usamos para evaluar la validez de constructo y comparar modelos.

El AFC, realizado con una muestra independiente ($n = 424$), se utilizó para probar la estructura de cinco factores identificada en el AFE. El modelo inicial se refinó basándose en las cargas estandarizadas y los índices de modificación. Los ítems con

cargas bajas o aquellos que creaban una covarianza significativa entre factores se eliminaron iterativamente para mejorar el ajuste del modelo, manteniendo al mismo tiempo la coherencia teórica. Este proceso dio como resultado el modelo final de 22 ítems. La Tabla 4 resume los principales índices de ajuste de los análisis de AFC.

Tabla 4
Resultados del AFC

Índices clave	Criterios de Hu y Bentler (1999)	Nuestros resultados
CFI	> .95 excelente, > .90 bueno, > .80 a veces aceptable	0.96
RMSEA	< .05 bueno, < .10 moderado, > .10 no bueno	0.04
CMIN/DF	< 3 bueno, < 5 aceptable	2.82
GFI	> .95 excelente, > .90 bueno	0.94
AGFI	> .80	0.92
SRMR	< .09	0.04
TLI	> .95 excelente, > .90 aceptable	.95

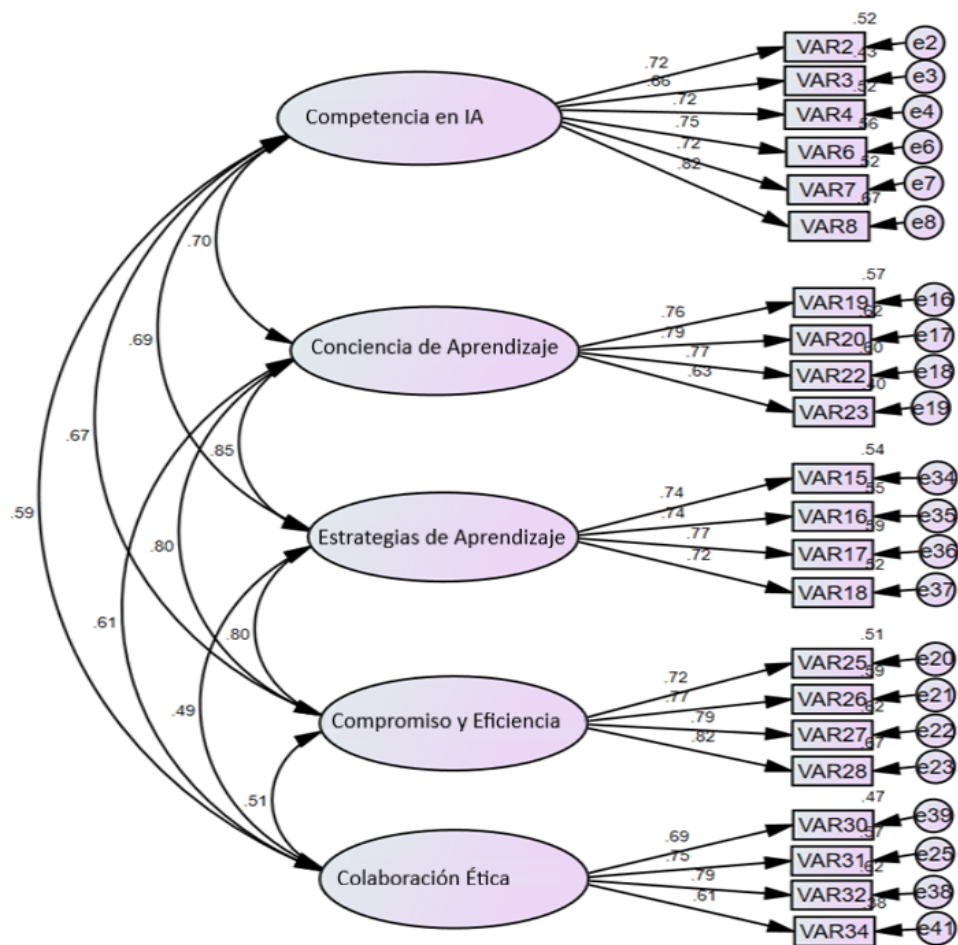
Como indica la Tabla 4, el modelo final demostró un buen ajuste a los datos, evaluado según los criterios de ajuste convencionales (Hu y Bentler, 1999). Los índices de ajuste fueron: χ^2/gl (CMIN/DF) = 2.82, Índice de Ajuste Comparativo (CFI) = .97, Índice de Tucker-Lewis (TLI) = .95, Error Cuadrático Medio de Aproximación (RMSEA) = .04 (IC 90 % [.035, .045]), y Raíz del Residuo Cuadrático Medio Estandarizado (SRMR) = .04. Dado que el CFI y el TLI fueron superiores a .95 y el RMSEA y el SRMR inferiores a .05, el modelo de cinco factores se confirmó como una representación válida de los datos.

Modelo de Análisis Factorial Confirmatorio (AFC)

La siguiente figura ilustra el modelo de AFC y los coeficientes entre variables e ítems.

Figura 2

Análisis factorial confirmatorio de la escala AI-SRL con coeficientes de trayectoria estandarizados



La figura confirma el modelo final, que consta de 5 factores y 22 ítems. Los cinco factores confirmados mediante el AFC representan dimensiones distintas del aprendizaje autorregulado potenciado por IA con claras implicaciones pedagógicas.

La Competencia en IA (Factor 1, 6 ítems) refleja la autoeficacia y la competencia técnica del aprendiz al utilizar herramientas de IA con fines educativos, reflejando su confianza para adaptarse a diversas interfaces, comprender instrucciones generadas por IA e interpretar eficazmente la información que esta produce para su

aprendizaje. La Conciencia del Aprendizaje (Factor 2, 4 ítems) abarca procesos de monitorización metacognitiva donde los aprendices supervisan de forma consciente su comprensión, progreso y eficacia al interactuar con sistemas de IA, lo que incluye la reflexión crítica acerca de su dependencia de estas herramientas. Las Estrategias de Aprendizaje (Factor 3, 4 ítems) implica la selección, adaptación e implementación de estrategias cognitivas apoyadas por IA, incluyendo cómo los aprendices usan la IA para encontrar métodos de estudio efectivos, mejorar la resolución de problemas y fortalecer el pensamiento crítico. La Implicación y Eficiencia (Factor 4, 4 ítems) evalúa la regulación conductual y motivacional del aprendiz para mantener la participación, lograr resultados de aprendizaje permanentes y sostener el interés al usar herramientas de IA adaptadas a su estilo. La Colaboración Ética (Factor 5, 4 ítems) representa una dimensión novedosa que aborda la conciencia del aprendiz sobre el uso responsable de la IA, incluyendo consideraciones sobre responsabilidad personal, privacidad, sesgos potenciales y seguridad de la información en contextos de aprendizaje asistidos por IA.

Estos factores, en conjunto, operacionalizan cómo los aprendices contemporáneos afrontan el complejo panorama de la educación potenciada por IA manteniendo sus capacidades de autorregulación. La versión validada completa de la Escala AI-SRL, compuesta por 22 ítems organizados en torno a estos cinco factores, se presenta en el Apéndice A. En las secciones siguientes se detalla la fiabilidad y validez de la escala.

Análisis de fiabilidad

Realizamos varias pruebas con el propósito de garantizar la fiabilidad de la escala AI-SRL. Estas pruebas se explican a continuación.

Consistencia interna

Se calculó el Alfa de Cronbach para evaluar la consistencia interna de la escala general y de cada factor, utilizado como indicador del grado de homogeneidad de los ítems. Valores de Alfa de Cronbach superiores a .90 se consideran de consistencia excelente; entre .80 y .90 reflejan una buena consistencia; entre .70 y .80 indican una consistencia aceptable; entre .60 y .70 son cuestionables; de .50 a .60 son pobres; y valores por debajo de .50 son inaceptables (George y Mallery, 2003). La escala general mostró una consistencia interna excelente ($\alpha = .94$, IC 95 % [.93, .95]). Los subfactores individuales también mostraron una fiabilidad de buena a excelente: Competencia en IA ($\alpha = .91$), Conciencia del Aprendizaje ($\alpha = .88$), Estrategias de Aprendizaje ($\alpha = .84$), Implicación y Eficiencia ($\alpha = .86$), y Colaboración Ética ($\alpha = .87$). Los estadísticos detallados, incluidas las correlaciones ítem-factor, se presentan en el Apéndice D.

Fiabilidad de dos mitades (split-half)

Se realizó un análisis de fiabilidad de dos mitades en la escala AI-SRL para evaluar adicionalmente la consistencia de los ítems. La Tabla 5 muestra los resultados de este análisis.

Tabla 5
Resultados del análisis de fiabilidad de dos mitades (split-half)

	Número de ítems	Medias Máx – Mín	Medias	Alfa de Cronbach	Correlación entre formas	Coefficiente de dos mitades de Guttman	Coefficiente de Spearman-Brown
Parte 1	11 ^a	3.56 ± 3.16	3.39	0.90	0.80	0.89	0.89
Parte 2	11 ^b	3.59 ± 3.19	3.41	0.89			
Total	22	3.59 ± 3.16	3.40	0.94			

Notas: ^a Los ítems son: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21.
^b Los ítems son: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22.

Como indica la Tabla 5, el coeficiente de Spearman-Brown fue de .89 (IC 95 % [.87, .91]) y el coeficiente de Dos Mitades de Guttman fue de .89 (IC 95 % [.87, .91]), lo que muestra una excelente fiabilidad de dos mitades. La correlación entre formas fue de .80 (IC 95 % [.77, .83]).

Validez

Según Cheung et al. (2024), la validez convergente requiere una Fiabilidad Compuesta (FC) ≥ 0.70, cargas estandarizadas ≥ 0.50 y una Varianza Media Extraída (VME) ≥ 0.50, criterios que nuestro modelo cumple. La Tabla 6 muestra las puntuaciones de las escalas AI-SRL.

Tabla 6
FC, VME, MVC y correlaciones entre factores

	FC	VME	MVC	Colaboración ética	Competencia en IA	Conc. del aprend.	Implic. y efíc.	Estrateg. de aprend.
Colaboración ética	0.80	0.51	0.37	0.71				
Competencia en IA	0.87	0.54	0.49	0.59	0.73			

Conciencia del aprendizaje	0.83	0.55	0.73	0.61	0.70	0.74		
Implicación y eficiencia	0.85	0.59	0.65	0.51	0.67	0.80	0.77	
Estrategias de aprendizaje	0.83	0.55	0.73	0.49	0.68	0.85	0.80	0.74

Nota: Los intervalos de confianza del 95% para las correlaciones están disponibles en los materiales suplementarios. Todas las correlaciones son significativas a $p < .001$. FC = Fiabilidad Compuesta; VME = Varianza Media Extraída; MVC = Varianza Máxima Compartida.

La validez discriminante se probó a través de múltiples criterios: (a) ningún ítem presentó cargas cruzadas entre constructos, (b) la VME de cada constructo superó las correlaciones inter-constructo al cuadrado, y (c) las correlaciones entre factores estuvieron por debajo de los valores recomendados —ninguna superó 0.85— (Cheung et al., 2024; Hair et al., 2010). Aunque se encontraron algunas correlaciones altas entre Conciencia del Aprendizaje, Implicación y Eficiencia y Estrategias de Aprendizaje, estos constructos mantuvieron su distinción teórica dentro del marco AI-SRL. Por tanto, su retención se consideró psicométricamente justificable y conceptualmente significativa.

Los análisis estadísticos se realizaron con IBM SPSS Statistics 22 y AMOS 22. Todos los intervalos de confianza se calcularon al nivel del 95 % usando el método bootstrap acelerado y con corrección de sesgo (BCa) con 5000 remuestreos. El análisis paralelo se empleó con la sintaxis de SPSS de O'Connor (2000), generando 1000 matrices de correlación aleatorias para la comparación de autovalores. Se fijó un nivel de significancia de $\alpha = .05$ para todas las pruebas, aplicando correcciones de Bonferroni para comparaciones múltiples cuando fue apropiado. La escala validada ofrece una medida fiable y válida del AAR en el contexto de la IA, con potenciales aplicaciones tanto en la investigación como en la práctica educativa. Futuras investigaciones podrían ampliar el uso de la escala en diversas poblaciones y entornos.

DISCUSIÓN

Este estudio desarrolló y validó con éxito la Escala AI-SRL, proporcionando un instrumento psicométricamente robusto diseñado específicamente para evaluar las capacidades de AAR en entornos educativos apoyados por IA. Para los análisis preliminares, se elaboró una herramienta de 34 ítems basada en la literatura y la opinión de expertos. El análisis factorial mediante KMO reveló una adecuación excelente (.96), y la Prueba de Esfericidad de Bartlett confirmó la distribución normal multivariante de los datos. Tras el AFE y el ACP para establecer factores

significativos y revelar sus correlaciones, se eliminaron 12 ítems para preservar la coherencia teórica y la integridad psicométrica de la escala. Con las respuestas de 928 participantes, se realizaron los análisis de AFE y AFC. Los resultados de validez de contenido, validez de constructo, consistencia interna y fiabilidad de dos mitades confirmaron que la Escala AI-SRL es psicométricamente sólida para evaluar el AAR con IA en entornos educativos. La versión final de la escala consta de 22 ítems bajo 5 factores: Competencia en IA, Conciencia del Aprendizaje, Estrategias de Aprendizaje, Implicación y Eficiencia, y Colaboración Ética.

Esta estructura de cinco factores muestra que el uso de la IA en entornos educativos profundiza y contribuye al constructo tradicional de la autorregulación al detallar la competencia, la conciencia del aprendizaje, las estrategias, la implicación y eficiencia, y la colaboración ética. Esta nueva estructura se alinea con la literatura reciente al destacar cómo la integración de la IA en entornos educativos avanza las teorías convencionales del AAR propuestas por Zimmerman (1989) y Pintrich (1999). Además, el factor Competencia en IA de este estudio contribuye a la literatura. Por ejemplo, Bećirović et al. (2025) hallaron una correlación significativa entre el rendimiento académico y la autoeficacia en IA. Este factor también puede vincularse con las estrategias de regulación cognitiva de Pintrich (1999), ya que el estudio actual destacó que las herramientas de IA pueden apoyar las estrategias de elaboración. Por otra parte, el factor Conciencia del Aprendizaje evoca el proceso de autoobservación de la fase de ejecución de Zimmerman (Panadero, 2017; Zimmerman, 1989), señalando el potencial de la IA para mejorar la monitorización metacognitiva. Gkintoni et al. (2025) revelaron que la retroalimentación de las herramientas de IA, al permitir evaluar el rendimiento simultáneamente, promueve la conciencia metacognitiva de los aprendices, ayudándoles a seguir sus procesos de aprendizaje de forma más consciente. Adicionalmente, el factor Colaboración Ética contribuye a la literatura previa. Cuestiones como la privacidad de datos, la vigilancia, el sesgo algorítmico y la integridad académica fueron señaladas como preocupaciones éticas que pueden afectar directamente al AAR en contextos educativos con IA (Gkintoni et al., 2025; Walter, 2024). Así, este factor subraya las responsabilidades éticas de los estudiantes al usar herramientas de IA.

El factor Colaboración Ética representa una extensión paradigmática más allá de los marcos clásicos del AAR, introduciendo dimensiones que ni Zimmerman ni Pintrich anticiparon. Este factor operacionaliza lo que Pintrich (1999) denominó “regulación del contexto”, pero lo amplía para abarcar la transparencia algorítmica, la soberanía de los datos y la ética colaborativa en las interacciones humano-IA. La aparición de este factor, estadística y conceptualmente distinta, sugiere que las consideraciones éticas en el aprendizaje mediado por IA no son un mero añadido, sino un dominio regulatorio fundamental que requiere atención teórica. Este hallazgo coincide con las llamadas recientes a la “alfabetización algorítmica” como una competencia central en la educación digital (Walter, 2024). En cuanto al factor de Implicación y Eficiencia, las tecnologías de IA pueden presentar un

potencial inestimable para apoyar las dimensiones cognitivas y metacognitivas del AAR, aunque podrían resultar limitadas en la regulación motivacional (Lan y Zhou, 2025; Seo et al., 2021; Wang y Lin, 2023). Considerando las teorías de Zimmerman (1989) y Pintrich (1999), la regulación motivacional está estrechamente asociada con la confianza, la afinidad y el apoyo afectivo individualizado que proporcionan las herramientas educativas basadas en IA. Así, este factor emerge como un área crítica para nuevas investigaciones. Por último, el factor Estrategias de Aprendizaje se vincula tanto con la fase de control cognitivo de Pintrich (1999) como con la fase de ejecución de Zimmerman (1989), y muestra cómo los aprendices seleccionan, adaptan e implementan estrategias cognitivas apoyadas por IA. Este factor refleja los hallazgos de la literatura de que las herramientas de IA apoyan eficazmente la adaptación y selección de estrategias, siendo utilizadas para encontrar métodos de estudio efectivos y mejorar la resolución de problemas (Liu et al., 2024; Seo et al., 2021). Su incorporación reconoce que los aprendices contemporáneos integran activamente las recomendaciones de la IA en su repertorio estratégico.

CONCLUSIÓN Y FUTURAS DIRECCIONES

Los rápidos avances en las herramientas de IA han generado la necesidad de que los entornos educativos las adapten adecuadamente a sus procesos. Junto con otros constructos psicológicos, se ha hecho evidente la necesidad de contar con un instrumento de medición del AAR en el uso de la IA en contextos educativos. Con este propósito, el presente estudio tuvo como objetivo realizar el estudio de validez y fiabilidad de la escala AI-SRL para responder a esta necesidad en la literatura. El estudio propuso la escala AI-SRL, la cual puede usarse como una herramienta psicométricamente sólida para evaluar las habilidades de AAR de los estudiantes en el uso de la IA en entornos educativos. Esta escala puede hacer una contribución original a la literatura y tiene potencial como un instrumento eficaz para su implementación en contextos educativos mediados por IA.

La Escala AI-SRL representa una contribución pionera a la medición educativa al ser uno de los primeros instrumentos validados diseñados específicamente para ver la naturaleza multidimensional del AAR en entornos potenciados por IA. A diferencia de las medidas genéricas de AAR o las escalas de aceptación de la tecnología, este instrumento une de forma única los fundamentos teóricos de la psicología educativa con las realidades contemporáneas de la integración de la IA, y proporciona una herramienta psicométricamente robusta para comprender cómo los aprendices manejan las complejidades de la educación mediada por IA. El potencial de la escala va más allá de la evaluación individual, pues también puede orientar políticas institucionales e innovación pedagógica. Las instituciones educativas que implementan sistemas de aprendizaje con IA pueden utilizar la escala para evaluar la preparación de los estudiantes, identificar necesidades de apoyo y valorar la eficacia de sus iniciativas. La estructura de cinco factores proporciona ideas prácticas para los

diseñadores curriculares, indicando que los programas de alfabetización en IA deben abordar no solo la competencia técnica, sino también la conciencia metacognitiva, la adaptación estratégica, la implicación motivacional y el razonamiento ético.

Para futuras investigaciones, se pueden realizar estudios de validación transcultural de este instrumento con el fin de analizar su aplicabilidad en diferentes entornos culturales. Además, dado que los datos del estudio actual se recolectaron en contextos de educación superior, se podrían recolectar datos futuros en contextos de educación secundaria o profesionales para observar si se obtienen resultados diferentes. Los estudios futuros de validación deberían priorizar investigaciones transculturales, particularmente en contextos educativos no occidentales, donde las orientaciones de aprendizaje colectivistas podrían interactuar de manera distinta con las herramientas de IA. También la aplicabilidad de la escala en entornos de educación primaria y secundaria (K-12) merece ser investigada, ya que los aprendices más jóvenes pueden presentar patrones distintos de autorregulación. Los contextos de formación profesional, incluidos los entornos de aprendizaje corporativo, representan otra frontera donde la escala podría iluminar cómo los profesionales integran la IA en sus prácticas de aprendizaje permanente. Por último, los estudios longitudinales que sigan los perfiles de AI-SRL a lo largo del tiempo proporcionarían información valiosa sobre las trayectorias de desarrollo y el impacto de las intervenciones educativas.

IMPLICACIONES PRÁCTICAS Y LIMITACIONES

La escala AI-SRL puede implementarse como una herramienta eficaz y fiable en entornos educativos, desde la planificación hasta la evaluación. Puede aplicarse para medir los niveles de AAR con IA de los estudiantes; se pueden diseñar u observar analíticas del aprendizaje basadas en sus informes; y puede proporcionar información para la elaboración de diseños instruccionales. Además, la escala puede usarse en la educación presencial, en línea e híbrida.

La escala AI-SRL ofrece aplicaciones concretas para mejorar las prácticas docentes. En la evaluación diagnóstica, los docentes pueden administrarla al inicio del curso para identificar a estudiantes que podrían tener dificultades con los componentes de aprendizaje potenciados por IA, permitiendo intervenciones de apoyo específicas. Para las analíticas del aprendizaje, las puntuaciones pueden integrarse con datos de rastro digital para crear perfiles de aprendizaje completos, revelando patrones entre las capacidades de AAR con IA autoinformadas y los comportamientos de uso reales. Esta integración es valiosa en cursos en línea masivos y abiertos (MOOCs), donde el apoyo personalizado a gran escala es un desafío. En el diseño instruccional, su estructura de cinco factores orienta el desarrollo de estrategias de andamiaje adaptadas: por ejemplo, los estudiantes con baja Competencia en IA podrían beneficiarse de módulos de orientación técnica, mientras que aquellos con una Colaboración Ética débil podrían trabajar con estudios de caso sobre el uso

responsable de la IA. En entornos de aprendizaje híbridos, la escala puede informar sobre cuándo desplegar el apoyo de la IA frente a la instrucción humana. Asimismo, los programas de educación a distancia pueden usarla para evaluar si los estudiantes poseen las habilidades de AAR con IA necesarias para el éxito en entornos de aprendizaje autónomos.

Sin embargo, el presente estudio presenta varias limitaciones. En primer lugar, la mayoría de los participantes fueron seleccionados de departamentos de enseñanza de lenguas extranjeras y de literatura inglesa en universidades de Türkiye, lo que puede limitar la generalizabilidad de los hallazgos a otros contextos de educación superior. En segundo lugar, se utilizaron métodos de muestreo por conveniencia e intencional, lo que, al ser un muestreo no aleatorio, puede dificultar la generalizabilidad a otras poblaciones. En tercer lugar, los ítems de la escala pueden ser interpretados de manera diferente por individuos de distintos contextos lingüísticos. Por último, los sesgos disciplinarios pueden influir negativamente en la implementación del instrumento. Otras limitaciones a considerar: el desarrollo de la escala en el contexto de la educación superior turca, con participantes predominantemente de disciplinas lingüísticas, puede haber limitado la generalizabilidad transcultural y transdisciplinaria. Los ítems en inglés podrían haber introducido sesgos de respuesta entre los hablantes no nativos, afectando potencialmente la invarianza de la medición entre grupos lingüísticos. Además, la rápida evolución de las tecnologías de IA significa que la escala requerirá actualizaciones periódicas para mantener su relevancia.

Aprobación ética

Esta investigación recibió aprobación ética de la Universidad de Ciencias Sociales de Ankara, según el número de decisión 98038.

Uso ético de herramientas de IA

De acuerdo con los principios de transparencia, declaramos el uso de herramientas de IA en la preparación de este manuscrito. Utilizamos Claude Opus principalmente para el refinamiento lingüístico, específicamente para mejorar estructuras gramaticales, la fluidez de las oraciones y asegurar la consistencia en la terminología académica. Esta herramienta se empleó exclusivamente como asistente editorial para el pulido del lenguaje, no para generar contenido de investigación, análisis de datos o contribuciones teóricas. Todo el contenido sustantivo, incluyendo el diseño de la investigación, la recolección de datos, los análisis estadísticos y la interpretación de los resultados, representa el trabajo original de los autores. Mantuvimos el control editorial total, revisando y modificando críticamente todas las sugerencias de la IA. Los autores asumen la plena responsabilidad por la exactitud, originalidad e integridad académica del manuscrito final.

Disponibilidad de los datos

Los datos y materiales utilizados en este estudio están disponibles previa solicitud al autor de correspondencia.

Financiación

Los autores declaran no haber recibido financiación.

REFERENCIAS

- Adams, D., Chuah, K.-M., Devadason, E. y Azzis, M. S. A. (2024). From novice to navigator: Students' academic help-seeking behaviour, readiness, and perceived usefulness of ChatGPT in learning. *Education and Information Technologies*, 29(11), 13617-13634. <https://doi.org/10.1007/s10639-023-12427-8>
- Anggoro, K. J. y Pratiwi, D. I. (2023). Fostering self-assessment in English learning with a generative AI platform: A case of Quizizz AI. *Studies in Self-Access Learning Journal*, 14(4), 489-501. <https://doi.org/10.37237/140406>
- Baker, R. S. (2023). AI and self-regulated learning theory: What could be on the horizon? *Computers in Human Behavior*, 147, 107849. <https://doi.org/10.1016/j.chb.2023.107849>
- Bauer, E., Greiff, S., Graesser, A. C., Scheiter, K. y Sailer, M. (2025). Looking beyond the hype: Understanding the effects of AI on learning. *Educational Psychology Review*, 37(2). <https://doi.org/10.1007/s10648-025-10020-8>
- Bećirović, S., Polz, E. y Tinkel, I. (2025). Exploring students' AI literacy and its effects on their AI output quality, self-efficacy, and academic performance. *Smart Learning Environments*, 12, 29. <https://doi.org/10.1186/s40561-025-00384-3>
- Biju, N., Abdelrasheed, N. S. G., Bakiyeva, K., Prasad, K. D. V. y Jember, B. (2024). Which one? AI-assisted language assessment or paper format: An exploration of the impacts on foreign language anxiety, learning attitudes, motivation, and writing performance. *Language Testing in Asia*, 14, 45. <https://doi.org/10.1186/s40468-024-00322-z>
- Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quinonez, H. R. y Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health*, 6. <https://doi.org/10.3389/fpubh.2018.00149>
- Boguslawski, S., Deer, R. y Dawson, M. G. (2025). Programming education and learner motivation in the age of generative AI: Student and educator perspectives. *Information and Learning Sciences*, 126(1/2), 91-109. <https://doi.org/10.1108/ILS-10-2023-0163>
- Büyüköztürk, Ş. (2018). *Sosyal bilimler için veri analizi el kitabı*. Pegem Akademi Yayıncılık. <https://doi.org/10.14527/9789756802748>
- Chang, D. H., Lin, M. P.-C., Hajian, S. y Wang, Q. Q. (2023). Educational design principles of using AI chatbot that supports self-regulated learning in education: Goal setting, feedback, and personalization. *Sustainability*, 15(17), 12921. <https://doi.org/10.3390/su151712921>

- Cheung, G. W., Cooper-Thomas, H. D., Lau, R. S. y Wang, L. C. (2024). Reporting reliability, convergent and discriminant validity with structural equation modeling: A review and best-practice recommendations. *Asia Pacific Journal of Management*, 41(2), 745-783. <https://doi.org/10.1007/s10490-023-09871-y>
- Chiu, T. K. F. (2024). A classification tool to foster self-regulated learning with generative artificial intelligence by applying self-determination theory: A case of ChatGPT. *Educational Technology Research and Development*, 72(4), 2401-2416. <https://doi.org/10.1007/s11423-024-10366-w>
- Chiu, T. K. F., Moorhouse, B. L., Chai, C. S. y Ismailov, M. (2023). Teacher support and student motivation to learn with artificial intelligence (AI) based chatbot. *Interactive Learning Environments*, 1-17. <https://doi.org/10.1080/10494820.2023.2172044>
- Creswell, J. W. (2014). *Research design: Qualitative, quantitative and mixed methods approaches* (4ª ed.). SAGE.
- Dahri, N. A., Yahaya, N., Al-Rahmi, W. M., Aldraiweesh, A., Alturki, U., Almutairy, S., Shutaleva, A. y Soomro, R. B. (2024). Extended TAM based acceptance of AI-powered ChatGPT for supporting metacognitive self-regulated learning in education: A mixed-methods study. *Heliyon*, 10(8), e29317. <https://doi.org/10.1016/j.heliyon.2024.e29317>
- DeVellis, R. F. (1991). *Scale development: Theory and applications*. SAGE.
- Eisinga, R., te Grotenhuis, M. y Pelzer, B. (2013). The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown? *International Journal of Public Health*, 58(4), 637-642. <https://doi.org/10.1007/s00038-012-0416-3>
- Fabrigar, L. R. y Wegener, D. T. (2012). *Exploratory factor analysis*. Oxford University Press. <https://doi.org/10.1093/acprof:osobl/9780199734177.001.0001>
- Field, A. P. (2018). *Discovering statistics using IBM SPSS statistics* (5ª ed.). SAGE.
- Floyd, F. J. y Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment*, 7(3), 286-299. <https://doi.org/10.1037/1040-3590.7.3.286>
- George, D. y Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference, 11.0 update* (4ª ed.). Allyn & Bacon.
- Gerlich, M. (2025). AI tools in society: Impacts on cognitive offloading and the future of critical thinking. *Societies*, 15(1), 6. <https://doi.org/10.3390/soc15010006>
- Gkintoni, E., Antonopoulou, H., Sortwell, A. y Halkiopoulos, C. (2025). Challenging cognitive load theory: The role of educational neuroscience and artificial intelligence in redefining learning efficacy. *Brain Sciences*, 15(2), 203. <https://doi.org/10.3390/brainsci15020203>
- Glick, D., Miedijensky, S. y Zhang, H. (2024). Examining the effect of AI-powered virtual-human training on STEM majors' self-regulated learning behavior. *Frontiers in Education*, 9, 1465207. <https://doi.org/10.3389/educ.2024.1465207>
- Grüneke, T., Guggenberger, T., Hofmeister, S. y Stoetzer, J. C. (2024). AI-enabled self-regulated learning: A multi-layer taxonomy development. In *Proceedings of the 32nd European Conference on Information Systems (ECIS 2024)*, Paphos, Cyprus. (ISBN 978-1-958200-10-0). https://aisel.aisnet.org/ecis2024/track13_learning_teach/track13_learning_teach/4
- Hair, J. F., Black, W. C., Babin, B. J. y Anderson, R. E. (2010). *Multivariate data analysis* (7ª ed.). Pearson.
- Hertzog, M. A. (2008). Considerations in determining sample size for pilot studies. *Research in Nursing & Health*, 31, 180-191. <https://doi.org/10.1002/nur.20247>

- Hill, R. (1998). What sample size is 'enough' in Internet survey research? *Interpersonal Computing and Technology: An Electronic Journal for the 21st Century*, 6, 1–10.
- Hilpert, J. C., Greene, J. A. y Bernacki, M. (2023). Leveraging complexity frameworks to refine theories of engagement: Advancing self-regulated learning in the age of artificial intelligence. *British Journal of Educational Technology*, 54(5), 1204–1221. <https://doi.org/10.1111/bjet.13340>
- Hu, L. T. y Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Huang, A. Y. Q., Lu, O. H. T. y Yang, S. J. H. (2023). Effects of artificial intelligence-enabled personalized recommendations on learners' learning engagement, motivation, and outcomes in a flipped classroom. *Computers & Education*, 194, 104684. <https://doi.org/10.1016/j.compedu.2022.104684>
- Isaac, S. y Michael, W. B. (1995). *Handbook in research and evaluation: A collection of principles, methods, and strategies useful in the planning, design, and evaluation of studies in education and the behavioral sciences* (3ª ed.). EdITS Publishers.
- Jin, S.-H., Im, K., Yoo, M., Roll, I. y Seo, K. (2023). Supporting students' self-regulated learning in online learning using artificial intelligence applications. *International Journal of Educational Technology in Higher Education*, 20, 37. <https://doi.org/10.1186/s41239-023-00406-5>
- Johanson, G. A. y Brooks, G. P. (2009). Initial scale development: Sample size for pilot studies. *Educational and Psychological Measurement*, 70(3), 394–400. <https://doi.org/10.1177/0013164409355692>
- Jolliffe, I. (2014). Principal component analysis. En N. Balakrishnan, T. Colton, B. Everitt y W. W. Piegorsch (Eds.), *Wiley StatsRef: Statistics Reference Online*. John Wiley & Sons. <https://doi.org/10.1002/9781118445112.stat06472>
- Kaiser, H. F. y Rice, J. (1974). Little jiffy, mark IV. *Educational and Psychological Measurement*, 34(1), 111–117. <https://doi.org/10.1177/001316447403400115>
- Kline, P. (2014). *An easy guide to factor analysis*. Routledge. <https://doi.org/10.4324/9781315788135>
- Kong, S. C. y Yang, Y. (2024). A human-centered learning and teaching framework using generative artificial intelligence for self-regulated learning development through domain knowledge learning in K-12 settings. *IEEE Transactions on Learning Technologies*, 17, 1562–1573. <https://doi.org/10.1109/TLT.2024.3392830>
- Lai, J. W. (2024). Adapting self-regulated learning in an age of generative artificial intelligence chatbots. *Future Internet*, 16(6), 218. <https://doi.org/10.3390/fi16060218>
- Lan, M. y Zhou, X. (2025). A qualitative systematic review on AI empowered self-regulated learning in higher education. *npj Science of Learning*, 10(1). <https://doi.org/10.1038/s41539-025-00319-0>
- Li, Y., Sadiq, G., Qambar, G. y Zheng, P. (2025). The impact of students' use of ChatGPT on their research skills: The mediating effects of autonomous motivation, engagement, and self-directed learning. *Education and Information Technologies*, 30(4), 4185–4216. <https://doi.org/10.1007/s10639-024-12981-9>
- Liang, J., Wang, L., Luo, J., Yan, Y. y Fan, C. (2023). The relationship between student interaction with generative artificial intelligence and learning achievement: Serial mediating roles of self-efficacy and cognitive engagement. *Frontiers in*

- Psychology*, 14, 1285392. <https://doi.org/10.3389/fpsyg.2023.1285392>
- Liu, Z. M., Hwang, G. J., Chen, C. Q., Chen, X. D. y Ye, X. D. (2024). Integrating large language models into EFL writing instruction: Effects on performance, self-regulated learning strategies, and motivation. *Computer Assisted Language Learning*, 1-25. <https://doi.org/10.1080/09588221.2024.2389923>
- Mohamed, A. M., Shaaban, T. S., Bakry, S. H., Guillén-Gámez, F. D. y Strzelecki, A. (2025). Empowering the faculty of education students: Applying AI's potential for motivating and enhancing learning. *Innovative Higher Education*, 50(2), 587-609. <https://doi.org/10.1007/s10755-024-09747-z>
- Molenaar, I., Mooij, S. de, Azevedo, R., Bannert, M., Järvelä, S. y Gašević, D. (2023). Measuring self-regulated learning and the role of AI: Five years of research using multimodal multichannel data. *Computers in Human Behavior*, 139, 107540. <https://doi.org/10.1016/j.chb.2022.107540>
- Mooney, C. Z. y Duval, R. D. (1993). *Bootstrapping: A nonparametric approach to statistical inference*. SAGE. <https://doi.org/10.4135/9781412983532>
- Morgado, F. F. R., Meireles, J. F. F., Neves, C. M., Amaral, A. C. S. y Ferreira, M. E. C. (2017). Scale development: Ten main limitations and recommendations to improve future research practices. *Psicologia: Reflexão e Crítica*, 30(1). <https://doi.org/10.1186/s41155-016-0057-1>
- Muthén, L. K. y Muthén, B. O. (2017). *Mplus: Statistical analysis with latent variables: User's guide* (Version 8). Muthén & Muthén.
- Ng, D. T. K., Tan, C. W. y Leung, J. K. L. (2024). Empowering student self-regulated learning and science education through ChatGPT: A pioneering pilot study. *British Journal of Educational Technology*, 55(4), 1328-1353. <https://doi.org/10.1111/bjet.13454>
- O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, & Computers*, 32(3), 396-402. <https://doi.org/10.3758/BF03200807>
- Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology*, 8, 422. <https://doi.org/10.3389/fpsyg.2017.00422>
- Pintrich, P. R. (1999). The role of motivation in promoting and sustaining self-regulated learning. *International Journal of Educational Research*, 31(6), 459-470. [https://doi.org/10.1016/S0883-0355\(99\)00015-4](https://doi.org/10.1016/S0883-0355(99)00015-4)
- Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. En M. Boekaerts, P. R. Pintrich y M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 451-502). Academic Press. <https://doi.org/10.1016/B978-012109890-2/50043-3>
- Schunk, D. H. y Zimmerman, B. J. (2012). Self-regulation and learning. En I. B. Weiner, G. E. Miller y W. M. Reynolds (Eds.), *Handbook of psychology: Vol. 7. Educational psychology* (2ª ed., pp. 59-78). John Wiley & Sons. <https://doi.org/10.1002/9781118133880.hop207003>
- Seo, K., Tang, J., Roll, I., Fels, S. y Yoon, D. (2021). The impact of artificial intelligence on learner-instructor interaction in online learning. *International Journal of Educational Technology in Higher Education*, 18(1), 54. <https://doi.org/10.1186/s41239-021-00292-9>
- Sun, J. (2005). Assessing goodness of fit in confirmatory factor analysis. *Measurement and Evaluation in Counseling and Development*, 37(4), 240-256. <https://doi.org/10.1080/07481756.2005.11909764>

- Teeluckdharry, N., Teeroovengadam, V. y Seebaluck, A. (2021). Scale development in marketing research. En R. Nunkoo, V. Teeroovengadam y C. M. Ringle (Eds.), *Handbook of research methods for marketing management* (pp. 17-50). Edward Elgar. <https://doi.org/10.4337/9781788976954.00007>
- Tinajero, C., Mayo, M. E., Villar, E. y Martínez-López, Z. (2024). Classic and modern models of self-regulated learning: Integrative and componential analysis. *Frontiers in Psychology*, 15, 1307574. <https://doi.org/10.3389/fpsyg.2024.1307574>
- Walter, Y. (2024). Embracing the future of artificial intelligence in the classroom: The relevance of AI literacy, prompt engineering, and critical thinking in modern education. *International Journal of Educational Technology in Higher Education*, 21(1), 15. <https://doi.org/10.1186/s41239-024-00448-3>
- Wang, C. Y. y Lin, J. J. H. (2023). Utilizing artificial intelligence to support analyzing self-regulated learning: A preliminary mixed-methods evaluation from a human-centered perspective. *Computers in Human Behavior*, 144, 107721. <https://doi.org/10.1016/j.chb.2023.107721>
- Wu, X. Y. y Chiu, T. K. F. (2025). Integrating learner characteristics and generative AI affordances to enhance self-regulated learning: A configurational analysis. *Journal of New Approaches in Educational Research*, 14(1). <https://doi.org/10.1007/s44322-025-00028-x>
- Xia, Q., Chiu, T. K. F., Chai, C. S. y Xie, K. (2023). The mediating effects of needs satisfaction on the relationships between prior knowledge and self-regulated learning through artificial intelligence chatbot. *British Journal of Educational Technology*, 54(4), 967–986. <https://doi.org/10.1111/bjet.13305>
- Yot-Domínguez, C. y Marcelo, C. (2017). University students' self-regulated learning using digital technologies. *International Journal of Educational Technology in Higher Education*, 14, 38. <https://doi.org/10.1186/s41239-017-0076-8>
- Zimmerman, B. J. (1989). Models of self-regulated learning and academic achievement. En B. J. Zimmerman y D. H. Schunk (Eds.), *Self-regulated learning and academic achievement: Theory, research, and practice* (pp. 1-25). Springer. https://doi.org/10.1007/978-1-4612-3618-4_1

APÉNDICES

Apéndice A: Ítems de la Escala de Aprendizaje Autorregulado Potenciado por IA (Escala AI-SRL)

Factor/ Ítem	Descripción	Escala de respuesta 1 = Totalmente en desacuerdo 5 = Totalmente de acuerdo
Competencia en IA		1 - 2 - 3 - 4 - 5
1.	Uso fácilmente herramientas de IA con diferentes interfaces.	1 - 2 - 3 - 4 - 5
2.	El uso activo de herramientas de IA mejora mi capacidad de aprendizaje.	1 - 2 - 3 - 4 - 5
3.	Puedo utilizar eficientemente las herramientas de IA para encontrar información relevante para mis estudios.	1 - 2 - 3 - 4 - 5
4.	Confío en mi capacidad para aprender a través de nuevas herramientas de IA.	1 - 2 - 3 - 4 - 5
5.	Las instrucciones de las herramientas de IA son fáciles de entender.	1 - 2 - 3 - 4 - 5
6.	Interpreto fácilmente la información generada por IA en mi proceso de aprendizaje.	1 - 2 - 3 - 4 - 5
Conciencia del aprendizaje		1 - 2 - 3 - 4 - 5
7.	La retroalimentación de las herramientas de IA me ayuda a comprender mis fortalezas de aprendizaje.	1 - 2 - 3 - 4 - 5
8.	La retroalimentación proporcionada por la IA me ayuda a identificar áreas de mejora.	1 - 2 - 3 - 4 - 5
9.	Evalúo la eficacia de las herramientas de IA para ayudarme a alcanzar mis objetivos de aprendizaje.	1 - 2 - 3 - 4 - 5
10.	Reflexiono críticamente sobre mi dependencia de las herramientas de IA para aprender.	1 - 2 - 3 - 4 - 5
Estrategias de aprendizaje		1 - 2 - 3 - 4 - 5
11.	Adapto mis métodos de estudio según las recomendaciones de las herramientas de IA.	1 - 2 - 3 - 4 - 5
12.	Uso herramientas de IA para encontrar las formas más eficaces de aprender temas nuevos.	1 - 2 - 3 - 4 - 5
13.	Utilizo la información generada por IA para mejorar mis habilidades de resolución de problemas.	1 - 2 - 3 - 4 - 5

14.	Al usar herramientas de IA para analizar diferentes perspectivas, mejoro mis habilidades de pensamiento crítico.	1 - 2 - 3 - 4 - 5
Implicación y eficiencia		1 - 2 - 3 - 4 - 5
15.	Mi participación en las actividades de aprendizaje aumenta con el uso de la IA.	1 - 2 - 3 - 4 - 5
16.	Las herramientas de IA me ayudan a lograr un aprendizaje más permanente.	1 - 2 - 3 - 4 - 5
17.	Creo que aprendo mejor cuando uso herramientas de IA.	1 - 2 - 3 - 4 - 5
18.	Creo que las herramientas de IA que se adaptan a mi estilo de aprendizaje aumentan mi implicación en el aprendizaje.	1 - 2 - 3 - 4 - 5
Colaboración ética		1 - 2 - 3 - 4 - 5
19.	Sé que debo asumir toda la responsabilidad cuando uso herramientas de IA para aprender.	1 - 2 - 3 - 4 - 5
20.	Soy consciente de las preocupaciones de privacidad en las herramientas de IA.	1 - 2 - 3 - 4 - 5
21.	Soy consciente de los posibles sesgos en las herramientas de IA.	1 - 2 - 3 - 4 - 5
22.	Soy consciente de la seguridad de la información que proporcione a las herramientas de IA.	1 - 2 - 3 - 4 - 5

Apéndice B: Ítems eliminados del modelo final con justificación empírica y conceptual

Fase	Ítem (redacción completa)	Factor original	Carga	Motivo de eliminación
AFE	Uso herramientas de IA para comprobar la eficacia de mi aprendizaje.	Probablemente Conciencia del Aprendizaje	.463/.439	Carga cruzada con diferencia < 0.10; con cargas de .463 y .439 en dos factores.
AFE	Reflexionar profundamente sobre las razones para usar herramientas de IA me ayuda a aplicarlas mejor en mis estudios.	Probablemente Estrategias de Aprendizaje	.285/.463	Carga cruzada en dos ítems; similar a otros ítems de integridad académica, también disminuyó la VME en el AFC.
AFE	Sé cómo usar éticamente la información generada por IA al aprender con IA.	Probablemente Colaboración Ética	.329/- .260/.491	Carga cruzada con diferencia < 0.10; con cargas de .329, -.260 y .491 en tres factores.

Fase	Ítem (redacción completa)	Factor original	Carga	Motivo de eliminación
AFC	Me resulta fácil adaptarme a las características cambiantes de las herramientas de IA.	Competencia en IA	.66	Disminuyó la VME; concepto cubierto por ítems más fuertes de Competencia en IA.
AFC	Soy capaz de usar las funciones específicas de las herramientas de IA (como análisis de datos o recomendación de contenido) para aprender.	Competencia en IA	.64	Disminuyó la validez discriminante; probablemente vago.
AFC	Las herramientas de IA me ayudan a priorizar mis objetivos de aprendizaje.	Conciencia del Aprendizaje	.64	Carga baja, disminuyó la VME y la validez discriminante.
AFC	Las herramientas de IA me ayudan a acceder a una amplia gama de recursos de aprendizaje.	Conciencia del Aprendizaje	.62	Disminuyó la VME; probablemente vago.
AFC	Las herramientas de IA me hacen más autónomo (independiente) en mi proceso de aprendizaje.	Conciencia del Aprendizaje	.59	Carga baja; representación débil del constructo.
AFC	Las herramientas de IA me ayudan a gestionar mejor mi tiempo de aprendizaje.	Conciencia del Aprendizaje	.63	Carga baja; disminuyó la VME.
AFC	Las herramientas de IA me ayudan a desarrollar estrategias para las tareas de aprendizaje.	Conciencia del Aprendizaje	.65	Carga baja; concepto cubierto por otros ítems de autosupervisión.
AFC	Las herramientas de IA me ayudan a entender temas complejos más fácilmente.	Conciencia del Aprendizaje	.64	Disminuyó la validez discriminante; mejor representado por otros ítems de supervisión del aprendizaje.
AFC	Soy escéptico/a sobre la capacidad de las herramientas de IA para proporcionar información precisa.	Colaboración Ética	.65	Carga marginal; probablemente vago.

Apéndice C: Cargas factoriales del modelo de AFE inicial

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Ítem 1	.860				
Ítem 2	.853				
Ítem 3	.555				
Ítem 4	.698				
Ítem 5	.693				
Ítem 6	.690				
Ítem 7	.672				
Ítem 8	.628				
Ítem 9		.665			
Ítem 10		.464			
Ítem 11		.676			
Ítem 12		.776			
Ítem 13		.489			
Ítem 14		.449			
Ítem 15				.725	
Ítem 16				.751	
Ítem 17				.835	
Ítem 18				.589	
Ítem 19		.684			
Ítem 20		.571			
Ítem 22		.498			
Ítem 23		.669			
Ítem 25					.636
Ítem 26					.916
Ítem 27					.861
Ítem 28					.669
Ítem 30			.625		
Ítem 31			.867		
Ítem 32			.789		
Ítem 33			.729		
Ítem 34			.850		

Nota: Todas las cargas son significativas a $p < .001$. Los errores estándar oscilaron entre .028 y .045.

Apéndice D: Alfa de Cronbach para cada factor y correlaciones ítem-factor

Factores/Descripción del ítem	Corre- lación ítem-total corregida	Alfa de Cronbach si se elimi- na el ítem	Alfa de Cronbach
Factor 1. Competencia en IA			.91
1. Uso fácilmente herramientas de IA con diferentes interfaces.	.60	.94	
2. El uso activo de herramientas de IA mejora mi capacidad de aprendizaje.	.60	.94	
3. Utilizo eficientemente las herramientas de IA para encontrar información relevante para mis estudios.	.66	.93	
4. Confío en mi capacidad para aprender a través de nuevas herramientas de IA.	.66	.93	
5. Las instrucciones de las herramientas de IA son fáciles de entender.	.64	.93	
6. Interpreto fácilmente la información generada por IA en mi proceso de aprendizaje.	.72	.93	
Factor 2. Conciencia del Aprendizaje			.88
7. La retroalimentación de las herramientas de IA me ayuda a comprender mis fortalezas de aprendizaje.	.61	.94	
8. La retroalimentación proporcionada por la IA me ayuda a identificar áreas de mejora.	.62	.93	
9. Evalúo la eficacia de las herramientas de IA para ayudarme a alcanzar mis objetivos de aprendizaje.	.65	.93	
10. Reflexiono críticamente sobre mi dependencia de las herramientas de IA para aprender.	.67	.93	
Factor 3. Estrategias de Aprendizaje			.84
11. Adapto mis métodos de estudio según las recomendaciones de las herramientas de IA.	.66	.93	
12. Uso herramientas de IA para encontrar las formas más eficaces de aprender temas nuevos.	.71	.93	
13. Utilizo la información generada por IA para mejorar mis habilidades de resolución de problemas.	.70	.93	
14. Al usar herramientas de IA para analizar diferentes perspectivas, mejoro mis habilidades de pensamiento crítico.	.56	.94	
Factor 4. Implicación y Eficiencia			.86
15. Mi participación en las actividades de aprendizaje aumenta con el uso de la IA.	.64	.93	

Factores/Descripción del ítem	Corre- lación ítem-total corregida	Alfa de Cronbach si se elimi- na el ítem	Alfa de Cronbach
16. Las herramientas de IA me ayudan a lograr un aprendizaje más permanente.	.60	.94	
17. Creo que aprendo mejor cuando uso herramientas de IA.	.63	.93	
18. Creo que las herramientas de IA que se adaptan a mi estilo de aprendizaje aumentan mi implicación en el aprendizaje.	.74	.93	
Factor 5. Colaboración Ética			.87
19. Sé que debo asumir toda la responsabilidad cuando uso herramientas de IA para aprender.	.58	.94	
20. Soy consciente de las preocupaciones de privacidad en las herramientas de IA.	.54	.94	
21. Soy consciente de los posibles sesgos en las herramientas de IA.	.60	.94	
22. Soy consciente de la seguridad de la información que proporciono a las herramientas de IA.	.55	.94	
Escala Completa			.94

Nota: IC del 95% calculados usando 5000 muestras bootstrap. Todas las correlaciones ítem-total son significativas a $p < .001$.

Fecha de recepción del artículo: 1 de junio de 2025

Fecha de aceptación del artículo: 27 de agosto de 2025

Fecha de aprobación para maquetación: 25 de septiembre de 2025

Fecha de publicación en OnlineFirst: 16 de octubre de 2025

Fecha de publicación: 1 de enero de 2026

Influencia de la motivación hacia el aprendizaje y la procrastinación en la dependencia a ChatGPT

The influence of learning motivation and procrastination on ChatGPT dependence



- ID Héctor Galindo-Domínguez - Universidad del País Vasco, UPV/EHU (España)
- ID Martín Sainz-de-la-Maza - Universidad del País Vasco, UPV/EHU (España)
- ID Lucía Campo - Universidad del País Vasco, UPV/EHU (España)
- ID Daniel Losada-Iglesias - Universidad del País Vasco, UPV/EHU (España)

RESUMEN

En poco tiempo, las inteligencias artificiales generativas como ChatGPT se han convertido en herramientas de apoyo para el aprendizaje, transformando la forma en que muchos estudiantes afrontan las tareas académicas. Su inmediatez y accesibilidad las hacen especialmente atractivas frente a los recursos tradicionales, hasta el punto de que algunos estudiantes pueden desarrollar cierta dependencia a ellas. Este trabajo investiga el impacto de la motivación hacia el aprendizaje y la procrastinación en la dependencia a ChatGPT. 467 estudiantes universitarios del área de la Educación participaron completando una serie de escalas validadas para evaluar diferentes tipos de motivación definidas en la *Teoría de la Autodeterminación* (motivación extrínseca, intrínseca y amotivación), procrastinación y dependencia a ChatGPT. Los análisis de mediación mostraron que los estudiantes con menor motivación intrínseca ($\beta = -.076$; $LI = -.121$; $LS = -.037$) y mayor amotivación ($\beta = .090$; $LI = .041$; $LS = .144$) tendían a procrastinar con mayor frecuencia, siendo la procrastinación un factor clave que aumentaba la dependencia a ChatGPT. Además, el alumnado con alta motivación extrínseca (incluso sin la mediación de la procrastinación) resultó ser más propenso a desarrollar una mayor dependencia a ChatGPT ($\beta = .122$; $p = .022$). Estos hallazgos destacan la importancia de implementar estrategias que fomenten la motivación intrínseca y la autorregulación, ayudando a los estudiantes a utilizar adecuadamente las herramientas basadas en la IA generativa mientras desarrollan competencias esenciales que podrían estar en riesgo por el uso excesivo de estas herramientas, como el pensamiento crítico y la resolución de problemas.

Palabras clave: motivación; dependencia; ChatGPT; procrastinación; inteligencia artificial; educación superior.

ABSTRACT

In a short period of time, generative artificial intelligences such as ChatGPT have become a widely used support tool for students' learning processes, altering how many approach academic tasks. Their immediacy and accessibility make them attractive alternatives to traditional study resources, to the extent that some students may have even developed a certain degree of dependence on these tools. The present study investigates the impact of learning motivation and procrastination on ChatGPT dependence. A total of 467 university students from the field of education participated by completing a series of validated scales measuring different types of motivation defined in Self-Determination Theory (extrinsic motivation, intrinsic motivation, and amotivation), procrastination, and ChatGPT dependence. The indirect effects of the mediation analyses revealed that students with lower levels of intrinsic motivation ($\beta = -.076$; $LL = -.121$; $UL = -.037$) and higher levels of amotivation ($\beta = .090$; $LL = .041$; $UL = .144$) were more likely to procrastinate frequently, with procrastination emerging as a significant factor contributing to greater ChatGPT dependence. Similarly, results indicated that students with high extrinsic motivation (without procrastination serving as a mediator) were more prone to develop greater dependence on ChatGPT ($\beta = .122$; $p = .022$). These findings highlight the importance of implementing strategies that foster intrinsic motivation and self-regulation, helping students use generative AI-based tools appropriately while developing essential competencies that could be at risk from excessive use of these tools, such as critical thinking and problem-solving.

Keywords: motivation; dependence; ChatGPT; procrastination; artificial intelligence; higher education.

INTRODUCCIÓN

En este estudio, el marco teórico se diseña para poner a prueba un modelo de mediación en el que la motivación actúa como variable independiente, la procrastinación como variable mediadora y la dependencia a ChatGPT como variable dependiente. Con el fin de ofrecer al lector una hoja de ruta clara, la revisión de la literatura se organiza de manera lógica. En primer lugar, se abordan las teorías fundamentales sobre motivación y procrastinación, analizando la relación previamente establecida entre estos dos constructos. En segundo lugar, se define la dependencia a la inteligencia artificial en el contexto de la educación superior, centrándose en la excesiva confianza en estos sistemas para la realización de tareas académicas. Finalmente, el marco integra estos tres conceptos al vincular la motivación con la dependencia a ChatGPT y, a su vez, la procrastinación con dicha dependencia, alineando el flujo teórico con el modelo propuesto.

Motivación y procrastinación en los procesos de aprendizaje

La motivación constituye un constructo complejo dentro de la psicología y el comportamiento humano, ejerciendo influencia en la forma en que las personas deciden distribuir su tiempo, el nivel de esfuerzo que invierten en una tarea, los pensamientos y emociones que desarrollan respecto a la actividad, así como en su perseverancia para completarla (Filgona et al., 2020).

Aunque se han propuesto numerosas teorías influyentes para comprender la motivación individual, como la Teoría de la Atribución, la Teoría de la Expectativa-Valor, el modelo ARCS, la Teoría Social Cognitiva y la Teoría de Orientación a Metas, entre otras, la *Teoría de la Autodeterminación* (del inglés, *Self-Determination Theory*, en adelante SDT) ha cobrado recientemente mayor relevancia al explicar los cambios de actitudes hacia la Inteligencia Artificial (p. ej., Chai et al., 2023; Chiu, 2024; Chiu y Chai, 2020; Xia et al., 2022). La taxonomía asociada a esta teoría explica cómo las distintas formas de motivación varían según el grado de autodeterminación, desde la ausencia de motivación, pasando por motivaciones controladas por factores externos, hasta formas más autónomas vinculadas a intereses personales (Ryan y Deci, 2020).

En este sentido, la amotivación se refiere a la ausencia de intención de actuar, a menudo derivada de una falta percibida de competencia o de valor en la actividad. La motivación extrínseca, en cambio, está impulsada por recompensas o presiones externas, aunque varía en su nivel de autonomía según el grado de internalización. La forma más autónoma de motivación es la motivación intrínseca, que surge del disfrute y la satisfacción inherentes a la propia actividad, promoviendo así tanto el compromiso como la persistencia (Ryan y Deci, 2020).

Asimismo, la SDT sostiene que las diferentes formas de motivación (intrínseca, extrínseca y amotivación) están estrechamente relacionadas con el grado en que se

satisfacen tres necesidades psicológicas básicas: competencia, autonomía y relación (Ryan y Deci, 2020).

En primer lugar, la competencia puede entenderse como la sensación de ser eficaz y capaz en las interacciones con el entorno. Implica dominar tareas y alcanzar los resultados deseados, fomentando un sentido de logro y crecimiento. Sentirse competente fortalece tanto la motivación intrínseca como las formas más autodeterminadas de motivación extrínseca, mientras que la falta de competencia puede conducir a la amotivación. Diversos estudios han mostrado cómo una fuerte competencia, a menudo reflejada en un alto rendimiento en la ejecución de tareas, es un predictor del incremento de la motivación intrínseca en los estudiantes (p. ej., Garon-Carrier et al., 2016; Van Bergen et al., 2022).

En segundo lugar, la autonomía puede definirse como la sensación de controlar las propias acciones y decisiones. Permite a los individuos actuar de manera voluntaria y perseguir metas de forma autoaprobada. La motivación intrínseca se potencia cuando las personas perciben libertad y capacidad de elección, mientras que la falta de autonomía puede contribuir a la amotivación o a formas controladas de motivación extrínseca. En este sentido, estudios previos han mostrado que los estudiantes que aprenden en contextos docentes que apoyan la autonomía presentan una mayor motivación intrínseca, mayor compromiso en el aula y desarrollo de habilidades, así como una menor probabilidad de abandono escolar (p. ej., Cheon et al., 2020; Ljubin-Golub et al., 2020; Reeve y Cheon, 2021).

En tercer lugar, el concepto de relación alude a la necesidad fundamental de establecer vínculos con los demás, experimentar pertenencia y desarrollar lazos interpersonales significativos. Implica la percepción de ser cuidado, apoyado y comprendido en contextos sociales. Cuando esta necesidad se satisface, los individuos tienden a mostrar formas más autónomas de motivación, involucrándose de manera más plena en actividades que nutren su sentido de conexión. Evidencia previa ha demostrado que fomentar la relación en el aula influye parcialmente en el incremento de la motivación intrínseca del alumnado (p. ej., Ganotice et al., 2024; Guay et al., 2019).

Como se observa, desde la perspectiva de la SDT, las necesidades psicológicas básicas tienen un impacto significativo en la motivación de los estudiantes. Además, de acuerdo con la evidencia previa, contar con esta motivación es clave para reducir ciertos constructos como la procrastinación, un fenómeno que se analiza con mayor detalle en este trabajo.

La procrastinación se describe comúnmente como una tendencia irracional a posponer tareas o responsabilidades esenciales, a pesar de las consecuencias negativas que este retraso pueda tener tanto para las personas como para las organizaciones (Klingsieck, 2013). La procrastinación representa un déficit de autorregulación asociado tanto a disposiciones personales como a factores ambientales (Hen y Goroshit, 2018). En concreto, diversos estudios señalan que los factores relacionados con la tarea (como instrucciones poco claras, momento de las

recompensas y castigos, y el carácter aversivo de la tarea), los rasgos de personalidad (p. ej., el modelo de los cinco grandes, la motivación y la cognición) y las influencias contextuales (p. ej., tentaciones, incentivos y mecanismos de rendición de cuentas) son elementos clave que contribuyen a la procrastinación (Johnson y Bloom, 1995; Wypych et al., 2018).

La procrastinación puede actuar como una barrera para el logro, ejerciendo efectos perjudiciales sobre el bienestar emocional y contribuyendo a mayores niveles de ansiedad, síntomas depresivos y disminución de la autoestima (Duru y Balkis, 2017). Además, quienes procrastinan de manera crónica suelen mostrar un rendimiento más reducido, manifestado en resultados académicos más bajos, retrasos en el desarrollo profesional y consecuencias negativas para la salud (Legood et al., 2018; Bolden y Fillauer, 2020).

En lo que respecta a la relación entre motivación y procrastinación, esta ha sido ampliamente explorada en la literatura científica, mostrando que dicha relación es compleja y parece ser principalmente unidireccional: la motivación influye en la procrastinación más que a la inversa (p. ej., Grund y Fries, 2018; Li et al., 2021; Wu y Fan, 2016). En concreto, la evidencia previa ha revelado que niveles bajos de valor de logro (Wypych et al., 2018; Wu y Fan, 2016), autodeterminación (Grund y Fries, 2018), perseverancia (Wypych et al., 2018) y motivación intrínseca (Kok, 2016), así como niveles elevados de coste percibido (Wu y Fan, 2016), descuento por demora (Wypych et al., 2018), amotivación y motivación externa (Kok, 2016), conducen a mayores niveles de procrastinación académica.

Dependencia a la inteligencia artificial en la educación superior

La inteligencia artificial (IA) se define como la capacidad de las máquinas o sistemas computacionales para emular y llevar a cabo funciones tradicionalmente atribuidas a la inteligencia humana, incluyendo el razonamiento, el aprendizaje y la resolución de problemas (Morandín-Ahuerma, 2022). Si bien la integración de la IA en contextos educativos constituye un fenómeno relativamente reciente, la investigación existente subraya varias ventajas potenciales; entre ellas, el apoyo en la redacción de trabajos escritos, la provisión de retroalimentación automatizada y la asistencia en el análisis de datos e información (Chen et al., 2020; Delgado et al., 2024; Wang et al., 2018). No obstante, los estudios también destacan limitaciones notables, como la reducción de la interacción humana en el proceso de aprendizaje, la comprensión contextual limitada y la dependencia excesiva de la IA, que puede disminuir habilidades esenciales como el pensamiento crítico y la resolución de problemas (Ahmad et al., 2021; Jara y Ochoa, 2020; Korteling et al., 2021). Este estudio se centra precisamente en esta última preocupación: la dependencia a la IA.

En el ámbito clínico, la dependencia se ha vinculado históricamente al consumo de sustancias como el alcohol y las drogas, habitualmente identificada mediante indicadores como la tolerancia y el síndrome de abstinencia (American Psychiatric

Association, 2013). Sin embargo, con la llegada de la era de la inteligencia artificial, esta comprensión se ha ampliado para abarcar formas de dependencia a carácter conductual, como el juego patológico y la adicción a internet, que muestran patrones neurobiológicos y conductuales comparables (p. ej., Bucur et al., 2021; Siste et al., 2019).

En un contexto académico, la dependencia a la IA se define como la confianza excesiva depositada en sistemas automatizados para la toma de decisiones o la ejecución de tareas (Morales-García et al., 2024). Este fenómeno se manifiesta en una falta de confianza en las propias capacidades académicas y en la preferencia por recurrir a la IA como fuente principal, e incluso exclusiva, de apoyo (Del Cisne et al., 2024). De hecho, la evidencia reciente sugiere que el simple hecho de saber que un consejo proviene de una IA puede llevar a los estudiantes a confiar en él en exceso, incluso cuando contradice la información disponible o su propio juicio, lo que potencialmente genera resultados subóptimos tanto para ellos mismos como para otros (Klingbeil et al., 2024). Una de las principales razones, tal como señalan Baird y Maruping (2021), es que los sistemas de IA y los chatbots replican cada vez más las interacciones humanas de una forma natural, lo que conduce a que los estudiantes depositen una mayor confianza en este tipo de sistemas.

Este fenómeno puede justificarse a través de la *Teoría Unificada de Aceptación y Uso de la Tecnología* (UTAUT, por sus siglas en inglés) (Venkatesh et al., 2003), la cual ofrece un marco para analizar la dependencia a la IA al identificar los principales factores que afectan su adopción y uso. La teoría enfatiza que constructos como la expectativa de rendimiento, la expectativa de esfuerzo, la influencia social y las condiciones facilitadoras son determinantes en la configuración del comportamiento del usuario. La dependencia a la IA puede surgir cuando los usuarios tienen una elevada expectativa de rendimiento (al percibir que la IA mejora sustancialmente la eficiencia o precisión de las tareas) y una baja expectativa de esfuerzo, dado que la IA simplifica procesos que de otro modo serían complejos. Asimismo, la influencia social, incluyendo la presión de compañeros para adoptar y utilizar estas herramientas, puede amplificar la dependencia hacia estos sistemas. Finalmente, las condiciones facilitadoras, como la infraestructura y los recursos disponibles, crean un entorno propicio para el uso frecuente de la IA, lo que potencialmente podría conducir a sobreconfiar en ella.

En la actualidad, algunos estudios revelan que la prevalencia de la dependencia a herramientas de inteligencia artificial como ChatGPT es moderadamente baja, situándose entre un 16 % y un 25 % en estudiantes universitarios (Hong y Chen, 2024; Huang et al., 2024; Stojanov, 2024). Algunos autores incluso sugieren que este fenómeno aún no debería ser motivo de una alarma social excesiva (Huang et al., 2024). No obstante, los rápidos avances de la inteligencia artificial en el ámbito educativo, junto con su amplia gama de funcionalidades, han generado nuevas oportunidades, pero también plantean el riesgo de una dependencia excesiva que requiere de una regulación institucional (Liu et al., 2023; Stojanov, 2024). En

este sentido, las universidades podrían desempeñar un papel crucial guiando al estudiantado hacia un uso equilibrado de estas herramientas, promoviendo su integración de manera crítica y asegurando al mismo tiempo que los aprendices continúen fortaleciendo sus propias competencias (Lee et al., 2024).

Esta visión se alinea con modelos de educación abierta y a distancia, como el marco de la *Comunidad de Indagación*, que enfatizan la necesidad de la presencia cognitiva, social y docente para apoyar el aprendizaje autorregulado y prevenir una excesiva dependencia a las herramientas tecnológicas (Carroll et al., 2025).

Impacto de la motivación y la procrastinación en la dependencia a ChatGPT

ChatGPT (*Chat Generative Pre-Trained Transformer*), desarrollado por OpenAI, es un sofisticado modelo de lenguaje basado en inteligencia artificial diseñado para gestionar tareas lingüísticas complejas. Reconocido por su capacidad de generar texto con apariencia humana, facilita una amplia variedad de aplicaciones, como la generación automática de textos, la respuesta a preguntas, la elaboración de resúmenes o la creación de imágenes. Al ofrecer respuestas que emulan de manera cercana la comunicación humana, ChatGPT ha revolucionado la interacción con chatbots y asistentes virtuales, alcanzando una notable popularidad global y un uso extendido desde su lanzamiento en noviembre de 2022 (Yu et al., 2024).

Pese a su impacto, la literatura existente que analiza la relación entre motivación, procrastinación y dependencia a ChatGPT es aún extremadamente limitada debido a la novedad del fenómeno. No obstante, algunos estudios comienzan a aportar hallazgos relevantes.

En primer lugar, en lo que respecta a la relación entre motivación y dependencia a ChatGPT, diversas investigaciones han mostrado que altos niveles de *lazy thinking* o pensamiento vago (que reflejarían amotivación hacia las tareas), elevados niveles de estrés emocional (que podrían indicar agotamiento frente a las tareas), baja autoeficacia académica y escasas habilidades de toma de decisiones se asocian con un incremento en la dependencia a la IA generativa (Ahmad et al., 2023; Estrada-Araoz et al., 2025; Naseer et al., 2025; Ye et al., 2024). Asimismo, la literatura reciente subraya el efecto de distintos tipos de motivación sobre la dependencia a ChatGPT. En particular, bajos niveles de motivación de logro (Hong y Chen, 2024) y de orientación de dominio-aproximación (Stojanov, 2024), así como altos niveles de motivación de escape, motivación social (Huang et al., 2024), orientación de dominio-evitación y orientación de rendimiento-aproximación (Stojanov, 2024) se han relacionado con mayores niveles de dependencia hacia esta herramienta. Tal como afirman Ye et al. (2024), las personas que han tenido experiencias positivas utilizando ChatGPT pueden verse más motivadas a evitar las tareas de aprendizaje y a confiar en las respuestas proporcionadas por el sistema.

En segundo lugar, en relación con la procrastinación y la dependencia a ChatGPT, algunos estudios han comenzado a explorar esta conexión. Por ejemplo, Feng et al. (2023) formularon la hipótesis de que la procrastinación podría influir en la dependencia a ChatGPT, una idea que ha recibido cierto respaldo empírico. Stojanov (2024) observó que los estudiantes procrastinadores, casi una cuarta parte de la población universitaria, eran altamente dependientes a ChatGPT para completar tareas asignadas. De manera similar, Swargiary (2024), a través de un diseño pre-post con grupo control, llevó a cabo una intervención en la que el grupo experimental utilizó ChatGPT para resolver tareas. Los resultados mostraron que dicho grupo presentó un aumento en los niveles de procrastinación y un descenso en el rendimiento académico respecto al grupo de comparación. Por otro lado, estudios con metodologías cualitativas (p. ej., Ayele, 2024; Paraso et al., 2024) reportaron que los estudiantes universitarios consideran la IA como una herramienta extremadamente útil, especialmente entre aquellos que suelen dejar las tareas para el último momento.

En conjunto, estos hallazgos sugieren que la procrastinación podría desempeñar un papel mediador en la relación entre los distintos tipos de motivación y la dependencia a ChatGPT. Esto implicaría que los estudiantes con baja motivación interna o con desafección hacia las tareas académicas tenderían a posponerlas con mayor frecuencia, recurriendo posteriormente a herramientas de IA como estrategia compensatoria. Esta dinámica sería particularmente relevante en contextos de educación abierta y a distancia, donde la autorregulación del aprendizaje resulta crucial para evitar una sobreconfianza en la tecnología.

El presente estudio

El impacto de la inteligencia artificial en el ámbito educativo ha despertado un creciente interés, lo que ha dado lugar a una proliferación de artículos de opinión. Sin embargo, la investigación empírica sobre este tema sigue siendo limitada (Ye et al., 2024). Una de las principales preocupaciones para docentes e investigadores son las posibles consecuencias negativas de ChatGPT, especialmente derivadas de su uso problemático (Ye et al., 2024). A pesar de ello, como señalan diversos estudios, existe una notable falta de análisis exhaustivos sobre los usos inapropiados, como la dependencia a la IA, lo que justifica una mayor exploración (Yu et al., 2024). Por ello, algunos autores subrayan la importancia de investigar los factores psicológicos que contribuyen al desarrollo de esta dependencia (Ahmad et al., 2023; Ye et al., 2024), dado que el uso intensivo de estas herramientas podría afectar negativamente la autorregulación, el pensamiento crítico y las habilidades de resolución de problemas de los estudiantes (Ali et al., 2024; Kamalov et al., 2023; Zhai et al., 2024).

Del mismo modo, como se ha observado previamente, existe cierta evidencia que vincula la motivación y la procrastinación con la dependencia a ChatGPT. Sin embargo, hasta donde alcanzan los conocimientos de los autores, la evidencia actual

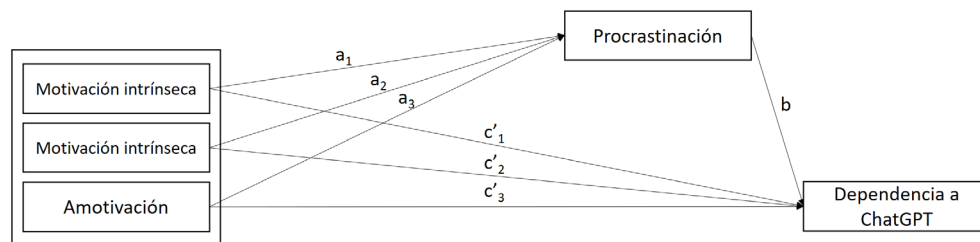
que emplea análisis más complejos, como modelos de ecuaciones estructurales, análisis de mediación o de moderación, es extremadamente limitada. Esta escasez dificulta una comprensión integral de la relación entre diversos constructos de interés, como la motivación o la procrastinación. Además, esta limitación se ve reforzada por la falta de estudios empíricos sobre la dependencia a la IA abordados desde marcos teóricos sólidos, como la *Teoría de la Autodeterminación*, que ha ganado relevancia en el campo de la IA (p. ej., Chai et al., 2023; Chiu, 2024; Chiu y Chai, 2020; Xia et al., 2022). Cabe destacar, además, que en los últimos años ha aumentado el número de artículos académicos que examinan la procrastinación en estudiantes universitarios, evidenciando un interés científico creciente en esta temática (Yan y Zhang, 2022).

En consecuencia, este estudio busca responder a tres preguntas de investigación (RQ):

- RQ₁: ¿En qué medida diferentes tipos de motivación (amotivación, intrínseca y extrínseca) influyen en la dependencia a ChatGPT?
- RQ₂: ¿En qué medida la procrastinación impacta en la dependencia a ChatGPT?
- RQ₃: ¿En qué grado la procrastinación media la relación entre las distintas formas de motivación y la dependencia a ChatGPT?

La Figura 1 presenta el diagrama del modelo que se pretende testar en los análisis subsiguientes.

Figura 1
Análisis de mediación propuesto



METODOLOGÍA

Diseño y participantes

El estudio empleó un diseño cuantitativo, transversal y no experimental, siguiendo un enfoque descriptivo-correlacional. Se llevaron a cabo análisis de mediación para examinar los efectos indirectos de la motivación para el aprendizaje y la procrastinación sobre la dependencia a ChatGPT.

En este diseño participaron 467 estudiantes universitarios españoles del área de educación de la Universidad del País Vasco ($M = 20,44$; $DT = 2,83$). De ellos, 292 se identificaron como hombres, 173 como mujeres y 2 como otro género. Respecto a los programas académicos, 40 estudiantes estaban matriculados en el Grado en Educación Infantil, 379 en el Grado en Educación Primaria y 48 en el Máster en Profesorado de Educación Secundaria. En términos de curso académico, 274 participantes cursaban el segundo año, 103 el tercer año, 45 el cuarto año y 45 el Máster (equivalente al quinto año).

Los participantes fueron seleccionados mediante muestreo no probabilístico por conveniencia, basado en la proximidad al equipo de investigación, ya que todos los estudiantes habían sido estudiantes de uno o más miembros del equipo. Cabe destacar que estudios previos sugieren que los métodos de muestreo no probabilístico suelen producir resultados comparables a los obtenidos mediante técnicas de muestreo aleatorio (Coppock et al., 2018).

Asimismo, se realizó un análisis de potencia a priori utilizando *G*Power 3.1* para un modelo de regresión múltiple lineal con cuatro predictores (tres variables independientes y un mediador). Asumiendo un tamaño de efecto medio ($f^2 = 0,15$) y $\alpha = 0,05$, el análisis indicó que se requería un mínimo de 55 participantes para alcanzar un poder estadístico de 0,80, valor comúnmente usado en estudios correlacionales en educación y psicología (Cohen, 1988). Con la muestra actual de 467 participantes, el estudio cuenta con un nivel robusto de potencia estadística para detectar los efectos anticipados.

Instrumentos

- **Motivación para el aprendizaje**

Motivación intrínseca: se utilizaron las dimensiones Motivación Intrínseca por Conocimiento y Motivación Intrínseca hacia la Realización de la versión corta de la Escala de Motivación Académica (Kotera et al., 2023). Cada dimensión comprende cuatro ítems tipo Likert de 5 puntos, que capturan la intención interna del participante de estudiar con el objetivo de ampliar conocimientos (p. ej., “Estudio porque mis estudios me permiten seguir aprendiendo sobre muchas cosas que me interesan”) y

mejorar como estudiante (p. ej., “Estudio por el placer que siento mientras me supero en alguno de mis logros personales”). Según Kotera et al. (2023), la confiabilidad de estas dimensiones fue $\alpha = .84$ y $.79$, respectivamente.

Motivación extrínseca: para evaluar la motivación extrínseca se empleó la dimensión Regulación Externa de la misma escala. Esta dimensión consta de dos ítems tipo Likert de 5 puntos que miden la intención del participante de estudiar por recompensas externas (p. ej., “Estudio para obtener un trabajo más prestigioso en el futuro”). La confiabilidad reportada fue $\alpha = .82$ (Kotera et al., 2023).

Amotivación: se midió utilizando la dimensión Amotivación de la misma escala, que incluye dos ítems tipo Likert de 5 puntos destinados a capturar la falta de interés por estudiar (p. ej., “No entiendo por qué voy a la universidad y, francamente, me da igual”). La confiabilidad reportada fue $\alpha = .79$ (Kotera et al., 2023).

- **Procrastinación**

Para evaluar el nivel de procrastinación se utilizó la Versión Corta de la Escala de Procrastinación Académica (Yockey, 2016). Esta escala consta de cinco ítems tipo Likert de 5 puntos agrupados en un solo factor, diseñados para medir hasta qué punto los estudiantes retrasan y posponen la realización de sus tareas hasta el último momento (p. ej., “Cuando me asignan una tarea, generalmente la dejo de lado y me olvido de ella hasta que casi vence el plazo”). Según Yockey (2016), la confiabilidad de esta escala es $\alpha = .87$.

- **Dependencia a ChatGPT**

Para evaluar el grado de dependencia a ChatGPT se empleó la Escala de Dependencia a ChatGPT (Ye et al., 2024). Esta escala consta de una sola dimensión con cinco ítems tipo Likert de 5 puntos, que miden hasta qué punto los individuos sienten la necesidad de usar ChatGPT para tareas académicas (p. ej., “Me sentiría incómodo sin ChatGPT”). La confiabilidad reportada fue $\alpha = .71$ (Ye et al., 2024).

Procedimiento

El procedimiento comenzó con una revisión de diversas bases de datos para identificar las escalas más apropiadas para el estudio, las cuales fueron posteriormente digitalizadas mediante Google Forms. A continuación, se definió la muestra, incluyendo a todos los estudiantes que habían sido instruidos por los investigadores participantes y que estaban matriculados en programas relacionados con la educación. Como se señaló previamente, aunque se trató de una muestra no probabilística, investigaciones anteriores indican que las muestras por conveniencia

suelen producir resultados comparables a los obtenidos mediante muestreos aleatorios (Coppock et al., 2018).

Los participantes completaron el cuestionario de *Google Forms* durante las horas de clase, utilizando ordenadores individuales, previo consentimiento de participación en el estudio. Las condiciones de participación cumplieron con los principios establecidos en la Declaración de Helsinki, garantizando el anonimato y la privacidad de las respuestas, el derecho a retirarse antes o durante la participación, así como la gestión y disposición adecuada de los datos durante y después de la investigación.

Al finalizar el estudio, los participantes recibieron un informe resumen con los principales hallazgos y sus implicaciones.

Análisis de datos

En primer lugar, para evaluar la validez de los datos, se examinó el ajuste del modelo utilizando los índices X^2/gl , CFI, RMSEA y AIC mediante el software SPSS AMOS 25, esperando que cumplieran los siguientes umbrales: $X^2/\text{gl} < 5$; CFI $> .90$; RMSEA $< .080$; AIC = valor más bajo posible (Hooper et al., 2008).

Posteriormente, en *SPSS Statistics 27*, se realizó un análisis descriptivo, estudiando la media aritmética y la desviación estándar de las variables. A continuación, se llevaron a cabo otros análisis, incluyendo diversos análisis correlacionales mediante el coeficiente de Pearson (r) entre las dimensiones principales y diversos análisis de confiabilidad utilizando el alfa de Cronbach. Finalmente, también en *SPSS Statistics 27*, se utilizó la macro *PROCESS* para realizar un análisis de mediación, como se muestra en la Figura 1. Este análisis se llevó a cabo empleando 10,000 muestras de *bootstrapping*. Se prestó especial atención a los efectos directos e indirectos obtenidos. Específicamente, para responder a las preguntas de investigación RQ₁ y RQ₂, se examinaron los efectos directos, observando los valores de R^2 para determinar el tamaño del efecto, lo que indica la proporción de varianza de la variable dependiente explicada por la variable independiente.

Para responder a la pregunta de investigación RQ₃, se estudiaron los efectos indirectos, calculando la Proporción de Mediación (PM), que indica la proporción del efecto total de la variable independiente sobre la dependiente que es explicado por el mediador.

Todos los análisis paramétricos se justificaron por la distribución normal de todas las variables medidas, evaluada mediante asimetría y curtosis, con valores comprendidos en el rango ± 2 para la asimetría y ± 7 para la curtosis (Hair et al., 2010). Se evaluó también la multicolinealidad, con valores dentro del rango aceptable de VIF < 10 y Tolerancia $> .10$ (Field, 2013). También se examinó la independencia a los errores mediante el estadístico Durbin–Watson, obteniéndose un valor de 1,636, estando este dentro del rango 1,5 a 2,5, lo que indica una ligera autocorrelación positiva pero generalmente aceptable (Field, 2013).

RESULTADOS

En primer lugar, para determinar si eran necesarios ajustes preliminares, se evaluó el ajuste del modelo propuesto mediante los índices X^2/df , CFI, RMSEA y AIC. El modelo resultante mostró un buen ajuste ($X^2/df = 2.328$; CFI = .955; RMSEA = .053; AIC = 382.995), ya que los valores se encontraban dentro de los umbrales aceptables establecidos para este estudio: $X^2/df < 5$; CFI $> .90$; RMSEA $< .080$; AIC = valor más bajo posible (Hooper et al., 2008).

Posteriormente, se analizaron los estadísticos descriptivos, correlaciones e índices de fiabilidad de las dimensiones principales. Como se muestra en la Tabla 1, los participantes presentaron en general altos niveles de motivación intrínseca ($M = 3.48$; $DT = .843$) y extrínseca ($M = 3.95$; $DT = .849$), junto con bajos niveles de amotivación ($M = 1.63$; $DT = .884$). Además, se observaron niveles moderados de procrastinación ($M = 2.57$; $DT = .947$) y dependencia a ChatGPT ($M = 2.37$; $DT = .996$).

Los análisis de correlación indicaron que la motivación intrínseca se relacionaba negativamente con la procrastinación ($r = -.287$, $p < .001$), mientras que la amotivación mostró una asociación positiva con la procrastinación ($r = .403$, $p < .001$). Por otro lado, el único tipo de motivación significativamente asociado con la dependencia a ChatGPT fue la amotivación ($r = .162$; $p < .001$), mostrando también una correlación positiva. De manera similar, la procrastinación presentó una correlación positiva con la dependencia a ChatGPT ($r = .231$; $p < .001$).

Finalmente, todos los índices de confiabilidad superaron el umbral de $\alpha = .70$ (Tavakol y Dennick, 2011), lo que indica una buena consistencia interna de las dimensiones utilizadas en el estudio.

Tabla 1
Estadísticos descriptivos, correlaciones e índices de confiabilidad de las dimensiones principales

	M	DT	1	2	3	4	5
MI	3.48	.843	(.865)	.176***	-.200***	-.287***	-.084
ME	3.95	.849		(.785)	-.046	-.091	.082
AM	1.63	.884			(.813)	.403***	.162***
PRC	2.57	.947				(.858)	.231***
GPT	2.37	.996					(.858)

Nota: * $p < .05$; ** $p < .01$; *** $p < .001$; MI, Motivación Intrínseca; ME, Motivación Extrínseca; AM, Amotivación; PRC, Procrastinación; GPT, Dependencia a ChatGPT. Se muestran los valores de consistencia interna Alfa de Cronbach en la diagonal principal de la matriz.

Para dar respuesta a las preguntas de investigación, se llevó a cabo un análisis de mediación, el cual permite evaluar en qué medida el efecto de una variable independiente sobre una dependiente opera de manera indirecta a través de una variable mediadora. En este modelo, las variables independientes fueron los distintos tipos de motivación, la variable dependiente fue la dependencia a ChatGPT, y la variable mediadora fue la procrastinación.

Los principales resultados de los efectos directos, presentados en la Tabla 2, indican que tanto la motivación intrínseca ($R^2 = .082$; $\beta = -.322$; $p < .001$) como la amotivación ($R^2 = .162$; $\beta = .432$; $p < .001$) se asocian significativamente con la procrastinación. Estos valores de R^2 indican que la motivación intrínseca por sí sola explica aproximadamente un 8 % de la varianza en procrastinación, mientras que la amotivación explica un 16 %. Esto sugiere que menor motivación intrínseca y mayor amotivación se asocian con niveles más elevados de procrastinación, evidenciando que la ausencia de razones autodeterminadas para realizar tareas académicas puede resultar particularmente perjudicial.

Además, en relación con las preguntas de investigación RQ₁ y RQ₂, la procrastinación se encontró significativamente asociada con la dependencia a ChatGPT ($R^2 = .053$; $\beta = .237$; $p < .001$), lo que indica que la procrastinación explica alrededor del 5 % de la varianza en la dependencia a ChatGPT. Entre los tres tipos de motivación analizados, la **única directamente asociada con la** dependencia a ChatGPT fue la motivación extrínseca ($R^2 = .026$; $\beta = .122$; $p = .022$). Aunque este efecto fue estadísticamente significativo, la proporción de varianza explicada es relativamente baja, lo que sugiere que otros factores probablemente contribuyen de manera más relevante a la dependencia a ChatGPT.

Tabla 2
Resultados del análisis de mediación

	Ruta	R ²	PM	β	SE	LL	UL	p
Efectos directos								
MI → PRC	a ₁	.082	-	-.322	.050	-.420	-.224	< .001
ME → PRC	a ₂	.008	-	-.101	.052	-.203	.000	.054
AM → PRC	a ₃	.162	-	.432	.045	.342	.521	< .001
PRC → GPT	b	.053	-	.237	.050	.140	.334	< .001
MI → GPT	c' ₁	.007	-	-.023	.056	-.133	.086	.677
ME → GPT	c' ₂	.007	-	.122	.053	.018	.226	.022
AM → GPT	c' ₃	.026	-	.093	.055	-.016	.202	.096
Efectos indirectos								
MI → PRC → GPT	a ₁ * b	-	.764	-.076	.022	-.121	-.037	***
ME → PRC → GPT	a ₂ * b	-	-.270	-.026	.015	-.058	.001	(ns)

	Ruta	R ²	PM	β	SE	LL	UL	p
AM → PRC → GPT	a ₃ * b	-	.493	.090	.026	.041	.144	***

Nota: MI, Motivación Intrínseca; ME, Motivación Extrínseca; AM, Amotivación; PRC, Procrastinación; GPT, Dependencia a ChatGPT.

La Tabla 2 también presenta los efectos indirectos, que muestran si la variable independiente influye sobre la variable dependiente a través de la variable mediadora. La significancia se establece cuando el intervalo de confianza no incluye el cero (MacKinnon et al., 2002).

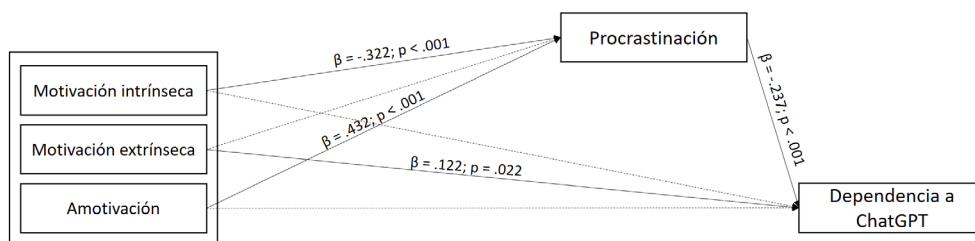
En este estudio, y dando respuesta a la pregunta de investigación RQ₃, se encontró que la procrastinación media las relaciones desde motivación intrínseca ($\beta = -.076$; $LL = -.121$; $UL = -.037$) y amotivación ($\beta = .090$; $LL = .041$; $UL = .144$) hacia la dependencia a ChatGPT. En otras palabras, una alta motivación intrínseca y baja amotivación se asocia con menor procrastinación, lo que a su vez contribuye a reducir la dependencia a ChatGPT. Una representación de estos resultados principales se puede observar en la Figura 2.

Asimismo, se calcularon los valores de Proporción de Mediación (PM) para determinar el tamaño del efecto de estas relaciones indirectas. Para la relación desde motivación intrínseca hacia dependencia a ChatGPT mediado por la procrastinación el PM fue .764, indicando que aproximadamente el 76,4 % del efecto total está mediado por la procrastinación. Esto sugiere una mediación casi completa, con una relevancia muy alta de la procrastinación. Similarmente, para la relación desde amotivación hacia dependencia a ChatGPT mediado por la procrastinación, el PM fue .493, mostrando que la procrastinación media casi la mitad (49,3 %) del efecto total.

En ambos casos, la procrastinación resulta clave para comprender la relación entre motivación intrínseca, amotivación y dependencia a ChatGPT, destacando su papel como mecanismo mediador central.

Figura 2

Figura representante de los resultados del análisis de mediación



DISCUSIÓN

El objetivo del presente estudio fue triple: primero, analizar el impacto de la motivación sobre la dependencia a ChatGPT; segundo, examinar el impacto de la procrastinación sobre la dependencia a ChatGPT; y, finalmente, determinar si la asociación entre los diferentes tipos de motivación y la dependencia a ChatGPT está mediada por los niveles de procrastinación.

Los resultados revelaron que el único tipo de motivación que tuvo un efecto directo sobre la dependencia a ChatGPT fue la motivación extrínseca, aunque la proporción de varianza explicada fue relativamente baja, lo que sugiere que otros factores pueden ejercer una influencia más fuerte sobre la dependencia a ChatGPT. En cuanto a la motivación intrínseca y la amotivación, se observó que su relación con la dependencia a ChatGPT está significativamente mediada por los niveles de procrastinación de los estudiantes, de manera que altos niveles de motivación intrínseca y bajos niveles de amotivación se asocian con una reducción de la procrastinación, lo que a su vez disminuye la dependencia a ChatGPT.

Aunque estos hallazgos son novedosos, ya que, hasta donde conocen los autores, no existen análisis de mediación previos que vinculen estas variables, son consistentes con la literatura existente. Por un lado, diversos estudios indican que una mayor motivación intrínseca hacia el aprendizaje se asocia con una menor dependencia a ChatGPT (p. ej., Hong y Chen, 2024; Stojanov, 2024), mientras que altos niveles de motivación impulsada por factores externos (p. ej., Huang et al., 2024; Stojanov, 2024), la falta de motivación o interés por el aprendizaje (p. ej., Ahmad et al., 2023; Ye et al., 2024) se relacionan con un aumento de la dependencia a ChatGPT. Asimismo, estudios cuantitativos y cualitativos previos señalan que los estudiantes con tendencia a procrastinar son más propensos a depender de ChatGPT para sus tareas académicas (p. ej., Ayele, 2024; Feng et al., 2023; Paraso et al., 2024; Stojanov, 2024; Swargiary, 2024).

Este estudio integra estos hallazgos mediante análisis de mediación, revelando que una motivación intrínseca fuerte y baja amotivación son clave para reducir la procrastinación de los estudiantes, la cual, a su vez, constituye un factor crítico para disminuir la dependencia a ChatGPT. Esto puede explicarse desde la *Teoría de la Autodeterminación* que propone un continuo de internalización (Ryan y Deci, 2020), de manera que la motivación extrínseca opera de forma directa, ya que los estudiantes utilizan ChatGPT como una solución rápida para cumplir demandas externas con el mínimo esfuerzo, mientras que la motivación intrínseca y la amotivación influyen en el coste cognitivo percibido de las tareas, lo que activa un mecanismo de afrontamiento: la procrastinación. Nuestros hallazgos sugieren que esta procrastinación actúa como el verdadero mediador, facilitando el uso de ChatGPT para evitar la tarea aversiva, de modo que la motivación intrínseca influye indirectamente sobre la dependencia a través de la procrastinación, mientras que la motivación extrínseca parece conducir a un uso directo.

Posteriormente, el presente estudio tiene implicaciones relevantes para diversos interesados. En primer lugar, estos hallazgos pueden ser valiosos para los investigadores, ya que contribuyen a un mejor entendimiento y consolidación de un marco teórico reciente sobre el impacto de factores emocionales en la dependencia a la inteligencia artificial generativa, como ChatGPT.

Asimismo, estos hallazgos pueden ser útiles para el profesorado universitario, ya que demuestran que incrementar la motivación intrínseca y reducir la amotivación y la procrastinación puede disminuir significativamente la dependencia a los estudiantes universitarios hacia ChatGPT. Como se señaló en el marco teórico de este trabajo, según la SDT, la motivación intrínseca está influida por las necesidades psicológicas básicas de competencia, autonomía y relación (Ryan y Deci, 2020). Dado que los resultados muestran que la procrastinación media las relaciones tanto de la motivación intrínseca como de la amotivación con la dependencia a ChatGPT, existe una amplia literatura que aporta estrategias para fortalecer cada una de estas necesidades, con efectos consecuentes sobre la motivación intrínseca de los estudiantes.

Entre las estrategias que el profesorado podría aplicar para fomentar las necesidades psicológicas básicas se incluyen: promover la retroalimentación positiva y constructiva (García et al., 2019; Koike y Pearson, 2005), utilizar técnicas de andamiaje o apoyo gradual (Mulatsih, 2011; Neitzel y Stright, 2003), fomentar el uso de funciones ejecutivas como la planificación de metas a corto y largo plazo (Benavides-Nieto et al., 2017; Ganesalingam et al., 2011; Roebbers et al., 2012), impulsar habilidades metacognitivas como la autorreflexión (Roebbers et al., 2012; Jho y Chae, 2014), promover técnicas de aprendizaje cooperativo, como 1-2-4 o Jigsaw (Fernández-Río et al., 2017; Namaziandost et al., 2019), y usar estrategias de aprendizaje que ayuden a organizar la información y vincularla con conocimientos previos, como aprendizaje distribuido, recuperación activa de información (*practice testing*), práctica intercalada, resúmenes, mapas conceptuales o subrayado de información relevante (Donoghue y Hattie, 2021; Maldonado-Sánchez et al., 2019).

Estas estrategias resultan especialmente relevantes en entornos de aprendizaje en línea y educación a distancia, ya que pueden adaptarse fácilmente a plataformas digitales. Por ejemplo, los sistemas de aprendizaje en línea pueden integrar funciones que proporcionen retroalimentación automática e instantánea, los módulos de los cursos pueden diseñarse con una estructura escalonada que fomente la autonomía, y los proyectos grupales virtuales pueden facilitar el aprendizaje cooperativo, ayudando a mitigar la sensación de aislamiento que puede favorecer la amotivación y la dependencia a herramientas externas. En este contexto, mediante estrategias digitales de aprendizaje y promoción de la autorregulación, los educadores pueden guiar a los estudiantes a percibir la IA como un apoyo complementario y no como un reemplazo de sus propias capacidades (Darvishi et al., 2024).

De manera similar, autores como Kachgal et al. (2001) proponen estrategias efectivas para reducir la procrastinación, tales como fomentar la autoconciencia

de cómo los estudiantes gestionan su procrastinación, estimular la reflexión sobre los beneficios obtenidos al evitar tareas, realizar evaluaciones formales que permitan a los estudiantes comprender sus niveles de procrastinación (por ejemplo, utilizando la escala PASS), promover estrategias de autorregulación con impacto en la gestión del tiempo (como preparar calendarios que permitan hacer seguimiento del progreso) y enseñar estrategias de manejo del estrés. Estas recomendaciones abordan directamente el hallazgo de que la procrastinación es un mediador clave en la relación entre amotivación y dependencia a ChatGPT, sugiriendo que atacar la procrastinación es esencial para reducir esta forma de dependencia.

En entornos *e-learning*, estas estrategias pueden implementarse mediante herramientas digitales, como aplicaciones de gestión del tiempo, cuestionarios integrados para autoevaluación y foros de discusión entre pares sobre manejo del estrés académico, proporcionando a los estudiantes la estructura y el apoyo necesarios para combatir la procrastinación y, con ello, disminuir su dependencia a la IA generativa.

Finalmente, este estudio presenta varias limitaciones que deben considerarse al interpretar los resultados. En primer lugar, la muestra proviene íntegramente del ámbito de las ciencias sociales, con la mayoría de los participantes matriculados en el programa de Educación Primaria, mientras que una proporción menor corresponde al programa de Educación Infantil o al Máster en Formación del Profesorado de Educación Secundaria. Este desequilibrio, junto con el uso de un método de muestreo por conveniencia, puede haber sesgado parcialmente los resultados y limita la generalización de los hallazgos. En esta línea, estudios futuros podrían realizar análisis comparativos entre diferentes programas de grado y universidades, lo que permitiría examinar posibles variaciones en el uso educativo de la IA. Replicar estos hallazgos en estudiantes de otras disciplinas, como ciencias experimentales o de la salud, sería valioso y podría confirmar la robustez de los resultados en distintos contextos educativos. Además, el diseño transversal restringe las inferencias causales, subrayando la necesidad de estudios longitudinales para fortalecer las conclusiones. Es por ello que los diseños que combinen enfoques longitudinales y métodos mixtos podrían explorar con mayor profundidad la evolución del uso educativo de la IA a lo largo del tiempo. Asimismo, los datos también se recopilaron mediante autoinformes, lo que pudo introducir sesgos como la varianza común de método y la deseabilidad social, afectando potencialmente a las relaciones observadas. Finalmente, aunque el estudio actual evaluó un único modelo de mediación teórico, investigaciones futuras podrían aportar información más precisa sobre las condiciones en que el modelo propuesto opera de manera más efectiva mediante la incorporación de posibles variables moderadoras. Por ejemplo, la falta de confianza en las propias habilidades podría llevar a que estudiantes con baja autoeficacia académica perciban la IA como un recurso que proporciona respuestas inmediatas, reduciendo así la necesidad de enfrentar desafíos de forma autónoma. Este comportamiento, especialmente bajo presión académica o estrés, podría reforzar su dependencia a la tecnología (Rodríguez

et al., 2024). Por ende, variables como la autoeficacia académica y la carga de trabajo podrían servir como factores de control o moderadores en futuros modelos, ayudando a explicar las condiciones específicas bajo las cuales los estudiantes son más propensos a desarrollar dependencia a la IA. No obstante, a pesar de estas limitaciones, este trabajo proporciona una base sólida para futuras investigaciones que exploren el impacto de variables emocionales en la dependencia a la inteligencia artificial generativa, ejemplificada por ChatGPT.

CONCLUSIÓN

Este estudio demuestra que una alta motivación intrínseca y una baja amotivación reducen la procrastinación de los estudiantes, lo que a su vez disminuye la dependencia a ChatGPT. Estos hallazgos sugieren que fomentar las necesidades psicológicas básicas de los estudiantes e implementar estrategias para mejorar la motivación y la autorregulación podría ayudar a mitigar la dependencia a la IA generativa en contextos educativos. Con base en estos hallazgos, investigaciones futuras podrían replicar estos resultados en distintas disciplinas, empleando diseños longitudinales, reemplazando las medidas de autoinforme y explorando factores adicionales moderadores o mediadores que influyan en la dependencia de los estudiantes a las herramientas basadas en la IA.

REFERENCIAS

- Ahmad, S. F., Han, H., Alam, M. M., Rehmat, M. K., Irshad, M., Arraño-Muñoz, M. y Ariza-Montes, A. (2023). Impact of artificial intelligence on human loss in decision making, laziness and safety in education. *Humanities & Social Sciences Communications*, 10(1), 311. <https://doi.org/10.1057/s41599-023-01842-4>
- Ahmad, S. F., Rahmat, M. K., Mubarik, M. S., Alam, M. M. y Hyder, S. I. (2021). Artificial intelligence and its role in education. *Sustainability*, 13(22), 12902. <https://doi.org/10.3390/su132212902>
- Ali, K., García, A. y Vadsariya, A. (2024). Impact of the AI dependency revolution on both physical and mental health. *Journal of Strategic Innovation and Sustainability*, 19(2), 70-81. <https://doi.org/10.33423/jsis.v19i2.7006>
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5^a ed.). American Psychiatric Publishing.
- Ayele, N. (2024). *The usage of new AI technologies by college students and its influence on learning and dependence* [Manuscrito, Caldwell University Archives]. JSTOR Community Collections. <https://www.jstor.org/stable/community.37312365>
- Baird, A. y Maruping, L. M. (2021). The next generation of research on IS use: A theoretical framework of delegation to and from agentic IS artifacts. *MIS Quarterly*, 45(1), 315-341. <https://doi.org/10.25300/MISQ/2021/15882>
- Benavides-Nieto, A., Romero-López, M., Quesada-Conde, A. B. y Alba, G. (2017). Basic executive functions in early childhood education and their relationship with social competence. *Procedia - Social and Behavioral*

- Sciences, 237, 471-478. <https://doi.org/10.1016/j.sbspro.2017.02.092>
- Bolden, J. y Fillauer, J. P. (2020). "Tomorrow is the busiest day of the week": Executive functions mediate the relation between procrastination and attention problems. *Journal of American College Health*, 68(8), 854-863. <https://doi.org/10.1080/07448481.2019.1626399>
- Bucur, A. M., Cosma, A. y Dinu, L. P. (2021). *Early risk detection of pathological gambling, self-harm and depression using BERT* [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2106.16175>
- Carroll, N., Lang, M. y Connolly, C. (2025). An extended community of inquiry framework supporting students in online and digital education. *Innovations in Education and Teaching International*, 62(2), 369-385. <https://doi.org/10.1080/14703297.2024.2326658>
- Chai, C. S., Chiu, T. K. F., Wang, X., Jiang, F. y Lin, X.-F. (2023). Modeling Chinese secondary school students' behavioral intentions to learn artificial intelligence with the theory of planned behavior and self-determination theory. *Sustainability*, 15, 605. <https://doi.org/10.3390/su15010605>
- Chen, N.-S., Yin, C., Isaias, P. y Psotka, J. (2020). Educational big data: Extracting meaning from data for smart education. *Interactive Learning Environments*, 28(2), 142-147. <https://doi.org/10.1080/10494820.2019.1635395>
- Cheon, S. H., Reeve, J. y Vansteenkiste, M. (2020). When teachers learn how to provide classroom structure in an autonomy-supportive way: Benefits to teachers and their students. *Teaching and Teacher Education*, 90, 103004. <https://doi.org/10.1016/j.tate.2019.103004>
- Chiu, T. K. F. (2024). A classification tool to foster self-regulated learning with generative artificial intelligence by applying self-determination theory: A case of ChatGPT. *Educational Technology Research & Development*. <https://doi.org/10.1007/s11423-024-10366-w>
- Chiu, T. K. F. y Chai, C.-S. (2020). Sustainable curriculum planning for artificial intelligence education: A self-determination theory perspective. *Sustainability*, 12(14), 5568. <https://doi.org/10.3390/su12145568>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2^a ed.). Lawrence Erlbaum Associates.
- Coppock, A., Leeper, T. J. y Mullinix, K. J. (2018). Generalizability of heterogeneous treatment effect estimates across samples. *Proceedings of the National Academy of Sciences*, 115(49), 12441-12446. <https://doi.org/10.1073/pnas.1808083115>
- Darvishi, A., Khosravi, H., Sadiq, S., Gašević, D. y Siemens, G. (2024). Impact of AI assistance on student agency. *Computers & Education*, 210, 104967. <https://doi.org/10.1016/j.compedu.2023.104967>
- Del Cisne, M., Romero, J., Sancho, D. y Romero, A. (2024). Consecuencias de la dependencia a la inteligencia artificial en habilidades críticas y aprendizaje autónomo en los estudiantes. *Ciencia Latina Revista Científica Multidisciplinar*, 8(2), 2368-2382. https://doi.org/10.37811/cl_rcm.v8i2.10678
- Delgado, N., Campo Carrasco, L., Sainz de la Maza, M. y Etxabe-Urbietta, J. M. (2024). Application of artificial intelligence (AI) in education: Benefits and limitations of AI as perceived by primary, secondary, and higher education teachers. *Revista Electrónica Interuniversitaria de Formación del Profesorado*, 27(1), 207-224. <https://doi.org/10.6018/reifop.577211>
- Donoghue, G. M. y Hattie, J. A. C. (2021). A meta-analysis of ten learning techniques. *Frontiers in Education*, 6, 581216. <https://doi.org/10.3389/educ.2021.581216>
- Duru, E. y Balkis, M. (2017). Procrastination, self-esteem, academic performance,

- and well-being: A moderated mediation model. *International Journal of Educational Psychology*, 6(2), 97-119. <https://doi.org/10.17583/ijep.2017.2584>
- Estrada-Araoz, E. G., Mamani-Roque, M., Quispe-Aquise, J., Manrique-Jaramillo, Y. V. y Cruz-Laricano, E. O. (2025). Academic self-efficacy and dependence on artificial intelligence in a sample of university students. *Sapienza: International Journal of Interdisciplinary Studies*, 6(1), e25008. <https://doi.org/10.51798/sijis.v6i1.916>
- Feng, M., Xia, A. y Xia, X. (2023). The association between ChatGPT usage and college students' online learning burnout: The mediating role of self-control. En X. Chen y Q. Liu (Eds.), *Proceedings of the International Conference of Educational Innovation through Technology* (pp. 209-212). IEEE. <https://doi.org/10.1109/EITT61659.2023.00046>
- Fernández-Río, J., Sanz, N., Fernández-Cando, J. y Santos, L. (2017). Impact of a sustained cooperative learning intervention on student motivation. *Physical Education and Sport Pedagogy*, 22(1), 89-105. <https://doi.org/10.1080/17408989.2015.1123238>
- Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4ª ed.). Sage.
- Filgona, J., Sakiyo, J., Gwany, D. M. y Okoronka, A. U. (2020). Motivation in learning. *Asian Journal of Education and Social Studies*, 10(4), 16-37. <https://doi.org/10.9734/ajess/2020/v10i430273>
- Ganesalingam, K., Yeates, K. O., Taylor, H. G., Chertkoff, N., Stancin, T. y Wade, S. (2011). Executive functions and social competence in young children 6 months following traumatic brain injury. *Neuropsychology*, 25(4), 466-476. <https://doi.org/10.1037/a0022768>
- Ganotice, F. A., Mendoza, N. B., Dizon, J. I. W. T., Shen, X., Lee, J. C.-Y., Chan, E., Luk, P., Manio, M. M., He, Q., Khoo, U. S., Lam, M. P. S., Chan, S. C., Chow, A. Y. M., Wang, N. y Tipoe, G. L. (2024). Students' motivation and engagement in interprofessional education: The mediating role of peer relatedness. *Medical Education Online*, 29(1), 2430593. <https://doi.org/10.1080/10872981.2024.2430593>
- García, J. A., Carcedo, R. J. y Castaño, J. L. (2019). The influence of feedback on competence, motivation, vitality, and performance in a throwing task. *Research Quarterly for Exercise and Sport*, 90(2), 172-179. <https://doi.org/10.1080/02701367.2019.1571677>
- Garon-Carrier, G., Boivin, M., Guay, F., Kovas, Y., Dionne, G., Lemelin, J. P., Séguin, J. R., Vitaro, F. y Tremblay, R. E. (2016). Intrinsic motivation and achievement in mathematics in elementary school: A longitudinal investigation of their association. *Child Development*, 87(1), 165-175. <https://doi.org/10.1111/cdev.12458>
- Grund, A. y Fries, S. (2018). Understanding procrastination: A motivational approach. *Personality and Individual Differences*, 121, 120-130. <https://doi.org/10.1016/j.paid.2017.09.035>
- Guay, F., Stupnisky, R., Boivin, M., Japel, C. y Dionne, G. (2019). Teachers' relatedness with students as a predictor of students' intrinsic motivation, self-concept, and reading achievement. *Early Childhood Research Quarterly*, 48(3), 215-225. <https://doi.org/10.1016/j.ecresq.2019.03.005>
- Hair, J. F., Black, W. C., Babin, B. J. y Anderson, R. E. (2010). *Multivariate data analysis* (7ª ed.). Pearson Education International.
- Hen, M. y Goroshit, M. (2018). General and life-domain procrastination in highly educated adults in Israel. *Frontiers in Psychology*, 9, 1173. <https://doi.org/10.3389/fpsyg.2018.01173>
- Hong, J.-C. y Chen, N. (2024). *Effects of achievement motivation and*

- ChatGPT self-efficacy on ChatGPT addiction and ChatGPT continuance intention* [Preprint]. SSRN. <https://doi.org/10.2139/ssrn.5064583>
- Hooper, D., Coughlan, J. y Mullen, M. R. (2008). Structural equation modeling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods*, 6(1), 53-60.
- Huang, S., Lai, X., Ke, L., Li, Y., Wang, H., Zhao, X., Dai, X. y Wang, Y. (2024). AI technology panic—Is AI dependence bad for mental health? A cross-lagged panel model and the mediating roles of motivations for AI use among adolescents. *Psychology Research and Behavior Management*, 17, 1087-1102. <https://doi.org/10.2147/PRBM.S440889>
- Jara, I. y Ochoa, J. M. (2020). Usos y efectos de la inteligencia artificial en educación. *Banco Interamericano de Desarrollo*. <https://doi.org/10.18235/0002380>
- Jho, M. Y. y Chae, M. O. (2014). Impact of self-directed learning ability and metacognition on clinical competence among nursing students. *Journal of Korean Academic Society of Nursing Education*, 20(4), 513-522. <https://doi.org/10.5977/jkasne.2014.20.4.513>
- Johnson, J. L. y Bloom, A. M. (1995). An analysis of the contribution of the five factors of personality to variance in academic procrastination. *Personality and Individual Differences*, 18(1), 127-133. [https://doi.org/10.1016/0191-8869\(94\)00109-6](https://doi.org/10.1016/0191-8869(94)00109-6)
- Kachgal, M. M., Hansen, L. S. y Nutter, K. J. (2001). Academic procrastination prevention/intervention: Strategies and recommendations. *Journal of Developmental Education*, 25(1), 14-24. <https://www.jstor.org/stable/42775842>
- Kamalov, F., Santandreu Calonge, D. y Gurrib, I. (2023). New era of artificial intelligence in education: Towards a sustainable multifaceted revolution. *Sustainability*, 15(16), 12451. <https://doi.org/10.3390/su151612451>
- Klingbeil, A., Grützner, C. y Schreck, P. (2024). Trust and reliance on AI—An experimental study on the extent and costs of overreliance on AI. *Computers in Human Behavior*, 160, 108352. <https://doi.org/10.1016/j.chb.2024.108352>
- Klingsieck, K. B. (2013). Procrastination: When good things don't come to those who wait. *European Psychologist*, 18, 24-34. <https://doi.org/10.1027/1016-9040/a000138>
- Koike, D. A. y Pearson, L. (2005). The effect of instruction and feedback in the development of pragmatic competence. *System*, 33(3), 481-501. <https://doi.org/10.1016/j.system.2005.06.008>
- Kok, J. K. (2016). The relationships between procrastination and motivational aspects of self-regulation. *Jurnal Psikologi Malaysia*, 30(1), 30-39.
- Korteling, J. E., van de Boer-Visschedijk, G. C., Blankendaal, R. A. M., Boonekamp, R. C. y Eikelboom, A. R. (2021). Human-versus artificial intelligence. *Frontiers in Artificial Intelligence*, 4, 622364. <https://doi.org/10.3389/frai.2021.622364>
- Kotera, Y., Conway, E. y Green, P. (2023). Construction and factorial validation of a short version of the Academic Motivation Scale. *British Journal of Guidance & Counselling*, 51(2), 274-283. <https://doi.org/10.1080/03069885.2021.1903387>
- Lee, D., Arnold, M., Srivastava, A., Plastow, K., Strelan, P., Ploeckl, F., Lekkas, D. y Palmer, E. (2024). The impact of generative AI on higher education learning and teaching: A study of educators' perspectives. *Computers and Education: Artificial Intelligence*, 6, 100221. <https://doi.org/10.1016/j.caeai.2024.100221>
- Legood, A., Lee, A., Schwarz, G. y Newman, A. (2018). From self-defeating to other defeating: Examining the effects of leader procrastination on follower work outcomes. *Journal of Occupational and*

- Organizational Psychology*, 91, 430-439. <https://doi.org/10.1111/joop.12205>
- Li, Y., Zhang, L., Zhang, R., Xu, T. y Feng, T. (2021). The neural basis linking achievement motivation with procrastination: Left precuneus connectivity with right anterior cingulate cortex. *Personality and Social Psychology Bulletin*, 48, 1382-1392. <https://doi.org/10.1177/01461672211040677>
- Liu, M., Ren, Y., Nyagoga, L. M., Stonier, F., Wu, Z. y Yu, L. (2023). Future of education in the era of generative artificial intelligence: Consensus among Chinese scholars on applications of ChatGPT in schools. *Future in Educational Research*, 1(1), 72-101. <https://doi.org/10.1002/fer3.10>
- Ljubin-Golub, T., Rijavec, M. y Olčar, D. (2020). Student flow and burnout: The role of teacher autonomy support and student autonomous motivation. *Psychological Studies*, 65, 145-156. <https://doi.org/10.1007/s12646-019-00539-6>
- MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G. y Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods*, 7(1), 83-104. <https://doi.org/10.1037/1082-989X.7.1.83>
- Maldonado-Sánchez, M., Aguinaga-Villegas, D., Nieto-Gamboa, J., Fonseca-Arellano, F., Shardin-Flores, L. y Cadenillas-Albornoz, V. (2019). Learning strategies for the development of the autonomy of secondary school students. *Propósitos y Representaciones*, 7(2), 415-439. <https://doi.org/10.20511/pyr2019.v7n2.290>
- Morales-García, W. C., Sairitupa-Sánchez, L. Z., Morales-García, S. B. y Morales-García, M. (2024). Development and validation of a scale for dependence on artificial intelligence in university students. *Frontiers in Education*, 9, 1323898. <https://doi.org/10.3389/feduc.2024.1323898>
- Morandín-Ahuérma, F. (2022). What is artificial intelligence? *International Journal of Research Publication and Reviews*, 3(12), 1947-1951. <https://doi.org/10.55248/gengpi.2022.31261>
- Mulatsih, S. (2011). The use of scaffolding technique to improve the students' competence in writing genre-based texts. *Parole: Journal of Linguistics and Education*, 2(1), 101-109.
- Namaziandost, E., Shatalebi, V. y Nasri, M. (2019). The impact of cooperative learning on developing speaking ability and motivation toward learning English. *Journal of Language and Education*, 5(3), 83-101. <https://doi.org/10.17323/jle.2019.9809>
- Naseer, A., Rafaqat, N. y Amjad, M. (2025). Psychological impacts of AI dependence: Assessing the cognitive and emotional costs of intelligent systems in daily life. *Review of Applied Management and Social Sciences*, 8(1), 291-307. <https://doi.org/10.47067/ramss.v8i1.458>
- Neitzel, C. y Stright, A. D. (2003). Mothers' scaffolding of children's problem solving: Establishing a foundation of academic self-regulatory competence. *Journal of Family Psychology*, 17(1), 147-159. <https://doi.org/10.1037/0893-3200.17.1.147>
- Paraso, Y. M., Sedon, S. B. y Mahilum, D. Z. E. (2024). Student's artificial intelligence (AI) dependency: The lived experience of STEM students at Tongantongan National High School. *International Journal of All Research Writings*, 6(4), 34-43.
- Reeve, J. y Cheon, S. H. (2021). Autonomy-supportive teaching: Its malleability, benefits, and potential to improve educational practice. *Educational Psychologist*, 56(1), 54-77. <https://doi.org/10.1080/00461520.2020.1862657>
- Rodríguez, J., Marín, I. y Espejo, R. (2024). Is artificial intelligence use related to

- self-control, self-esteem, and self-efficacy among university students? *Education and Information Technologies*, 30, 2507-2524. <https://doi.org/10.1007/s10639-024-12906-6>
- Roebbers, C. M., Cimeli, P., Röthlisberger, M. y Neunschwander, R. (2012). Executive functioning, metacognition, and self-perceived competence in elementary school children: An explorative study on their interrelations and their role for school achievement. *Metacognition and Learning*, 7, 151-173. <https://doi.org/10.1007/s11409-012-9089-9>
- Ryan, R. M. y Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Contemporary Educational Psychology*, 61, 101860. <https://doi.org/10.1016/j.cedpsych.2020.101860>
- Siste, K., Assandi, P., Irawati, R., Setiawan, M. W. y Wiguna, T. (2019). Internet addiction: A new addiction? *Medical Journal of Indonesia*, 28(1), 82-91. <https://doi.org/10.13181/mji.v28i1.2752>
- Stojanov, A. (2024). Achievement goal motivation and reliance on ChatGPT for learning. En C. Pracana y M. Wang (Eds.), *Psychological Applications and Trends* (pp. 547-551). inScience Press. <https://doi.org/10.36315/2024inpact127>
- Swargiary, K. (2024). *The impact of ChatGPT on student learning outcomes: A comparative study of cognitive engagement, procrastination, and academic performance* [Preprint]. SSRN. <https://doi.org/10.2139/ssrn.4914743>
- Tavakol, M. y Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53-55. <https://doi.org/10.5116/ijme.4dfb.8dfd>
- Van Bergen, E., Hart, S. A., Latvala, A., Vuoksima, E., Tolvanen, A. y Torppa, M. (2022). Literacy skills seem to fuel literacy enjoyment, rather than vice versa. *Developmental Science*, 26(3), e13325. <https://doi.org/10.1111/desc.13325>
- Venkatesh, V., Morris, M. G., Davis, G. B. y Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3), 425-478. <https://doi.org/10.2307/30036540>
- Wang, Z., Liu, J. y Dong, R. (2018). Intelligent auto-grading system. En *Proceedings of the 5th IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS)* (pp. 430-435). IEEE. <https://doi.org/10.1109/CCIS.2018.8691244>
- Wu, F. y Fan, W. (2016). Academic procrastination in linking motivation and achievement-related behaviours: A perspective of expectancy-value theory. *Educational Psychology*, 37(6), 1-17. <https://doi.org/10.1080/01443410.2016.1202901>
- Wypych, M., Matuszewski, J. y Dragan, W. Ł. (2018). Roles of impulsivity, motivation, and emotion regulation in procrastination—Path analysis and comparison between students and non-students. *Frontiers in Psychology*, 9, 891. <https://doi.org/10.3389/fpsyg.2018.00891>
- Xia, Q., Chiu, T. K. F., Lee, M., Sanusi, I. T., Dai, Y. y Chai, C. S. (2022). A self-determination theory (SDT) design approach for inclusive and diverse artificial intelligence (AI) education. *Computers & Education*, 189, 104582. <https://doi.org/10.1016/j.compedu.2022.104582>
- Yan, B. y Zhang, X. (2022). What research has been conducted on procrastination? Evidence from a systematic bibliometric analysis. *Frontiers in Psychology*, 13, 809044. <https://doi.org/10.3389/fpsyg.2022.809044>
- Ye, J. H., Zhang, M., Nong, W., Wang, L. y Yang, X. (2024). The relationship between inert thinking and ChatGPT dependence: An I-PACE model perspective. *Education*

- and Information Technologies. <https://doi.org/10.1007/s10639-024-12966-8>
- Yockey, R. D. (2016). Validation of the short form of the Academic Procrastination Scale. *Psychological Reports*, 118(1), 171-179. <https://doi.org/10.1177/0033294115626825>
- Yu, S.-C., Chen, H.-R. y Yang, Y.-W. (2024). Development and validation the Problematic ChatGPT Use Scale: A preliminary report. *Current Psychology*, 43, 26080–26092. <https://doi.org/10.1007/s12144-024-06259-z>
- Zhai, C., Wibowo, S. y Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. *Smart Learning Environments*, 11, 28. <https://doi.org/10.1186/s40561-024-00316-7>

Fecha de recepción del artículo: 1 de junio de 2025

Fecha de aceptación del artículo: 9 de septiembre de 2025

Fecha de aprobación para maquetación: 2 de octubre de 2025

Fecha de publicación en OnlineFirst: 17 de octubre de 2025

Fecha de publicación: 1 de enero de 2026

Retroalimentación de aprendizajes con inteligencia artificial generativa en estudiantes universitarios

Feedback on learning with generative artificial intelligence in university students

-----  -----

📖 María Verónica Leiva-Guerrero - Pontificia Universidad Católica de Valparaíso, PUCV (Chile)
📖 Ignacio Araya Zamorano - Pontificia Universidad Católica de Valparaíso, PUCV (Chile)
📖 Rafael Escobar Collins - Pontificia Universidad Católica de Valparaíso, PUCV (Chile)
📖 Francisca Silva Castro - Pontificia Universidad Católica de Valparaíso, PUCV (Chile)

RESUMEN

La evaluación para el aprendizaje ha adquirido creciente relevancia en la docencia universitaria, especialmente el proceso de retroalimentación. Sin embargo, persiste una percepción de insatisfacción en estudiantes sobre la calidad de retroalimentación brindada por el profesorado, lo que evidencia la necesidad de innovar en sus estrategias. Esta investigación exploró la pertinencia pedagógica y tecnológica de integrar la Escalera de Retroalimentación de Wilson con inteligencia artificial generativa, específicamente GPT-4o, para fortalecer la retroalimentación formativa en estudiantes universitarios. El estudio se desarrolló bajo un enfoque cualitativo y exploratorio en dos fases. En primer lugar, se diseñó y validó un prompt mediante el método Delphi con la participación de ocho expertos en evaluación e inteligencia artificial, aplicándolo a siete modelos de lenguaje de última generación. En la segunda fase, el prompt validado se implementó en dos cursos universitarios de distinta naturaleza, Evaluación para el Aprendizaje y Estructura de Datos, integrando la retroalimentación automática en la plataforma Moodle. Los resultados mostraron que los expertos coincidieron en la idoneidad de la Escalera de Wilson mediada por IA y destacaron el desempeño superior de GPT-4o. A nivel de aula, los estudiantes valoraron la claridad, utilidad e inmediatez de la retroalimentación, aunque identificaron limitaciones en la falta de contextualización y tono impersonal de la herramienta. Se concluye que la integración de la Escalera de Wilson con inteligencia artificial generativa representa una innovación prometedora, pero que requiere ajustes disciplinares, supervisión docente y resguardo de la dimensión humana en los procesos de retroalimentación en entornos abiertos y a distancia.

Palabras clave: retroalimentación; evaluación formativa; inteligencia artificial generativa; ChatGPT; Escalera de Wilson; estudiantes universitarios.

ABSTRACT

Assessment for learning has become increasingly important in university teaching, particularly regarding the feedback process. However, there is still a perception of student dissatisfaction with the quality of feedback provided by faculty, highlighting the need to innovate in feedback strategies. This study aimed to explore the pedagogical and technological relevance of integrating Wilson's Feedback Ladder with generative artificial intelligence, specifically GPT-4o, to strengthen formative feedback in university students. The study was conducted using a qualitative and exploratory approach in two phases. First, a prompt was designed and validated using the Delphi method with the participation of eight experts in assessment and artificial intelligence, applying it to seven state-of-the-art language models. In the second phase, the validated prompt was implemented in two university courses of different nature, Assessment for Learning and Data Structures, integrating automatic feedback into the Moodle platform. The results showed that the experts agreed on the suitability of AI-mediated Wilson's Ladder and highlighted the superior performance of GPT-4o. At the classroom level, students valued the clarity, usefulness, and immediacy of the feedback, although they identified limitations in the tool's lack of contextualization and impersonal tone. It is concluded that the integration of Wilson's Ladder with generative artificial intelligence represents a promising innovation, but one that requires disciplinary adjustments, teacher supervision, and careful attention to the human dimension of feedback in e-learning contexts.

Keywords: feedback; formative evaluation; generative artificial intelligence; ChatGPT; Ladder of Feedback; university students.

INTRODUCCIÓN

La evaluación para el aprendizaje con un enfoque formativo ha adquirido creciente relevancia a nivel global, tanto en contextos escolares como universitarios, al centrarse en el acompañamiento y mejora continua del proceso educativo (Black y Wiliam, 2018; Andrade y Brookhart, 2019). En este marco, la retroalimentación se posiciona como un componente clave, al utilizar evidencia del desempeño estudiantil para orientar la mejora de los aprendizajes (Carless y Boud, 2018).

Sin embargo, diversos estudios reportan una persistente insatisfacción del estudiantado universitario respecto a la calidad de la retroalimentación proporcionada por el profesorado, especialmente en lo que respecta a su oportunidad, especificidad y utilidad (Quezada y Salinas, 2021; Galindo-Domínguez et al., 2023). Esta situación ha sido evidenciada en investigaciones realizadas en países como Australia, Reino Unido, México y Chile, lo que sugiere una problemática de alcance internacional.

En varios sistemas educativos, las prácticas de retroalimentación no han evolucionado significativamente, lo que limita su impacto formativo. Estrategias como la Escalera de Retroalimentación de Wilson (2013) o el Panel de Retroalimentación (Booth et al., 2008) han sido propuestas como alternativas efectivas para mejorar los procesos evaluativos, aunque su incorporación en la práctica docente sigue siendo limitada.

Frente a este escenario, se vuelve necesario renovar las estrategias de evaluación universitaria, incorporando tecnologías digitales emergentes que puedan apoyar la retroalimentación formativa (Puertas y Cano, 2024). En particular, la Inteligencia Artificial Generativa (IAGen) ofrece oportunidades relevantes para transformar la retroalimentación en tareas de escritura y pensamiento complejo (Dai et al., 2023; Al-Azawei et al., 2023).

Entre las herramientas basadas en esta tecnología, destaca ChatGPT (OpenAI, 2024), que integra modelos como GPT-4o, GPT-4.1 o GPT-4.5 y se caracteriza por generar textos coherentes y contextualizados a partir de prompts diseñados con criterios pedagógicos. Aunque su incorporación al ámbito educativo es reciente, estudios preliminares evidencian un uso creciente en enseñanza, aprendizaje y evaluación, lo que abre nuevas líneas de investigación sobre su aplicabilidad y eficacia (Galindo-Domínguez et al., 2023; García-Peñalvo, 2023).

En este marco, el presente estudio se plantea tres preguntas centrales: ¿es adecuada la Escalera de Retroalimentación de Wilson para su aplicación mediante inteligencia artificial?, ¿qué herramienta de IAGen, programada con prompts según estos criterios, resulta más pertinente para retroalimentar aprendizajes universitarios?, y ¿cómo perciben los estudiantes la retroalimentación recibida a través de esta estrategia aplicada con GPT-4o?

Inteligencia artificial en educación

La integración de la inteligencia artificial (IA) en el campo educativo ha redefinido los marcos tradicionales de enseñanza, constituyéndose en un eje fundamental de la innovación pedagógica contemporánea. El rol docente, históricamente centrado en la transmisión de contenidos, ha experimentado una transición sustantiva hacia funciones de facilitación del aprendizaje, gracias a las potencialidades de la IA para automatizar procesos, analizar datos y personalizar trayectorias formativas (Bonales-Daimiel et al., 2025). Esta reconfiguración permite a los educadores concentrarse en tareas de mayor valor pedagógico, tales como la retroalimentación cualitativa y el acompañamiento diferenciado (García Peñalvo et al., 2024).

Además, la IA ha demostrado ser una aliada estratégica en la mejora de la experiencia educativa, tanto en contextos presenciales como virtuales, mediante la personalización de recursos, la evaluación automatizada, la retroalimentación inmediata y el monitoreo en tiempo real del desempeño estudiantil (López Regalado et al., 2024; Romero Alonso et al., 2025). La implementación de sistemas adaptativos y chatbots educativos ha permitido no solo optimizar la eficiencia docente, sino también enriquecer los procesos de aprendizaje con enfoques centrados en el estudiante.

Inteligencia Artificial Generativa

La inteligencia artificial generativa (IAGen) abarca técnicas y modelos diseñados para crear contenido nuevo (texto, imágenes, audio, video) que imita la distribución estadística de grandes conjuntos de datos. En el núcleo de muchas soluciones de IAGen están los transformers, arquitecturas de atención que, tras un pre-entrenamiento masivo, aprenden a prever la siguiente unidad (palabra, píxel, nota) y así generan salidas coherentes y contextualizadas (Vaswani et al., 2017; OpenAI, 2023). Gracias a algoritmos de ajuste fino y a la incorporación de múltiples modalidades de entrada y salida estos sistemas pueden adaptarse a tareas diversas desde redacción creativa hasta simulaciones conversacionales manteniendo fluidez y sentido en sus respuestas (Weng et al., 2024).

Desde 2024, el ecosistema de LLMs se ha diversificado con propuestas comerciales GPT-4o y GPT-4.5 (OpenAI, 2024), Claude 3.7 Sonnet (Anthropic, 2024) y Gemini 2.5 (Google DeepMind, 2024) y de código abierto LLaMa 4 (LLaMa, 2024), Mistral (Mistral AI, 2024). Además, algunos modelos, como DeepSeek R1 (DeepSeek, 2024), destacan por exponer su razonamiento interno frente a problemas complejos, y xAI Grok 3 (xAI, 2024) dedica fases de “reflexión” profundas para matemáticas y lógica avanzadas.

En educación, estas herramientas no solo sirven para automatizar contenidos, sino que ayudan a personalizar la enseñanza y la retroalimentación, adaptándose a cada estudiante y aliviando tareas repetitivas (corrección de trabajos), siempre que

se consideren aspectos éticos y metodológicos. Su integración responsable puede potenciar significativamente el aprendizaje (Weng et al., 2024).

Prompts y técnicas de prompting

En el contexto de la inteligencia artificial generativa, un prompt es la entrada textual o estructurada que se entrega a un modelo de lenguaje para guiar su respuesta. Su diseño no es trivial: la forma, el contenido y el formato del prompt influyen directamente en la calidad, precisión y utilidad de la salida generada. De ahí surge el campo del prompt engineering, dedicado a la elaboración de estrategias específicas para construir prompts efectivos según la tarea o el contexto educativo (Liu et al., 2023; Reynolds y McDonell, 2021).

Entre las principales técnicas de prompting se encuentran el zero-shot prompting, en el que el modelo responde únicamente a una instrucción directa sin ejemplos; y el few-shot prompting, que incorpora ejemplos representativos antes de la tarea para modelar el tipo de respuesta esperada (Brown et al., 2020). Estas técnicas permiten modular el comportamiento del modelo y anticipar mejor el tipo de contenido que se generará.

Otra técnica fundamental es chain-of-thought, que guía al modelo a razonar paso a paso antes de emitir una respuesta final, mejorando su desempeño en tareas complejas de lógica o evaluación estructurada (Wei et al., 2022). A esto se suma el role prompting, que asigna explícitamente un rol al modelo —como “evaluador experto” o “docente universitario”— para controlar el tono, la profundidad y el estilo de la retroalimentación (Sahoo et al., 2024). Finalmente, el structured prompting especifica el formato de salida requerido, como un diccionario o una tabla, lo que resulta especialmente útil cuando la información generada debe ser procesada automáticamente por sistemas externos, como plataformas educativas (Schulhoff et al., 2025).

En educación, las técnicas de prompting permiten estructurar retroalimentación automatizada de forma precisa y pedagógicamente alineada. El uso combinado de role prompting, chain-of-thought y structured prompting guía al modelo a razonar como un evaluador experto, generando respuestas organizadas y coherentes. Esto facilita su integración en plataformas educativas y potencia una retroalimentación más útil, trazable y centrada en el aprendizaje.

Retroalimentación formativa y su mediación con inteligencia artificial

La retroalimentación es un pilar del aprendizaje formativo, concebida no solo como corrección de errores, sino como un proceso interactivo que promueve reflexión, ajuste de estrategias y construcción activa del conocimiento. Winstone et al. (2022) la describen como una interacción dialógica entre docente y estudiante, o entre pares, que favorece la comprensión profunda siempre que sea oportuna,

específica y con metas claras. En la misma línea, Carless y Boud (2018) subrayan que debe entenderse como un proceso de influencia mutua y de agencia estudiantil, mientras que Molloy et al. (2020) destacan la necesidad de que los estudiantes desarrollen competencias para interpretarla y aplicarla en contextos autónomos.

No obstante, en la práctica persisten barreras estructurales. En Chile, el Ministerio de Educación (2019) advierte que, pese al reconocimiento teórico de la retroalimentación, las prácticas evaluativas siguen siendo mayoritariamente unidireccionales y correctivas, con poco énfasis en la metacognición y la autorregulación. Para responder a esta brecha, se requieren enfoques estructurados que potencien la dimensión formativa. En este marco, la Escalera de Retroalimentación de Wilson (Goodrich, 2011) ofrece una guía clara con cuatro niveles progresivos (Aclarar: formular preguntas que apunten a resolver ambigüedades o aspectos omitidos en el trabajo del estudiante, fomentando la comprensión conceptual; Valorar: destacar logros y avances, reconociendo de manera auténtica los aspectos positivos del desempeño; Expresar inquietudes: identificar dificultades observadas con un tono respetuoso y constructivo, evitando juicios descalificadores; y, Sugerir mejoras: ofrecer orientaciones específicas que orienten al estudiante hacia una comprensión más profunda y acciones de mejora viables) orientados al pensamiento crítico y la autorregulación. Su valor radica en transformar la retroalimentación en una experiencia empática, participativa y adaptable a diversos contextos educativos, presenciales o virtuales.

En este estudio, la Escalera de Wilson se adopta como base conceptual para diseñar retroalimentación automatizada mediada por inteligencia artificial. Desde esta perspectiva, la IA se plantea como recurso para personalizar y ampliar la cobertura, en línea con lo señalado por Wiliam (2011) y Shute y Rahimi (2017) respecto al potencial de las tecnologías emergentes en la evaluación para el aprendizaje.

Inteligencia artificial para la retroalimentación educativa

La incorporación de IA en los procesos de retroalimentación educativa responde a la necesidad de transformar prácticas tradicionales hacia modelos más interactivos, personalizados y efectivos. Diversas investigaciones han propuesto innovar en esta dimensión mediante el uso de IA, destacando su potencial para mejorar la calidad del aprendizaje y reducir la carga evaluativa docente (Puertas y Cano, 2024; Ossa y Willatt, 2023). El informe *Teaching with AI – Assessment, Feedback and Personalisation*, elaborado por European Commission (2023), profundiza en cómo la IA puede intervenir positivamente en los procesos de evaluación y retroalimentación desde cuatro niveles complementarios: social, institucional, docente y estudiantil.

En este contexto, la IA ofrece herramientas capaces de adaptar la retroalimentación a las necesidades individuales del estudiantado mediante la generación automatizada de comentarios, la evaluación en tiempo real y el seguimiento del progreso académico (Ayeni et al., 2024; Holmes et al., 2019). Este

tipo de apoyo permite a los docentes identificar con mayor precisión las dificultades específicas de cada estudiante y ofrecer una enseñanza más focalizada y diferenciada (Luckin y Holmes, 2016). Asimismo, la retroalimentación inmediata proporcionada por sistemas inteligentes incide positivamente en la motivación estudiantil, ya que entrega respuestas oportunas que facilitan la comprensión de errores y refuerzan aprendizajes claves.

En particular, las herramientas de chat basadas en modelos IAGen han cobrado relevancia como asistentes pedagógicos eficaces. Estos sistemas permiten automatizar la corrección y la retroalimentación a partir de criterios previamente establecidos, incrementando así la eficiencia y coherencia del proceso evaluativo (Carless y Winstone, 2023). Para el profesorado, esto significa una reducción en la carga operativa; para los estudiantes, una oportunidad de recibir orientación formativa inmediata, promoviendo la autoevaluación y el pensamiento crítico.

La efectividad de estas herramientas ha sido documentada en diversos estudios empíricos. Zhang et al. (2024) evaluaron el uso de ChatGPT para generar retroalimentación en tareas de programación con estudiantes de un curso introductorio de Ciencias de la Computación. La mayoría de los participantes valoró positivamente la claridad y utilidad de los comentarios generados por la IA, aunque algunos expresaron que habrían preferido una retroalimentación más rica en ejemplos específicos. Por su parte, Jauhiainen y Garagorry Guerra (2024) analizaron la aplicación de ChatGPT-4 en la evaluación de respuestas abiertas escritas por estudiantes universitarios. Utilizando una rúbrica con cinco criterios, relevancia, exactitud, exhaustividad, coherencia y corrección lingüística, observaron que el modelo generaba evaluaciones detalladas y consistentes, aunque insistieron en la necesidad de supervisión docente para garantizar su precisión final.

Complementariamente, Baral et al. (2024) compararon ChatGPT-4 con otros modelos (como LLaMa y SBERT-Canberra) en la corrección de respuestas matemáticas de nivel secundario. A partir de 500 respuestas evaluadas con una rúbrica común, encontraron que ChatGPT-4 alcanzó un 92 % de concordancia con las calificaciones de docentes humanos, además de generar retroalimentación cualitativa de alta calidad, evaluada mediante métricas lingüísticas y juicios de expertos. No obstante, los autores advierten que, aunque la IA puede aproximarse al criterio humano, su uso debe enmarcarse en un sistema que contemple revisión experta y reflexión pedagógica.

METODOLOGÍA

Diseño

El estudio adoptó un enfoque cualitativo de carácter exploratorio, orientado a diseñar, validar e implementar un prompt de retroalimentación formativa automatizada mediante inteligencia artificial generativa (IAGen), sustentado en la

Escalera de Retroalimentación de Wilson (2013). El proceso se desarrolló en dos fases.

En la Fase 1, se elaboró el prompt a partir de una revisión documental sobre estrategias de retroalimentación y se validó mediante el método Delphi (Landeta, 1999). El procedimiento incluyó dos rondas sucesivas de juicio experto. En la primera, se evaluó la adecuación del prompt y de los cuatro peldaños de la Escalera (aclarar, valorar, expresar inquietudes, sugerir), aplicados a siete modelos de lenguaje (LLM) representativos. En la segunda, se devolvió una síntesis de resultados y se solicitó re-juicio para convergencia. Se utilizó una escala Likert 0–3 anclada en desempeño y se definió como criterio de consenso una desviación intercuartílica ≤ 1 . Además, se calculó el coeficiente W de Kendall como medida de consistencia global (criterio de aceptabilidad $\geq 0,70$).

En la Fase 2, el prompt validado se integró en la plataforma Moodle¹ en dos asignaturas universitarias. Las respuestas abiertas de los estudiantes fueron procesadas automáticamente con GPT-4o para generar retroalimentación inmediata estructurada según la Escalera de Wilson. Finalmente, se recogieron percepciones estudiantiles mediante entrevistas semiestructuradas para complementar los datos técnicos con la experiencia del alumnado.

Procedimiento

En la primera fase, el objetivo fue asegurar que el prompt diseñado generara retroalimentación estructurada y coherente con los cuatro niveles de la Escalera de Wilson (2013) antes de ser aplicado con estudiantes. Este proceso incluyó la selección de modelos de lenguaje, la elaboración iterativa del prompt y su validación. Se eligieron siete modelos de lenguaje representativos por su diversidad técnica y accesibilidad: GPT-4o², GPT-4o-mini³ (OpenAI, 2024), Microsoft Copilot, Google Gemini (2024), Claude (Anthropic, 2024), Perplexity AI y LLaMA (2024), accedidos desde sus canales oficiales.

El prompt fue diseñado como plantilla estructurada con: (i) contexto teórico, (ii) enunciado de la pregunta, (iii) criterios de evaluación y/o respuesta esperada, (iv) respuesta del estudiante y (v) instrucciones explícitas para orientar la retroalimentación siguiendo los peldaños de Wilson. En la evaluación piloto se utilizaron dos actividades: una pregunta abierta basada en un caso de estudio y una tarea de elaboración de rúbrica. Las respuestas de cada modelo fueron analizadas y posteriormente evaluadas por un panel de ocho expertos —tres en evaluación, tres en IA y dos docentes universitarios— usando una escala de 0 a 3 y comentarios cualitativos, en tres rondas según el enfoque Delphi (Steurer, 2011).

La segunda fase tuvo como propósito implementar el prompt validado en dos asignaturas de pregrado: Evaluación para el aprendizaje y Estructura de Datos. La evaluación (prueba escrita) se realizó en Moodle mediante el plugin Coderunner⁴, automatizando el flujo entre respuestas estudiantiles y GPT-4o. Cada estudiante

respondió preguntas abiertas que, una vez enviadas, eran procesadas en tiempo real generando retroalimentación inmediata estructurada según la Escalera de Wilson, junto con un puntaje preliminar de 0 a 10 basado en rúbricas.

El estudiantado fue informado sobre la naturaleza experimental del sistema y se ofreció una opción de apelación, revisada manualmente por el equipo docente, como medida de equidad. Además, se realizó un análisis cualitativo posterior mediante entrevistas semiestructuradas a estudiantes voluntarios, indagando en la claridad, utilidad y confiabilidad de la retroalimentación automatizada. Las entrevistas fueron analizadas utilizando categorías deductivas (basadas en los peldaños de Wilson) y subcategorías emergentes Mayring (2000). ChatGPT asistió en la sistematización del análisis, apoyando la codificación temática y la extracción de patrones significativos en los testimonios estudiantiles.

Instrumentos

Se utilizaron instrumentos y técnicas que se detallan a continuación:

1. *Protocolo de Validación del Prompt*

Se elaboró un protocolo para validar el desempeño del prompt en siete modelos de lenguaje (LLM). Este contenía las retroalimentaciones generadas por cada modelo, organizadas según los cuatro peldaños de la Escalera de Wilson (2013): Aclarar, Valorar, Expresar inquietudes y Sugerir mejoras. El panel de expertas evaluó cada peldaño con escala ordinal (0–3) y emitió observaciones cualitativas cuando la puntuación fue igual o inferior a 1.

La validación se llevó a cabo mediante el método Delphi con un panel interdisciplinario de ocho expertos. El proceso contempló dos rondas sucesivas, utilizando una escala Likert 0–3 y criterios de consenso basados en desviación intercuartílica (≤ 1) y coeficiente de Kendall W. Esta dinámica permitió ajustar el instrumento antes de su aplicación, confirmando su coherencia con los peldaños de la Escalera de Wilson y respaldando la selección de GPT-4o como modelo preferente para la fase piloto.

2. *Prueba Escrita para Estudiantes (Evaluación Sumativa)*

Las evaluaciones aplicadas en ambas asignaturas incluyeron preguntas abiertas diseñadas para activar habilidades cognitivas de orden superior. En ‘Estructura de Datos’, se abordaron explicaciones, análisis de estructuras algorítmicas y justificaciones técnicas. En ‘Evaluación del y para el aprendizaje’, las preguntas demandaban respuestas argumentadas sobre conceptos pedagógicos a partir de

clases observadas. Las preguntas fueron acompañadas de rúbricas específicas, lo cual permitió aplicar el prompt de manera coherente.

3. *Prompt Estructurado*

El instrumento central del estudio fue un prompt diseñado conforme a principios de prompt engineering (Brown et al., 2020; Wei et al., 2022; Hao et al., 2022; Liu et al., 2023; Schulhoff et al., 2025), integrado por seis componentes: (i) Contexto teórico: sintetiza los fundamentos conceptuales y metodológicos que enmarcan la tarea, permitiendo situar la evaluación dentro de un marco disciplinar y pedagógico sólido. (ii) Enunciado de la pregunta: plantea la tarea evaluativa con claridad, asegurando coherencia con los objetivos de aprendizaje. (iii) Criterios de evaluación/respuesta esperada: establecen los parámetros de calidad y pertinencia de las respuestas, descritos mediante una rúbrica analítica con todos sus niveles de desempeño. (iv) Respuesta del estudiante: contiene la producción a evaluar. (v) Instrucciones para retroalimentar: orientan la formulación de juicios y comentarios siguiendo los peldaños de la Escalera de Wilson (valoraciones, aclaraciones, inquietudes, sugerencias), con límites de extensión y tono constructivo. (vi) Formato de salida en diccionario Python: organiza los resultados en los siguientes campos: análisis y puntaje por cada criterio de evaluación, más los cuatro peldaños de la Escalera de Wilson. Se escribió en español, empleando formato markdown para favorecer la legibilidad y se asignó el rol de “evaluador experto” para asegurar aspectos técnicos. Tras su validación, se decidió incluir en el punto (iii) la descripción completa de todos los niveles de la rúbrica analítica y agregar el punto (vi) para facilitar la automatización de la respuesta en la plataforma Moodle. La versión validada se adjunta como anexo para su replicabilidad. Es importante resaltar que los puntos (v) y (vi) pueden permanecer sin modificación, mientras que los apartados anteriores se deben adecuar a la evaluación correspondiente.

4. *Sistema Automatizado de Retroalimentación*

Se desarrolló un sistema integrado en Moodle mediante el plugin Coderunner, que conecta los datos con el modelo GPT-4o. Cada vez que una respuesta era enviada por un o una estudiante, se generaba un prompt estructurado que era procesado sincrónicamente, devolviendo retroalimentación inmediata. La salida era formateada automáticamente en Moodle junto al puntaje preliminar. Además, se almacenaron los datos para su posterior análisis. El modelo se configuró con una temperatura de 0.1, priorizando así respuestas más deterministas y consistentes con la rúbrica de evaluación establecida.

Técnica

Entrevista Semiestructurada

Con el fin de explorar la percepción estudiantil sobre el sistema, se diseñó una entrevista con preguntas abiertas organizadas en torno a seis dimensiones: los cuatro peldaños de Wilson, la usabilidad del sistema y el impacto percibido en el aprendizaje. Antes de su aplicación, cada participante firmó un consentimiento informado, resguardando los principios de voluntariedad, confidencialidad y anonimato. Las entrevistas fueron grabadas en audio con autorización de los estudiantes y transcritas íntegramente para su análisis. Se garantizó en todo momento que la información sería utilizada únicamente con fines académicos, cumpliendo las normas éticas institucionales vigentes.

El análisis se realizó mediante el enfoque de contenido cualitativo de Mayring (2000), combinando categorías deductivas (derivadas de los peldaños de la Escalera de Wilson y de las dimensiones complementarias de usabilidad, impacto en el aprendizaje y juicio crítico con recomendaciones) e inductivas, surgidas de la propia información. Dos codificadores trabajaron de manera independiente en la lectura, segmentación y codificación de las entrevistas, aplicando la matriz previamente consensuada. Posteriormente, revisaron conjuntamente los resultados para resolver discrepancias y consolidar la codificación definitiva, lo que permitió asegurar la fiabilidad intercodificador y dotar de mayor solidez a los hallazgos.

Las categorías se enriquecieron con subcategorías emergentes basadas en las experiencias de los participantes, diferenciando entre valoraciones favorables (utilidad, claridad, motivación, guía formativa) y problemáticas (rigidez, descontextualización, automatismo, falta de empatía). El proceso contó además con el apoyo de ChatGPT como herramienta metodológica complementaria, utilizada para organizar y contrastar fragmentos textuales, explorar patrones temáticos y verificar la coherencia del sistema categorial. Su uso se limitó a facilitar tareas operativas, sin sustituir el juicio analítico de los investigadores, y contribuyó a fortalecer la trazabilidad y consistencia del proceso. La Tabla 1 resume las categorías y subcategorías empleadas en el análisis.

Tabla 1
Categorías y subcategorías usadas para el análisis

Categoría	Valoraciones Favorables	Valoraciones Problemáticas
I. Experiencia general	• Valoración de la innovación. • Rapidez y precisión del sistema.	• Extrañeza ante evaluación automatizada. • Ausencia de vínculo humano.
II. Aclaraciones (Escalera)	• Señala omisiones útiles. • Ejemplos clarificadores.	• Correcciones innecesarias. • Sugerencias confusas.

Categoría	Valoraciones Favorables	Valoraciones Problemáticas
III. Valoración (Escalera)	• Refuerzo positivo genera confianza. • Estimula el aprendizaje.	• Impacto emocional limitado por ser de IA. • Tono mecánico o forzado.
IV. Inquietudes (Escalera)	• Críticas claras y constructivas. • Impulso a la mejora.	• Redundancia o irrelevancia en algunas críticas.
V. Sugerencias (Escalera)	• Claridad y aplicabilidad. • Mejora futura.	• Sugerencias fuera del contenido visto. • Poca pertinencia contextual.
VI. Usabilidad y confianza	• Confianza en criterios. técnicos estables.	• Falta de transparencia evaluativa. • Dudas sobre sensibilidad pedagógica.
VII. Impacto en el aprendizaje	• Mejora de comprensión. • Retroalimentación útil para el estudio.	• Exigencias percibidas como excesivas
VIII. Juicio crítico y recomendaciones	• Disposición a seguir usando la IA. • Aporte al desarrollo del sistema.	• Necesidad de personalización. • Críticas al ajuste de nivel.
IX. Autonomía y control percibido (emergente)	• Claridad sobre expectativas. • Retroalimentación como guía. • Ajuste estratégico a la IA(sin conflicto explícito).	• Evaluación rígida. • Penalización por estilo personal o creatividad. • Desmotivación por falta de reconocimiento a respuestas válidas alternativas.

Fuente: Elaboración propia.

RESULTADOS

A continuación, se presentan los principales resultados en torno a las preguntas de investigación: ¿Es adecuada la Escalera de Retroalimentación de Wilson para ser aplicada mediante herramientas de inteligencia artificial?; ¿Cuál de las herramientas de IAGen, con prompts diseñados según escalera Wilson, resulta más pertinente para retroalimentar aprendizajes en estudiantes universitarios?; y ¿Cómo percibe el estudiantado universitario la retroalimentación recibida a través de GPT-4o?

Escalera de Retroalimentación de Wilson con IA

El 100 % de las valoraciones del panel de expertos concordó en que la estrategia de la Escalera de Wilson para retroalimentar aprendizajes con el uso de IA era apropiada, ya que brinda al estudiante claridad, valor, reflexión y sugerencias para

impulsar su aprendizaje, fortaleciendo la relación entre profesor(a) y alumno(a) al estar claramente definidos los pasos de retroalimentación. Al respecto, los expertos señalaron:

Entrega un nivel de descripción y detalle, que no solo se centra en señalar aspectos positivos y negativos, sino que va más allá, ofreciendo más posibilidades para la reflexión, corrección y mejora de aprendizajes a través de los pasos aclarar, valorar, señalar inquietudes y entregar sugerencias. (Experto 6).

Es interesante y apropiada la estrategia de retroalimentación Escala de Wilson, proporciona un marco para que el *feedback* sea empático, descriptivo y con foco en el aporte para la construcción de aprendizajes, en lugar de la crítica. (Experto 2).

Herramientas de IA más apropiadas de acuerdo con la Escalera de Retroalimentación de Wilson

La Tabla 2 presenta los promedios de valoraciones entregadas por expertos sobre retroalimentaciones generadas por distintas herramientas de IA, evaluadas en una escala de 0 (muy bajo) a 3 (muy alto) según los cuatro peldaños de la Escalera de Wilson. Cada herramienta fue evaluada en términos de claridad, valoración, expresión de inquietudes y sugerencias, incluyendo además su media y desviación estándar. De acuerdo con estos datos, GPT-4o destaca como la herramienta con mejor desempeño general, al obtener puntajes consistentemente altos en todos los criterios y una media de 2,87, junto con una baja desviación estándar (0,18), lo que refleja tanto efectividad como estabilidad en la calidad de sus retroalimentaciones.

Tabla 2
Promedios de valoraciones expertos de retroalimentaciones con IA

Herramienta IA	Promedio Retroalimentación Escalera de Wilson Ejemplos 1 y 2				Media	Desviación estándar
	Aclarar	Valorar	Expresar inquietudes	Ofrecer Sugerencias		
GPT-4o	2,9	2,75	2,9	2,95	2,87	0,18
GPT-4o mini	2,2	2,1	2,6	2,35	2,30	0,35
Microsoft Copilot	1,55	1,6	1,75	1,95	1,70	0,49
Gemini	1,65	1,45	1,7	1,95	1,67	0,46
Claude	1,6	1,5	1,1	1,35	1,36	0,62
Perplexity	2,45	1,95	2,3	2,5	2,28	0,25
LlaMa 3.2	2	1,9	2	2,3	2,04	0,27

Fuente: Elaboración propia.

La concordancia entre los ocho jueces al evaluar las siete herramientas (GPT-4, GPT Gratuito, Copilot, Gemini, Claude, Perplexity y LLaMa 3.2) fue medida mediante el coeficiente W de Kendall, que alcanzó un valor de 0,43, lo cual indica un nivel moderado de acuerdo entre los jueces. La prueba de significancia asociada arrojó un valor de $\chi^2 = 20,67$ con 6 grados de libertad y un $p < 0,01$, lo que permite concluir que la concordancia observada es estadísticamente significativa, indicando que los jueces mostraron un consenso consistente y validable estadísticamente respecto a la clasificación de las herramientas según los criterios de retroalimentación de Wilson.

Aplicación de la herramienta en asignatura Estructura de Datos

La herramienta fue aplicada a un total de 107 estudiantes de segundo año. Durante la prueba, se procesaron 846 respuestas abiertas, cada una evaluada automáticamente mediante el modelo GPT-4o, utilizando el prompt estructurado previamente validado.

Se habilitó un canal de apelación voluntaria a través del correo institucional, permitiendo a los estudiantes solicitar una revisión de la retroalimentación o del puntaje asignado. Se recibieron 34 apelaciones, de las cuales 27 fueron consideradas válidas, lo que representa un 3,19 % del total de respuestas, tasa relativamente baja de disconformidad con el sistema automatizado.

La Tabla 3 presenta un análisis de los principales patrones de error detectados en el modelo GPT-4o, sobre la base del total de respuestas procesadas ($N = 846$). Para cada patrón identificado, se incluye un ejemplo representativo de retroalimentación generada, un análisis interpretativo del tipo de error, y la frecuencia (Frec.) con que este fue observado, expresada tanto en número de casos como en porcentaje del total.

Tabla 3

Patrones de error observado en respuestas del modelo

Patrón de error	Ejemplo retroalimentación GPT-4o	Análisis del error	Casos	Frec.
La retroalimentación del modelo sugiere incluir detalles que ya estaban presentes en la respuesta del estudiante.	“Sería útil considerar una descripción más clara de las operaciones específicas que se realizan en la lista, como la adición de elementos al final y la eliminación del primer elemento”.	La retroalimentación sugiere que faltan detalles sobre las operaciones de agregar y eliminar elementos en la lista. Sin embargo, el estudiante ya lo había mencionado en su respuesta.	31	3,7 %

Patrón de error	Ejemplo retroalimentación GPT-4o	Análisis del error	Casos	Frec.
La retroalimentación del modelo pide detalles o aspectos que no estaban solicitados en la pregunta.	“Sería útil considerar la importancia de la función hash y cómo afecta la eficiencia de la tabla”.	La retroalimentación sugiere incluir información adicional (como la importancia de la función hash) que no se pedía en la pregunta original. Esto puede confundir al estudiante porque extiende el alcance de la pregunta y lo hace sentir que su respuesta estaba incompleta aunque en realidad respondió adecuadamente a lo solicitado.	127	15 %
La retroalimentación es poco clara o ambigua limitando la orientación a la mejora del estudiante	“tu respuesta no aborda cómo manejar el rango específico de búsqueda ni cómo recolectar los elementos dentro de ese rango”.	La retroalimentación de la IA sugiere que el estudiante no explicó nada sobre la recolección de elementos, cuando en realidad sí lo hizo, pero su enfoque no era el más adecuado. Esto puede desorientar al estudiante porque no queda claro si lo que escribió estaba mal o si simplemente necesitaba mejorarlo.	47	5,6 %

Fuente: Elaboración propia.

Aplicación de la herramienta en asignatura Evaluación para el Aprendizaje

La herramienta fue aplicada a un total de 37 estudiantes. Durante la prueba, se procesaron 592 respuestas abiertas, cada una evaluada automáticamente mediante el modelo GPT-4o, utilizando el mismo prompt estructurado previamente validado.

Al igual que en la experiencia anterior, se habilitó un canal de apelación voluntaria a través del correo institucional, con el fin de que los estudiantes pudieran solicitar una revisión de la retroalimentación o del puntaje asignado. En este caso, no se recibieron apelaciones, lo que sugiere una alta aceptación de las evaluaciones generadas automáticamente.

La Tabla 4 presenta un análisis de los principales patrones de error detectados en este segundo experimento, sobre la base del total de respuestas procesadas ($N = 592$).

Tabla 4

Patrones de error observado en respuestas del modelo

Patrón de error	Ejemplo retroalimentación GPT-4o	Análisis del error	Casos	Frec.
La retroalimentación del modelo pide detalles o aspectos que no estaban solicitados en la pregunta.	“¿Qué ejemplos específicos de evaluación formativa observaste durante la clase?”	La retroalimentación no resulta pertinente respecto de la respuesta de la o el estudiante ni del caso presentado en la pregunta, el cual no refleja una aplicación explícita de los principios de la evaluación formativa.	3	0,5 %
La retroalimentación contiene errores conceptuales o afirmaciones incorrectas.	La respuesta destaca correctamente la función pedagógica de la evaluación y su enfoque formativo, lo cual es fundamental en un contexto de aprendizaje de idiomas. Además, se identifica adecuadamente el uso de la heteroevaluación, lo que muestra una comprensión de los roles de los agentes evaluativos en la clase.	La retroalimentación sugiere que la respuesta de la o el estudiante es correcta y adecuada; sin embargo, esta no aborda todos los aspectos requeridos por la pregunta. Esta incongruencia puede generar confusión, especialmente si no se asigna el puntaje máximo y la retroalimentación no especifica claramente qué elementos deben ser mejorados.	4	0,7 %

Patrón de error	Ejemplo retroalimentación GPT-4o	Análisis del error	Casos	Frec.
La retroalimentación ocupa conceptos negativos o rígidos que no se condicen con una retroalimentación constructiva.	La respuesta carece de una discusión sobre cómo las pruebas estandarizadas podrían influir...	Se reemplazó la palabra “carece” por considerarse poco apropiada en el contexto de una retroalimentación constructiva, ya que podría resultar rígida o negativa. En su lugar, se utilizó una formulación más formativa y orientada a la mejora, como “faltaría incorporar...” o “aún no se aborda...”.	26	4,4 %
La retroalimentación es poco clara o ambigua limitando la orientación a la mejora del estudiante.	La respuesta destaca correctamente la naturaleza formativa de la evaluación en la clase.	Se emplea una doble valoración al utilizar términos como “destaca” y “correctamente”; sin embargo, no se entrega una retroalimentación constructiva desde la perspectiva del logro alcanzado. La respuesta no especifica qué aspecto de la evaluación formativa se valora en el contexto de la clase, lo que limita su utilidad para orientar mejoras.	12	2,0 %

Fuente: Elaboración propia.

Percepción de estudiantes sobre retroalimentación recibida

A partir de este sistema de categorías, se procedió al análisis sistemático de las entrevistas, codificando las respuestas de los estudiantes en función de las categorías y subcategorías definidas. Esta codificación permitió detectar patrones de sentido, recurrencias temáticas y contrastes en las valoraciones individuales, así como matices relevantes en la forma en que los participantes interpretaron la retroalimentación automatizada. El análisis se realizó combinando una lectura inductiva del discurso con la clasificación estructurada por dimensiones, lo que facilitó una comprensión

holística y matizada de la experiencia estudiantil. Cada testimonio fue interpretado a la luz de su coherencia interna, su relación con las categorías del marco analítico y su conexión con los demás casos.

En términos generales, los estudiantes valoran positivamente la experiencia de recibir retroalimentación automatizada, destacando especialmente su rapidez, claridad e innovación en comparación con la retroalimentación tradicional de los docentes. La sorpresa ante el uso de inteligencia artificial es recurrente: “es sorprendente que esta te revise la evaluación, ya que uno como estudiante está acostumbrado a que esta retroalimentación la realice el profesor” (estudiante 3). Esta innovación es vista como un cambio significativo en la práctica evaluativa, percibido como algo novedoso y prometedor. No obstante, esta valoración técnica convive con matices emocionales: algunos describen la experiencia como “un poco estresante” (estudiantes 1 y 8), porque “te hace pensar que tienes que estar mucho más preparada y ser más precisa” (Estudiantes, 1, 4, 6). En conjunto, los estudiantes reconocen el valor funcional del sistema, pero también notan la ausencia del “factor humano” (estudiantes 1, 2, 5, 7 y 8) y una cierta distancia afectiva en el estilo de la retroalimentación.

Respecto al peldaño de aclaraciones, los participantes aprecian que la IA sea capaz de identificar omisiones o ambigüedades que ellos mismos no habían detectado. Se valora, por ejemplo, cuando la IA indica que “falta detallar el proceso completo de ajuste de punteros para que el nuevo nodo se integre correctamente en la lista” (estudiante 7). Estas observaciones son percibidas como útiles y formativas, pero no están exentas de crítica. Algunos estudiantes cuestionan que “me corrige y me explica algo que yo ya puse en mi respuesta” (estudiante 10), lo cual puede generar confusión y una sensación de evaluación injusta. Esta tensión entre estructura y subjetividad se refuerza cuando la “IA se vuelve más crítica cuando uno da una respuesta con libertad” (estudiante 9), lo que pone en cuestión la flexibilidad interpretativa del sistema frente a respuestas no literales pero válidas.

En cuanto al peldaño de valoración, muchos estudiantes reconocen el refuerzo positivo como un estímulo para continuar aprendiendo. Afirmaciones como “resalta lo que hiciste bien y te invita a seguir estudiando” (estudiante 6) o “me confirma que lo que sé no es tan erróneo” (estudiante 2) muestran que el reconocimiento, incluso automatizado, puede cumplir una función pedagógica significativa. Sin embargo, su impacto emocional es limitado para algunos: “si los comentarios vinieran de un profesor, probablemente me sentiría más reconocida” (estudiante 4). Esta ambivalencia revela que, aunque el refuerzo valida el conocimiento y orienta al estudiante, carece de la carga afectiva y la empatía que normalmente aporta un docente humano, lo cual plantea consideraciones éticas sobre la calidad del vínculo e interacción pedagógico mediado por IA.

El peldaño de inquietudes muestra que las críticas de la IA son, en general, bien recibidas gracias a su tono claro y respetuoso. El estudiante 1 señala que “estas críticas son constructivas, lo cual refuerza tus errores”, mientras otro comenta que

“me sentí bien criticado y al grano, sin dar vueltas” (estudiante 3). No obstante, aparecen cuestionamientos cuando la IA exige explicaciones adicionales que no fueron requeridas por la consigna. Por ejemplo, “aunque me dio todo el puntaje, creo que la parte donde menciona incluir ejemplos es un poco innecesaria” (estudiante 9). Estas observaciones sugieren que la retroalimentación sería más efectiva si ajustara su nivel de exigencia al contenido realmente trabajado y a los criterios explícitos de la evaluación.

En relación con el peldaño de sugerencias, los estudiantes consideran que la IA ofrece recomendaciones “claras, concisas y fáciles de aplicar”, lo que contribuye a su aprendizaje autónomo. “Me ayudó a tenerlo en cuenta para futuras ocasiones” (estudiante 10), comenta uno, mientras otro destaca que “me recalco el orden de redacción al explicar mi algoritmo” (estudiante 3). Sin embargo, la utilidad de estas sugerencias depende del grado de alineación con el contexto del curso. Algunos critican que “no habíamos visto complejidad temporal, por lo cual confundía la retroalimentación” (estudiante 5), evidenciando que el sistema puede perder eficacia si opera desconectado del temario. A pesar de ello, la mayoría valora la estructura y claridad de las sugerencias como un recurso para mejorar progresivamente.

Finalmente, en cuanto al impacto en el aprendizaje, los estudiantes coinciden en que el uso de inteligencia artificial ha fortalecido su proceso formativo, contribuyendo significativamente a la reflexión sobre sus propios aprendizajes. La retroalimentación inmediata fue valorada por su capacidad de activar conocimientos previos, “recordar detalles que se me olvidaban” (estudiante 6) y permitir una revisión oportuna: “retroalimentar al momento” (estudiante 2); además, fue descrita como “de utilidad para poder reflexionar sobre mi proceso de enseñanza-aprendizaje” (estudiante 4).

Algunos participantes reconocen el aporte de esta herramienta como un recurso estratégico: “me hizo mejorar en temas que no tenía muy bien profundizados” (estudiante 8) y “me ayudó a identificar elementos importantes no considerados antes” (estudiante 1). Estas percepciones reflejan una apropiación activa de la retroalimentación automatizada como apoyo al desarrollo de competencias reflexivas y metacognitivas.

Por otra parte, en la categoría de juicio crítico y recomendaciones, se observa una actitud proactiva y analítica. Los estudiantes no solo valoran el uso de la IA, sino que expresan propuestas concretas para mejorar su funcionamiento. Entre ellas destacan el reorganizar los pasos de la Escalera de Wilson colocando valorar antes que aclarar, incorporar una sección de discusión para profundizar contenidos, entrenar previamente la IA con materiales del curso, y evitar penalizaciones injustificadas no bajar puntaje por falta de ejemplos si no son requeridos o están en los criterios de evaluación.

Estas propuestas evidencian que el estudiantado no se posiciona como receptor pasivo del sistema, sino como agente reflexivo y comprometido con el perfeccionamiento de una herramienta pedagógica que consideran útil, pero que desean ver más contextualizada, justa y alineada con los objetivos formativos del curso.

DISCUSIÓN Y CONCLUSIONES

Los resultados de este estudio evidencian el potencial pedagógico de integrar retroalimentación automatizada mediada por inteligencia artificial generativa, en particular a través de GPT-4o estructurado mediante la Escalera de Wilson. Esta propuesta se alinea con la necesidad de transformar las prácticas de retroalimentación en la educación universitaria, que los estudiantes suelen percibir como tardías, superficiales y poco específicas (Carless y Boud, 2018; Winstone et al., 2022; Quezada y Salinas, 2021). En entornos abiertos y a distancia, donde la inmediatez y la claridad son factores críticos, la IAGen demuestra ser capaz de reducir tiempos de respuesta y de entregar comentarios estructurados y focalizados, potenciando la dimensión formativa de la evaluación (Puertas y Cano, 2024; Holmes et al., 2019).

La Escalera de Wilson ofreció un andamiaje útil para organizar la retroalimentación en los cuatro peldaños (aclarar, valorar, expresar inquietudes y sugerir), lo que permitió generar observaciones más comprensibles, específicas y orientadas a la mejora, incluso en un formato automatizado. Los estudiantes destacaron esta experiencia por su rapidez, claridad e innovación, reconociendo en ella una herramienta capaz de complementar el rol docente. No obstante, también señalaron tensiones asociadas a sentimientos de estrés, mayor exigencia en la preparación de sus respuestas y la ausencia del factor humano que confiere cercanía y calidez al proceso de retroalimentación. Estos matices refuerzan que la tecnología no puede sustituir al profesor, sino que debe concebirse como un apoyo complementario a la interacción pedagógica. Esta interpretación coincide con lo planteado por Valenzuela y Pérez (2025), quienes subrayan el carácter insustituible del profesorado en términos de interacción humana y apoyo personalizado, recordando que la inteligencia artificial puede constituir un recurso valioso, pero siempre subordinado a la dimensión relacional y afectiva propia de la enseñanza.

Un hallazgo relevante fue la observación estudiantil respecto del peldaño Aclarar. En la práctica actual, la herramienta no permite al estudiante aportar información adicional antes de recibir comentarios, lo que reduce el alcance formativo de este paso. Incluir un mecanismo interactivo, como un chat o ventana de aclaración, podría enriquecer la personalización de la retroalimentación, alineando los comentarios con las necesidades reales de los estudiantes (Luckin y Holmes, 2016; López Regalado et al., 2024). Esta posibilidad adquiere mayor importancia en contextos masivos o de educación a distancia, donde los espacios de interacción suelen ser limitados.

En el plano tecnológico, la validación por expertos y la percepción positiva del estudiantado coinciden con lo planteado por Zhang et al. (2024) y Baral et al. (2024), quienes destacan que GPT puede generar retroalimentación coherente y funcional siempre que los prompts estén cuidadosamente diseñados. Este aspecto resulta crítico en escenarios que requieren estandarización y escalabilidad (Carless y Winstone, 2023). No obstante, como advierten Jauhiainen y Garagorry Guerra (2024), persisten limitaciones vinculadas a la empatía, la contextualización y la

sensibilidad afectiva, lo que reafirma la necesidad de supervisión docente para asegurar la pertinencia pedagógica y el impacto formativo.

El estudio también muestra que estudiantes de distintas áreas y niveles académicos (segundo y cuarto año) valoraron de manera similar la retroalimentación automatizada, lo que sugiere que la Escalera de Wilson, aplicada con IA generativa, constituye un marco adaptable y replicable en diversos contextos. Sin embargo, su implementación requiere incorporar criterios éticos, revisión docente y un diseño riguroso de prompts que orienten adecuadamente la producción de la IA.

Finalmente, emergen desafíos que invitan a profundizar en investigaciones futuras, particularmente en el análisis de sesgos algorítmicos, la evaluación de riesgos éticos vinculados a la privacidad y la equidad, así como en la necesidad de garantizar la sostenibilidad técnica de estas herramientas. Resulta igualmente clave explorar su adaptación a diversas disciplinas académicas, asegurando la pertinencia pedagógica y la alineación con estándares de calidad en educación superior. En esta línea, se recomienda avanzar hacia una integración responsable y contextualizada de la inteligencia artificial en los procesos evaluativos universitarios, de manera que sus beneficios técnicos se complementen con la mediación pedagógica docente. A futuro, resulta pertinente ampliar el alcance de los estudios a distintas disciplinas y explorar nuevas dimensiones de su impacto en la educación, considerando variables como la calidad argumentativa, la resolución de problemas, el pensamiento crítico, la autonomía estudiantil y la transformación de las prácticas evaluativas. Asimismo, se sugiere desarrollar investigaciones longitudinales que permitan analizar los efectos sostenidos del uso de IAGen en el aprendizaje y en la práctica docente. En paralelo, se destaca la necesidad de fortalecer la formación del profesorado en el diseño de prompts, la reflexión ética y la evaluación crítica del uso de estas tecnologías, con el fin de asegurar una implementación más consciente y formativa.

NOTAS

- ¹ Moodle es una plataforma de gestión del aprendizaje (LMS, por sus siglas en inglés) de código abierto ampliamente utilizada en contextos educativos para la creación, entrega y seguimiento de cursos en línea.
- ² Versión gpt-4o-2024-08-06
- ³ Versión gpt-4o-mini-2024-07-18
- ⁴ CodeRunner es un plugin para Moodle que permite la evaluación automática de código informático. Se integra con un motor de ejecución (por ejemplo, Python) que procesa las respuestas en tiempo real, generando retroalimentación y puntajes inmediatos de acuerdo con criterios de evaluación preestablecidos.

REFERENCIAS

- Al-Azawei, A., Abdullah, A. A., Mohammed, M. K. y Abod, Z. A. (2023). Predicting online learning success based on learners' perceptions: The integration of the information system success model and the security triangle framework. *International Review of Research in Open and Distributed Learning*, 24(2), 7295. <https://doi.org/10.19173/irrodl.v24i2.6895>
- Andrade, H. L. y Brookhart, S. M. (2019). Classroom assessment as the co-regulation of learning. *Assessment in Education: Principles, Policy & Practice*, 26(1), 103-117. <https://doi.org/10.1080/0969594X.2019.1571992>
- Anthropic. (2024). *Claude 3.7* [Modelo de lenguaje de gran tamaño]. <https://www.anthropic.com>
- Ayeni, O. O., Al Hamad, N. M., Chisom, O. N., Osawaru, B. y Adewusi, O. E. (2024). AI in education: A review of personalized learning and educational technology. *GSC Advanced Research and Reviews*, 18(2), 261-271. <https://doi.org/10.30574/gscarr.2024.18.2.0062>
- Baral, S., Worden, E., Lim, W.-C., Luo, Z., Santorelli, C., Gurung, A. y Heffernan, N. (2024). *Automated feedback in math education: A comparative analysis of LLMs for open-ended responses*. arXiv. <https://doi.org/10.48550/arXiv.2411.08910>
- Black, P. y Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice*, 25(6), 551-575. <https://doi.org/10.1080/0969594X.2018.1441807>
- Bonales-Daimiel, G., Martínez-Estrella, E. C. y Sierra-Sánchez, J. (2025). Evolución del perfil docente y surgimiento de nuevos roles profesionales en la era de la inteligencia artificial (IA). *Pixel-Bit. Revista de Medios y Educación*, 73, art. 3. <https://doi.org/10.12795/pixelbit.109085>
- Booth, W. C., Colomb, G. G. y Williams, J. M. (2008). *The craft of research* (3ª ed.). University of Chicago Press. <https://doi.org/10.7208/chicago/9780226062648.001.0001>
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., ... Amodei, D. (2020). *Language models are few-shot learners*. arXiv. <https://doi.org/10.48550/arXiv.2005.14165>
- Carless, D. y Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, 43(8), 1315-1325. <https://doi.org/10.1080/02602938.2018.1463354>
- Carless, D. y Winstone, N. (2023). Teacher feedback literacy and its interplay with student feedback literacy. *Teaching in Higher Education*, 28(1), 150-163. <https://doi.org/10.1080/13562517.2020.1782372>
- Dai, W., Lin, J., Jin, F., Li, T., Tsai, Y. S., Gašević, D. y Chen, G. (2023). *Can large language models provide feedback to students? A case study on ChatGPT*. EdArXiv. <https://doi.org/10.35542/osf.io/hcgzj>
- DeepSeek. (2024). *DeepSeek R1* [Modelo de IA generativa]. <https://www.deepseek.com/>
- European Commission. (2023). *Teaching with AI – Assessment, feedback and personalisation. Briefing report No. 7 (European Digital Education Hub)*. Erasmus+ Programme. https://resitve.sio.si/wp-content/uploads/sites/7/2023/11/AI-squad-output_briefing-report-7.pdf
- Galindo-Domínguez, H., Delgado, N., Losada, D. y Etxabe, J. M. (2023). An

- analysis of the use of artificial intelligence in education in Spain: The in-service teacher's perspective. *Journal of Digital Learning in Teacher Education*, 40(1), 41-56. <https://doi.org/10.1080/21532974.2023.2284726>
- García-Peñalvo, F. J. (2023). La percepción de la inteligencia artificial en contextos educativos tras el lanzamiento de ChatGPT: ¿Disrupción o pánico? *Education in the Knowledge Society (EKS)*, 24, e31279-e31279. <https://doi.org/10.14201/eks.31279>
- García Peñalvo, F. J., Llorens-Largo, F. y Vidal, J. (2024). La nueva realidad de la educación ante los avances de la inteligencia artificial generativa. *RIED-Revista Iberoamericana de Educación a Distancia*, 27(1), 9-39. <https://doi.org/10.5944/ried.27.1.37716>
- Goodrich, H. (2011). Escalera de la retroalimentación [Adaptación basada en la Escalera de Retroalimentación de Ron Berger]. En A. Goodrich (Coord.), *Protocols in the classroom*. Harvard Project Zero.
- Google DeepMind. (2024). *Gemini 2.5* [Multimodal AI model]. <https://deepmind.google/>
- Hao, Y., Sun, Y., Dong, L., Han, Z., Gu, Y. y Wei, F. (2022). *Structured prompting: Scaling in-context learning to 1,000 examples*. arXiv. <https://doi.org/10.48550/arXiv.2212.06713>
- Holmes, W., Bialik, M. y Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning*. Center for Curriculum Redesign.
- Jauhainen, J. S. y Garagorry Guerra, A. (2024). Generative AI in education: ChatGPT-4 in evaluating students' written responses. *Innovations in Education and Teaching International*, 1-18. <https://doi.org/10.1080/14703297.2024.2422337>
- Landeta, J. (1999). *El método Delphi: Una técnica de previsión para la incertidumbre*. Ariel.
- Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H. y Neubig, G. (2023). Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Computing Surveys*, 55(9), 1-35. <https://doi.org/10.1145/3560815>
- LLaMa. (2024). *LLaMA4* [Large language model]. <https://www.llama.com/>
- López Regalado, O., Núñez-Rojas, N., López Gil, O. R. y Sánchez-Rodríguez, J. (2024). El análisis del uso de la inteligencia artificial en la educación universitaria: Una revisión sistemática. *Pixel-Bit. Revista de Medios y Educación*, 70, 97-122. <https://doi.org/10.12795/pixelbit.106336>
- Luckin, R. y Holmes, W. (2016). *Intelligence unleashed: An argument for AI in education*. Pearson. <https://www.pearson.com/content/dam/one-dot-com/one-dot-com/global/Files/about-pearson/innovation/open-ideas/IntelligenceUnleashedSPANISH.pdf>
- Mayring, P. (2000). Qualitative content analysis. *Forum: Qualitative Social Research*, 1(2), 1-10. <https://doi.org/10.17169/fqs-1.2.1089>
- Ministerio de Educación de Chile. (2019). *Orientaciones para la implementación del Decreto 67/2018 de evaluación, calificación y promoción*. MINEDUC. <https://bibliotecadigital.mineduc.cl/bitstream/handle/20.500.12365/14279/orientaciones%20decreto%2067.pdf>
- Mistral AI. (2024). *Mistral* [Large language model]. <https://mistral.ai/>
- Molloy, E., Boud, D. y Henderson, M. (2020). Developing a learning-centred framework for feedback literacy. *Assessment & Evaluation in Higher Education*, 45(4), 527-540. <https://doi.org/10.1080/02602938.2019.1667955>
- OpenAI. (2023). *Best practices for prompt engineering with the OpenAI API*. <https://>

- help.openai.com/en/articles/6654000-best-practices-for-prompt-engineering-with-the-openai-api
- OpenAI. (2024). *GPT-4o, GPT-4.5 model cards*. <https://openai.com>
- Ossa, C. y Willatt, C. (2023). Uso de inteligencia artificial generativa para retroalimentar escritura académica en procesos de formación inicial docente. *European Journal of Education and Psychology*, 16(2), 1-16. <https://doi.org/10.32457/ejep.v16i2.2412>
- Puertas, E. y Cano, E. (2024). ¿Puede la inteligencia artificial proporcionar un feedback más sostenible? *Digital Education Review*, 45(1), 50-58.
- Quezada, S. y Salinas, C. (2021). Modelo de retroalimentación para el aprendizaje: Una propuesta basada en la revisión de literatura. *Revista Mexicana de Investigación Educativa*, 26(88), 225-251.
- Reynolds, L. y McDonell, K. (2021, mayo). Prompt programming for large language models: Beyond the few-shot paradigm. En *Extended abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1-7). <https://doi.org/10.1145/3411763.3451760>
- Romero Alonso, R., Araya Carvajal, K. y Reyes Acevedo, N. (2025). Rol de la inteligencia artificial en la personalización de la educación a distancia: Una revisión sistemática. *RIED-Revista Iberoamericana de Educación a Distancia*, 28(1), 9-36. <https://doi.org/10.5944/ried.28.1.41538>
- Sahoo, S. S., Plasek, J. M., Xu, H., Uzuner, Ö., Cohen, T., Yetisgen, M. y Wang, Y. (2024). Large language models for biomedicine: Foundations, opportunities, challenges, and best practices. *Journal of the American Medical Informatics Association*, 31(9), 2114-2124. <https://doi.org/10.1093/jamia/ocae074>
- Schulhoff, S., Ilie, M., Balepur, N., Kahadze, K., Liu, A., Si, C., Li, Y., Gupta, A., Han, H., Schulhoff, S., Dulepet, P. S., Vidyadhara, S., Ki, D., Agrawal, S., Pham, C., Kroiz, G., Li, F., Tao, H., Srivastava, A., ... Resnik, P. (2025). *The prompt report: A systematic survey of prompt engineering techniques*. arXiv. <https://doi.org/10.48550/arXiv.2406.06608>
- Shute, V. J. y Rahimi, S. (2017). Review of computer-based assessment for learning in elementary and secondary education. *Journal of Computer Assisted Learning*, 33(1), 1-19. <https://doi.org/10.1111/jcal.12172>
- Steurer, J. (2011). The Delphi method: An efficient procedure to generate knowledge. *Skeletal Radiology*, 40, 959-961. <https://doi.org/10.1007/s00256-011-1145-z>
- Valenzuela Caico, R. y Pérez Carvajal, A. (2025). Inteligencia artificial en educación superior: ¿Un reemplazo para los profesores o una herramienta de apoyo? *Revista Iberoamericana de Investigación en Educación*, (9). <https://doi.org/10.58663/ried.vi9.221>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. y Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30.
- Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q. y Zhou, D. (2023). *Chain-of-thought prompting elicits reasoning in large language models*. arXiv. <https://doi.org/10.48550/arXiv.2201.11903>
- Weng, X., Xia, Q., Gu, M., Rajaram, K. y Chiu, T. K. (2024). Assessment and learning outcomes for generative AI in higher education: A scoping review on current research status and trends. *Australasian Journal of Educational Technology*, 40(6), 37-55. <https://doi.org/10.14742/ajet.9540>
- William, D. (2011). What is assessment for learning? *Studies in Educational*

- Evaluation*, 37(1), 3-14. <https://doi.org/10.1016/j.stueduc.2011.03.001>
- Wilson, D. (2013). *Ladder of feedback* [Documento de trabajo]. Project Zero, Harvard Graduate School of Education. <https://pz.harvard.edu/resources/ladder-of-feedback>
- Winstone, N. E., Boud, D., Dawson, P. y Heron, M. (2022). From feedback-as-information to feedback-as-process: A linguistic analysis of the feedback literature. *Assessment & Evaluation in Higher Education*, 47(2), 213-230. <https://doi.org/10.1080/02602938.2021.1902467>
- xAI. (2024). *Grok 3* [Large language model]. <https://grok.com/>
- Zhang, Z., Dong, Z., Shi, Y., Price, T., Matsuda, N. y Xu, D. (2024). Students' perceptions and preferences of generative artificial intelligence feedback for programming. En *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(21), 23250-23258. <https://doi.org/10.1609/aaai.v38i21.30372>

ANEXO

Prompt replicable

i. Contexto teórico

En este apartado se debe describir de manera breve y precisa el instrumento de evaluación utilizado. La descripción debe incluir: propósito general, público objetivo, objetivos de aprendizaje evaluados, criterios de evaluación, puntaje total, condiciones de aprobación, formato de aplicación y lineamientos para responder. También debe detallar la estructura general (tipos de ítems y cantidad aproximada), el tipo de habilidades o conocimientos que busca medir y el modo en que se espera que el estudiante desarrolle sus respuestas. La redacción debe ser clara, suficiente y comprensible para permitir la replicabilidad del instrumento. (Borrar este texto después de completar el contexto)

ii. Enunciado de la pregunta:

Aquí se coloca el enunciado de la pregunta.

iii. Criterios de evaluación/respuesta esperada:

En este espacio se debe incluir una rúbrica que contenga los criterios de evaluación, los niveles de desempeño esperados y las descripciones correspondientes a cada nivel. Los criterios deben estar alineados con los objetivos de aprendizaje y permitir una valoración objetiva de la calidad, pertinencia y profundidad de la respuesta.

La escala de puntuación debe ser explícita y consistente, y debe incluirse el rango de puntajes para cada nivel. (Borrar este texto después de completar la rúbrica)

iv. Respuesta del estudiante:

Aquí se coloca la respuesta del estudiante.

v. Instrucciones para retroalimentar:

Asigna puntuación a la Respuesta del Estudiante utilizando la rúbrica (puntuación para cada criterio).

Justifica cada uno de los puntajes de la rúbrica.

Luego, redacta una retroalimentación breve en español utilizando el marco de la Escalera de Wilson (valoraciones, aclaraciones, inquietudes, sugerencias).

La retroalimentación debe tener los 4 componentes de 100 palabras cada uno (máximo), enfocados en cada componente de la Escalera:

VALORACIONES: Reconoce de forma genuina los aspectos positivos de la respuesta. Destaca fortalezas, ideas correctas o puntos bien expresados para reforzar buenas prácticas.

ACLARACIONES: Invita a precisar o desarrollar aspectos faltantes de la respuesta para mejorar su claridad y profundidad, sin emitir juicio.

INQUIETUDES: Señala errores en la respuesta utilizando un tono respetuoso y constructivo. Prefiere expresiones sugerentes y amables como “sería útil considerar...”, “podrías revisar...”, o “tal vez convenga precisar...”.

SUGERENCIAS: Ofrece sugerencias específicas de mejora para futuras respuestas. No pidas ejemplos ni códigos.

vi. Formato de salida en diccionario Python:

Como evaluador, sé generoso: si la respuesta cumple parcialmente con el criterio, asigna el puntaje máximo para ese nivel.

Se espera un diccionario en Python como respuesta: {"analisis1": "(nombre criterio 1) análisis criterio 1", "score1": score1, "analisis2": "(nombre criterio 2) análisis criterio 2", "score2": score2, ..., "valoraciones": "valoraciones", "aclaraciones": "aclaraciones", "inquietudes": "inquietudes", "sugerencias": "sugerencias"}

Fecha de recepción del artículo: 1 de junio de 2025

Fecha de aceptación del artículo: 2 de septiembre de 2025

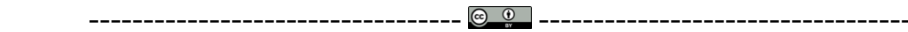
Fecha de aprobación para maquetación: 1 de octubre de 2025

Fecha de publicación en OnlineFirst: 17 de octubre de 2025

Fecha de publicación: 1 de enero de 2026

Aplicación de PathRAG en aprendizaje adaptativo con IA generativa para una educación inclusiva y sostenible

PathRAG application in adaptive learning with generative AI for inclusive and sustainable education



 Rubén Juárez Cádiz - *Universidad Alfonso X El Sabio, UAX (España)*

RESUMEN

Este estudio presenta la aplicación del modelo PathRAG en un entorno de aprendizaje adaptativo, híbrido e inclusivo, mediante el uso de inteligencia artificial generativa. Enmarcado en los Objetivos de Desarrollo Sostenible (ODS), el trabajo se orienta a personalizar el aprendizaje en contextos universitarios a través de rutas educativas dinámicas y equitativas. La propuesta busca responder a la diversidad del alumnado y reducir la brecha de acceso y participación mediante tecnología educativa avanzada. Se implementó un diseño cuasi experimental con una muestra de 52 estudiantes del Máster en Educación Inclusiva, perteneciente a una universidad española. La intervención se desarrolló en un entorno híbrido, integrando el modelo PathRAG con herramientas basadas en IA generativa (GPT-3.5 turbo). Se evaluaron indicadores de participación, desarrollo competencial, percepción de equidad e inclusión, así como la satisfacción general del alumnado. Los resultados indican mejoras significativas en la participación activa, el logro competencial y la percepción de inclusión, especialmente entre estudiantes con necesidades educativas específicas o dificultades de acceso tecnológico. La satisfacción alcanzó niveles elevados, destacando la utilidad de las rutas personalizadas. Se concluye que el modelo PathRAG potencia un aprendizaje más equitativo y adaptativo, aunque se reconoce la necesidad de futuros estudios con diseños más robustos, instrumentos validados y muestras más amplias. Este trabajo evidencia el potencial transformador de la IA generativa aplicada a contextos educativos sostenibles e inclusivos.

Palabras clave: inteligencia artificial generativa; aprendizaje adaptativo; inclusión educativa; sostenibilidad; educación híbrida; personalización del aprendizaje.

ABSTRACT

This study presents the implementation of the PathRAG model within an adaptive, hybrid, and inclusive learning environment, supported by generative artificial intelligence. Aligned with the Sustainable Development Goals (SDGs), the proposal aims to personalize university-level learning through dynamic and equitable educational pathways. The objective is to address student diversity while reducing access and participation gaps through advanced educational technology. A quasi-experimental design was applied to a sample of 52 students enrolled in a Master's program in Inclusive Education at a Spanish university. The intervention was developed in a hybrid format, combining the PathRAG algorithm with generative AI tools (GPT-3.5 turbo). Key indicators such as active participation, competence development, perceived inclusion and equity, and overall student satisfaction were assessed. Findings show significant improvements in active engagement, skill acquisition, and inclusive perception, especially among students with special educational needs or limited technological access. Overall satisfaction was high, particularly regarding the usefulness of personalized learning paths. The study concludes that PathRAG fosters more equitable and adaptive learning processes. Nevertheless, limitations such as the absence of a control group, short duration, and lack of validated instruments are acknowledged. Future research should involve controlled designs, broader samples, and longitudinal approaches. This work highlights the transformative potential of generative AI in promoting sustainable and inclusive educational models.

Keywords: generative artificial intelligence; adaptive learning; educational inclusion; sustainability; hybrid education; personalized learning.

INTRODUCCIÓN

La transformación digital en la educación superior ha propiciado el desarrollo de modelos pedagógicos que promueven la personalización del aprendizaje y la inclusión educativa. En este contexto, el aprendizaje adaptativo se ha consolidado como una estrategia eficaz para atender a la diversidad del alumnado, mediante el diseño de trayectorias formativas ajustadas a sus características, ritmos y necesidades. Sin embargo, su implementación en entornos reales aún enfrenta desafíos significativos, entre ellos la falta de herramientas técnicas accesibles, la limitada validación empírica de modelos adaptativos, y la escasa integración de enfoques inclusivos y sostenibles.

En paralelo, la reciente incorporación de modelos de inteligencia artificial generativa (IA-G) —como GPT-3.5 turbo— ofrece nuevas oportunidades para facilitar el diseño de rutas personalizadas, accesibles y contextualizadas. No obstante, también conlleva riesgos asociados, como el sesgo algorítmico en la generación de contenidos o la dependencia tecnológica de plataformas privativas, lo que requiere una implementación crítica y pedagógicamente supervisada (Birhane, 2023; Bender et al., 2021).

Este trabajo presenta la aplicación del modelo PathRAG, basado en un grafo semántico de recursos educativos, en un entorno híbrido universitario. El diseño fue cuasi experimental sin grupo de control, con una muestra intencionada y limitada de estudiantes. La intervención se enmarca en los Objetivos de Desarrollo Sostenible (ODS), particularmente en la meta 4.5, orientada a eliminar las desigualdades en el acceso a la educación y promover una formación de calidad, inclusiva y equitativa (UNESCO, 2022).

MARCO TEÓRICO

La integración de tecnología avanzada en contextos educativos ha intensificado el debate sobre la personalización del aprendizaje, la equidad digital y el rol de la inteligencia artificial en la construcción de trayectorias formativas sostenibles. Este marco teórico se articula en tres ejes: el aprendizaje adaptativo, el potencial y los límites de la IA generativa, y los principios de inclusión educativa en entornos híbridos, incorporando hallazgos recientes.

Aprendizaje adaptativo y modelos basados en grafos

El aprendizaje adaptativo se define como un enfoque pedagógico que ajusta dinámicamente los contenidos, ritmos y niveles de dificultad en función del progreso del estudiante (Conati y Kardan, 2013; Ifenthaler y Yau, 2020). En los últimos años, los grafos de conocimiento han sido reconocidos como estructuras eficaces para

representar conceptos y relaciones semánticas en sistemas educativos personalizados (Zhou y Wang, 2020; Zhang et al., 2021; Ruipérez-Valiente et al., 2022).

Modelos recientes como PathRAG aprovechan la estructura de grafos dirigidos para priorizar caminos formativos relevantes según una consulta inicial del estudiante, favoreciendo así la coherencia y progresión del aprendizaje (Chen et al., 2024). Este enfoque ha demostrado ser eficaz en disciplinas técnicas y entornos híbridos, donde el soporte automatizado de la IA permite ajustar las recomendaciones a perfiles diversos (Lu et al., 2024; Zhang y Wang, 2025).

Inteligencia artificial generativa: oportunidades y desafíos

La IA generativa, especialmente a través de modelos de lenguaje como GPT-3.5-turbo, ha generado un impacto sustancial en la educación superior, facilitando tutorías automatizadas, retroalimentación en tiempo real y generación de contenidos adaptativos (Kasneji et al., 2023; van de Sande et al., 2023). Su integración en plataformas educativas ha demostrado mejoras en la eficiencia y motivación del alumnado, especialmente cuando se orienta a tareas de apoyo y no de reemplazo del docente (Susnjak, 2022; Luckin et al., 2022).

No obstante, esta tecnología no está exenta de desafíos éticos y pedagógicos. Investigaciones recientes alertan sobre la reproducción de sesgos algorítmicos presentes en los datos de entrenamiento, lo cual puede reforzar desigualdades preexistentes (Bender et al., 2021; Birhane, 2023; Selwyn et al., 2023). Además, la dependencia de proveedores tecnológicos cerrados limita la transparencia, la trazabilidad y el control pedagógico, comprometiendo la soberanía educativa y generando riesgos de dependencia estructural (Zawacki-Richter et al., 2019; Chen et al., 2024).

En este sentido, autores como Floridi (2019) y Holmes et al. (2021) enfatizan la necesidad de aplicar marcos éticos sólidos para la implementación de IA en contextos educativos. La gobernanza algorítmica, la auditoría semántica y la formación docente en alfabetización digital avanzada son elementos clave para minimizar la opacidad de los sistemas y asegurar una IA centrada en el bienestar del estudiante.

Fundamento matemático del algoritmo de poda por flujo

El algoritmo de PathRAG se fundamenta en un modelo de propagación de recursos sobre un grafo dirigido $G = (V, E)$, donde V representa el conjunto de nodos (entidades educativas como competencias, conceptos o recursos) y E el conjunto de aristas (relaciones semánticas entre dichos nodos, como prerrequisitos o dependencias temáticas).

El objetivo del algoritmo es identificar los caminos relacionales más relevantes desde un nodo inicial v_{start} relacionado con la consulta del estudiante, hacia un

conjunto de nodos objetivo V_q previamente recuperados por un sistema de búsqueda o selección semántica.

Inicialización y propagación de recursos

Cada nodo $v_i \in V$ recibe un valor de recurso acumulado $S(v_i)$, que se inicializa como sigue:

$$S(v_{start}) = 1; S(v_i) = 0 \forall i \neq start$$

La propagación del recurso sigue una dinámica de tipo iterativo, inspirada en modelos de difusión. El valor de recurso en cada nodo v_i se actualiza como:

$$S(v_i) = \sum_{v_j \in N_{in}(v_i)} [\alpha \cdot S(v_j) / |N_{out}(v_j)|]$$

donde:

- $\alpha \in (0,1]$ es un factor de decaimiento que penaliza trayectorias más largas.
- $N_{in}(v_i)$ es el conjunto de nodos con aristas entrantes hacia v_i .
- $N_{out}(v_j)$ es el conjunto de nodos hacia los cuales v_j tiene aristas salientes.

Este esquema de propagación garantiza que los nodos más cercanos y más transitados en el grafo reciban una mayor acumulación de recurso, lo que refleja su relevancia estructural.

Cálculo de confiabilidad y poda

Para cada camino $P = (v_0 \rightarrow \dots \rightarrow v_n)$, la confiabilidad se define como el promedio del recurso acumulado en sus nodos:

$$S(P) = (1 / |E_P|) \sum_{v_i \in V_P} S(v_i)$$

donde E_P es el número de aristas del camino y V_P el conjunto de nodos que lo conforman. Esta métrica permite identificar las rutas semánticamente más sólidas, aquellas por las que fluye más recurso y, por tanto, más valor explicativo para el modelo generativo.

Los caminos que no superan un umbral mínimo de confiabilidad θ son descartados tempranamente:

$$\text{Prune if } S(v_i) / |N_{out}(v_i)| < \theta$$

Esta operación de poda por flujo permite reducir drásticamente el espacio de búsqueda y centrar la atención en los caminos más prometedores, evitando combinaciones redundantes o poco informativas.

Selección final

Finalmente, se seleccionan los K caminos con mayor confiabilidad para ser transformados en entradas estructuradas que alimentan el modelo de lenguaje. Cada camino se convierte en una cadena textual compuesta por los nodos visitados, sus relaciones y contexto, que guía la generación de explicaciones o contenidos adaptativos en función del perfil del estudiante.

Este proceso de selección, basado en estructuras relacionales dinámicas, aporta ventajas significativas frente a los métodos de recuperación plana. En entornos educativos, permite construir itinerarios personalizados, mejorar la coherencia de los contenidos generados por IA y asegurar una alineación semántica con los objetivos de aprendizaje definidos.

Educación inclusiva, equidad y sostenibilidad

La inclusión educativa implica garantizar que todos los estudiantes, independientemente de sus condiciones personales, sociales o tecnológicas, accedan en igualdad de oportunidades a procesos de calidad (Ainscow, 2020). La equidad no solo se refiere al acceso, sino también a la personalización efectiva y a la participación significativa, especialmente en contextos híbridos o digitales (Anderson et al., 2020).

En este sentido, la tecnología educativa adaptativa —cuando es bien diseñada— puede actuar como herramienta de compensación de desigualdades (García-Peñalvo, 2022). Modelos como PathRAG, al incorporar IA generativa y representación semántica, permiten ajustar los itinerarios a las limitaciones tecnológicas del estudiante, sus estilos de aprendizaje y sus necesidades específicas, favoreciendo un entorno de aprendizaje más justo y sostenible.

La inclusión educativa, entendida como la capacidad de atender de forma equitativa a la diversidad del alumnado, se ha visto fortalecida mediante el uso de tecnologías adaptativas que permiten personalizar los recursos y reducir barreras de acceso (UNESCO, 2023; Cobo et al., 2022). Estudios recientes subrayan que la personalización inteligente, cuando se implementa con principios de accesibilidad universal, puede mejorar los resultados de estudiantes con necesidades específicas y promover trayectorias más sostenibles (Pedro et al., 2021; Liu et al., 2023).

Además, las plataformas que integran recuperación aumentada por grafos y generación de contenidos ofrecen oportunidades significativas para el diseño de rutas formativas contextualizadas, respetando ritmos, estilos y capacidades individuales. Estas estrategias se alinean con el ODS 4 de la Agenda 2030, fomentando una educación de calidad, inclusiva y equitativa.

METODOLOGÍA

Este estudio adopta un enfoque metodológico mixto de tipo cuantitativo-cuasi experimental sin grupo de control, con el apoyo de técnicas cualitativas de corte interpretativo. Esta estrategia se eligió por su idoneidad para evaluar, de manera integral, tanto la eficacia del sistema PathRAG —basado en inteligencia artificial generativa y recuperación aumentada por grafos— como la experiencia subjetiva del alumnado en contextos virtuales de aprendizaje personalizados. La combinación de métodos cuantitativos y cualitativos buscó aportar solidez interpretativa a los resultados, mediante la triangulación de datos y la convergencia de perspectivas.

El objetivo general fue analizar el impacto de la implementación de PathRAG en educación superior a distancia, prestando atención a tres dimensiones clave: (1) personalización adaptativa del aprendizaje, (2) mejora de la experiencia de estudiantes con necesidades educativas diversas, y (3) alineación con principios de sostenibilidad educativa y equidad digital.

El diseño se estructuró en tres fases secuenciales:

- Fase 1 (Pretest): Aplicación de una prueba diagnóstica y un cuestionario inicial para evaluar conocimientos previos, autonomía y motivación hacia el aprendizaje virtual.
- Fase 2 (Intervención): Desarrollo de tres ejercicios prácticos en el aula virtual, asistidos por PathRAG. El sistema generó rutas personalizadas de aprendizaje y contenidos adaptados según el perfil cognitivo de cada estudiante.
- Fase 3 (Postest): Evaluación de la evolución en comprensión conceptual, habilidades específicas y percepción de inclusión, complementada con preguntas abiertas sobre la claridad, utilidad y pertinencia de la intervención con IA generativa.

El diseño cuasi-experimental sin grupo de control responde a las características propias del contexto institucional y del acceso a un único grupo consolidado. Aunque este diseño limita la inferencia causal directa, permite aproximarse a los efectos del sistema en condiciones ecológicas reales, ofreciendo evidencia preliminar relevante para futuras investigaciones.

Participantes

La muestra estuvo compuesta por 350 estudiantes del Máster Online en Ciberseguridad, impartido de forma 100 % virtual en modalidad asincrónica con tutorización en línea. El programa, de carácter interuniversitario y profesionalizante, ofrece una formación avanzada en ciberseguridad en un entorno distribuido, multicultural y tecnológicamente diverso.

Los estudiantes presentaban un perfil heterogéneo: edades entre 22 y 54 años, distintos niveles de experiencia laboral y procedencia de más de 10 países hispanohablantes (principalmente España y Latinoamérica). Un 12 % de la muestra declaró alguna necesidad educativa específica, desglosada en tres categorías:

- Trastornos del aprendizaje: dislexia, TDAH u otros (6 %).
- Discapacidades sensoriales leves: visuales o auditivas (4 %).
- Limitaciones tecnológicas: conectividad deficiente o carencia de equipos (2 %).

El muestreo fue no probabilístico por conveniencia, aprovechando un grupo ya consolidado, accesible y alineado con los objetivos exploratorios del estudio. Si bien esta estrategia limita la generalización, se justifica por la viabilidad del acceso, la diversidad de perfiles presentes y la pertinencia temática del programa. Para estudios futuros, se recomienda la implementación de diseños controlados con muestreo probabilístico y grupos comparativos, que permitan ampliar la validez externa de los resultados y realizar inferencias causales más sólidas.

Entorno de implementación y variables

La intervención se desarrolló en un entorno virtual de aprendizaje (EVA) basado en Moodle 4.1, al que se integró un microservicio en Python (FastAPI) conectado a un modelo generativo (GPT-3.5-turbo). PathRAG se apoyó en un grafo de conocimiento estructurado con más de 12.000 nodos y 19.000 aristas, construido a partir de materiales docentes del máster (PDFs, vídeos transcritos, SCORM, glosarios, etc.) y tratado con técnicas NLP (tokenización, extracción de entidades, análisis semántico).

- La intervención consistió en tres actividades prácticas:
- Análisis de vulnerabilidades en sistemas críticos.
- Simulación de ataques OSINT.
- Diseño de defensas adaptativas.

Cada ejercicio fue precedido por un pretest y seguido por un posttest. PathRAG generó rutas personalizadas y contenidos adaptativos (resúmenes, retroalimentación, ejercicios dirigidos) que se entregaban dinámicamente mediante un dashboard individualizado, en función de errores, interacciones y patrones de navegación.

Integración de enfoques mixtos

Este estudio se fundamenta en un diseño metodológico mixto de tipo cuantitativo-cuasi experimental sin grupo de control, complementado por un análisis cualitativo

de tipo interpretativo, con el objetivo de triangular los resultados y enriquecer la comprensión del impacto de PathRAG desde múltiples dimensiones del aprendizaje.

Justificación del enfoque mixto

La combinación de métodos se justifica por la naturaleza compleja del fenómeno educativo evaluado: la personalización algorítmica en entornos virtuales. Mientras que los datos cuantitativos permitieron medir la evolución objetiva de indicadores como participación, comprensión o habilidades mediante pruebas estandarizadas y registros de actividad, los datos cualitativos permitieron acceder a la experiencia vivida del alumnado, capturando aspectos subjetivos como la percepción de inclusión, claridad de las rutas personalizadas o barreras cognitivas.

Fase cualitativa: instrumentos y procedimiento

Durante la tercera fase del estudio (postest), se incluyeron ítems abiertos en los instrumentos de evaluación, diseñados para recoger impresiones del alumnado sobre:

- Utilidad de las rutas personalizadas.
- Claridad y pertinencia de los contenidos generados.
- Satisfacción con la experiencia de aprendizaje mediada por IA.
- Percepción de equidad, esfuerzo cognitivo y accesibilidad.

Estas respuestas fueron codificadas utilizando un análisis temático inductivo (Braun y Clarke, 2006), siguiendo este procedimiento:

- Codificación abierta: dos investigadores analizaron de forma independiente el corpus de respuestas abiertas.
- Categorización por consenso: se agruparon las unidades de significado en categorías emergentes comunes.
- Triangulación con resultados cuantitativos: se cruzaron las categorías con los resultados estadísticos para detectar correspondencias o contradicciones.

El índice de concordancia intercodificador fue superior al 90 %, asegurando fiabilidad en la codificación.

Triangulación e interpretación integradora

Los hallazgos cualitativos permitieron explicar patrones numéricos observados en los análisis cuantitativos. Por ejemplo:

- El aumento de la participación activa detectado en los registros de logs se vio reforzado por comentarios que destacaban la utilidad de las rutas personalizadas y el carácter motivador del *feedback* automatizado.
- Las mejoras en comprensión conceptual coincidieron con percepciones positivas sobre la organización lógica de los contenidos generados.
- Los subgrupos con limitaciones tecnológicas o sensoriales (12 % de la muestra) manifestaron sentirse mejor atendidos gracias a la entrega progresiva y ajustada de los contenidos, lo que se alineó con sus mejoras superiores en los indicadores de inclusión y equidad.

Un ejemplo representativo fue la declaración de un estudiante con dislexia: “Por primera vez, el sistema me entregaba lo justo y necesario, sin saturarme. Era como si supiera lo que más me costaba”. Este tipo de evidencia cualitativa refuerza la conclusión de que PathRAG contribuyó a reducir la sobrecarga cognitiva y mejorar la equidad educativa.

Fortalecimiento del modelo

Esta integración metodológica permitió no solo validar los resultados estadísticos desde una perspectiva más humana y contextualizada, sino también detectar aspectos de mejora y sugerencias para futuras implementaciones. Así, los datos cualitativos no fueron meramente ilustrativos, sino complementarios y explicativos, cumpliendo plenamente con los principios de los enfoques mixtos orientados a la mejora educativa.

Instrumentos e indicadores

La recolección de datos se realizó mediante un conjunto de instrumentos diseñados ad hoc o adaptados de escalas previas, validados por un panel de seis expertos en tecnología educativa, accesibilidad y evaluación pedagógica. Estos instrumentos fueron ajustados mediante una prueba piloto con 30 estudiantes de otro máster afín, lo que permitió asegurar la claridad, pertinencia y coherencia de los ítems.

Los instrumentos empleados fueron validados por un panel de expertos y ajustados en una prueba piloto. La Tabla 1 resume las variables e instrumentos aplicados:

Tabla 1

Instrumentos de recogida de datos por variable

Variable	Instrumento	Tipo
Participación activa	Registro de logs en Moodle (clics, accesos, permanencia)	Cuantitativa
Comprensión conceptual	Test pre/post de conocimientos específicos	Cuantitativa
Desarrollo de habilidades	Rúbrica docente + autoevaluación (basada en DigCompEdu)	Cuantitativa
Inclusión y percepción de equidad	Cuestionario Likert de 7 ítems + 2 ítems abiertos	Mixta
Satisfacción	Escala validada de 10 ítems ($\alpha = 0.89$, test-retest = 0.81)	Cuantitativa

Las respuestas abiertas se analizaron con análisis temático inductivo (Braun y Clarke, 2006), mediante doble codificación independiente (índice de concordancia > 90 %).

Clasificación del alumnado

Con el fin de evitar agrupaciones excesivamente amplias bajo el término “necesidades educativas específicas”, los participantes que reportaron alguna condición fueron clasificados en tres grupos diferenciados, siguiendo criterios de categorización internacional (UNESCO, 2020):

- Trastornos del aprendizaje (6 %): incluyen dislexia, TDAH u otras condiciones que afectan el procesamiento cognitivo del aprendizaje formal.
- Discapacidades sensoriales (4 %): incluyen limitaciones auditivas o visuales leves, no discapacitantes, que requerían ajustes de formato.
- Limitaciones tecnológicas (2 %): incluyen falta de conectividad estable o acceso exclusivo a dispositivos móviles no compatibles con el entorno virtual.

Esta clasificación permitió analizar con mayor precisión el efecto de la personalización adaptativa sobre perfiles diversos, y evaluar la capacidad inclusiva del sistema en tiempo real.

Participación activa

Para medir la participación activa se utilizó el registro de actividad en Moodle, incluyendo número total de accesos, clics en recursos clave, participación en foros y tiempo de permanencia en el EVA. Se definió “seguimiento completo” como haber

accedido a los tres ejercicios prácticos, completado al menos el 80 % de las tareas propuestas, y haber participado en los cuestionarios pre y post.

Los datos fueron anonimizados y tratados de forma agregada para preservar la privacidad de los participantes, siguiendo los criterios del RGPD.

Comprensión conceptual

La comprensión conceptual se evaluó mediante un test específico de 15 ítems de opción múltiple, administrado antes y después de cada ejercicio. Estos ítems fueron desarrollados por el equipo docente y revisados por expertos externos para asegurar su validez de contenido. El índice de fiabilidad interna obtenido fue $\alpha = 0.81$.

Desarrollo de habilidades

Se aplicó una rúbrica docente basada en el marco DigCompEdu (European Framework for the Digital Competence of Educators), que abarca seis dimensiones: alfabetización digital, diseño instruccional, implementación, evaluación, empoderamiento y desarrollo profesional. Cada dimensión se valoró en cinco niveles (A1-C2). La rúbrica fue aplicada por dos evaluadores con experiencia docente en tecnologías digitales, y complementada por una autoevaluación del estudiante con la misma escala.

La consistencia interevaluador fue superior al 92 % y el índice de fiabilidad interna de la escala fue $\alpha = 0.86$.

Inclusión y percepción de equidad

Para evaluar la inclusión percibida, se diseñó un cuestionario mixto compuesto por:

- 7 ítems tipo Likert (escala 1-5): accesibilidad, percepción de justicia algorítmica, adecuación del lenguaje y adaptación a necesidades específicas.
- 2 ítems abiertos: “¿En qué medida sentiste que el sistema se adaptó a tus características?” y “¿Qué mejorarías en términos de inclusión?”

El cuestionario fue validado por cinco especialistas en inclusión digital y educación accesible. Su consistencia interna fue $\alpha = 0.83$, y las respuestas cualitativas fueron codificadas mediante análisis temático inductivo (Braun y Clarke, 2006), con doble codificación independiente (índice de concordancia > 90 %).

Satisfacción

Se utilizó una escala validada de satisfacción con sistemas de IA educativa, compuesta por 10 ítems tipo Likert (1-5) que abordan claridad, utilidad, adecuación, confiabilidad, motivación generada, entre otros aspectos. La escala fue adaptada de Conati et al. (2021), con ajustes terminológicos y de contexto digital.

El análisis de fiabilidad arrojó $\alpha = 0.89$ y test-retest = 0.81, lo que garantiza la estabilidad y consistencia de los resultados.

Análisis de datos

Con el objetivo de ofrecer una comprensión holística de los efectos del sistema PathRAG en entornos virtuales de aprendizaje, se aplicó una estrategia mixta de análisis que combina técnicas cuantitativas de inferencia estadística con análisis cualitativo temático.

Análisis cuantitativo

Incluyó:

- Estadística descriptiva: medias, DE, frecuencias.
- Inferencial: t de Student para muestras relacionadas y ANOVA de medidas repetidas.
- Tamaños de efecto: Cohen's d , η^2 parcial.
- Correlaciones entre variables: coeficiente r de Pearson.

Verificación de supuestos: Shapiro-Wilk (normalidad), Levene (homogeneidad) y Mauchly (esfericidad).

El análisis cualitativo permitió identificar patrones relacionados con accesibilidad, claridad del sistema, esfuerzo cognitivo y percepción de justicia educativa. La triangulación de resultados fortaleció la interpretación integradora de los hallazgos.

Análisis cuantitativo

Se emplearon análisis descriptivos e inferenciales, priorizando el rigor metodológico mediante la verificación de supuestos y el cálculo de medidas de tamaño del efecto e intervalos de confianza, como se detalla a continuación:

Estadística descriptiva:

- Cálculo de medias aritméticas, desviaciones estándar (DE), frecuencias absolutas (n) y relativas (%).

Contrastes inferenciales:

- t de Student para muestras relacionadas (pretest-posttest), en los tres ejercicios secuenciales.
- ANOVA de medidas repetidas para evaluar la evolución a lo largo de los ejercicios en variables continuas (comprensión, habilidades, satisfacción).

Tamaños del efecto:

- d de Cohen para comparaciones de dos momentos (pre/post).
- η^2 parcial para análisis multivariados (ANOVA).
- Los valores se interpretaron según Cohen (1988): pequeño (0.2), mediano (0.5), grande (0.8).

Correlaciones bivariadas:

Cálculo del coeficiente de correlación de Pearson (r) entre variables clave (participación activa, comprensión conceptual, satisfacción), para explorar relaciones lineales significativas ($p < .05$).

Verificación de supuestos estadísticos

Para garantizar la validez de los análisis inferenciales, se verificaron los supuestos siguientes:

- Normalidad: prueba de Shapiro-Wilk, con $p > .05$ como criterio de aceptación.
- Homogeneidad de varianzas: prueba de Levene, especialmente en la comparación entre ejercicios.
- Esfericidad: verificada mediante la prueba de Mauchly; en caso de violación, se aplicó la corrección de Greenhouse-Geisser.

Resultados ilustrativos

En la Tabla 2 se resumen los resultados del ANOVA de medidas repetidas en comprensión conceptual:

Tabla 2*ANOVA de medidas repetidas: comprensión conceptual (n = 350)*

Ejercicio	Media	DE	F	gl	η^2	Parcial
Pretest	5.28	1.12				
Post 1	6.42	1.03				
Post 2	6.89	0.98	45.21	2, 698	< .001	0.179

El resultado indica un incremento significativo en comprensión conceptual a lo largo de los tres ejercicios ($p < .001$), con un tamaño de efecto considerable ($\eta^2 = .179$).

Además, se observó una correlación positiva moderada entre participación activa y satisfacción ($r = .47, p < .01$), lo que indica que los estudiantes más comprometidos percibieron mayor utilidad del sistema.

Análisis cualitativo

El análisis cualitativo se realizó sobre los ítems abiertos del cuestionario y los comentarios de los ejercicios, empleando un análisis temático inductivo (Braun y Clarke, 2006). El proceso incluyó:

- Codificación abierta realizada por dos investigadores de forma independiente.
- Categorización por consenso y triangulación con resultados cuantitativos.
- Obtención de temas emergentes relacionados con:
 - Claridad de las rutas personalizadas.
 - Reducción del esfuerzo cognitivo percibido.
 - Mejora en accesibilidad y equidad en la entrega de contenidos.
 - Valoración crítica del uso de IA en la formación.

El índice de concordancia intercodificador fue superior al 90 %, asegurando fiabilidad en la categorización. La integración de datos cualitativos y cuantitativos permitió explicar fenómenos como el mayor impacto del sistema en estudiantes con dificultades tecnológicas o sensoriales.

Entorno de Implementación

Construcción del grafo educativo

Para desarrollar la base semántica del sistema, se construyó un grafo de conocimiento educativo a partir de los contenidos del programa formativo. Este grafo fue generado mediante técnicas de procesamiento del lenguaje natural (NLP)

aplicadas sobre textos estructurados e instructivos del máster (PDFs, SCORMs, glosarios, cuestionarios y vídeos transcritos). Las entidades conceptuales (nodos) se extrajeron utilizando el modelo spaCy, combinado con heurísticas personalizadas basadas en patrones léxicos, gramaticales y pedagógicos. Las relaciones entre entidades (aristas) fueron identificadas mediante coocurrencias frecuentes, dependencias sintácticas y vínculos explícitos del tipo “prerrequisito de”, “amplía a”, o “se resuelve con”.

El grafo final incorporó aproximadamente 12.000 nodos y 19.400 aristas semánticas, representando un mapa estructurado de competencias, conceptos clave y conexiones lógicas del currículo. Esta base sirvió como eje para las tareas de recuperación aumentada e inferencia de rutas formativas personalizadas.

Extracción de rutas adaptativas

El motor PathRAG utilizó este grafo para calcular, en tiempo real, las rutas semánticas más relevantes para cada estudiante, a partir de su interacción con el EVA. Se analizaron patrones de comportamiento (por ejemplo, recursos consultados, errores frecuentes, omisiones de materiales clave) y se generaron consultas implícitas personalizadas. La ejecución del algoritmo se realizó con los siguientes parámetros:

- Número de nodos recuperados: $N = 40$
- Número de caminos generados: $K = 15$
- Tasa de decaimiento: $\alpha = 0.8$
- Umbral de poda: $\theta = 0.05$

Este procedimiento permitió identificar rutas altamente confiables y didácticamente pertinentes, priorizando conceptos fundamentales y trayectorias de aprendizaje coherentes con los objetivos del curso y el perfil del estudiante. El resultado fue un grafo de conocimiento compuesto por más de 12.000 nodos y aproximadamente 19.000 aristas relacionales, que reflejaban las interdependencias pedagógicas entre los conceptos del programa. Este grafo fue utilizado por PathRAG para guiar la generación de rutas adaptativas, proporcionando a cada estudiante una secuencia personalizada de contenidos y actividades, generada en tiempo real a partir de sus interacciones previas, errores comunes y áreas de menor rendimiento.

La integración de este sistema en el EVA permitió mantener la trazabilidad de las acciones del estudiante, facilitar la generación automática de recursos educativos coherentes con su nivel y trayectoria, y evaluar el impacto de la IA generativa en el aprendizaje personalizado, todo ello en un entorno controlado, replicable y vinculado a los objetivos curriculares del programa.

Generación adaptativa de contenidos

Los caminos seleccionados fueron transformados en prompts estructurados, que sirvieron como entrada al modelo generativo GPT-3.5-turbo (OpenAI). Este modelo generó contenidos adaptativos como:

- Resúmenes explicativos personalizados
- Ejercicios prácticos dirigidos
- Retroalimentación automatizada en función del rendimiento

Estos recursos fueron entregados al estudiante a través de un dashboard individual integrado en la plataforma Moodle, actualizándose dinámicamente en función de su desempeño, consulta de recursos y evolución durante cada actividad.

Personalización en tiempo real

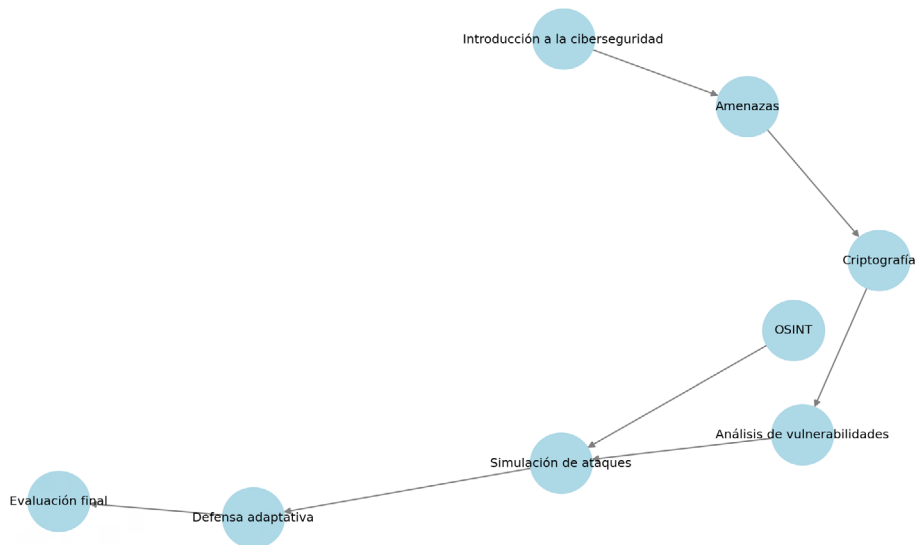
Cada ruta seleccionada fue convertida en una secuencia de actividades didácticas personalizadas —resúmenes, ejemplos, tareas prácticas, evaluaciones formativas— generadas en tiempo real y presentadas en un dashboard individual. La generación de contenidos tenía en cuenta:

- Áreas de menor rendimiento detectadas (según errores recurrentes y bajo tiempo de permanencia),
- Competencias no alcanzadas en ejercicios previos,
- Preferencias de formato (visual, textual, práctico),
- Nivel de profundidad requerido según progreso acumulado.

Este enfoque permitió una personalización semántica dinámica, orientada a maximizar la comprensión conceptual, reducir la carga cognitiva innecesaria y fomentar la progresión autónoma en entornos virtuales como se demuestra en la Figura 1.

Figura 1

Ejemplo de ruta personalizada en el grafo de conocimiento



Consideraciones éticas

Todo el procedimiento se desarrolló bajo el marco del Reglamento General de Protección de Datos (RGPD, UE 2016/679) y la legislación educativa española vigente. Se garantizó la anonimización completa de los datos, la voluntariedad de la participación y el acceso a un consentimiento informado digital, aceptado previamente por los estudiantes. El equipo investigador veló por mantener la trazabilidad de los contenidos generados, la explicabilidad de los procesos automatizados y la supervisión humana durante toda la intervención, en consonancia con los principios de uso ético de la IA en contextos educativos.

La inteligencia artificial fue aplicada en dos momentos clave:

- Recuperación aumentada con PathRAG, que actuó como sistema de filtrado semántico para construir las rutas más relevantes desde el grafo de conocimiento, aplicando poda por flujo acumulado para garantizar calidad y eficiencia informativa.
- Generación de contenido adaptativo, que utilizó las rutas seleccionadas como base para la producción textual mediante IA generativa. El modelo fue ajustado

para limitar la generación creativa excesiva y asegurar precisión, brevedad y alineación con los ODS y el marco curricular.

El uso de IA fue supervisado por el equipo investigador, garantizando la aplicación de principios de transparencia, explicabilidad y equidad algorítmica. Se tomaron medidas específicas para asegurar la anonimización de datos, el consentimiento informado de los participantes y la trazabilidad de los contenidos generados. Estos procedimientos buscaron salvaguardar la integridad del proceso formativo, evitar sesgos y asegurar la coherencia ética y pedagógica de la intervención.

Limitaciones del estudio

Este estudio presenta diversas limitaciones que es necesario considerar al interpretar los hallazgos:

- El diseño cuasi-experimental sin grupo de control impide establecer relaciones causales concluyentes.
- El muestreo intencional y focalizado en un único máster restringe la representatividad y dificulta la extrapolación a otros niveles educativos o disciplinas.
- La duración acotada de la intervención (septiembre-diciembre de 2024) limita la evaluación de impactos a largo plazo.
- Algunos instrumentos empleados fueron desarrollados ad hoc y, aunque validados por juicio experto y prueba piloto, no fueron sometidos a procesos de validación estandarizados.
- No se realizó un cálculo previo de poder estadístico ni se estimaron tamaños de efecto esperados, lo cual afecta la precisión de las conclusiones inferenciales.

Estas limitaciones metodológicas afectan principalmente a la validez interna (posible influencia de variables no controladas) y a la validez externa (limitada generalización). Sin embargo, el carácter exploratorio, el uso de triangulación metodológica y la solidez del diseño instruccional compensan parcialmente estas debilidades, y sientan las bases para futuras investigaciones más robustas.

RESULTADOS

Los resultados obtenidos tras la implementación del sistema PathRAG con apoyo de IA generativa en el entorno del Máster Online en Ciberseguridad presentan mejoras estadísticamente significativas en todas las dimensiones evaluadas. Se analizaron cinco variables clave: participación, comprensión conceptual, desarrollo de habilidades, percepción de inclusión y satisfacción general. Los análisis incluyeron

comparativas pretest/posttest, evolución longitudinal y desglose por subgrupos. Además, se trianguló con evidencias cualitativas obtenidas a través de respuestas abiertas.

Mejora global y progresión formativa

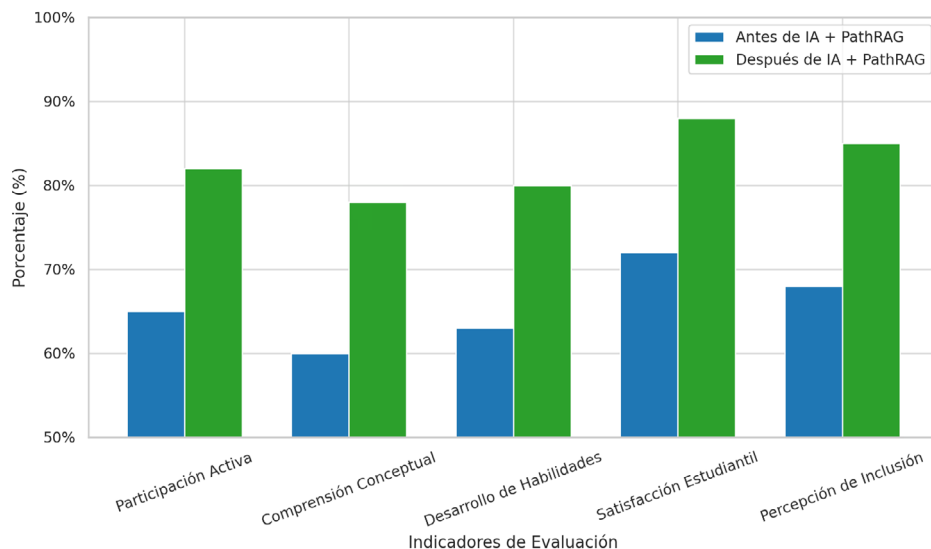
Los indicadores principales experimentaron un incremento estadísticamente significativo tras la intervención (ver Figura 2). Las mejoras observadas oscilan entre el 15 % y el 20 %, con tamaños de efecto moderados a grandes (Cohen's $d = 0.63 - 0.71$), lo que indica un impacto relevante en los indicadores clave de aprendizaje. Estas mejoras fueron consistentes a lo largo de los tres ejercicios prácticos, como confirma la prueba ANOVA de medidas repetidas: comprensión conceptual ($F(2,698) = 19.6$, $p < .001$, $\eta^2 = .22$), habilidades ($F(2,698) = 18.1$, $p < .001$, $\eta^2 = .19$), y participación ($F(2,698) = 15.7$, $p < .001$, $\eta^2 = .17$).

Esta evolución apunta a que la generación adaptativa de contenidos, apoyada en rutas personalizadas del grafo educativo, favorece un aprendizaje progresivo, ajustado al ritmo del alumnado. Estos hallazgos coinciden con estudios recientes como los de Chen et al. (2024) y van de Sande et al. (2023), que destacan el potencial de la IA generativa en la tutoría personalizada y la mejora de la motivación en contextos híbridos.

Los resultados indican una mejora global de entre el 15 % y el 20 %, con tamaños de efecto moderados a grandes. La Figura 2 sintetiza visualmente esta evolución.

Figura 2

Resultados del aprendizaje adaptativo con IA y PathRAG



Representa cinco indicadores clave:

Participación activa

La participación activa, medida a través del número de accesos, permanencia en la plataforma y finalización de tareas, mostró un incremento significativo. El porcentaje de estudiantes con seguimiento completo aumentó del 65 % al 82 %, con diferencias estadísticamente significativas ($t(349) = 11.78, p < .001$). En la Figura 2 se muestra la evolución de la participación antes y después de cada ejercicio.

Comprensión conceptual

El test de conocimientos aplicado en cada ejercicio reflejó una mejora media del 60 % al 78 %. El mayor incremento se observó en el segundo ejercicio (+18 %), donde los contenidos personalizados generados por la IA incluían simulaciones prácticas contextualizadas. La Tabla 2 presenta la evolución por ejercicio e indicador.

Desarrollo de habilidades

A partir de las rúbricas docentes y la autoevaluación, se identificó un aumento en habilidades clave (análisis crítico, resolución de problemas, pensamiento sistémico) con una media global del 63 % al 80 %. El 87 % del alumnado mejoró en al menos dos dimensiones de la rúbrica entre el primer y tercer ejercicio. En la Figura 3 se observa la progresión acumulada.

Satisfacción estudiantil

El subgrupo de estudiantes con necesidades educativas específicas mostró mejoras sustanciales. En ellos, la participación aumentó del 55 % al 75 %, la comprensión del 50 % al 70 %, y la satisfacción pasó del 60 % al 82 %.

Los datos de la escala validada de satisfacción ($\alpha = 0.89$) mostraron un nivel medio-alto de aceptación del sistema PathRAG + IA generativa. El **89 % de los estudiantes valoró positivamente** la integración de contenidos personalizados, mientras que el **92 % indicó que el sistema les ayudó a comprender mejor los temas complejos**. Los ítems relacionados con la utilidad, claridad, contextualización y grado de personalización obtuvieron puntuaciones superiores a 4,2 en una escala de 5 puntos.

Percepción de inclusión

El 89 % de los estudiantes valoró positivamente el uso de inteligencia artificial para personalizar recursos, indicando que mejoró su motivación y comprensión. La percepción general de la utilidad del sistema (media = 4.3/5) fue mayor entre quienes recibieron más prompts generados. La satisfacción total aumentó del 72 % al 88 % tras la intervención.

Resultados por ejercicio práctico

La Tabla 3 muestra la evolución desglosada por ejercicio. Se observa un patrón consistente de mejora, con incrementos significativos en cada indicador a lo largo de los tres ejercicios. Por ejemplo, en comprensión conceptual, la media pasó del 58 % (Ej. 1) al 78 % (Ej. 3); en habilidades, del 60 % al 80 %; y en participación, del 63 % al 84 %.

La prueba ANOVA de medidas repetidas confirmó diferencias significativas en los tres momentos para comprensión ($F(2,698) = 19.6, p < .001, \eta^2 = .22$), habilidades ($F(2,698) = 18.1, p < .001, \eta^2 = .19$) y participación ($F(2,698) = 15.7, p < .001, \eta^2 = .17$).

Tabla 3
Resultados por ejercicio

Ejercicio	Indicador	Antes	Después
Ejercicio 1	Comprensión (%)	58	76
Ejercicio 1	Habilidades (%)	60	78
Ejercicio 1	Participación (%)	63	80
Ejercicio 2	Comprensión (%)	61	77
Ejercicio 2	Habilidades (%)	62	79
Ejercicio 2	Participación (%)	66	83
Ejercicio 3	Comprensión (%)	62	78
Ejercicio 3	Habilidades (%)	64	80
Ejercicio 3	Participación (%)	67	84

Estas mejoras reflejan una progresión pedagógica y una adaptación efectiva del contenido a lo largo del proceso formativo.

Impacto en estudiantes con necesidades educativas específicas

La Tabla 4 muestra los resultados del subgrupo de estudiantes que reportaron limitaciones de acceso o necesidades educativas especiales. Este grupo partía de niveles más bajos (por ejemplo, comprensión inicial del 50 % y participación del 55 %), pero tras la intervención se alcanzaron valores de hasta el 75 % y 70 %, respectivamente.

Tabla 4
Impacto en estudiantes con necesidades educativas

Indicador	Antes	Después
Participación (%)	55	75
Comprensión (%)	50	70
Habilidades (%)	52	72
Satisfacción (%)	60	82

Este hallazgo subraya el valor inclusivo del enfoque PathRAG, especialmente al ofrecer rutas personalizadas y contenidos adaptativos generados automáticamente en función del perfil y la trayectoria del estudiante.

Impacto en estudiantes con necesidades educativas

El subgrupo con necesidades educativas específicas mostró **una evolución especialmente destacada**, mejorando desde niveles iniciales más bajos (50-55 %) hasta valores comparables al resto de la muestra (70-75 %). Este comportamiento sugiere que el sistema **compensa desigualdades de partida**, en línea con los principios de equidad educativa del ODS 4 y los hallazgos de Cobo et al. (2022) sobre IA inclusiva.

Las evidencias cualitativas refuerzan este análisis. Declaraciones como “el contenido se adaptaba a mi ritmo” (Est. 42, con dislexia) o “las explicaciones eran más claras cuando se ajustaban a mis errores” (Est. 117, con problemas de conectividad) ilustran el efecto subjetivo positivo de la personalización algorítmica, ya documentado también por Selwyn et al. (2023).

Las respuestas cualitativas nos ponen de manifiesto que la personalización adaptativa fue clave en su progresión. A continuación, se muestran ejemplos representativos de la codificación temática:

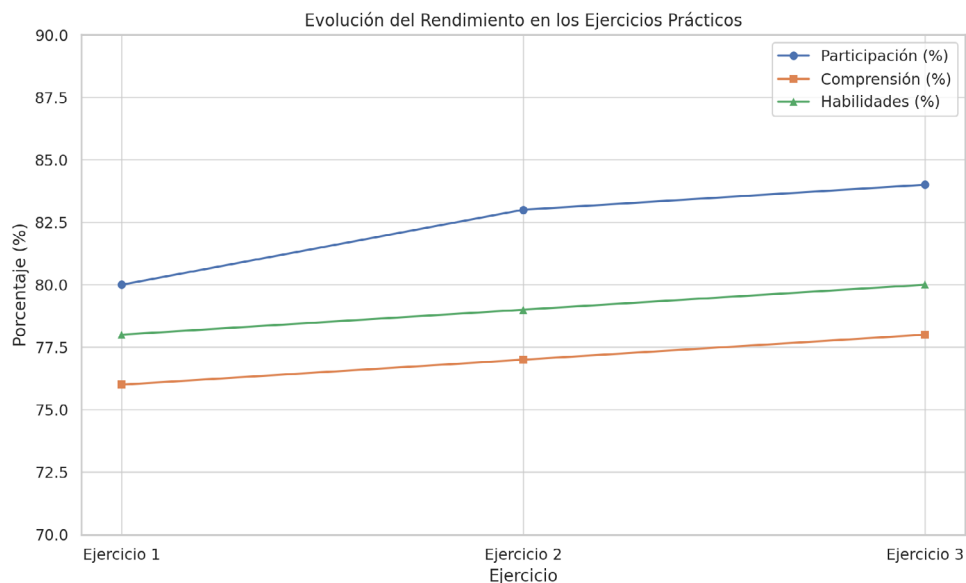
“Por primera vez, sentí que el contenido se adaptaba a mi ritmo y no al revés” (Est. 42, con dislexia)

“Las explicaciones eran más claras cuando se ajustaban a mis errores” (Est. 117, acceso limitado)

Evolución progresiva y análisis longitudinal

La Figura 3 presenta la evolución continua de los tres indicadores principales, mostrando una mejora progresiva por ejercicio. La comprensión aumentó del 58 % al 78 %, habilidades del 60 % al 80 %, y participación del 63 % al 84 %. Estos incrementos presentan una curva de aprendizaje sostenida.

Figura 3
Evolución del rendimiento en los ejercicios prácticos



La consistencia en la mejora respalda la utilidad del sistema como herramienta de aprendizaje adaptativo más allá de intervenciones puntuales.

Percepción de inclusión y satisfacción

La satisfacción general aumentó del 72 % al 88 %. Los cuestionarios Likert ($\alpha = 0.89$, test-retest = 0.81) reflejan una **alta aceptación del sistema PathRAG**, con valoraciones superiores a 4.2/5 en utilidad, claridad, contextualización y personalización. El 92 % del alumnado señaló que la IA generativa les ayudó a **comprender mejor contenidos complejos**, lo que se alinea con los resultados de Luckin et al. (2022) y Pedro et al. (2021) sobre la eficacia de los sistemas adaptativos mediados por IA.

- El análisis temático identificó patrones como:
- Claridad en explicaciones personalizadas
- Reducción de carga cognitiva
- Mayor motivación y autonomía

Distribución geográfica y contexto multicultural

La diversidad geográfica y profesional de la muestra ($n = 350$, con representación de más de 10 países) aporta cierta validez externa, pero la limitación a un único máster técnico, el muestreo no probabilístico por conveniencia, y la ausencia de grupo de control impiden generalizar los resultados sin reservas. La aplicabilidad a niveles no universitarios o a otras áreas disciplinares debería verificarse empíricamente.

No obstante, el marco técnico de PathRAG —basado en grafos semánticos y prompts dirigidos— podría ser adaptado fácilmente a otros dominios curriculares, siempre que se disponga de una base documental bien estructurada y de supervisión docente experta.

La Tabla 5 muestra una distribución geográfica diversa, con predominancia de estudiantes de España (34 %), México (17 %) y Colombia (14 %). Esta heterogeneidad refuerza la validez contextual del estudio y su potencial aplicación en entornos educativos globales y sostenibles.

Tabla 5
Distribución geográfica de la muestra

País	Nº de estudiantes
España	120
México	60
Colombia	50
Argentina	40
Perú	30
Ecuador	25
Chile	15
Otros	10

Esta diversidad refuerza la aplicabilidad del modelo PathRAG en contextos educativos globales, sostenibles y multiculturales, característica especialmente relevante para programas a distancia.

Riesgos potenciales: sesgo algorítmico y dependencia tecnológica

No obstante, también se identifican riesgos que limitan la generalización y transferibilidad del modelo. En primer lugar, la estructura del grafo educativo, aunque curada manualmente, podría incorporar sesgos epistemológicos, derivados del sesgo en los materiales docentes o de decisiones humanas no auditadas. Este

fenómeno ha sido señalado por Birhane (2023) y Holmes et al. (2021) como una fuente crítica de desigualdad algorítmica en entornos educativos.

Además, la dependencia del modelo GPT-3.5-turbo de OpenAI, aunque funcional en esta investigación, reduce la replicabilidad y transparencia, al tratarse de una solución cerrada y no explicable en su totalidad (Bender et al., 2021; Liu et al., 2023). Estas limitaciones técnicas comprometen la soberanía educativa y podrían afectar la escalabilidad del sistema en contextos con restricciones de acceso tecnológico o normativas más estrictas.

DISCUSIÓN

Los resultados obtenidos sugieren que el sistema PathRAG, que combina recuperación aumentada basada en grafos de conocimiento con generación de contenidos vía IA, indica un impacto positivo en la personalización y equidad del aprendizaje en contextos de educación superior en línea. Las mejoras observadas en participación, comprensión, habilidades e inclusión, tanto a nivel global como en subgrupos vulnerables, apuntan a la eficacia pedagógica del modelo, aunque no permiten establecer relaciones causales directas por el diseño cuasi-experimental empleado.

A diferencia de enfoques tradicionales centrados en búsquedas de contenido estático, PathRAG genera rutas semánticas personalizadas en tiempo real. Estas rutas están adaptadas al perfil del estudiante y optimizadas mediante poda por flujo acumulado. Este enfoque, respaldado por estudios recientes (Liang et al., 2023; Luckin et al., 2022), sugiere una mejora en la progresión autónoma del aprendizaje y una reducción de la sobrecarga cognitiva, dos factores clave en entornos virtuales masivos y diversos.

Particularmente significativo fue el aumento de rendimiento en estudiantes con dificultades de aprendizaje, discapacidades sensoriales o barreras tecnológicas. Este subgrupo partía de niveles inferiores y logró mejoras comparables al conjunto de la muestra, lo cual indica un efecto democratizador del sistema. Estas observaciones están alineadas con investigaciones recientes (UNESCO, 2023; Cobo et al., 2022) que reconocen el papel de la IA bien diseñada en la reducción de brechas de aprendizaje.

Desde el punto de vista contextual, el máster analizado —con estudiantes de más de diez países y niveles heterogéneos de competencia digital— ofreció un laboratorio natural para evaluar la capacidad de adaptación intercultural del sistema. El análisis no mostró diferencias significativas por nacionalidad o edad, lo que refuerza la potencial transferibilidad de PathRAG a entornos internacionales, multiculturales y distribuidos.

No obstante, deben reconocerse varias limitaciones. La dependencia de GPT-3.5-turbo como motor generativo plantea riesgos importantes: sesgos algorítmicos, opacidad del modelo y falta de control pedagógico sobre las respuestas (Bender et al., 2021; OECD, 2023). Aunque se implementaron filtros y ajustes, persiste la necesidad

de utilizar modelos auditables, entrenados en corpus educativos validados, y con supervisión humana sistemática.

También se detecta una posible fuente de sesgo en la construcción del grafo educativo, dado que se basa en materiales docentes existentes que pueden reflejar limitaciones o desequilibrios en el currículo. La validación manual parcial y el enfoque de curación instruccional intensiva, aunque rigurosos, requieren recursos técnicos y humanos difíciles de replicar en otros contextos con menor madurez digital.

Metodológicamente, la ausencia de grupo de control, el muestreo no probabilístico y la duración limitada de la intervención afectan la validez interna y la capacidad de generalización. Sin embargo, la triangulación entre resultados cuantitativos y evidencias cualitativas permitió interpretar patrones más allá de los promedios: por ejemplo, estudiantes con baja puntuación inicial relataron cómo los resúmenes personalizados generados por IA “les permitieron retomar desde lo básico sin desmotivarse”, mientras que perfiles avanzados valoraron la posibilidad de profundizar en contenidos complejos.

Estos hallazgos se inscriben en una línea creciente de investigación sobre IA adaptativa e inclusiva en educación. Trabajos recientes como los de Jivet et al. (2021) y Zawacki-Richter et al. (2019) refuerzan la idea de que la personalización algorítmica puede fomentar la autorregulación del aprendizaje y promover trayectorias más justas cuando se diseñan con base en principios pedagógicos, éticos y de equidad digital.

Líneas de investigación futura

A partir de los hallazgos y limitaciones detectadas, se proponen varias líneas de investigación para futuros trabajos:

- Estudios longitudinales que analicen el impacto sostenido del sistema a lo largo de un curso académico completo.
- Diseños experimentales con grupo de control, que permitan aislar efectos y estimar mejor la causalidad.
- Comparación entre diferentes modelos generativos (e.g., GPT-4, Claude, LLaMA) para evaluar la calidad, sesgo y eficiencia de los contenidos generados.
- Auditorías semánticas del grafo educativo, enfocadas en detectar sesgos cognitivos, culturales o de género.
- Aplicaciones en otros contextos disciplinares o educativos, como educación secundaria, FP o programas para adultos con baja alfabetización digital.

CONCLUSIONES Y CONSIDERACIONES FINALES

Este estudio indica que la combinación de inteligencia artificial generativa con recuperación semántica basada en grafos, implementada mediante PathRAG, constituye una propuesta metodológica innovadora con potencial para fortalecer el aprendizaje adaptativo en entornos virtuales, híbridos e inclusivos. Al generar rutas personalizadas sobre una base curricular estructurada, el sistema facilita una progresión coherente del aprendizaje, ajustada a los errores, necesidades y trayectorias del estudiante.

Sin embargo, las evidencias recogidas deben interpretarse dentro de los límites del diseño metodológico. La ausencia de grupo de control, el muestreo por conveniencia, la duración breve de la intervención, y el uso de instrumentos parcialmente ad hoc limitan la validez interna y externa de los resultados. En este sentido, los hallazgos deben considerarse exploratorios, aunque prometedores, y no como prueba definitiva de eficacia.

A pesar de estas limitaciones, los resultados sugieren que el sistema beneficia especialmente a estudiantes con necesidades educativas diferenciadas, como dificultades de aprendizaje, discapacidades sensoriales o barreras tecnológicas, promoviendo un aprendizaje más equitativo. Esto refuerza su alineación con los Objetivos de Desarrollo Sostenible, en particular con el ODS 4 (educación de calidad) y el ODS 10 (reducción de desigualdades), y subraya su potencial para contribuir a modelos educativos más inclusivos y sostenibles.

Desde el punto de vista técnico, el éxito del sistema depende en gran medida de la calidad y coherencia del grafo de conocimiento. Su construcción requiere procesos exigentes de curación, diseño instruccional y aplicación de técnicas de procesamiento del lenguaje natural. Asimismo, el modelo depende actualmente del uso exclusivo de GPT-3.5-turbo, lo que introduce riesgos de dependencia tecnológica y sesgo algorítmico, así como limitaciones en cuanto a transparencia y control pedagógico.

En cuanto a su aplicabilidad futura, el enfoque PathRAG es escalable y adaptable a otros niveles educativos (como secundaria o formación profesional) y a diferentes áreas disciplinares (humanidades, ciencias sociales o salud), siempre que se cuente con una base documental adecuada y un equipo docente capacitado en alfabetización algorítmica. También puede aplicarse en programas de formación continua, contextos multiculturales y entornos con alta diversidad estudiantil.

En el plano ético, el estudio fue desarrollado bajo los principios del RGPD y la supervisión docente, sin automatizar decisiones evaluativas ni comprometer la agencia del profesorado. No obstante, el avance hacia una educación con IA exige no solo herramientas técnicas, sino marcos claros de gobernanza, auditorías de equidad y formación docente crítica y especializada.

En conclusión, PathRAG no debe entenderse como una solución cerrada ni como un sustituto del rol docente, sino como una herramienta estratégica de apoyo al diseño pedagógico personalizado. Su valor reside en aumentar la capacidad de respuesta de

los sistemas educativos ante la diversidad, facilitar trayectorias formativas ajustadas, y contribuir a un modelo de educación más ético, justo y resiliente. La inteligencia artificial generativa, utilizada de forma crítica y supervisada, puede convertirse en un motor clave para democratizar el acceso al conocimiento y construir ecosistemas educativos más inclusivos y sostenibles.

REFERENCIAS

- Ainscow, M. (2020). *Promoting inclusion and equity in education: Lessons from international experiences*. UNESCO. <https://doi.org/10.1080/20020317.2020.1729587>
- Anderson, T., Borup, J. y West, R. E. (2020). Learning analytics in higher education: A review of the literature. *Journal of Research on Technology in Education*, 52(4), 433-451. <https://doi.org/10.1080/15391523.2020.1739846>
- Bender, E. M., Gebru, T., McMillan-Major, A. y Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610-623). <https://doi.org/10.1145/3442188.3445922>
- Birhane, A. (2023). Algorithmic injustice: A relational ethics approach. *Patterns*, 4(2), 100693. <https://doi.org/10.1016/j.patter.2023.100693>
- Braun, V. y Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101. <https://doi.org/10.1191/1478088706qp0630a>
- Chen, B., Guo, Z., Yang, Z., Chen, Y., Chen, J., Liu, Z., Shi, C. y Yang, C. (2024). PathRAG: Pruning graph-based retrieval augmented generation with relational paths. *arXiv*. <https://doi.org/10.48550/arXiv.2502.14902>
- Chen, X., Wang, L. y Li, Z. (2024). Generative AI in higher education: Ethical and pedagogical considerations for ChatGPT. *AI & Society*. <https://doi.org/10.1007/s00146-024-01720-9>
- Cobo, C., Morales, A. y Navarro, C. (2022). Artificial intelligence in education: Equity and ethics in Latin America. *International Journal of Educational Technology in Higher Education*, 19, 32. <https://doi.org/10.1186/s41239-022-00312-3>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2ª ed.). Lawrence Erlbaum Associates.
- Conati, C., Holstein, K. y Roll, I. (2021). The future of intelligent tutoring systems: A survey. *IEEE Transactions on Learning Technologies*, 14(4), 464-484. <https://doi.org/10.1109/TLT.2021.3071740>
- Conati, C. y Kardan, S. (2013). User modeling in intelligent tutoring systems: A survey. En P. Brusilovsky, G. D. Verbert y P. De Bra (Eds.), *Adaptive and adaptable learning* (pp. 45-63). Springer. https://doi.org/10.1007/978-3-642-40866-0_3
- Floridi, L. (2019). *The ethics of artificial intelligence*. Oxford University Press.
- García-Peñalvo, F. J. (2022). *Tecnología y Educación: Hacia una visión transformadora*. Ediciones Universidad de Salamanca.
- Holmes, W., Porayska-Pomsta, K., Holstein, K. y Rodrigo, M. M. T. (2021). Ethics of AI in education: Towards a community-wide framework. *European Journal of Education*, 56(4), 541-558. <https://doi.org/10.1111/ejed.12481>
- Ifenthaler, D. y Yau, J. Y.-K. (2020). *Utilising learning analytics to support study success in higher education*. Springer. <https://doi.org/10.1007/978-3-319-64792-0>

- Jivet, I., Scheffel, M., Drachsler, H. y Specht, M. (2021). Awareness is not enough: Pitfalls of learning analytics dashboards in the educational practice. *British Journal of Educational Technology*, 52(6), 1864-1880. <https://doi.org/10.1111/bjet.13168>
- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Liang, W., Tadesse, G. A., Ho, D., Li, F., Liu, S., Yu, H. y Sun, J. (2023). Personalized education in the era of big data. *IEEE Transactions on Knowledge and Data Engineering*, 35(3), 2245-2261. <https://doi.org/10.1109/TKDE.2021.3117424>
- Liu, Y., Tamkin, A., Geng, X. y Ganguli, D. (2023). Lost in the middle: How LLMs use long contexts. *Transactions of the Association for Computational Linguistics*, 11, 1200-1218. https://doi.org/10.1162/tacl_a_00606
- Lu, X., van de Sande, B. y Marsh, J. A. (2024). Hybrid learning environments and adaptive support. *Computers & Education*, 210, 104975. <https://doi.org/10.1016/j.compedu.2023.104975>
- Luckin, R., Holmes, W., Griffiths, M. y Forcier, L. B. (2022). *Artificial intelligence and human learning: Partners in progress*. OECD Publishing.
- OECD. (2023). *The OECD AI principles: An overview*. OECD Publishing.
- Pedro, F., Subosa, M., Rivas, A. y Valverde, P. (2021). *Artificial intelligence in education: Challenges and opportunities for sustainable development*. UNESCO.
- Ruipérez-Valiente, J. A., Cobos, R. y Muñoz-Merino, P. J. (2022). Knowledge graphs in education: A systematic literature review. *IEEE Transactions on Learning Technologies*, 15(4), 485-500. <https://doi.org/10.1109/TLT.2022.3190858>
- Selwyn, N., Macgilchrist, F. y Williamson, B. (2023). EdTech and equity: Critical perspectives on the promises of digital education. *Learning, Media and Technology*, 48(1), 1-11. <https://doi.org/10.1080/17439884.2022.2156041>
- Susnjak, T. (2022). ChatGPT: The end of online exam integrity? *arXiv*. <https://doi.org/10.48550/arXiv.2212.09292>
- UNESCO. (2020). *Global education monitoring report 2020: Inclusion and education: All means all*. UNESCO. <https://doi.org/10.54676/JJNK6989>
- UNESCO. (2022). *Reimagining our futures together: A new social contract for education*. UNESCO.
- UNESCO. (2023). *AI and education: Guidance for policy-makers*. <https://unesdoc.unesco.org/ark:/48223/pf0000384803>
- van de Sande, B., Marsh, J. A. y Lu, X. (2023). Large language models in education: Opportunities and challenges. *Computers & Education: Artificial Intelligence*, 4, 100132. <https://doi.org/10.1016/j.caeai.2023.100132>
- Zawacki-Richter, O., Marín, V. I., Bond, M. y Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—Where are the educators? *International Journal of Educational Technology in Higher Education*, 16, 39. <https://doi.org/10.1186/s41239-019-0171-0>
- Zhang, D. y Wang, X. (2025). Adaptive learning systems: A review and future directions. *Journal of Educational Technology & Society*, 28(1), 1-15.
- Zhang, D., Wang, X. y Chen, X. (2021). Adaptive learning technologies in higher education: A meta-analysis of effectiveness. *Educational Research*

Review, 33, 100390. <https://doi.org/10.1016/j.edurev.2021.100390>
Zhou, M. y Wang, D. (2020). Ethical considerations in AI-supported education: A perspective from responsible innovation. *AI & Society*, 35(3), 601-611. <https://doi.org/10.1007/s00146-019-0910-1>

Fecha de recepción del artículo: 1 de junio de 2025

Fecha de aceptación del artículo: 11 de agosto de 2025

Fecha de aprobación para maquetación: 15 de septiembre de 2025

Fecha de publicación en OnlineFirst: 12 de octubre de 2025

Fecha de publicación: 1 de enero de 2026

Nivelación inteligente: estrategia de aprendizaje adaptativo con IA en educación superior

Smart leveling: an AI-driven adaptive learning strategy in higher education



-  Patricia Aldape-Valdes - *Tecnológico de Monterrey, ITESM (México)*
-  Elvira G. Rincon-Flores - *Tecnológico de Monterrey, ITESM (México)*
-  Leticia Castano - *Tecnológico de Monterrey, ITESM (México)*
-  Sadie Guerrero - *Tecnológico de Monterrey, ITESM (México)*

RESUMEN

La nivelación académica es uno de los principales retos de la educación superior, particularmente en las disciplinas STEM, un desafío que se ha intensificado tras el confinamiento académico y ante el cual los cursos remediales tradicionales no han mostrado los resultados esperados. Este estudio evaluó el impacto de una estrategia de aprendizaje adaptativo (EAA) orientada a reforzar conocimientos previos en 1,309 estudiantes de primer año inscritos en las unidades formativas de Pensamiento Computacional, Modelación Matemática Fundamental, Razonamiento Matemático y Pensamiento Matemático de una universidad privada en México. A diferencia de los enfoques remediales convencionales, la EAA ofrece una experiencia de aprendizaje flexible y centrada en el estudiante mediante breves módulos adaptativos integrados en los cursos regulares. Estos módulos, respaldados por una plataforma de inteligencia artificial, generan rutas personalizadas de aprendizaje, adaptan contenidos, producen analíticas y facilitan la toma de decisiones docentes basadas en datos. La investigación adoptó un enfoque mixto (CUAN > CUAL) y un diseño cuasiexperimental con muestras emparejadas para asegurar la comparabilidad entre grupos. Los resultados mostraron diferencias estadísticamente significativas en el desempeño académico a favor de los grupos experimentales respecto a los grupos control. Además, estudiantes y profesores evaluaron positivamente la EAA en términos de utilidad y experiencia de aprendizaje. Los hallazgos sugieren que una estrategia adaptativa, integral y mediada por tecnología constituye una alternativa eficaz para abordar la nivelación académica en cursos universitarios iniciales, ofreciendo ventajas sustanciales sobre los métodos remediales tradicionales.

Palabras clave: aprendizaje adaptativo; inteligencia artificial; nivelación; educación superior.

ABSTRACT

Academic leveling is one of the main challenges in higher education, particularly in STEM disciplines—a challenge that has intensified following academic lockdowns and for which traditional remedial courses have not yielded the expected results. This study evaluated the impact of an adaptive learning strategy (ALS) designed to reinforce prior knowledge in 1,309 first-year students enrolled in the courses of Computational Thinking, Fundamental Mathematical Modeling, Mathematical Reasoning, and Mathematical Thinking at a private university in Mexico. Unlike conventional remedial approaches, the ALS offers a flexible, student-centered learning experience through brief adaptive modules integrated into regular courses. These modules, supported by an artificial intelligence platform, generate personalized learning pathways, adapt content, produce analytics, and facilitate data-driven instructional decision-making. The research adopted a mixed-methods approach (QUAN > QUAL) and a quasi-experimental design with matched samples to ensure group comparability. Results revealed statistically significant differences in academic performance in favor of the experimental groups compared to the control groups. Additionally, both students and instructors evaluated the ALS positively in terms of usefulness and learning experience. Findings suggest that an adaptive, integrated, and technology-mediated strategy is a promising alternative to address academic leveling in introductory university courses, offering substantial advantages over traditional remedial methods.

Keywords: adaptive learning; artificial intelligence; academic leveling; higher education.

INTRODUCCIÓN

En las últimas décadas, el Aprendizaje Adaptativo (AA) ha emergido como una estrategia educativa innovadora orientada a personalizar el proceso de enseñanza-aprendizaje, respondiendo a las necesidades, ritmos y estilos de cada estudiante (Ochukut et al., 2023). Su desarrollo ha estado estrechamente vinculado a los avances en tecnología educativa, analítica de datos y al desarrollo de la Inteligencia Artificial (IA), lo que ha permitido diseñar experiencias de aprendizaje más dinámicas y personalizadas. Estas innovaciones ofrecen la posibilidad de afrontar uno de los grandes desafíos de la educación contemporánea: transformar las prácticas de enseñanza y aprendizaje para acelerar el progreso educativo (UNESCO, 2019).

En este contexto, el AA se presenta como una alternativa sólida y prometedora para abordar uno de los retos más persistentes en la educación superior: la nivelación académica (Dagunduro et al., 2024; du Plooy et al., 2024; Shi y Liu, 2025). Al ingresar a la universidad, un número considerable de estudiantes presentan deficiencias en conocimientos fundamentales particularmente en disciplinas STEM (Ciencia, Tecnología, Ingeniería y Matemáticas), donde los vacíos en conceptos clave, el razonamiento lógico o las habilidades cuantitativas suelen limitar su desempeño académico (Ross et al., 2022). Los enfoques tradicionales de nivelación, basados en cursos remediales estandarizados, han mostrado limitaciones al no considerar las particularidades de cada estudiante. Por ello, se requiere una estrategia educativa diferente, que se enfoque en los requerimientos individuales y en el desempeño real de cada alumno. Antes de abordar la Estrategia de Aprendizaje Adaptativo (EAA) propuesta en este estudio, se expondrán algunos fundamentos teóricos que la sustentan.

Nivelación, una problemática persistente

La falta de conocimientos previos en los cursos iniciales de las carreras profesionales se ha agudizado tras el confinamiento causado por la pandemia de Covid-19. Las teorías clásicas del aprendizaje destacan la relevancia del conocimiento previo para facilitar la comprensión de nuevos conceptos. Wood et al. (1976) planteó que los estudiantes requieren un apoyo estructurado para construir sobre sus conocimientos existentes; sin embargo, cuando este conocimiento previo es insuficiente, el proceso de andamiaje se vuelve más intensivo y complejo.

El aprendizaje ocurre de manera más efectiva cuando se trabaja ligeramente por encima del nivel actual del estudiante, siempre que exista algún conocimiento previo que sirva como base (Vygotsky, 1978). En ausencia de este conocimiento, se requiere un enfoque remedial para facilitar el aprendizaje. Más adelante, Sweller (1988) indicó que la falta de conocimientos previos incrementa la carga cognitiva, lo cual dificulta el proceso de aprendizaje. Este planteamiento coincide con la teoría de

Vygotsky en cuanto a la necesidad de aplicar estrategias de nivelación para apoyar al estudiante.

En el metaanálisis realizado por Hattie (2008), que recopila más de 800 estudios sobre los factores que influyen en el aprendizaje de los estudiantes, se destaca que el conocimiento previo es fundamental, ya que proporciona la base sobre la cual se construyen nuevos aprendizajes. Los estudiantes con conocimientos previos relevantes tienden a comprender de manera más profunda y efectiva los nuevos conceptos. De hecho, Hattie (2017) encontró que la habilidad previa tiene un tamaño de efecto de 0.94 en el rendimiento académico, lo cual es considerado altamente significativo.

Diversas investigaciones han demostrado que los estudiantes de ingeniería con deficiencias en álgebra o cálculo presentan tasas más altas de abandono (Klingbeil et al., 2007; McKenna et al., 2000). Sweller et al. (1998), a través de la teoría de la carga cognitiva en ingeniería, señalaron que la falta de conocimientos previos sobrecarga a los estudiantes al enfrentar temas complejos como diseño o programación. Por lo tanto, la nivelación en estudios universitarios requiere la implementación de estrategias como cursos puente, programas de refuerzo, tutorías, cursos introductorios y enfoques pedagógicos diseñados para asegurar que los estudiantes dominen los prerrequisitos necesarios antes de avanzar en su formación (Felder y Brent, 2005; Seymour y Hewitt, 1997).

Algunas universidades han adoptado múltiples estrategias para apoyar la nivelación académica de los estudiantes de nuevo ingreso. Por ejemplo, la Universidad de California State cuenta con un programa de STEM *peer mentoring* en el que estudiantes de semestres avanzados orientan a los de primer ingreso, brindándoles información sobre las carreras y compartiendo su experiencia personal. Contar con un par que ofrezca guía y apoyo durante los primeros semestres ha demostrado ser una estrategia efectiva para mejorar las tasas de retención y facilitar una integración más rápida y positiva a la vida universitaria (Taha et al., 2015; Rockinson-Szapkiw et al., 2021).

Otras universidades, como Massachusetts Institute of Technology (MIT) y su programa MIT Introduction to Technology, Engineering and Science (MITES) o la Universidad Nacional Autónoma de México (UNAM) y su programa inspiración para mujeres STEM, entre muchos otros, han implementado programas específicos para fortalecer las carreras STEM. Estos programas no solo ofrecen asesorías académicas para nivelar a los estudiantes, sino que también brindan acompañamiento emocional a quienes inician su formación universitaria, con el objetivo de fortalecer sus conocimientos y prepararlos para enfrentar los desafíos propios de estas disciplinas.

Además de los factores estructurales que explican la falta de conocimientos previos en ciertas áreas, la pandemia por COVID-19 exacerbó significativamente esta problemática (Filatova et al., 2023; López del Puerto et al., 2021). Como respuesta, diversas instituciones desarrollaron programas remotos durante y después de la

pandemia, enfocados especialmente en apoyar la nivelación académica, con énfasis en los cursos STEM (Lasater et al., 2021; Schaal et al., 2021).

Sin embargo, algunas de las estrategias tradicionales utilizadas para la nivelación, como los cursos propedéuticos previos al inicio formal de las materias del programa, requieren ser repensadas. En el estudio realizado por González-Beltrán et al. (2018), se aplicaron técnicas de minería de datos para analizar si estos cursos de nivelación inciden de manera significativa en el desempeño o la aprobación de los estudiantes en las asignaturas de Cálculo. No obstante, los resultados no mostraron evidencia que respaldara dicha relación.

Si bien los cursos de nivelación previos al ingreso universitario han sido una de las estrategias más comunes implementadas por las universidades, no siempre resultan ser los más efectivos. Una de las principales razones es que estos cursos suelen abordar de manera general todos los contenidos del nivel educativo anterior, sin considerar que no todos los estudiantes presentan los mismos vacíos de conocimiento. Esto puede generar desmotivación y una percepción negativa sobre la utilidad de los contenidos. En este contexto, y gracias a los avances tecnológicos, el AA se presenta como una alternativa prometedora para la nivelación académica, ya que permite personalizar el proceso de aprendizaje, enfocándose únicamente en los temas que cada estudiante realmente necesita reforzar (du Plooy et al., 2024; Taylor et al., 2021).

Más allá de la enseñanza tradicional: aprendizaje adaptativo para recuperar y potenciar saberes.

El AA ha sido implementado desde hace algunos años, se sitúa bajo el paraguas del Aprendizaje personalizado, según Jing et al. (2023) el AA es un enfoque orientado a la personalización del aprendizaje y materializa el principio de una educación centrada en el alumno. En esta misma línea, Hernandez et al. (2022) definen el Aprendizaje Personalizado como un enfoque educativo que ajusta la experiencia de aprendizaje de cada estudiante en función de sus necesidades, fortalezas, capacidades e intereses individuales, brindando flexibilidad respecto a qué, cuándo, cómo y dónde aprender a lo largo de su plan de estudios y diversas experiencias formativas. De ese modo, diferentes estrategias y tecnologías pueden proporcionar distintos grados de autonomía y opciones, permitiendo que los estudiantes asuman un rol activo y protagónico de su propio aprendizaje dentro de estos entornos.

Para lograr la personalización en la educación se puede trabajar con diversas técnicas y metodologías con el apoyo de la tecnología; cuanto mayor sea el uso de herramientas tecnológicas, más precisa será dicha personalización. El Aprendizaje Adaptativo es una estrategia educativa que utiliza tecnología basada en la analítica de datos para adaptar la educación y crear un itinerario de aprendizaje personalizado cuyos contenidos han sido previamente diseñados por el profesor para ser eficaces y eficientes para cada alumno en función de su nivel de rendimiento,

perfil y necesidades de aprendizaje, permitiendo al docente identificar vacíos en la comprensión de contenidos para establecer acciones de mejora y ajustar su práctica educativa, optimizando el rendimiento de los alumnos (Rincon-Flores et al., 2024). De esta manera, el papel del profesor pasa de proveedor de contenidos a facilitador del aprendizaje. Sin embargo, para lograr estas eficiencias, los docentes deberán utilizar los sistemas tecnológicos de manera adecuada y eficiente (Cavanagh et al., 2020).

En una encuesta sobre informática y tecnología de la información en la educación superior, Simon y Zeng (2024) comentan que la mayoría de los líderes universitarios y docentes muestran una actitud positiva hacia el Aprendizaje Adaptativo (AA) y lo consideran una estrategia con alto potencial para mejorar el éxito estudiantil. Complementariamente, Rincon-Flores et al. (2023) demostraron que el AA asistido por tecnología constituye una herramienta eficaz para optimizar el aprendizaje. De forma similar, Sariyalçinkaya et al. (2021) señalan que el AA permite definir la trayectoria de enseñanza de cada alumno en función de su estilo de aprendizaje, a partir de la evaluación de sus conocimientos previos.

En el ámbito de los sistemas tecnológicos, es fundamental comprender la relevancia de los datos y su procesamiento. Long y Siemens (2011) señalan que, ante el fuerte impulso por incrementar la eficiencia y reformar la educación superior, ha resultado natural orientar el aprendizaje hacia nuevos formatos digitales. Esto ha derivado en una explosión de datos educativos recopilados, que incluyen flujos de actividad, interacciones de los estudiantes con sus pares y mentores, registros administrativos y otros valiosos rastros de información. La posibilidad de recopilar este volumen masivo de datos, analizarlos en tiempo real y aplicar técnicas de aprendizaje automático para identificar patrones se debe, en gran medida, al soporte que ofrece la inteligencia artificial (Miao et al., 2021). Estos elementos se encuentran hoy integrados en plataformas de aprendizaje adaptativo que incorporan IA como núcleo de su funcionamiento.

Así mismo, Fernández-Morante et al. (2022) destacan que la analítica del aprendizaje posee un enorme potencial para optimizar los procesos educativos; no obstante, para que dicho potencial se traduzca en resultados significativos, es imprescindible trasladar el enfoque de la tecnología en sí misma hacia un diseño instruccional centrado en el aprendizaje. En esta misma línea, Kara y Sevim (2013) coinciden en que las plataformas adaptativas emplean algoritmos, evaluaciones, retroalimentación, ajustes, intervención del instructor y diferentes medios para ofrecer nuevos materiales de aprendizaje a los estudiantes que han alcanzado el nivel esperado, o bien proporcionar actividades de refuerzo a los que aún no lo han logrado.

Una de las principales ventajas de las plataformas adaptativas radica en su capacidad para generar analíticas de aprendizaje y monitorear el progreso de los estudiantes tanto a nivel individual como grupal. Estas herramientas recaban datos en tiempo real sobre desempeño, los hábitos de estudio y las áreas de oportunidad,

lo que posibilita una toma de decisiones pedagógicas más oportuna y eficaz, tanto durante como después de la implementación. El uso estratégico de plataformas adaptativas con integración de IA no solo optimiza la intervención docente, sino que también fortalece la personalización y promueve la equidad en cursos STEM, caracterizados por una alta heterogeneidad académica (Koester et al., 2016; Li et al., 2023). Estas tecnologías facilitan rutas de aprendizaje diferenciadas y ajustadas a las necesidades específicas de cada estudiante, incrementando así las posibilidades de éxito académico.

En este sentido, Moskal et al. (2017) destacan que las plataformas adaptativas emplean un enfoque instruccional basado en datos y, en ciertos casos, no lineal, lo que favorece la nivelación académica y la adaptación continua del contenido. Esta perspectiva sitúa a las plataformas adaptativas como un recurso pedagógico de alto valor para enfrentar los retos de la enseñanza en entornos educativos complejos y diversos.

En relación con el diseño instruccional, el AA puede complementarse con otras metodologías como el Aprendizaje Autorregulado y el Microaprendizaje (Rincon-Flores et al., 2024). El primero requiere que el estudiante utilice estrategias de autorregulación, responda a la retroalimentación y mantenga procesos motivacionales interdependientes (Zimmerman, 1990), mientras que el segundo se caracteriza por ofrecer contenidos breves y enfocados, facilitando la asimilación de conceptos clave (Allela et al., 2020). Una EAA efectiva integra un diseño de contenidos estructurado, estrategias didácticas pertinentes, un modelo de implementación coherente y tecnologías que potencien la personalización (Rincon-Flores et al., 2024; Simon y Zeng, 2024; Shi y Liu, 2025).

Un ejemplo exitoso de la aplicación de una EAA en cursos STEM se observa en la Universidad Central de Florida (UCF), que rediseñó sus cursos de Física incorporando tecnología adecuada para atender a 4 500 estudiantes, logrando mejoras en el éxito estudiantil, incrementos en los resultados de aprendizaje y un mayor compromiso con el contenido (Dubey et al., 2023). De igual forma, en la India, el estado de Haryana, consciente del impacto positivo del AA, se convirtió en el primero en implementarlo a gran escala, seleccionando a un socio de tecnología educativa para proporcionar software y contenidos pertinentes a 500 000 tabletas distribuidas entre el alumnado de escuelas públicas en todo el país (Press Trust of India, 2022).

Por otra parte, los resultados del AA no siempre son exitosos o concluyentes. Un ejemplo es el estudio de Wang et al. (2023), donde, aunque los estudiantes del grupo experimental (aprendizaje adaptativo) mejoraron significativamente entre la preprueba y la postprueba, el grupo de control (enseñanza guiada por el profesor) no mostró diferencias significativas. Hallazgos similares fueron reportados por Eau et al. (2022), quienes encontraron beneficios únicamente en algunos estudiantes, particularmente en aquellos con alto desempeño. En la misma línea, Olmos-López et al. (2023) evaluaron un enfoque en el que los docentes personalizaban la enseñanza

a partir de un algoritmo predictivo sobre el rendimiento estudiantil; sin embargo, las predicciones resultaron útiles principalmente para estimar el desempeño grupal, por lo que la enseñanza no se personalizó a nivel individual, sino que se adaptó a las necesidades del grupo.

Modelo de nivelación inteligente: aprendizaje adaptativo con IA en programas STEM.

La EAA presentada en este estudio integra módulos adaptativos de nivelación dentro de cursos del primer año en programas STEM, abordando conocimientos básicos que los estudiantes deberían dominar al egresar del bachillerato. Cada módulo inicia con un examen diagnóstico obligatorio que define una ruta personalizada, permitiendo cursar solo los contenidos necesarios. Se desarrolla en las primeras semanas, pero permanece disponible para consulta durante todo el curso. La participación no es opcional y representa el 5 % de la calificación final, incentivando su aprovechamiento y asegurando que todos los estudiantes alcancen el nivel requerido.

La EAA contempla cuatro componentes clave para asegurar su efectividad pedagógica y tecnológica como son el diseño de contenidos, el modelo didáctico, el rol del profesor y la plataforma tecnológica.

Diseño de contenidos

El diseño de contenidos se basa en el microaprendizaje, dividiendo el contenido en pequeñas lecciones con objetivos específicos. Cada lección incluye contextualización, material de reconocimiento, ejemplos, ejercicios y un examen rápido final. Los recursos de reconocimiento son uno base y dos de reforzamiento opcionales. Los contenidos se presentan en diversos formatos como videos, infografías, lecturas y animaciones. El examen rápido tiene reactivos de diferente dificultad según el nivel del estudiante.

Modelo didáctico

Basado en principios de aprendizaje autorregulado y autoestudio, permite a los estudiantes controlar su aprendizaje y adaptarlo a sus necesidades. La estructura pedagógica está diseñada para el aprendizaje autónomo, gestionando tiempo, ritmo y necesidades. El estudiante empieza con un examen diagnóstico inicial; según los resultados, reflexiona y sigue una ruta personalizada que le sugiere lecciones a reforzar. Cada lección incluye contexto, videos, explicaciones, ejemplos, ejercicios y un examen rápido. Al finalizar todas las lecciones, se presenta un examen diagnóstico final y se registra la calificación obtenida.

Rol del profesor

Aunque el modelo de autoestudio promueve la autonomía, el rol del profesor es clave para el éxito. El profesor motiva al grupo, orienta sobre el módulo adaptativo y asegura que los exámenes diagnósticos se realicen en clase. Revisa la analítica de resultados generada con IA en la plataforma tecnológica para identificar y seguir necesidades específicas. El profesor es un acompañante estratégico en el proceso de nivelación.

Plataforma tecnológica

Para implementar la EAA fue necesario seleccionar una plataforma que permitiera a los docentes crear contenidos, integrar diversos materiales, y utilizar IA para análisis, seguimiento y generación de rutas personalizadas. Tras evaluar distintas opciones, se eligió **RealizeIT** por su integración con el LMS, variedad de formatos de contenido, analíticas detalladas, interfaz amigable y su motor de inteligencia adaptativa (*Adaptive Intelligence Engine*). Este motor emplea modelos matemáticos, estadísticos y algoritmos para registrar y analizar el progreso, fortalezas, debilidades y preferencias de cada estudiante (Howlin y Lynch, 2014), detectando patrones de aprendizaje y ofreciendo rutas personalizadas y adaptables en función de la interacción de cada estudiante con los contenidos, así como recomendaciones de recursos y prácticas para favorecer el dominio de conceptos.

METODOLOGÍA

El estudio adoptó un enfoque mixto, **CUAN > CUAL** (Creswell, 2015), donde los datos cualitativos complementaron los cuantitativos. Se aplicó un diseño cuasiexperimental, con grupos control y experimental en cuatro cursos de primer año: Pensamiento Computacional y Modelación Matemática Fundamental de la Escuela de Ingeniería y Ciencias (EIC), Razonamiento Matemático de la Escuela de Negocios (EN) y Pensamiento Matemático de la Escuela de Ciencias Sociales y Gobierno (ESCG). La investigación se desarrolló entre agosto y diciembre de 2024 en una universidad privada de modelo multicampus en México, con estudiantes de nivel socioeconómico medio y alto.

Objetivo de la investigación

Evaluar el impacto de la Estrategia Educativa de Aprendizaje Adaptativo (EAA), implementada a través de la plataforma tecnológica, en la **nivelación de conocimientos** previos de los estudiantes en las unidades formativas Pensamiento Matemático (MTC), Razonamiento Matemático (MRC), Modelación Matemática

(MMC) y Pensamiento Computacional (CTC). Asimismo, analizar la **experiencia de aprendizaje** de docentes y alumnos.

Muestra

El proceso de muestreo fue no probabilístico. La muestra total estuvo conformada por 1 309 estudiantes de las Escuelas: EIC, EN y ECSG. Las muestras por curso fueron emparejadas. Para ello, mediante el software SPSS (versión 29) se seleccionaron los participantes de los grupos experimental y de control con características equivalentes. Las variables consideradas para el emparejamiento fueron: calificación final de preparatoria (bachillerato), género, tipo de escuela de procedencia (pública o privada) y calificación obtenida en el pre-test. Como resultado, la muestra emparejada estuvo conformada por 734 estudiantes: 120 del curso *Pensamiento Matemático*, 50 de *Razonamiento Matemático*, 410 de *Modelación Matemática* y 154 de *Pensamiento Computacional*. En la Figura 1 se muestran algunas características sociodemográficas de la muestra total.

Figura 1

Resultados sociodemográficos en relación con el género y estatus de beca



Además, se encontró que el 88 % de los estudiantes provenían de escuelas privadas mientras que el 12 % de escuelas públicas. En cuanto al profesorado, participaron 40 docentes en total en ambos tratamientos; sin embargo, únicamente 11 respondieron la encuesta final. La edad promedio fue de 40 años y la experiencia docente media de 10 años.

Instrumentos

Se aplicó un pre y post-test de conocimientos previos en cada curso, estos instrumentos fueron desarrollados por los equipos de profesores encargados de elaborar los contenidos de los módulos de nivelación, luego fueron validados a través

de un pilotaje para asegurar que las pruebas contaran con los valores apropiados de los índices de dificultad, de discriminación y el alfa de Cronbach (ver Tabla 1).

Tabla 1

Resultados psicométricos de los instrumentos

Curso	Número de estudiantes	Índice de dificultad	Índice de discriminación	Alpha de Cronbach
Pensamiento Computacional	28	0.71	0.42	0.8
Razonamiento Matemático	36	0.61	0.48	0.86
Pensamiento Matemático	41	0.71	0.29	0.78
Modelación Matemática	34	0.57	0.33	0.74

Cada instrumento fue piloteado con el doble de reactivos previstos; tras el análisis, se seleccionaron los 16 con mejores resultados psicométricos, de modo que cada instrumento quedó conformado por 16 reactivos.

Al finalizar cada curso se aplicó una encuesta a estudiantes y profesores de los grupos control y experimental. El objetivo de este cuestionario fue recabar información sobre la experiencia de aprendizaje en relación con la EAA para la nivelación. Las dimensiones de este instrumento son: *Ruta de aprendizaje, Proceso de aprendizaje, Proceso de enseñanza, Contenido, Engagement, Usabilidad, Experiencia de usuario y Satisfacción*. El instrumento empleó una escala tipo Likert de intervalo continuo con rango de 0 a 100 puntos, diseñada para captar matices en la intensidad de la respuesta.

Proceso estadístico

Para analizar los resultados del pre y post-test se realizaron pruebas ANOVA, previa verificación de los supuestos de normalidad y homogeneidad de varianzas. Asimismo, se aplicó un análisis de Diferencia en Diferencias (DiD) para estimar el efecto real entre los tratamientos. En cuanto a la encuesta, dado que se utilizó una escala tipo Likert continua, se emplearon pruebas de ji cuadrada para identificar diferencias entre los docentes de los grupos experimentales. También, se efectuó un análisis del *uso de datos* proporcionado por la plataforma adaptativa. Por último, las preguntas abiertas incluidas en los cuestionarios fueron analizadas mediante categorización temática, con el fin de complementar y enriquecer los resultados cuantitativos.

Aspectos Éticos

El estudio se llevó a cabo de acuerdo con los principios éticos establecidos en la Declaración de Helsinki. Todos los estudiantes, profesores y autoridades académicas fueron informados sobre los objetivos y procedimientos del estudio. Se notificó expresamente a los participantes que toda la información recopilada sería tratada conforme a estrictos estándares de confidencialidad, garantizando el anonimato absoluto. Asimismo, todos los participantes otorgaron su consentimiento informado antes de iniciar su participación.

RESULTADOS

En la Tabla 2 se presentan los resultados del análisis comparativo entre los grupos control y experimental (adaptativo) en el post-test de conocimientos previos para cada curso. Se verificó el cumplimiento de los supuestos de normalidad (prueba de Shapiro–Wilk) e igualdad de varianzas (prueba de Levene). Dado que en el pre-test no se identificaron diferencias estadísticamente significativas entre los tratamientos, se optó por realizar un ANOVA de un factor sin covariables, prescindiendo del uso de ANCOVA.

Tabla 2

Prueba de diferencia de medias con muestras emparejadas

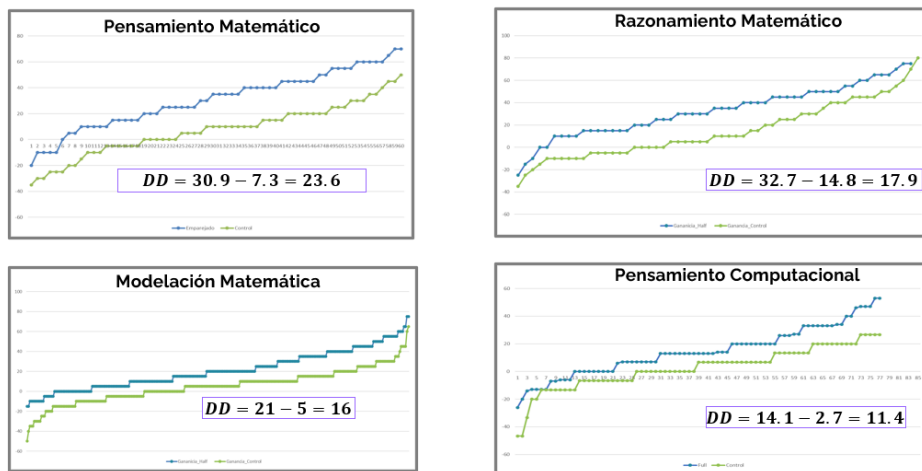
Curso	Tratamiento	n	Media (DE)	gl	F	p	η^2	Intervalo
Pensamiento Computacional	Adaptativo	74	84.7 (13.3)	146	32	.001	0.18	(.08, .29)
	Control	74	70.8 (16.2)					
Razonamiento Matemático	Adaptativo	25	72 (19.3)	48	7.9	.007	0.14	(.01, .32)
	Control	25	56.4 (19.4)					
Pensamiento Matemático	Adaptativo	44	68.3(24.1)	86	26	.001	0.23	(.09, .37)
	Control	44	45.3(17.6)					
Modelación Matemática	Adaptativo	196	73.8 (19.4)	394	45	.001	0.1	(.05, .16)
	Control	196	59.9 (21.7)					

Nota: DE = desviación estándar.

Los resultados evidenciaron diferencias estadísticamente significativas entre los grupos control y experimental, con un desempeño superior en los grupos experimentales ($\beta = 17.2$, IC95 % [11.4, 23.6], $p < .001$). La Figura 2 muestra las ganancias de aprendizaje estimadas mediante el modelo de Diferencia en Diferencias (DiD), donde la línea azul representa el desempeño de los grupos experimentales y la línea verde corresponde a los grupos control. La intervención adaptativa

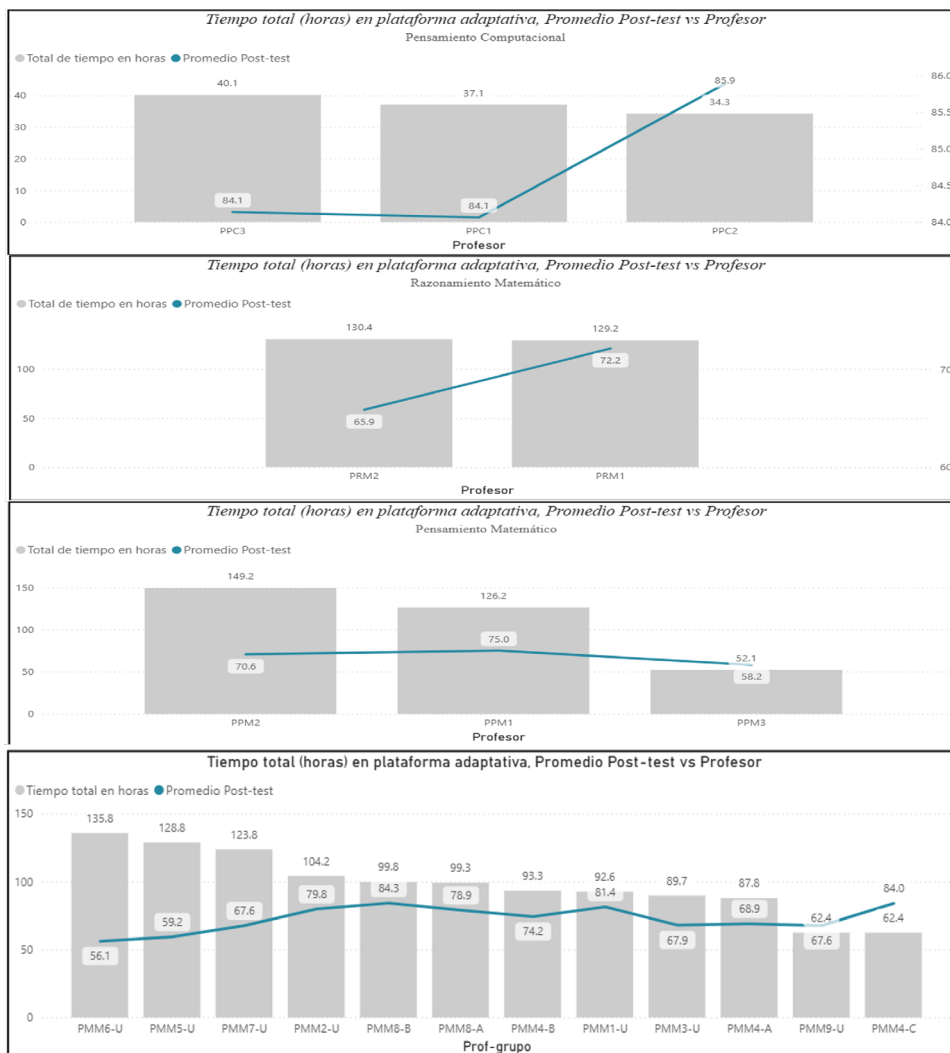
produjo incrementos netos (DiD) de 11.4, 17.9, 23.6 y 16.0 puntos en Pensamiento Computacional, Razonamiento Matemático, Pensamiento Matemático y Modelación Matemática, respectivamente. El promedio ponderado de la ganancia adicional fue de ≈ 16.2 puntos frente al control. Este patrón es consistente con un efecto atribuible a la intervención.

Figura 2
Resultados de la prueba DiD



En el análisis descriptivo de uso de datos no se identificó una relación clara entre el tiempo total de uso por grupo (en horas) y la calificación promedio obtenida en el post-test. Esto sugiere que la plataforma se adapta a las necesidades individuales de cada estudiante: mientras algunos requieren más tiempo para alcanzar el aprendizaje esperado, otros lo logran en menos tiempo. Por ejemplo, los estudiantes de *Razonamiento Matemático* y *Pensamiento Computacional* registraron menos horas de uso y obtuvieron calificaciones más altas, mientras que en *Modelación Matemática* y *Pensamiento Matemático* quienes dedicaron más tiempo lograron un mejor desempeño (ver Figura 3).

Figura 3
Uso de datos en la plataforma adaptativa



Para evaluar la EAA, se envió una encuesta a los estudiantes por correo electrónico. La participación fue voluntaria y, aunque los profesores realizaron esfuerzos para motivarlos, la tasa de respuesta obtenida fue menor a la esperada. Los resultados se presentan en la Tabla 3, en la encuesta se establecieron siete dimensiones relacionadas a la EAA, es decir, a todo el proceso involucrado, desde el contenido hasta la experiencia de usuario. La escala utilizada fue de tipo Likert

continua, la interpretación de los puntajes se estableció en cinco categorías: 0-19: totalmente en desacuerdo; 20-39: en desacuerdo; 40-59: neutral; 60-79: de acuerdo; y 80-100: totalmente de acuerdo. Esta codificación permitió un análisis más fino de la variabilidad en las percepciones de los participantes.

Tabla 3

Evaluación de la estrategia de aprendizaje adaptativo por parte de los estudiantes

Dimensiones	Pensamiento Computacional	Razonamiento Matemático	Pensamiento Matemático	Modelación Matemática
	n = 7 Media (DE)	n = 22 Media (DE)	n = 41 Media (DE)	n = 193 Media (DE)
Ruta de aprendizaje	75 (22)	78.9 (16.9)	76.6 (18.7)	75.9 (22.9)
Proceso de aprendizaje	73 (24)	78.2 (21.6)	75.2 (25.2)	73.5 (27)
Proceso de enseñanza	71 (26.6)	81.7 (19.7)	85.6 (19)	75.8 (27.4)
Contenido	75 (22)	81.2 (17.1)	76.8 (25)	78.7 (24.2)
Engagement	76 (23)	79.9 (19.3)	78.6 (23.9)	78 (24.5)
Usabilidad	81 (20)	85.8 (16.3)	83 (19.4)	82.9 (22.4)
Experiencia de usuario	77 (23)	83.3 (16.7)	78.2 (25.5)	79 (26)
Satisfacción general con la plataforma	77.6 (24.1)	80 (18.2)	77.9 (23.9)	78 (25.3)

Nota: DE = desviación estándar.

De acuerdo con la escala establecida, los resultados se ubicaron por encima del nivel “De acuerdo”. No se encontraron diferencias estadísticamente significativas entre los cursos, lo que indica que las evaluaciones fueron consistentes entre los estudiantes de las distintas asignaturas. Un hallazgo adicional, derivado de una pregunta de la encuesta sobre el uso de recursos externos a la plataforma adaptativa, fue que el 57 % de los estudiantes reportó no haber utilizado materiales adicionales para recuperar conocimientos previos.

Los comentarios más frecuentes se categorizaron en las siguientes fortalezas y áreas de oportunidad:

- Los estudiantes valoran la posibilidad de acceder a la plataforma en cualquier momento y avanzar a su propio ritmo.
- Destacan que los ejercicios están bien organizados, comenzando desde lo básico hacia lo avanzado.
- Los videos son señalados como una herramienta efectiva y fácil de entender.
- La plataforma es vista como práctica y comprensible, con buen material para repasar y reforzar conocimientos.

- Permite recordar y repasar temas olvidados de manera eficiente.
- Se valora el uso de videos, pero sugieren que sean más interactivos y con explicaciones paso a paso.
- Proponen una mejor explicación de los errores en los ejercicios para comprender cómo corregirlos.

Por otro lado, con respecto a los profesores, se obtuvieron las evaluaciones mostradas en la Tabla 4, la encuesta fue respondida solamente por 11 profesores ya que tampoco fue obligatoria.

Tabla 4

Evaluación de la Estrategia de Aprendizaje Adaptativo por parte del profesorado

Dimensiones	Evaluación
Usabilidad	90.9
Uso de analíticas	93.8
Enseñanza	95.3
Experiencia de usuario	95.8
Pre-lectura	91.6
Nivelación	90.6
Satisfacción en general	95.9

Como se observa en la Tabla 4, los profesores evaluaron positivamente a la Estrategia de Aprendizaje Adaptativo, destacándose las dimensiones Enseñanza y Experiencia de usuario. Los comentarios más frecuentes fueron los siguientes:

- Mejora el proceso de enseñanza-aprendizaje de manera innovadora.
- Fomenta mayor compromiso y participación de los estudiantes.
- Facilita el estudio previo a clase y la nivelación de conocimientos básicos, lo que implica un ahorro de tiempo en las clases.
- Permite a los estudiantes aprender desde casa.
- Ofrece análisis de datos para tomar decisiones sobre apoyo a estudiantes con menor desempeño y reconocimiento a los de mejor rendimiento.
- Incrementa la seguridad y confianza de los estudiantes al corroborar respuestas y obtener conocimientos sólidos.

DISCUSIÓN

Los resultados de este estudio muestran diferencias significativas en el desempeño académico en favor de los grupos experimentales que utilizaron la EAA, frente a los grupos control con un enfoque tradicional. En la Tabla 2 se observa que, en las cuatro unidades de formación, todos los grupos con módulos adaptativos superaron a los grupos control. Asimismo, estos grupos experimentales registraron ganancias netas de aprendizaje superiores, entre 11.4 y 23.6 puntos, como se ilustra en la Figura 2. Estos hallazgos sugieren que la EAA puede favorecer de manera efectiva la recuperación de conocimientos previos cuando se implementa de forma integral, incorporando un diseño de contenidos sólido, un modelo didáctico coherente, un rol docente activo y una plataforma tecnológica adecuada. Esto coincide con lo señalado por Rincon-Flores et al. (2024), Shi y Liu (2025) y Simon y Zeng (2024), quienes destacan que el impacto del AA se potencia al combinarse con estrategias de enseñanza pertinentes y plataformas alineadas con dichas estrategias.

La plataforma tecnológica con IA es un componente central de la EAA, ya que permite una personalización dinámica del aprendizaje. Mediante el diagnóstico preciso de los conocimientos previos de cada estudiante, la IA adapta en tiempo real las rutas de aprendizaje según sus necesidades específicas. Como señalan Moskal et al. (2017), las plataformas adaptativas emplean un enfoque basado en datos —y, en algunos casos, no lineal— para la instrucción y la nivelación, ajustándose de forma dinámica a las interacciones y niveles de rendimiento de los estudiantes. Esto les permite entregar contenidos en la secuencia y momento más apropiados para favorecer el progreso individual. La Figura 3 ilustra cómo la plataforma se ajusta a las particularidades de cada alumno: mientras algunos requieren más tiempo para alcanzar el aprendizaje esperado, otros lo logran en menor tiempo.

En la evaluación de la EAA, los resultados de la Tabla 3 muestran una percepción positiva en todas las dimensiones evaluadas, que incluyeron aspectos pedagógicos (contenidos, proceso de enseñanza y aprendizaje) y tecnológicos (*engagement*, experiencia de usuario y usabilidad), así como la satisfacción general. Todas superaron el valor de acuerdo, sin diferencias significativas entre profesores, lo que indica una percepción homogénea por parte de los estudiantes. Además, el 57 % de los estudiantes no requirió recursos externos para recuperar conocimientos previos, evidenciando la suficiencia del material. Los comentarios cualitativos destacaron la organización del contenido, la utilidad de los ejercicios y videos, y la eficacia de la plataforma para repasar y consolidar aprendizajes, en línea con los hallazgos de la Universidad Central de Florida, donde se observó mayor éxito estudiantil, mejores resultados de aprendizaje y más *engagement* (Dubey et al., 2023). Sin embargo, se sugirió aumentar la interactividad de los recursos audiovisuales y ofrecer retroalimentaciones más detalladas, coincidiendo con du Plooy et al. (2024), quienes señalan el potencial del AA para mejorar tanto el rendimiento académico como el compromiso estudiantil.

De manera similar, los docentes evaluaron positivamente la EAA, obteniendo en la Tabla 4 puntajes sobresalientes superiores a 90/100 en todas las dimensiones, destacando especialmente Enseñanza y Experiencia de usuario. Los comentarios cualitativos resaltaron que la estrategia mejora de forma innovadora el proceso de enseñanza-aprendizaje y fomenta un mayor compromiso y participación estudiantil. Estos hallazgos, tanto en percepciones como en resultados académicos, respaldan la efectividad de la EAA para la nivelación mediante módulos breves integrados en las unidades de formación. Asimismo, confirman que su éxito radica en un diseño integral que combine contenidos bien estructurados, estrategias didácticas pertinentes, un modelo docente adecuado y tecnología apropiada, tal como señalan Rincon-Flores et al. (2024) y sugieren Shi y Liu (2025).

Sin embargo, también es importante considerar estudios que muestran resultados divergentes. Por ejemplo, investigaciones como las de Wang et al. (2023) y Eau et al. (2022) evidencian que, si bien el sistema adaptativo favoreció la mejora académica, no se detectaron diferencias significativas entre los grupos en términos absolutos. De forma similar, en el estudio de Rincon-Flores et al. (2024) se observaron mejoras sustanciales en la ganancia de aprendizaje, los resultados académicos y otros indicadores en favor de los grupos adaptativos frente a los grupos control; no obstante, en el rubro relativo al logro de competencias disciplinares no se encontraron diferencias significativas entre ambos grupos, independientemente del curso o la etapa evaluada.

Aunque los resultados de esta investigación son prometedores, es fundamental reconocer sus limitaciones en cuanto a validez externa. El estudio se realizó en una universidad privada mexicana, lo que implica que los hallazgos podrían estar sesgados por las condiciones específicas de este entorno. Por ello, las generalizaciones a otros contextos educativos, especialmente aquellos donde los recursos y las condiciones difieren significativamente, deben hacerse con cautela. Asimismo, la disposición tanto de docentes como de estudiantes hacia el uso de herramientas tecnológicas podría ser mayor en este tipo de instituciones, que suelen ofrecer acceso a recursos avanzados y soporte especializado. Se recomienda que futuras investigaciones profundicen en los factores que potencian o limitan el impacto de la EAA en diferentes contextos educativos.

Por otro lado, la transferibilidad de esta EAA con módulos adaptativos a entornos virtuales y a la educación a distancia es plenamente viable gracias a su formato digital, que permite el acceso a los contenidos 24/7 desde cualquier lugar con conexión a internet. La plataforma tecnológica empleada, además, facilita su integración con distintos sistemas de gestión del aprendizaje (LMS), lo que amplía su aplicabilidad a diversos contextos educativos, siempre que se cuente con la infraestructura tecnológica necesaria. Cabe señalar que, aunque los módulos adaptativos también se aplicaron en cursos a distancia, estos no fueron incluidos en el presente estudio debido al reducido tamaño de la muestra, considerándose únicamente los grupos presenciales.

CONCLUSIONES

La EAA propuesta en este estudio se perfila como una alternativa prometedora para cerrar brechas en los conocimientos previos adquiridos en la educación media superior. Su implementación mediante módulos breves integrados en los cursos regulares permitió una recuperación significativa sin recurrir a cursos remediales paralelos o adicionales. Bajo este enfoque, cada estudiante aborda únicamente los contenidos que necesita, justo cuando los requiere, evitando la repetición innecesaria de temas ya dominados.

Los resultados sugieren que la articulación de los componentes clave de la estrategia —contenidos estructurados en microaprendizajes, un modelo didáctico centrado en el autoestudio y la autorregulación, un rol docente activo en motivar y dar seguimiento, y una plataforma tecnológica con IA para personalizar rutas y monitorear el progreso— influyó positivamente en el desempeño académico, la recuperación de conocimientos previos y la percepción favorable de estudiantes y docentes. Además, la incorporación formal de los módulos como parte de la evaluación incentivó la participación y el cumplimiento del proceso de nivelación.

Si bien los resultados son alentadores, se identificaron áreas de mejora, como aumentar la cantidad de videos, ofrecer retroalimentación más detallada en ejercicios y fomentar el uso docente de las analíticas de la plataforma para identificar patrones de aprendizaje y personalizar la retroalimentación. En algunos casos, la resistencia de ciertos profesores limitó la integración plena del módulo con la dinámica del curso y la aplicación de instrumentos de medición.

En conclusión, una EAA diseñada de forma integral a través de pequeños módulos integrados en los cursos regulares y mediada por una plataforma con IA puede favorecer la nivelación de conocimientos previos en estudiantes de primer año de programas STEM, constituyendo una alternativa valiosa frente a los modelos tradicionales de remediación. Como líneas futuras, se recomienda explorar su implementación en contextos diversos —instituciones públicas, entornos rurales y modalidades virtuales— y desarrollar estudios longitudinales que midan su impacto sostenido en el rendimiento académico, la permanencia y la autonomía en el aprendizaje.

Finalmente, resulta pertinente impulsar políticas educativas que integren estrategias de nivelación adaptativas al currículo y promover la formación docente continua en el uso de analíticas de aprendizaje para fortalecer el acompañamiento y la retroalimentación.

Agradecimientos

Los autores reconocen el apoyo financiero del Writing Lab, Institute for the Future of Education, Tecnológico de Monterrey, México, en la producción de este trabajo.

REFERENCIAS

- Allela, M. A., Ogange, B. O., Junaid, M. I. y Charles, P. B. (2020). Effectiveness of multimodal microlearning for in-service teacher training. *Journal of Learning for Development*, 7(3), 384-398. <https://doi.org/10.56059/jl4d.v7i3.387>
- Cavanagh, T., Chen, B., Ait Maalem Lahcen, R. y Paradiso, J. R. (2020). Constructing a design framework and pedagogical approach for adaptive learning in higher education: A practitioner's perspective. *The International Review of Research in Open and Distributed Learning*, 21(1), 173-195. <https://www.irrodl.org/index.php/irrodl/article/view/4557>
- Creswell, J. W. (2015). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (5^a ed.). Pearson.
- Dagunduro, A. O., Chikwe, C. F., Ajuwon, O. A. y Ediae, A. A. (2024). Adaptive learning models for diverse classrooms: Enhancing educational equity. *International Journal of Applied Research in Social Sciences*, 6(9), 2228-2240. <https://doi.org/10.51594/ijarss.v6i9.1588>
- du Plooy, E., Casteleijn, D. y Franzsen, D. (2024). Personalized adaptive learning in higher education: A scoping review of key characteristics and impact on academic performance and engagement. *Heliyon*, 10(21), e39630. <https://doi.org/10.1016/j.heliyon.2024.e39630>
- Dubey, A., Chen, B. y Lloyd, J. (2023, 4 de mayo). Personalized adaptive learning in algebra-based introductory physics courses at the University of Central Florida. *WCET Frontiers*. <https://wcet.wiche.edu/frontiers/2023/05/04/personalized-adaptive-learning-in-algebra-based-intro-physics-courses-at-ucf/>
- Eau, G., Hoodin, D. y Musaddiq, T. (2022). Testing the effects of adaptive learning courseware on student performance: An experimental approach. *Southern Economic Journal*, 88(3), 1086-1118. <https://doi.org/10.1002/soej.12547>
- Felder, R. M. y Brent, R. (2005). Understanding student differences. *Journal of Engineering Education*, 94(1), 57-72. <https://doi.org/10.1002/j.2168-9830.2005.tb00829.x>
- Fernández-Morante, C., Cebreiro-López, B., Rodríguez-Malmierca, M.-J. y Casal-Otero, L. (2022). Adaptive learning supported by learning analytics for student teachers' personalized training during in-school practices. *Sustainability*, 14(1), 124. <https://doi.org/10.3390/su14010124>
- Filatova, E., Chen, Y.-W. y Li, H. (2023, junio). Analysis of the COVID-19 impact on students' enrollment, performance, and retention. En *Proceedings of the 2023 ASEE Annual Conference & Exposition* (Baltimore, MD, United States). <https://doi.org/10.18260/1-2--42660>
- González-Beltrán, B. A., González-Brambila, S. B., Sánchez-Guerrero, L., Ardón-Pulido, I. y Figueroa-González, J. (2018). Estudio del impacto de un curso de nivelación en el desempeño de alumnos de ingeniería utilizando minería de datos educacional. *Research in Computing Science*, 147(5), 159-172. <https://doi.org/10.13053/rcs-147-5-12>
- Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge. <https://doi.org/10.4324/9780203887332>
- Hattie, J. (2017). *Visible Learning Plus: 252 influences and effect sizes related to student achievement* [PDF]. Visible Learning. <https://visible-learning.org/wp-content/uploads/2018/03/VLPLUS-252-Influences-Hattie-ranking-DEC-2017.pdf>

- Hernandez Cardenas, L. S., Castano, L., Cruz Guzman, C. y Nigenda Alvarez, J. P. (2022). Personalised learning model for academic leveling and improvement in higher education. *Australasian Journal of Educational Technology*, 38(2), 70-82. <https://doi.org/10.14742/ajet.7084>
- Howlin, C. y Lynch, D. (2014, octubre). Learning and academic analytics in the Realizeit system. En *Proceedings of the World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education (E-Learn 2014)* (New Orleans, LA, United States). Association for the Advancement of Computing in Education (AACE). <https://www.learntechlib.org/primary/p/148952/>
- Jing, Y., Zhao, L., Zhu, K., Wang, H., Wang, C. y Xia, Q. (2023). Research landscape of adaptive learning in education: A bibliometric study on research publications from 2000 to 2022. *Sustainability*, 15(4), 3115. <https://doi.org/10.3390/su15043115>
- Kara, N. y Sevim, N. (2013). Adaptive learning systems: Beyond teaching machines. *Contemporary Educational Technology*, 4(2), 108-120. <https://doi.org/10.30935/cedtech/6095>
- Klingbeil, N. W., Rattan, K. S., Raymer, M. L., Reynolds, D. B., Mercer, R., Kukreti, A. y Randolph, B. (2007, junio). A national model for engineering mathematics education. En *Proceedings of the 2007 ASEE Annual Conference & Exposition* (Honolulu, HI, United States). <https://doi.org/10.18260/1-2--2104>
- Koester, B. P., Grom, G. y McKay, T. A. (2016). *Patterns of gendered performance difference in introductory STEM courses* [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.1608.07565>
- Lasater, K., Smith, C., Pijanowski, J. y Brady, K. P. (2021). Redefining mentorship in an era of crisis: Responding to COVID-19 through compassionate relationships. *International Journal of Mentoring and Coaching in Education*, 10(2), 158-172. <https://doi.org/10.1108/IJMCE-11-2020-0078>
- Li, T., Kirk, C. y Oseguera, L. (2023). Heterogeneous academic achievement profiles of initially STEM-intending students over the college years. *Journal of College Student Development*, 64(6), 728-735. <https://doi.org/10.1353/csd.2023.a917027>
- Long, P. y Siemens, G. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review*, 46(5), 31-40. <https://er.educause.edu/articles/2011/9/penetrating-the-fog-analytics-in-learning-and-education>
- López del Puerto, C., Bellido, C. M., Suárez, O. M., Alfaro, M. y Jiménez, M. A. (2021, julio). Championing Hispanic student success following natural disasters in Puerto Rico. En *Proceedings of the 2021 ASEE Annual Conference & Exposition* (evento virtual). <https://doi.org/10.18260/1-2--36790>
- McKenna, A., Agogino, A. M. y McMartin, F. (2000, octubre). What students say about learning physics, math, and engineering. En *Proceedings of the 30th Annual Frontiers in Education Conference* (Kansas City, MO, United States). IEEE. <https://doi.org/10.1109/FIE.2000.897580>
- Miao, F., Holmes, W., Huang, R. y Zhang, H. (2021). *Inteligencia artificial y educación: Guía para las personas a cargo de formular políticas*. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000379376>
- Moskal, P., Carter, D. y Johnson, D. (2017, 4 de enero). *7 things you should know about... adaptive learning*. EDUCAUSE Learning Initiative. <https://library.educause.edu/resources/2017/1/7-things-you-should-know-about-adaptive-learning>
- Ochukut, S. A., Oboko, R. O., Miriti, E. y Maina, E. (2023). Research trends in

- adaptive online learning: Systematic literature review (2011–2020). *Technology, Knowledge and Learning*, 28, 431-448. <https://doi.org/10.1007/s10758-022-09615-9>
- Olmos-López, O., Rincon-Flores, E. G., Mena, J., Román, O. y Camacho-López, E. (2023). Artificial intelligence as a way to improve educational practices. En M. Cebral-Loureda, E. G. Rincon-Flores y G. Sanchez-Ante (Eds.), *What AI can do: Strengths and limitations of artificial intelligence* (pp. 135-151). Taylor & Francis. <https://doi.org/10.1201/b23345-10>
- Press Trust of India. (2022, 3 de mayo). Haryana govt to distribute over 5 lakh free tablets to Class 10, 12 students. *The Indian Express*. <https://indianexpress.com/article/education/haryana-govt-to-distribute-over-5-lakh-free-tablets-to-class-10-12-students-from-may-5-7898959/>
- Rincon-Flores, E. G., Castano Sánchez, L., Guerrero Solís, S. L., Olmos López, O., Rodríguez Hernández, C. F., Castillo Lara, L. A. y Aldape Valdes, L. P. (2024). Improving the learning-teaching process through adaptive learning strategy. *Smart Learning Environments*, 11, 27. <https://doi.org/10.1186/s40561-024-00314-9>
- Rincon-Flores, E. G., Rivera Vázquez, N., Castano Sánchez, L., Guerrero Solís, S. L., Aldape Valdes, L. P. y Castillo Lara, L. A. (2023, abril). Improving face-to-face learning using an adaptive technology. En *Proceedings of the 2023 IEEE Global Engineering Education Conference (EDUCON)* (Kuwait City, Kuwait). IEEE. <https://doi.org/10.1109/EDUCON54358.2023.10125276>
- Rockinson-Szapkiw, A., Herring Watson, J., Gishbaugher, J. y Wendt, J. L. (2021). A case for a virtual STEM peer-mentoring experience for racial and ethnic minority women mentees. *International Journal of Mentoring and Coaching in Education*, 10(3), 267-283. <https://doi.org/10.1108/IJMCE-08-2020-0053>
- Ross, P. M., Scanes, E., Poronnik, P., Coates, H. y Locke, W. (2022). Understanding STEM academics' responses and resilience to educational reform of academic roles in higher education. *International Journal of STEM Education*, 9(1), 11. <https://doi.org/10.1186/s40594-022-00327-1>
- Sarıyalçınkaya, A. D., Karal, H., Altınay, F. y Altınay, Z. (2021). Reflections on adaptive learning analytics: Adaptive learning analytics. En A. Azevedo, J. M. Azevedo, J. O. Uhomobhi y E. Ossianilsson (Eds.), *Advancing the power of learning analytics and big data in education* (pp. 61-84). IGI Global. <https://doi.org/10.4018/978-1-7998-7103-3.ch003>
- Schaal, N., Chan, S. E. y Saint Clair, J. K. (2021, julio). Insights gleaned from the GAIN peer-mentoring program pilot. En *Proceedings of the 2021 ASEE Annual Conference & Exposition* (evento virtual). <https://peer.asee.org/insights-gleaned-from-the-gain-peer-mentoring-program-pilot>
- Seymour, E. y Hewitt, N. M. (1997). *Talking about leaving: why undergraduates leave the sciences*. Westview Press.
- Shi, P. y Liu, W. (2025). Adaptive learning oriented higher educational classroom teaching strategies. *Scientific Reports*, 15, 15661. <https://doi.org/10.1038/s41598-025-00536-y>
- Simon, P. D. y Zeng, L. M. (2024). Behind the scenes of adaptive learning: A scoping review of teachers' perspectives on the use of adaptive learning technologies. *Education Sciences*, 14(12), 1413. <https://doi.org/10.3390/educsci14121413>
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257-285. https://doi.org/10.1207/s15516709cog1202_4
- Sweller, J., van Merriënboer, J. J. G. y Paas, F. G. W. C. (1998). Cognitive architecture and instructional

- design. *Educational Psychology Review*, 10(3), 251-296. <https://doi.org/10.1023/A:1022193728205>
- Taha, H., Hanim, F., Johar, M. y Shah, M. (2015). Peer mentoring module: The effect of an intervention of academic mentoring program towards motivation and self-esteem among foundation students in Malaysia. *Universal Journal of Psychology*, 3(3), 80-83. <https://doi.org/10.13189/ujp.2015.030304>
- Taylor, D. L., Yeung, M. y Bashet, A. Z. (2021). Personalized and adaptive learning. En J. Ryoo y K. Winkelmann (Eds.), *Innovative learning environments in STEM higher education: Opportunities, challenges, and looking forward* (pp. 17-34). Springer. https://doi.org/10.1007/978-3-030-58948-6_2
- UNESCO. (2019). *Consenso de Beijing sobre la inteligencia artificial y la educación* [Documento de programa]. <https://unesdoc.unesco.org/ark:/48223/pf0000368303>
- Vygotsky, L. S. (1978). *Mind in society: the development of higher psychological processes*. Harvard University Press.
- Wang, S., Christensen, C., Cui, W., Tong, R., Yarnall, L., Shear, L. y Feng, M. (2023). When adaptive learning is effective learning: Comparison of an adaptive learning system to teacher-led instruction. *Interactive Learning Environments*, 31(2), 793-803. <https://doi.org/10.1080/10494820.2020.1808794>
- Wood, D., Bruner, J. S. y Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, 17(2), 89-100. <https://doi.org/10.1111/j.1469-7610.1976.tb00381.x>
- Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist*, 25(1), 3-17. https://doi.org/10.1207/s15326985ep2501_2

Fecha de recepción del artículo: 1 de junio de 2025

Fecha de aceptación del artículo: 21 de agosto de 2025

Fecha de aprobación para maquetación: 7 de octubre de 2025

Fecha de publicación en OnlineFirst: 27 de octubre de 2025

Fecha de publicación: 1 de enero de 2026

Recomendador de evaluación para preguntas cortas utilizando modelos de lenguaje en propiedad intelectual

A language model-based recommender assessment system for short-answer questions in the intellectual property domain

David Bañeres Besora - *Universitat Oberta de Catalunya, UOC (España)*
Ana-Elena Guerrero Roldán - *Universitat Autònoma de Barcelona, UAB (España)*
M. Elena Rodríguez González - *Universitat Oberta de Catalunya, UOC (España)*

RESUMEN

El uso de la Inteligencia Artificial (IA) en educación está creciendo rápidamente, transformando el proceso de enseñanza-aprendizaje y también el proceso de evaluación. Este trabajo presenta SLASys, una herramienta para recomendar al profesorado la evaluación de preguntas cortas mediante técnicas de IA semántica, difiriendo de otros trabajos basados en IA generativa por el uso del modelo de lenguaje BERT que es más ligero, comprende mejor los conceptos en un contexto específico, mejora la eficiencia computacional y reduce los problemas éticos y de privacidad. SLASys implementa comparación semántica y modelos predictivos de clasificación de respuestas basados en BERT. Se ha seguido una metodología de investigación mixta, combinando investigación de acción con un enfoque de diseño y creación, para desarrollar y perfeccionar SLASys a lo largo de cuatro ediciones de un curso de nivel de máster sobre examen de patentes en el contexto de la propiedad intelectual. SLASys se ha integrado en Moodle, permitiendo su uso por parte de profesorado sin conocimientos técnicos, y ha sido probada por 120 estudiantes. Los resultados muestran su efectividad, tanto en el marco de la experiencia descrita como según la literatura existente, incluso con conjuntos de datos reducidos y un número limitado de participantes, y ha sido valorada positivamente por el profesorado y el estudiantado. Este trabajo contribuye a mostrar la viabilidad del uso de la IA en la educación superior, tanto en entornos híbridos como en línea, ofreciendo una solución para mejorar la evaluación y el *feedback* en preguntas cortas en contextos reales de aprendizaje.

Palabras clave: evaluación; *feedback*; Moodle; test; preguntas cortas; inteligencia artificial.

ABSTRACT

The use of Artificial Intelligence (AI) in education is growing rapidly, transforming the teaching-learning process as well as the assessment process. This work introduces SLASys, a tool to recommend the assessment of short-answer questions using semantic AI techniques. Unlike other works based on generative AI, SLASys uses the lightweight BERT language model, which better understands specific domain language concepts, improves computational efficiency, and reduces ethical and privacy concerns. SLASys implements semantic comparison and predictive classification models based on BERT. A mixed research methodology was followed, combining action research with a design and creation approach, to develop and refine SLASys over four editions of a master's-level course on patent examination within the intellectual property domain. SLASys has been integrated into Moodle, enabling its use by teachers without technical expertise, and has been tested by 120 students. The results demonstrate its effectiveness even with small datasets and limited participants within the described experience and according to existing literature. Additionally, it has been positively evaluated by both teachers and students. This work shows the feasibility of using AI in higher education, in both hybrid and online environments, offering a practical solution to improve assessment and feedback for short-answer questions in real learning contexts.

Keywords: assessment; feedback; Moodle; test; short-answer; artificial intelligence.

INTRODUCCIÓN

Hoy en día, las instituciones educativas (incluidas las tradicionales y las en línea) generalmente adoptan las tecnologías de la información y la comunicación para apoyar a su profesorado y estudiantado. La automatización de la tarea de calificación es una tendencia seguida por muchos Sistemas de Gestión del Aprendizaje (SGA) para evaluar de manera eficiente las actividades y proporcionar un *feedback* oportuno en el momento adecuado con finalidad formativa (György y Vajda, 2007). Sin embargo, muchos SGA limitan el tipo de preguntas que se califican automáticamente a aquellas basadas en la selección de opciones o texto simple, debido a la complejidad de evaluar preguntas cortas o de tipo ensayo.

Aunque la Calificación Automática de Preguntas de Respuesta Corta (CAPRC) ha sido ampliamente explorada en la literatura (Burrows et al., 2015), las contribuciones se han limitado a explorar su precisión en conjuntos de datos públicos sin considerar su efectividad en entornos educativos reales. Además, el renacimiento de la Inteligencia Artificial (IA) debido al acceso a recursos computacionales y a herramientas específicas para aplicar técnicas de IA, han permitido que los SGA clásicos evolucionen. Los algoritmos de IA pueden reforzar los SGA para apoyar el proceso de aprendizaje. Las recomendaciones, el *feedback* inmediato o la CAPRC son ejemplos de tareas que la IA puede mejorar.

Esta investigación tiene como objetivo proporcionar un punto de partida para un sistema de aprendizaje inteligente capaz de recomendar la evaluación y el *feedback* para preguntas cortas utilizando técnicas de IA. Este trabajo también muestra cómo se utiliza el sistema en un entorno educativo híbrido y presenta los resultados obtenidos. A pesar de que los modelos de lenguaje extenso –más conocidos por su denominación en inglés *Large Language Models* (LLM)– y las herramientas generativas derivadas, como ChatGPT (OpenAI, 2024), muestran actualmente su gran potencial en cualquier tarea académica, siguen teniendo limitaciones para tareas o dominios específicos como es el caso de la Propiedad Intelectual.

Por lo tanto, este trabajo propone una solución alternativa, utilizando BERT, un Modelo de Lenguaje (ML) ligero que comprende el significado de las respuestas. En lugar de delegar la decisión de la evaluación de una pregunta de respuesta corta en una herramienta de IA Generativa (IAGen) mediante la definición de criterios expresados como un conjunto complejo de instrucciones (*prompt*), se utiliza BERT. Este modelo comprende el significado de cualquier respuesta correcta e incorrecta, siendo más preciso para fines evaluativos. Al ser menos extenso, contribuye a un uso computacional más eficiente y sostenible, gestionable en instancias privadas (es decir, no se necesitan soluciones empresariales), reduciendo también los problemas éticos y de privacidad. Por lo tanto, este trabajo ha integrado esta solución de IA en un SGA (concretamente, Moodle) para que el profesorado pudiera proporcionar una evaluación significativa al estudiantado.

Se sigue una metodología de investigación de acción combinada con un enfoque de diseño y creación para crear el sistema automatizado de aprendizaje inteligente (denominado *Smart Learning Automated System*, SLASys), explícitamente adaptado a un curso de formación en Propiedad Intelectual, para definir un sistema de aprendizaje que facilite el éxito del estudiantado mediante el uso de técnicas de IA aplicadas a la enseñanza de la Propiedad Intelectual. El estudio se ha llevado a cabo en el área educativa de la Oficina Europea de Patentes (OEP).

Este trabajo se estructura de la siguiente manera. En primer lugar, se presentan los antecedentes teóricos de los modelos de lenguaje y sus usos educativos, centrándose en la CAPRC. A continuación, se detalla el diseño de la investigación, incluyendo la metodología, las preguntas de investigación, la herramienta SLASys y el análisis de datos. La sección de resultados destaca los principales hallazgos. El trabajo concluye con la discusión, limitaciones, contribuciones e investigación futura.

REVISIÓN DE LA LITERATURA

Modelos de Lenguaje Extenso, BERT y su aplicación en educación

Los LLM han surgido como una tecnología transformadora en los últimos años, demostrando capacidades notables en diversas tareas de Procesamiento del Lenguaje Natural (PLN). Su rápido progreso ha propiciado avances significativos en la traducción automática, la síntesis de textos e incluso la generación de código (Husein et al., 2025; Pang et al., 2025; Zhang et al., 2025). A pesar de su impresionante rendimiento, los LLM se enfrentan a limitaciones y desafíos en entornos educativos reales, como interpretaciones erróneas en lenguajes específicos de dominio (Arefeen et al., 2024) o su interpretabilidad y explicabilidad (Zhao et al., 2024). Las innovaciones recientes en computación han mejorado su rendimiento, permitiendo alcanzar el nivel de las herramientas de IAGen de última generación, como GPT-4o (OpenAI, 2024). Estas herramientas demuestran capacidades impresionantes de aprendizaje, aplicabilidad a tareas de razonamiento limitadas y capacidad de generalización en diversos dominios.

BERT (*Bidirectional Encoder Representations from Transformers*) es un modelo de PLN de última generación desarrollado por Google en 2018 (Devlin et al., 2019). BERT es un ML basado en transformadores previamente entrenado con una gran colección de documentos de texto. BERT se consideró inicialmente un LLM, pero actualmente se considera un ML en comparación con sus sucesores de IAGen. Su arquitectura se fundamenta en la representación del texto. Una frase está codificada de una manera especial (denominada *embeddings*) que almacena su significado (es decir, su semántica), lo que permite la aplicación directa en tareas de PLN. Su desarrollo y aplicación continuados en múltiples dominios de conocimiento demuestran su versatilidad y potencial para impulsar nuevos avances en IA y en la comprensión del lenguaje. BERT se ha utilizado para la clasificación de documentos

(Adhikari et al., 2019), el reconocimiento de categorías gramaticales (Souza et al., 2019), la implicación lógica entre pares de oraciones (Wang et al., 2018) o la síntesis de textos (Zhang, Cai et al., 2019).

Tanto BERT como las herramientas de IAGen se utilizan en educación afrontando desafíos similares. La escalabilidad y la integración con las infraestructuras existentes requieren de una inversión significativa y de conocimiento técnico (Xu y Zhu, 2023). Además, deben considerarse principios relacionados con la ética, la seguridad y la alineación con los objetivos educativos fundamentales (García-Peñalvo et al., 2024), y desarrollar la alfabetización en IA entre el profesorado es esencial para fomentar un uso crítico (Petridou y Lao, 2024). Asimismo, aunque parecen aplicables a cualquier tarea, las herramientas de IAGen son principalmente aplicables a la generación y el análisis de texto, la definición de conceptos, la creación de ejercicios o la mejora del estilo de escritura; mientras que BERT es válido para la clasificación de texto, el análisis de sentimientos, la comprensión de las dudas del estudiantado o la provisión de explicaciones detalladas (Qiu y Jin, 2024).

La evaluación también es una tarea relevante en ambos modelos. Por un lado, las herramientas de IAGen han empezado a explorarse en la evaluación como “cajas negras”, es decir, sin saber cómo funcionan internamente. Se han utilizado para producir *feedback*, pero también para la evaluación automática. La evaluación automática beneficia a algunos dominios, como el aprendizaje de idiomas (Escalante et al., 2023), la adquisición de habilidades de escritura (Banihashem et al., 2024) o el diseño multimedia (Almasre, 2024). Por otro lado, BERT se puede utilizar en tareas de evaluación específicas. El profesorado puede generar preguntas a partir de textos existentes, que se pueden utilizar en tests y exámenes (Nguyen et al., 2022). Los modelos basados en BERT también pueden evaluar la coherencia, la relevancia y la gramática de los ensayos del estudiantado, lo que ayuda a que este pueda mejorar sus habilidades de escritura (Wang et al., 2022). También se han diseñado aplicaciones basadas en BERT para comprender expresiones lingüísticas y matices culturales (Bahdanau et al., 2015), y recomendadores específicos que pueden ayudar al estudiantado a practicar la expresión oral y la comprensión de nuevos idiomas (Zhang, Zhang et al., 2019).

Aunque la evaluación automatizada parece ser una de las características futuras de la IAGen en educación, actualmente no está recomendada por las políticas institucionales ni gubernamentales (European Commission, 2024; Dai et al., 2025; González Fernández et al., 2025), porque las herramientas de IAGen son propensas a errores debido a alucinaciones (Jia et al., 2024) y su utilización podría plantear dilemas éticos. Además, las herramientas de IAGen pueden fallar en dominios con un lenguaje específico. Por ejemplo, algunas preguntas relacionadas con conceptos específicos en el examen de patentes dentro del dominio de la Propiedad Intelectual, como Novedad y Claridad, a veces son interpretadas por una herramienta de IAGen como si se tratara de evaluar si la respuesta es nueva y clara en lenguaje natural. Sin embargo, dichos conceptos especializados en Propiedad Intelectual

se refieren a cuán novedosa y clara está descrita una invención en una solicitud de patente (*reivindicación*). Las herramientas de IAGen podrían analizar la respuesta mejorando el contexto anterior (es decir, su *contextualización*), pero a veces mezclan directrices de patentes de diferentes fuentes con información obsoleta. Por lo tanto, se requiere añadir documentos específicos al modelo (es decir, es necesario realizar un *ajuste fino*) para mejorar la calidad de un modelo propietario de IAGen, lo que implica algunas preocupaciones acerca de la privacidad, y un debate sobre si es una solución rentable. En cambio, una aplicación basada en BERT aborda la tarea de evaluación comprendiendo las respuestas proporcionadas a una pregunta. BERT podría utilizarse para realizar una *comparación semántica* para evaluar lo cerca que está el significado de dos frases (la respuesta de un estudiante en comparación con la respuesta correcta del profesorado) o entrenar un modelo de IA de *clasificación de respuestas* para predecir si el significado de una frase se acerca a un conjunto correcto o incorrecto de frases (es decir, decidir si la respuesta de un estudiante es correcta o incorrecta en base a un conjunto de respuestas correctas e incorrectas). Estos enfoques evitan, además, el inconveniente de la variabilidad que las herramientas de IAGen exhiben actualmente.

Calificación automática de preguntas cortas y provisión de feedback

El *feedback* constituye uno de los componentes esenciales tanto del proceso de aprendizaje (Evans, 2013; Hattie y Timperley, 2007) como de la evaluación de la adquisición de conocimientos, habilidades y competencias. El *feedback* también es un elemento clave para fomentar la implicación del estudiantado y mejorar su aprendizaje (Winstone et al., 2017). Su objetivo es proporcionar información específica vinculada a una actividad de aprendizaje, con el fin de reducir la brecha entre la comprensión deseada y la real (Sadler, 1989). Además, el *feedback* es crucial en entornos en línea para garantizar un proceso de aprendizaje significativo y oportuno (Nicol y Macfarlane-Dick, 2006). La elaboración y entrega de *feedback* útil requiere de un esfuerzo considerable por parte del profesorado, lo que restringe su nivel de personalización cuando existen restricciones de tiempo o un gran número de estudiantes matriculados (Dhananjaya et al., 2024). Sin embargo, el *feedback* personalizado motiva al estudiantado en mayor medida, regula mejor su comportamiento, contribuye a una mejor adquisición de conocimientos y mejora el compromiso con los materiales de aprendizaje (Kim, 2023; Wang y Lehman, 2021).

La automatización de la generación de *feedback* se ha planteado como una solución tentativa para apoyar al profesorado. Aunque en la literatura se ha utilizado una terminología diversa para referirse a estas herramientas –por ejemplo, sistemas de tutoría inteligente, plataforma de enseñanza, sistema automatizado de *feedback*, herramientas de *feedback* en línea– todas ellas se centran en mejorar la eficiencia del profesorado mientras ayudan al estudiantado durante su proceso de aprendizaje (Xie y Li, 2018). Trabajos previos han presentado las técnicas aplicables en diferentes

contextos, como sería la generación de *feedback* para ensayos (Akçapinar, 2015), ejercicios de programación (Messer et al., 2024) o tareas colaborativas (Zheng et al., 2023).

La calificación automatizada de tests también ha sido ampliamente investigada debido a sus beneficios, tanto para proporcionar una evaluación automática como para facilitar los procesos de evaluación formativa (Burrows et al., 2015), así como su impacto positivo en el aprendizaje debido a la provisión de *feedback* (Gaona et al., 2018; Rezaei, 2015). Para simplificar su implementación, los tests han recurrido principalmente a preguntas de opción múltiple, mientras que las preguntas cortas y de ensayo han sido usualmente evitadas, ya que requieren de un mayor esfuerzo e intervención del profesorado durante la fase de evaluación, o bien se han utilizado métodos alternativos como la evaluación por pares (Huisman et al., 2017). A pesar de ello, existe trabajo previo relacionado con la CAPRC. Se ha utilizado la comparación de respuestas con una frase (Siddiqi y Harrison, 2008), con un conjunto de frases con un vocabulario similar (Klein et al., 2011) o con palabras clave principales (Saha et al., 2018) para detectar respuestas correctas. Aunque estas técnicas han demostrado buenos resultados, presentan algunas deficiencias, ya que no consideran la semántica (un patrón, un vocabulario similar o un conjunto de palabras clave pueden no abarcar todas las posibles respuestas correctas). Algunos autores han investigado la aplicabilidad de herramientas de IAGen para la CAPRC (Aggarwal et al., 2025; Grévisse, 2024). Aunque su aplicabilidad para la calificación de ensayos parece prometedora (Senthilnathan et al., 2025), los resultados para la CAPRC no son concluyentes para apoyar la evaluación del estudiantado. Las alucinaciones, la descontextualización, la información obsoleta y las limitaciones en el pensamiento crítico siguen afectando a su utilización (De La Cruz et al., 2024).

Dado que el PLN y los ML semánticos siguen siendo técnicas de vanguardia para la CAPRC, se utiliza BERT debido a sus capacidades, a su potencial precisión y a su mejor adaptación a entornos educativos reales. En la literatura se han explorado dos tipos de modelos. Algunos buscan predecir si la respuesta del estudiantado es correcta (Camus y Filighera, 2020; Liu et al., 2019; Lun et al., 2020; Padó et al., 2024; Schneider et al., 2023; Sung et al., 2019; Wang et al., 2019), mientras que otros van más allá al predecir la calificación, demostrando su potencial en conjuntos de datos públicos (Baral et al., 2021; del Gobbo et al., 2023; Gaddipati et al., 2020; Metzler et al., 2024; Soulimani et al., 2024). Sin embargo, hasta donde llega el conocimiento de los autores de esta investigación, ningún trabajo previo ha probado estas herramientas en entornos educativos reales debido a los problemas de aceptabilidad o a la necesidad de intervención humana para su uso (Hustad y Arntzen, 2013; Xavier et al., 2025). Por lo tanto, este trabajo propone SLASys como herramienta de recomendación para la CAPRC, la cual ha sido probada en un entorno educativo híbrido durante cuatro ediciones de un curso fundamental de examen de patentes a lo largo de un año, con 120 estudiantes, con el fin de obtener conocimiento sobre su uso. Así, el objetivo es responder a las siguientes preguntas de investigación:

- PI1. ¿Cómo de precisa es la clasificación de respuestas con respecto con la comparación semántica para la recomendación en la CAPRC?
- PI2. ¿Cómo de preciso es SLASys en un entorno educativo real?
- PI3. ¿Puede utilizarse SLASys sin respuestas previas del estudiantado?
- PI4. ¿Cuál es la opinión del estudiantado y del profesorado?

METODOLOGÍA

Diseño de la investigación

SLASys sigue una metodología de investigación mixta que combina investigación de acción (Oates, 2006) con un enfoque de diseño y creación (Kuechler y Vaishnavi, 2012). La investigación de acción es adecuada porque el sistema se desarrolla y se prueba en entornos educativos reales, con el objetivo de abordar desafíos prácticos en la enseñanza y el aprendizaje. Esta metodología enfatiza la resolución de problemas del mundo real a través de un ciclo colaborativo e iterativo de planificación, acción y reflexión, al tiempo que recopila datos para evaluar los resultados. El enfoque de diseño y creación complementa dicha metodología, al centrarse en el desarrollo de artefactos tecnológicos innovadores, siguiendo un proceso iterativo de resolución de problemas de cinco pasos (Kuechler y Vaishnavi, 2012) para guiar la creación y el perfeccionamiento del sistema.

Este trabajo se centra en las dos primeras iteraciones del desarrollo del recomendador para la CAPRC. La primera consistió en la creación del primer artefacto y la evaluación de la solución técnica, mientras que la segunda evaluó su integración en el proceso de enseñanza-aprendizaje. Esta última iteración implicó probar el artefacto en un curso fundamental de examen de patentes en el dominio de la Propiedad Intelectual. Este curso, con una duración de 6 semanas, forma parte de los cursos orientados a la formación del estudiantado como examinador de patentes, que, a su vez, consta de seis cursos con una duración total de 2 años. Existen requisitos específicos para ser elegible para esta capacitación. El estudiantado debe tener un máster en física, química, ingeniería o ciencias naturales, ya que se requiere conocimiento de campos técnicos específicos para examinar las propuestas de solicitud de patentes.

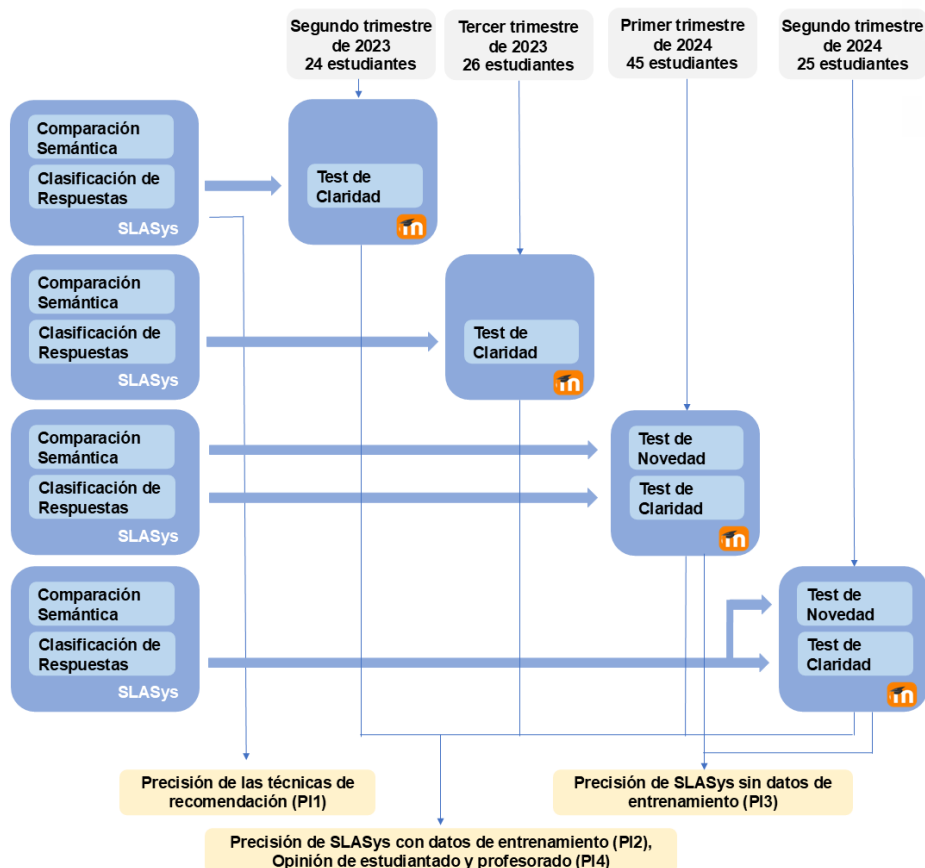
El objetivo del curso es preparar al estudiantado para el examen de patentes, introduciendo los conceptos fundamentales, las directrices de la Convención Europea de Patentes (CEP), los sistemas disponibles para dicha tarea y los ejercicios relacionados con el examen de patentes. La metodología de aprendizaje combina una estrategia de clase invertida (Bergmann y Sams, 2012) con aprendizaje justo a tiempo (Novak, 2012). Las actividades de autoaprendizaje y la lectura de recursos de aprendizaje se realizan de forma asincrónica. También hay sesiones en línea síncronas con la participación del estudiantado y del profesorado, donde se debaten

las actividades individuales realizadas previamente de forma asíncrona y se realizan ejercicios prácticos, preguntas y actividades grupales.

El profesorado informó de dificultades al aplicar la estrategia de aprendizaje de justo a tiempo, ya que no disponían de tiempo suficiente para evaluar las actividades de respuesta corta del estudiantado y proporcionar un *feedback* significativo antes de las sesiones en línea. Por lo tanto, SLASys responde a estas necesidades proporcionando una solución para evaluar y proporcionar *feedback* a cada estudiante de manera eficiente.

Figura 1

Diseño de la investigación



La Figura 1 muestra el diseño de la investigación. En primer lugar, los enfoques diseñados se comparan para ver su precisión para responder a la PI1. En segundo lugar, durante cuatro ediciones del curso, se probó el enfoque de clasificación de respuestas en un test relacionado con el concepto de Claridad de patentes para responder a la PI2. En total, 120 estudiantes realizaron el test. En tercer lugar, se diseñó un nuevo test sobre el concepto de Novedad en las dos últimas ediciones sin datos de entrenamiento (es decir, sin respuestas del estudiantado de ediciones anteriores). En este caso, se utilizaron los dos enfoques diseñados para evaluar el rendimiento del recomendador y responder a la PI3. Un total de 70 estudiantes respondieron este segundo test. Además, se recopilaron las opiniones del estudiantado y las experiencias del profesorado en todas las ediciones para responder a la PI4.

Las pruebas con estudiantado real se realizaron en línea de forma síncrona usando un Moodle adaptado como SGA que integraba SLASys para recomendar los resultados de la evaluación. Los tests de Claridad y Novedad incorporaron 8 y 11 preguntas, respectivamente, y el estudiantado dispuso de una hora para responderlos. Su objetivo fue mejorar su conocimiento, pero no sus calificaciones. El estudiantado es evaluado en dos ocasiones durante su formación mediante exámenes presenciales.

SLASys: un recomendador para la CAPRC

SLASys se diseñó para determinar si la respuesta de cada estudiante a una pregunta de respuesta corta es correcta, dependiendo de la información disponible relacionada con la pregunta. La Figura 2 muestra las dos técnicas utilizadas para detectar si una respuesta es correcta. La pregunta es sobre el concepto de Novedad en el dominio de la Propiedad Intelectual, preguntando por qué la reivindicación de la patente no es novedosa. La comparación se efectúa utilizando *embeddings* de BERT. Aunque un *embedding* es una matriz matemática compleja, se podría ver como una codificación del significado de la frase mediante un conjunto de símbolos que definen su significado.

Figura 2

Ejemplo de pregunta sobre Novedad

¿Por qué esta reivindicación no es novedosa?

Un teléfono (móvil) con medios de marcación, un micrófono y un altavoz, caracterizado porque los medios de marcación están hechos de aluminio...

Respuesta modelo

La reivindicación no está clara debido a la expresión entre paréntesis (móvil). No está claro si la función entre paréntesis debe ser limitante o no. Consulta las directrices F-IV, 4.19.

(, , ..., )

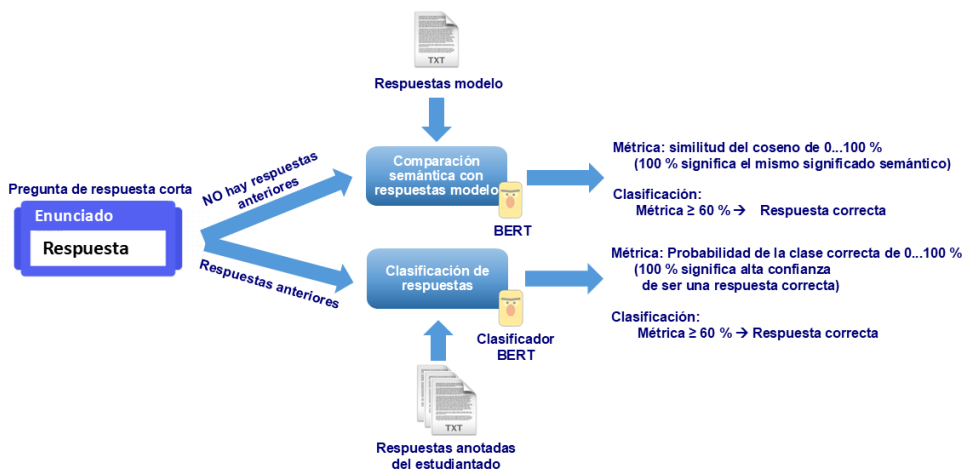
	<i>Embedding</i>	Similitud del coseno	Probabilidad de ser clasificado como correcto
Respuesta 1 (correcta) La característica entre paréntesis que no es un signo de referencia; no está claro si es opcional o no (F-IV, 4.19)	( ,  , ..., )	32,2 %	99,9 %
Respuesta 2 (incorrecta) No se permite la expresión entre paréntesis (no estoy seguro de si se encuentra dentro de la exención en este campo técnico).	($\sqrt{x}$,  , ..., )	23,3 %	0 %

La primera técnica usa el *embedding* de la respuesta correcta proporcionada por el profesorado (denominada *respuesta modelo*), para compararla con las respuestas proporcionadas por el estudiantado. Para evaluar lo similar que es la respuesta de cada estudiante con respecto a la respuesta modelo del profesorado (es decir, cuántos símbolos son iguales en ambos *embeddings*) se utiliza la métrica de *similitud del coseno*. El resultado de la comparación (es decir, la *puntuación de la métrica*) es una probabilidad de la similitud, donde el 100 % indica que el significado es el mismo que el de la respuesta modelo del profesorado. La Figura 2 muestra que este método falla para la respuesta correcta, ya que hay diferentes respuestas correctas y la respuesta modelo del profesorado no cubre todas ellas.

Sin embargo, este proceso podría mejorarse comparando las respuestas del estudiantado con un conjunto de respuestas correctas e incorrectas (extraídas del estudiantado real de ediciones anteriores del curso después de ser evaluadas por el profesorado), denominadas *respuestas anotadas*. Cada respuesta puede ser anotada como correcta o incorrecta. Aunque la comparación puede realizarse usando la similitud del coseno para cada respuesta anotada, existe una técnica más eficiente que implica el entrenamiento de un clasificador de IA mediante un *ajuste fino* basado en los *embeddings* de la información anotada. Dicho ajuste especializa el clasificador para centrarse en las respuestas específicas anotadas y recopiladas para una pregunta. El resultado del modelo es la probabilidad de que el *embedding* se clasifique como una respuesta correcta, donde un valor del 100 % significa que el modelo tiene una alta confianza en que la respuesta sea correcta.

Dado que una pregunta de respuesta corta puede crearse desde cero (es decir, se trata de una pregunta nueva) o reutilizarse de otros contextos de aprendizaje, el sistema considera estos dos posibles escenarios (Figura 3). La comparación semántica se utiliza cuando se crea la pregunta y no hay datos disponibles del estudiantado anterior. La comparación semántica utiliza la similitud del coseno con un conjunto de respuestas modelo correctas proporcionadas por el profesorado. La clasificación de respuestas entrena un clasificador de IA cuando hay respuestas anteriores disponibles, proporcionando una predicción sobre si la respuesta es correcta. Como ambas métricas se expresan en porcentajes, la identificación del umbral de calidad para decidir si la respuesta enviada por el estudiante es correcta se resolvió como un problema de optimización, ya que el objetivo es maximizar el umbral mientras que la clasificación correcta no empeora significativamente. Después de algunos experimentos y de la validación con profesorado experto en el dominio, se concluyó que un umbral superior al 60 % identificó principalmente respuestas correctas.

Figura 3
Diseño técnico de SLASys



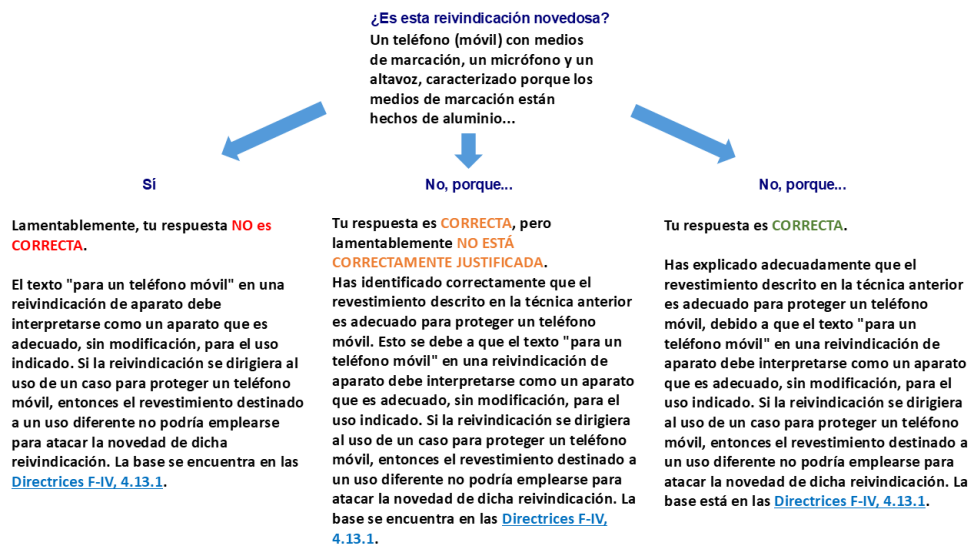
Integración de SLASys con Moodle para la evaluación y provisión de *feedback*

Dado que SLASys ofrece un proceso automático para recomendar la evaluación, el proceso de evaluación podría mejorarse proporcionando *feedback* automatizado. La Figura 4 muestra un ejemplo del *feedback* suministrado sobre la pregunta asociada a la reivindicación de la Figura 2. Es posible definir dos tipos de preguntas:

preguntas de respuesta corta simple o preguntas de opción con una justificación. Este segundo tipo de pregunta es relevante para el profesorado en el examen de patentes, ya que las preguntas suelen plantear alguna reivindicación relacionada con la definición de una patente, y el estudiantado debe examinar si la reivindicación cumple con las directrices de la CEP, como la novedad o la claridad. La Figura 4 muestra un ejemplo del segundo tipo de pregunta con el *feedback* proporcionado y el enlace a las directrices de la CEP según el resultado de la evaluación. El *feedback* está predefinido para cada opción y se ha configurado en el diseño de la pregunta, proporcionando información significativa para que el estudiantado comprenda el resultado de la evaluación y mejore su aprendizaje.

Figura 4

Ejemplo de feedback de una pregunta sobre Novedad



SLASys incorpora una API con diferentes servicios que se pueden integrar en cualquier SGA. En concreto, se utilizó una integración con Moodle mediante el desarrollo de un complemento (*plugin*) adaptado para acceder a las capacidades de SLASys. Esto permitió ocultar los aspectos más complejos, proporcionando una interfaz fácil de usar y de gestionar por el profesorado. Se integraron los nuevos tipos de pregunta en Moodle, adaptando la configuración de preguntas y la interfaz de evaluación para el profesorado, así como la interfaz del estudiantado para enviar las respuestas y revisar los resultados de la evaluación. El diseño minimiza la carga de trabajo del profesorado con una curva de aprendizaje extremadamente baja.

Se puede añadir una nueva pregunta a un test, proporcionando la información obligatoria siguiente: el tipo de pregunta, el enunciado y el conjunto de respuestas modelo del profesorado. Opcionalmente, se pueden agregar o recuperar las respuestas anotadas del estudiantado de ediciones anteriores del curso. Cuando no se proporciona esta última información, SLASys utiliza el enfoque de comparación semántica. En el caso contrario, el clasificador de IA de respuestas correspondiente se entrena automáticamente sin intervención técnica (Figura 3).

La particularidad más destacada está en la configuración de la evaluación y el *feedback*. La pregunta incorpora una rúbrica (un elemento no disponible para preguntas de respuesta corta en Moodle), permitiendo unificar los criterios de evaluación entre el profesorado y ofrecer el *feedback* correspondiente según los resultados de la evaluación.

La Figura 5.a) ilustra la interfaz del estudiantado para enviar una respuesta a una pregunta de opción habilitada. La figura también muestra los diferentes niveles de la rúbrica y las calificaciones asociadas. La respuesta entregada se envía a SLASys, y la recomendación de evaluación se devuelve a Moodle cuando se ha procesado. La recomendación se muestra en la interfaz de evaluación (Figura 6). El profesorado puede visualizar las respuestas modelo, la recomendación de SLASys, la información interpretable del modelo de clasificación de respuestas y la puntuación de la métrica. La información interpretable utiliza colores para indicar qué palabras impactan positiva o negativamente en las decisiones del clasificador. Aunque a veces se incluyen palabras irrelevantes, el profesorado aún puede identificar las palabras incorrectas que afectan la forma en que se clasifican las respuestas del estudiantado. La interfaz también sugiere la calificación, así como el *feedback* según la rúbrica que se selecciona automáticamente dependiendo de la recomendación. El profesorado debe revisar toda esta información y seleccionar el nivel de rúbrica correcto.

Después de ser evaluados, los resultados del test se pueden revisar en la interfaz del estudiantado (Figura 5.b), de manera que cada estudiante puede ver su calificación y el *feedback* con base en la opción de rúbrica seleccionada, validada por el profesorado.

Figura 5
Interfaz de la pregunta en Moodle para el estudiantado

Enunciado

In a telephone apparatus with an automatic dialler, a dial tone detector and a feature controller, the dial tone detector comprising ...

Respuesta de opción

Select answer

☒ Compliant with the requirements of Art. 84 EPC

☐ Not compliant with the requirements of Art. 84 EPC, because:

Respuesta del estudiante

Justification answer

Rubric (review, no manual comment)

Response	The option answer is correct, but the justification answer is incorrect	The option answer is incorrect	The option and justification answers are correct
	0 points	0 points	10 points

Evaluación de la respuesta de opción

Select answer

☒ Compliant with the requirements of Art. 84 EPC

☐ Not compliant with the requirements of Art. 84 EPC, because:

Feedback

Comment: Well done. The provided answer is correct.

The correct solution is The expression "in" makes it unclear whether protection is sought for the dial tone detector per se or a telephone apparatus (see Guidelines F-IV 4.15).

Evaluación mediante rúbrica

Response	The option answer is correct, but the justification answer is incorrect	The option answer is incorrect	The option and justification answers are correct
	0 points	0 points	10 points

Calificación

a) Interfaz del intento

b) Interfaz del *Feedback*

Figura 6
Interfaz de evaluación en Moodle para el profesorado

Model answer

The expression "in" makes it unclear whether protection is sought for the dial tone detector per se or a telephone apparatus (see Guidelines F-IV 4.15).

Respuesta modelo

Recomendación de evaluación

Assessment Recommendation

Well done. The provided answer is correct.

Información interpretable

Legend: ☒ Negative ☐ Neutral ☐ Positive

True Label	Predicted Label	Attribution Score	Word Importance
1	1 (1.00)	-0.25	[CLS] the claim does not clearly defines the subject-matter of the invention, as the term "in" makes it unclear whether the entire telephone apparatus is included in the scope of protection. [SEP]

Rúbrica

Response	The option answer is correct, but the justification answer is incorrect	The option answer is incorrect	The option and justification answers are correct
	0 points	0 points	10 points

Recomendación de feedback

Comment

Well done. The provided answer is correct.

The correct solution is The expression "in" makes it unclear whether protection is sought for the dial tone detector per se or a telephone apparatus (see Guidelines F-IV 4.15)

Calificación score **Métrica**

10.00 out of 10.00 1.000

Análisis de datos

Para responder a las preguntas de investigación se han utilizado dos fuentes de datos. En primer lugar, la información de SLASys se ha empleado para analizar el número de evaluaciones correctas sobre las respuestas utilizadas para el entrenamiento (*rendimiento*) (PI1). A continuación, usando el complemento de Moodle, se recuperaron las recomendaciones proporcionadas por SLASys y la evaluación realizada por el profesorado para analizar las recomendaciones correctas durante la prueba con los estudiantes (PI2 y PI3). El rendimiento durante la fase de entrenamiento y prueba con el estudiantado se puede calcular a partir de las siguientes métricas:

$$TNR = \frac{TN}{TN + FP} \quad ACC = \frac{TP + TN}{TP + FP + TN + FN}$$

$$TPR = \frac{TP}{TP + FN} \quad F_{1,5} = \frac{(1 + 1,5^2)TP}{(1 + 1,5^2)TP + 1,5^2FN + FP}$$

Donde TP denota el número de respuestas correctas correctamente identificadas, TN el número de respuestas incorrectas correctamente identificadas, FP el número de respuestas incorrectas no identificadas correctamente y FN el número de respuestas correctas no identificadas correctamente. Estas cuatro métricas permiten calcular la precisión en la detección de respuestas correctas e incorrectas (ACC), la precisión en la detección de respuestas correctas (*True Positive Rate*—TPR), la precisión en la detección de respuestas incorrectas (*True Negative Rate*—TNR) y el porcentaje de respuestas correctas con penalización por las identificaciones incorrectas (*F score* - $F_{1,5}$). Estas métricas se utilizan para evaluar el rendimiento de los enfoques de comparación semántica y de clasificación de respuestas (PI1) y para determinar si SLASys recomienda la evaluación esperada por el profesorado en un entorno educativo real (PI2 y PI3).

Por último, se recopiló el tiempo medio de acceso del estudiantado a la interfaz de *feedback* y sus valoraciones sobre los resultados proporcionados. Para ello, el complemento de Moodle añadió una escala Likert para recopilar las valoraciones del estudiantado en la interfaz de *feedback*. La escala osciló entre 1 y 5 (donde 5 significa que el *feedback* se considera muy útil). En cuanto a la experiencia del profesorado, solo se recopilaron sus comentarios sobre el uso del sistema porque únicamente estuvieron implicados tres profesores (PI4).

RESULTADOS

PI1: ¿Cómo de precisa es la clasificación de respuestas con respecto a la comparación semántica para la recomendación en la CAPRC?

Conocer la precisión de SLASys antes del primer piloto con el estudiantado es crucial para analizar la aplicabilidad del sistema en un entorno educativo real. Los resultados se detallan en la Tabla 1, donde se analizaron las ocho preguntas del test de Claridad.

Tabla 1
Métricas de rendimiento en el test de Claridad con los datos de entramiento

Ejercicio	Correcto	Incorrecto	Comparación semántica				Clasificación de respuestas			
	Entren. / Valid.	Entren. / Valid.	ACC	TPR	TNR	F1,5	ACC	TPR	TNR	F1,5
Pregunta 1	36 / 8	107 / 22	37 %	100 %	14 %	58 %	87 %	88 %	86 %	81 %
Pregunta 2	47 / 10	7 / 2	67 %	70 %	50 %	75 %	100 %	100 %	100 %	100 %
Pregunta 3	3 / 2	46 / 4	42 %	100 %	30 %	48 %	83 %	0 %	100 %	0 %
Pregunta 4	34 / 7	17 / 20	27 %	0 %	75 %	0 %	64 %	86 %	25 %	79 %
Pregunta 5	67 / 14	100 / 20	53 %	100 %	20 %	74 %	82 %	71 %	90 %	75 %
Pregunta 6	82 / 17	36 / 8	84 %	94 %	63 %	91 %	88 %	100 %	63 %	95 %
Pregunta 7	33 / 7	33 / 7	57 %	86 %	29 %	73 %	79 %	71 %	86 %	75 %
Pregunta 8	68 / 14	19 / 4	83 %	93 %	50 %	91 %	83 %	100 %	25 %	94 %
Promedio			56 %	80 %	41 %	64 %	83 %	77 %	72 %	75 %

La tabla resume el número de respuestas anotadas utilizadas para el ajuste fino de los clasificadores de IA (es decir, el *conjunto de datos de entrenamiento*, *entren.* en la tabla), para evaluar su precisión (es decir, el *conjunto de datos de validación*, *valid.* en la tabla) y los resultados de las métricas. Los clasificadores de IA se ajustaron usando el 80 % de las respuestas anotadas (nótese que existe un clasificador para cada pregunta). El 20 % restante se utilizó para realizar la validación y obtener los resultados de la tabla. El proceso de división de los datos considera respuestas correctas e incorrectas en ambos conjuntos de datos. Cabe destacar que no hay fase de ajuste de la precisión en el enfoque de comparación semántica. En este caso, se compararon tres respuestas correctas seleccionadas de forma aleatoria con las respuestas de validación.

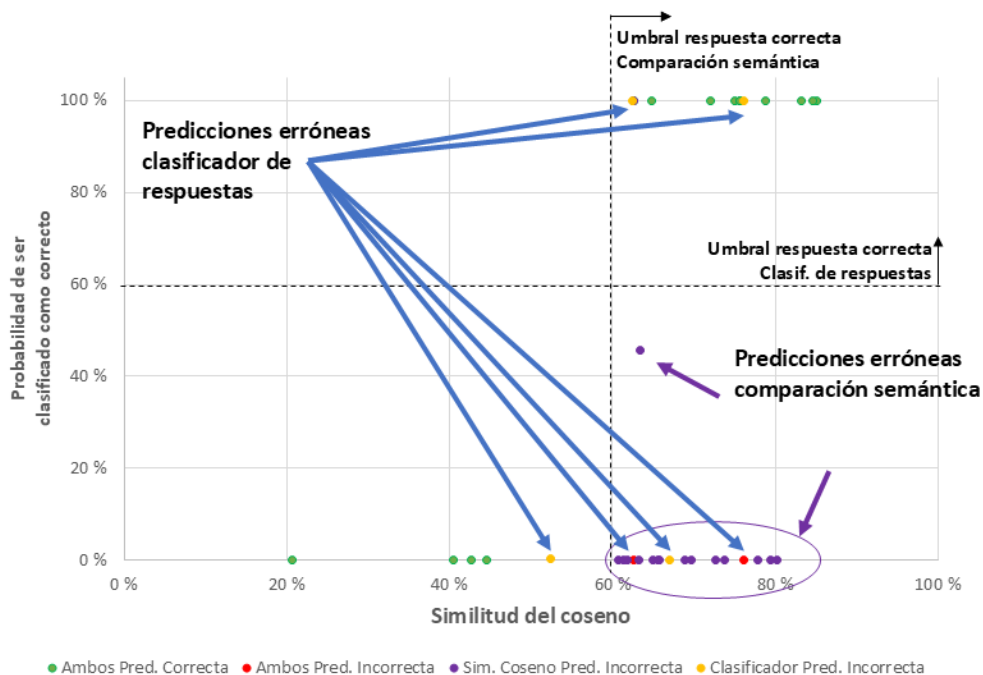
La clasificación de respuestas supera a la comparación semántica porque las respuestas anotadas proporcionan ejemplos de respuestas correctas e incorrectas.

La comparación semántica suele fallar cuando existen múltiples formas de responder a una pregunta, y las respuestas están semánticamente alejadas de las respuestas modelo del profesorado. Con respecto a las métricas para detectar respuestas correctas e incorrectas (es decir, el TPR y el TNR), la calidad de los resultados depende en gran medida del número de respuestas proporcionadas por el profesorado para el entrenamiento y del equilibrio entre respuestas correctas e incorrectas. Los clasificadores con pocas respuestas tienen dificultades para realizar recomendaciones. Se puede observar que la comparación semántica puede detectar más respuestas correctas en promedio (un 80 % en comparación con el 77 %). Sin embargo, este incremento se debe a que muchas respuestas incorrectas se identificaron incorrectamente como correctas (un 41 % en comparación con el 72 %).

Para comprender mejor el comportamiento de las técnicas, se analizaron las predicciones individuales. La Figura 7 muestra el análisis del conjunto de datos de validación para la Pregunta 5. Las métricas respectivas, la similitud del coseno y la probabilidad de que se predigan como correctas, se utilizan para la comparación semántica y la clasificación de respuestas, respectivamente. Se han utilizado colores diferentes para identificar qué predicciones son correctas utilizando ambas métricas (verde), cuáles se predicen incorrectamente mediante la comparación semántica (morado), cuáles se predicen incorrectamente mediante la clasificación de respuestas (amarillo) y cuáles no se pueden predecir por ninguno de los dos métodos (rojo).

Figura 7

Comparación del rendimiento de las métricas para la Pregunta 5



La mayoría de los errores aparecen en la comparación semántica. Las respuestas de buena parte del estudiantado se consideran correctas (es decir, semánticamente similares a la respuesta modelo proporcionada por el profesorado), con un rango del 60 % al 85 %, pero son incorrectas. Por lo tanto, pequeños cambios en el significado de la frase pueden generar predicciones incorrectas. La clasificación de respuestas también proporciona algunas predicciones erróneas en la detección de respuestas correctas, pero en menos casos. Además, existen casos en los que ambos métodos fallan. El profesorado observó que algunas de estas respuestas se predijeron incorrectamente porque no se cubrieron en las respuestas utilizadas para el ajuste fino ni en las respuestas modelo del profesorado.

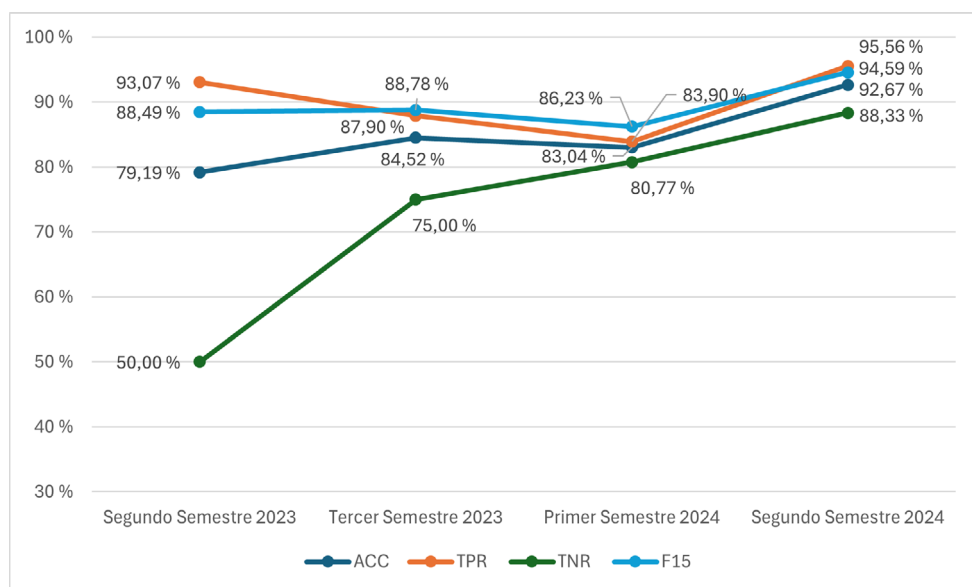
PI2: ¿Cómo de preciso es SLASys en un entorno educativo real?

La Figura 7 muestra algunos casos erróneos. Añadir dichas respuestas a los clasificadores de IA podría mejorar la identificación en nuevas ediciones del curso. Esta es la idea fundamental para mejorar los clasificadores: estos se reentrenan, incluyendo los resultados de la evaluación realizados en las ediciones anteriores.

De este modo, se perfeccionan con nuevas respuestas anotadas que mejoran la identificación. La PI2 analiza si la clasificación de evaluación de respuestas que SLASys recomienda al profesorado durante las cuatro ediciones del curso mejora, teniendo en cuenta el ajuste fino efectuado en cada edición. Los resultados agregados de las distintas métricas se han resumido en el gráfico de la Figura 8 para el test de Claridad. Los resultados de todas las preguntas se han agregado para simplificar la evaluación.

Figura 8

Rendimiento de las métricas en el test de Claridad



Se puede observar una mejora significativa en la segunda edición debido a las nuevas respuestas anotadas de la primera edición, que mejoraron la precisión en las preguntas con un número desequilibrado de respuestas correctas e incorrectas (es decir, las preguntas 2, 4 y 8 de la Tabla 1). Aunque hay una ligera reducción en la detección correcta de las respuestas correctas (es decir, el TPR), esto se debe al proceso de refinamiento.

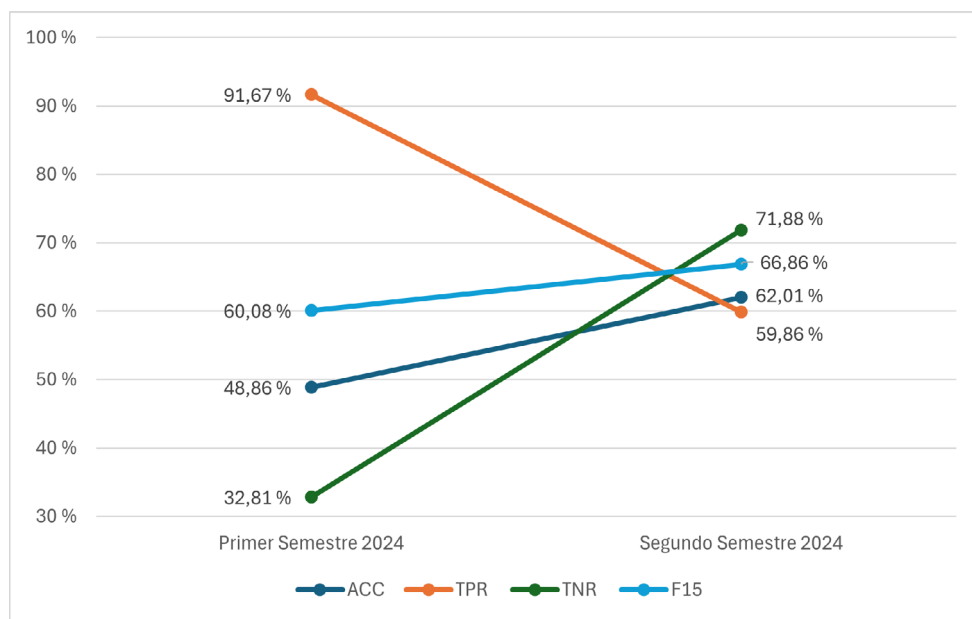
Las siguientes ediciones del curso contribuyeron a mejorar la precisión general, alcanzando valores superiores al 88 % en la última edición. Por lo tanto, incluir nuevos datos puede mejorar de forma eficaz la calidad del recomendador, lo que implica que puede utilizarse de forma efectiva con fines de evaluación. Una mayor precisión significa disponer de un recomendador con menos errores en el que el profesorado puede confiar durante el proceso de evaluación.

PI3: ¿Puede utilizarse SLASys sin respuestas previas del estudiantado?

Los buenos resultados mostrados en la sección anterior se obtuvieron porque el profesorado realizó una tarea inicial de anotación de las respuestas proporcionadas por el estudiantado anterior, lo que contribuyó a generar los clasificadores iniciales.

Sin embargo, se desea saber cómo funciona SLASys ante nuevas preguntas para las que no se dispone de respuestas anteriores. En este caso, solo se puede utilizar la comparación semántica con las respuestas modelo proporcionadas por el profesorado en la primera edición (se incluyeron tres respuestas modelo por pregunta). Sin embargo, en la segunda edición, la clasificación de respuestas puede aplicarse utilizando la evaluación realizada en la primera edición. El experimento con estudiantes reales se llevó a cabo en las dos últimas ediciones, es decir, en el primer y segundo trimestre de 2024, donde se diseñó un nuevo test para el concepto de Novedad. La Figura 9 muestra la mejora de SLASys al cambiar de la comparación semántica a la clasificación de respuestas en un entorno educativo real.

Figura 9
Rendimiento de las métricas en el test de Novedad



Se puede observar una diferencia significativa entre ambos enfoques. La precisión global (es decir, el ACC) mejora del 48,86 % al 62,01 %. Esta mejora se debe al incremento significativo en la detección de respuestas incorrectas (el TNR

aumentó del 32,81 % al 71,88 %). Sin embargo, la detección de respuestas correctas disminuye. La comparación semántica tiende a recomendar que la mayoría de las respuestas del estudiantado son correctas, ya que no puede distinguir las ligeras diferencias semánticas. Esto produce el efecto no deseado de que la comparación semántica no puede detectar correctamente la mayoría de las respuestas incorrectas. Por lo tanto, el cambio de la comparación semántica a la clasificación de respuestas tiene un impacto positivo en el recomendador.

El hallazgo relevante de este experimento es que los modelos de clasificación de respuestas pueden producir recomendaciones de alta calidad en pocas ediciones, incluso comenzando desde cero con un enfoque basado en comparación semántica.

PI4: ¿Cuál es la opinión del estudiantado y del profesorado?

Finalmente, se desea determinar si los nuevos tests y el *feedback* fueron útiles. La Tabla 2 resume la información recopilada del estudiantado, incluyendo el número de estudiantes que accedió al *feedback*, el tiempo promedio dedicado a la página y la valoración final otorgada al *feedback* recibido. No todo el estudiantado accedió al *feedback*, probablemente porque los tests no afectaron a la superación del curso. Además, el tiempo promedio de dedicación es relativamente bajo. Al comprobar el tiempo de acceso individual, se observa que este se correlaciona en gran medida con el número de respuestas incorrectas. En consecuencia, el estudiantado con respuestas incorrectas dedicó más tiempo a leer el *feedback*. Por lo tanto, se puede deducir que comprobó lo que debe aprender para mejorar. Por último, la valoración promedio es significativamente alta (superior al 70 %) en todas las ediciones y tests.

Tabla 2

Tiempo de acceso al feedback y valoración del estudiantado

Test	Claridad			Novedad		
	Segundo	Tercer	Primer	Segundo	Primer	Segundo
	Semestre	Semestre	Semestre	Semestre	Semestre	Semestre
Edición	2023	2023	2024	2024	2024	2024
Acceso estudiantado (%)	83,33	84,62	88,89	80,00	77,78	76,00
Tiempo medio de acceso (minutos)	2,33	1,73	3,51	2,91	4,74	4,68
Valoración (1-5)	3,94	3,84	3,63	3,85	3,55	3,77

Tabla 3

Comparación con trabajos relacionados

Referencia	Técnica	Predicción	Feedback	ACC
(Saha et al., 2018)	PLN	Corrección	No	66 %
(Soulimani et al., 2024)	Clasif. BERT	Calificación (0-4)*	No	71 %
(Padó et al., 2024)	Clasif. BERT	Corrección	No	72 %
(Wang et al., 2019)	Clasif. BERT	Corrección	No	80 %
(Camus y Filighera, 2020)	Clasif. BERT	Corrección	No	80 %
(Lun et al., 2020)	Clasif. BERT	Corrección	No	82 %
(Sung et al., 2019)	Clasif. BERT	Corrección	No	84 %
(Schneider et al., 2023)	Clasif. BERT	Corrección	No	86 %
(Liu et al., 2019)	Clasif. BERT	Corrección	No	89 %
(Grévisse, 2024)	IAGen (GPT4)	Calificación (0-10)*	Generado	64 %
(Aggarwal et al., 2025)	IAGen (Mistral)	Corrección	Generado	75 %
SLASys	Clasif. BERT	Corrección	Predefinido	83 %

(*) La precisión se calcula asumiendo como correcto cuando la calificación es superior a la mitad de los puntos máximos.

Además, el profesorado recopiló algunas opiniones durante las sesiones en línea. El estudiantado estuvo de acuerdo en que la calificación de la evaluación, el *feedback* sobre las respuestas y las directrices de la CEP a revisar fueron muy valiosos. Además, obtenerlos casi de forma inmediata fue positivo. Sin embargo, el estudiantado se quejó de que el *feedback* debería ser más personalizado, describiendo por qué su respuesta era incorrecta.

El profesorado también compartió su experiencia con el recomendador. Inicialmente, era reacio a usarlo, argumentando que confiar en la calificación automática podría dar lugar a evaluaciones erróneas. Sin embargo, la herramienta fue altamente aceptada como recomendador, puesto que la revisión final del profesorado era obligatoria. Después de evaluar las diferentes ediciones, el profesorado coincidió en que ganó eficiencia, ya que también adquirió experiencia con el tiempo. Con relación al experimento con el test de Novedad sin respuestas anotadas, se quejó en la primera edición (es decir, en el primer trimestre de 2024) puesto que no observó ninguna mejora en la reducción de su carga de trabajo, ya que el recomendador produjo muchas recomendaciones incorrectas. Sin embargo, constató el beneficio en la última edición cuando se pudo utilizar la clasificación de respuestas.

DISCUSIÓN

Los resultados obtenidos permiten responder a las preguntas de investigación. En relación con la PI1, como señalan otros trabajos (Burrows et al., 2015), los clasificadores de IA de BERT (es decir, la clasificación de respuestas) muestran una mayor precisión con respecto a la comparación semántica, y obtienen resultados similares a otros sistemas de CAPRC. La Tabla 3 compara SLASys con otros sistemas que usan conjuntos de datos públicos y que han sido reportados en la literatura. En concreto, se resume la técnica utilizada, el objetivo de predicción, el tipo de *feedback* proporcionado y la precisión obtenida. La clasificación de respuestas de SLASys alcanza una precisión similar a la obtenida por los trabajos relacionados que se centran en predecir la correctitud de respuestas, con una precisión que oscila entre el 71 % y el 89 %. Sin embargo, no es posible comparar SLASys con trabajos centrados en la predicción de calificaciones, ya que emplean métricas distintas para evaluar lo cerca que está la calificación predecida de la correcta (concretamente, se usa la *raíz cuadrada del error cuadrático medio*). Tales técnicas tienen tasas de error que oscilan entre moderadas y altas, con valores que van desde el 0,57 (Baral et al., 2021) hasta valores mayores que 1 (Gaddipati et al., 2020; Metzler et al., 2024), lo que dificulta su uso en un entorno educativo real. Cabe señalar que predecir la calificación es una tarea más compleja, ya que existe más variabilidad en el resultado (del Gobbo et al., 2023). Por lo tanto, el método propuesto en este trabajo simplifica el proceso, al proporcionar el resultado de la evaluación al profesorado, que decide la calificación en función de una rúbrica. Además, los resultados experimentales ofrecen información sobre el tamaño reducido del conjunto de datos necesario para obtener un clasificador ajustado para preguntas de respuesta corta (Mehrafarin et al., 2022). Por lo general, se recomienda tener un mínimo de 500 respuestas (es decir, *instancias*) para disponer de un clasificador adecuado. Sin embargo, la Tabla 1 muestra que se puede entrenar un buen clasificador incluso con menos respuestas. Además, se han comparado los métodos de clasificación semántica y de clasificación de preguntas, y sus métricas asociadas, para una pregunta específica (Figura 7). El profesorado puede acceder a esta información en Moodle para comprender mejor qué respuestas no están bien clasificadas. Cabe recordar que la interpretabilidad de la IA es una de las recomendaciones para las herramientas de IA (Zhao et al., 2024). La Tabla 3 también compara SLASys con trabajos basados en IAGen, que actualmente muestran un rendimiento inferior que los modelos basados en BERT. Aunque las herramientas IAGen generan un gran entusiasmo, todavía tienen algunos inconvenientes en el ámbito de la evaluación. Pueden incurrir en alucinaciones (Jia et al., 2024) o mostrar limitaciones conceptuales en dominios específicos (De La Cruz et al., 2024), lo que puede disminuir su eficacia para evaluar. Algunos de estos modelos de IAGen se pueden utilizar localmente. Sin embargo, las soluciones empresariales son ampliamente utilizadas. Estas operan de forma remota, ofreciendo una mayor capacidad de procesamiento, pero a mayores costes y con problemas de sostenibilidad

(van Wynsberghe, 2021). Además, requieren de un análisis de cuestiones éticas y el replanteamiento de las políticas educativas y de protección de datos. Es importante que las instituciones de educación superior revisen el manifiesto propuesto por García-Peñalvo et al. (2024) para obtener un sistema seguro y ético. SLASys opera localmente, reduciendo los costes de procesamiento, manteniendo su funcionalidad y no proporcionando datos a terceros. Además, no recopila datos confidenciales del estudiantado, dado que estos se mantienen seguros dentro del SGA.

Con respecto a la PI2, se ha mostrado la progresión de SLASys a lo largo de cuatro ediciones del curso. Como afirma la pregunta de investigación anterior, los conjuntos de respuestas anotadas (aunque sean reducidos) producen clasificadores de alta precisión, lo que hace disponible SLASys para su uso en un entorno educativo real. Al generar clasificadores precisos, SLASys es un ejemplo de una herramienta de IA que puede beneficiar a ambas partes. Por un lado, el profesorado puede beneficiarse de un recomendador para la CAPRC, lo que aumenta la eficiencia de la evaluación y unifica los criterios de evaluación (Xavier et al., 2025). Además, los beneficios para el profesorado son más claros a medida que se reduce su carga de trabajo, lo que le permite dedicar más tiempo a otras tareas cualitativas como la revisión del *feedback*, el diseño de nuevas preguntas o actividades de aprendizaje mejor diseñadas. Debido a la integración con Moodle, SLASys proporciona *feedback* formativo al estudiantado en comparación con los trabajos relacionados descritos en la Tabla 3. Estos trabajos únicamente proporcionan una recomendación sin *feedback*, que, en el caso de su aplicación en un entorno real, necesitaría de métodos adicionales para proporcionarlo. Según Gaddipati et al. (2021), una opción podría ser utilizar herramientas IAGen que muestran potencial para la generación de *feedback*. Por otro lado, con SLASys, el estudiantado obtiene *feedback* de calidad cuando se le hacen preguntas de razonamiento (Calimeris y Kosack, 2020). Con este *feedback*, el estudiantado puede aprender mejor con información más completa y detallada sobre su tarea de evaluación. En última instancia, este *feedback* se convierte en un componente personalizado que mejora el conocimiento del estudiantado (Abu Khurma et al., 2024).

En relación con la PI3, los resultados demuestran que SLASys se puede utilizar incluso sin datos de entrenamiento. En la primera edición que se proponga una nueva pregunta, el profesorado necesitará prestar más atención debido a las recomendaciones de baja precisión. Sin embargo, el clasificador de IA estará listo para su uso en la próxima edición. Es importante destacar también la simplicidad técnica del sistema. SLASys sigue las buenas prácticas de evaluación descritas en Petridou y Lao (2024).

Por último, en relación con la PI4, las opiniones y los resultados indican que las predicciones erróneas del recomendador no afectan al estudiantado, ya que este obtiene la evaluación revisada por el profesorado. En este sentido, la supervisión humana es obligatoria porque los errores en la evaluación pueden tener efectos negativos en el estudiantado (Li et al., 2023). Al igual que en Sangapu (2018), las

opiniones son principalmente positivas con la introducción de herramientas de IA en la educación. El profesorado no se siente abrumado por los problemas técnicos, ya que SLASys está integrado en Moodle. De forma similar, SLASys cambiará automáticamente de una técnica a otra sin la intervención del profesorado. Igualmente relevante, SLASys se ha diseñado como un recomendador de evaluación siguiendo las recomendaciones de la Ley Europea de IA (European Commission, 2024), evitando su uso como calificador automático basado en IA.

Este estudio presenta limitaciones relacionadas con el tamaño de la muestra y el muestreo no probabilístico que afectan a la validez externa. Debido a su aplicación en un entorno educativo real, se utilizó un conjunto de datos reducido de un dominio específico, lo que restringe la generalización de los hallazgos y puede generar sesgo muestral. Esto se puede observar principalmente en la PI3, donde el tamaño de la muestra para la última edición del curso es de solo 25 estudiantes. Sin embargo, los hallazgos que se derivan de la PI2 demuestran que SLASys mejora su precisión y sigue siendo aplicable y eficaz a lo largo del tiempo (es decir, en las cuatro ediciones del curso) incluso cuando los clasificadores de IA se entrenan con datos limitados. En cuanto al muestreo no probabilístico, todo el estudiantado participó en el estudio, lo que hizo inviable el análisis de adquisición de conocimientos con los nuevos tests. Cabe señalar que el primer examen dentro de la formación completa se realiza después del segundo curso. Un análisis longitudinal hasta ese examen enriquecería la comprensión del impacto de los nuevos tests sobre el conocimiento adquirido. El tamaño de la muestra también afecta a la PI4. Las opiniones positivas del estudiantado podrían estar sesgadas debido al tamaño de la muestra. Además, las opiniones del profesorado se recopilaban con entrevistas semiestructuradas simples. Un análisis cualitativo con grupos focales proporcionaría una visión más detallada y profunda.

Aunque los resultados experimentales obtenidos no se pueden generalizar, se destaca la facilidad de transferencia a distintos dominios con un lenguaje especializado, como Derecho, Historia, Filosofía o Ciencias Sociales. La transferibilidad dependería de la carga de trabajo inicial para diseñar las preguntas, añadir las respuestas modelo del profesorado correspondientes, y el esfuerzo de recopilar y revisar las respuestas anotadas. SLASys permite la creación de preguntas en las que el estudiantado debe aplicar el conocimiento adquirido, en lugar de responder basándose en conceptos teóricos, lo que conduce a una evaluación auténtica (Villarroel et al., 2018).

CONCLUSIONES

Esta investigación ofrece tres contribuciones principales. En primer lugar, se desarrolla un sistema de recomendación para la evaluación de preguntas de respuesta corta (SLASys) guiado por tres principios fundamentales: uso de técnicas gratuitas (BERT) y adecuadas para implementaciones privadas, ejecución en servidores con recursos limitados y facilidad de uso sin necesidad de conocimientos técnicos en

IA. En segundo lugar, una característica destacada es la integración con Moodle, que permite el seguimiento del estudiantado, la calificación y la visualización de los resultados de aprendizaje a través de SLASys. Los resultados han demostrado que, aunque no puedan generalizarse, la clasificación de respuestas puede conducir a mejores recomendaciones. Asimismo, SLASys permite refinar el recomendador para producir mejores resultados con nuevas cohortes de estudiantado. En tercer lugar, el trabajo subraya el papel de la IA como herramienta de apoyo en la educación, empoderando al profesorado para utilizar una herramienta diseñada para recomendar las evaluaciones y el *feedback* que puede aplicarse en diversos dominios. El profesorado puede emplear estrategias integradoras, destacando la convergencia de la tecnología y la pedagogía para mejorar la experiencia de aprendizaje del estudiantado.

Como trabajo futuro, se plantea extender la utilización del recomendador en otros tests del mismo curso fundamental de examen de patentes en el dominio de la Propiedad Intelectual. Esto permitirá evaluar la capacidad del sistema para abordar otros conceptos. Asimismo, se propone explorar cómo se puede personalizar el *feedback* en función de las respuestas incorrectas del estudiantado y de los recursos de aprendizaje disponibles. BERT se puede emplear para crear herramientas de búsqueda semántica basadas en los materiales del curso, capaces de identificar automáticamente los recursos necesarios para una pregunta específica, lo que podría contribuir a optimizar el esfuerzo del profesorado en la generación del *feedback*.

Agradecimientos

Este trabajo forma parte del proyecto PID2023-147592OB-I00 financiado por MCIU/AEI/10.13039/501100011033/ FEDER, UE, y del Programa de Investigación Académica de la Oficina Europea de Patentes, Acuerdo de Subvención N° 2021/8404.

REFERENCIAS

- Abu Khurma, O., Albahti, F., Ali, N. y Bustanji, A. (2024). AI ChatGPT and student engagement: Unraveling dimensions through PRISMA analysis for enhanced learning experiences. *Contemporary Educational Technology*, 16(2). <https://doi.org/10.30935/cedtech/14334>
- Adhikari, A., Ram, A., Tang, R. y Lin, J. (2019). DocBERT: BERT for document classification. arXiv. <https://doi.org/10.48550/arXiv.1904.08398>
- Aggarwal, D., Sil, P., Raman, B. y Bhattacharyya, P. (2025). "I understand why I got this grade": Automatic short answer grading with feedback. En *Lecture Notes in Computer Science*. https://doi.org/10.1007/978-3-031-98420-4_22
- Akçapınar, G. (2015). How automated feedback through text mining changes plagiaristic behavior in online assignments. *Computers & Education*, 87, 123-130. <https://doi.org/10.1016/j.compedu.2015.04.007>
- Almasre, M. (2024). Development and evaluation of a custom GPT for the assessment of students' designs in

- a typography course. *Education Sciences*, 14(2), Article 148. <https://doi.org/10.3390/educsci14020148>
- Arefeen, M. A., Debnath, B. y Chakradhar, S. (2024). LeanContext: Cost-efficient domain-specific question answering using LLMs. *Natural Language Processing Journal*, 7, 100065. <https://doi.org/10.1016/j.nlp.2024.100065>
- Bahdanau, D., Cho, K. y Bengio, Y. (2016). Neural machine translation by jointly learning to align and translate. *arXiv*. <https://doi.org/10.48550/arXiv.1409.0473>
- 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? *International Journal of Educational Technology in Higher Education*, 21(1), 23. <https://doi.org/10.1186/s41239-024-00455-4>
- Baral, S., Botelho, A. F., Erickson, J. A., Benachamardi, P. y Heffernan, N. T. (2021). Improving automated scoring of student open responses in mathematics. En *Proceedings of the 14th International Conference on Educational Data Mining (EDM 2021)*.
- Bergmann, J. y Sams, A. (2012). *Flip your classroom: Reach every student in every class every day*. International Society for Technology in Education.
- Burrows, S., Gurevych, I. y Stein, B. (2015). The eras and trends of automatic short answer grading. *International Journal of Artificial Intelligence in Education*, 25(1), 60-117. <https://doi.org/10.1007/s40593-014-0026-8>
- Calimeris, L. y Kosack, E. (2020). Immediate feedback assessment technique (IF-AT) quizzes and student performance in microeconomic principles courses. *Journal of Economic Education*, 51(3-4), 304-319. <https://doi.org/10.1080/00220485.2020.1804501>
- Camus, L. y Filighera, A. (2020). Investigating transformers for automatic short answer grading. En I. I. Bittencourt, M. Cukurova, K. Muldner, R. Luckin y E. Millán (Eds.), *Artificial intelligence in education* (Lecture Notes in Computer Science, Vol. 12164, pp. 43-48). Springer. https://doi.org/10.1007/978-3-030-52240-7_8
- Dai, Y., Lai, S., Lim, C. P. y Liu, A. (2025). University policies on generative AI in Asia: Promising practices, gaps, and future directions. *Journal of Asian Public Policy*, 18(2), 260-281. <https://doi.org/10.1080/17516234.2024.2379070>
- del Gobbo, E., Guarino, A., Cafarelli, B. y Grilli, L. (2023). GradeAid: A framework for automatic short answers grading in educational contexts-Design, implementation and evaluation. *Knowledge and Information Systems*, 65(10), 4479-4507. <https://doi.org/10.1007/s10115-023-01892-9>
- De La Cruz Martínez, G., Eslava-Cervantes, A.-L. y Ramírez, S. (2024, July 1). Analysis of solutions of ChatGPT to logic problems based on critical thinking. En *Proceedings of the 16th International Conference on Education and New Learning Technologies (EDULEARN24)* (pp. 10324-10331). IATED. <https://doi.org/10.21125/edulearn.2024.2525>
- Devlin, J., Chang, M.-W., Lee, K. y Toutanova, K. (2019). *BERT: Pre-training of deep bidirectional transformers for language understanding*. En J. Burstein, C. Doran y T. Solorio (Eds.), *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 4171-4186). Association for Computational Linguistics. <https://doi.org/10.18653/v1/N19-1423>
- Dhananjaya, G. M., Goudar, R. H., Kulkarni, A. A., Rathod, V. N. y Hukkeri, G. S. (2024). A digital recommendation system for

- personalized learning to enhance online education: A review. *IEEE Access*, 12, 33591–33615. <https://doi.org/10.1109/ACCESS.2024.3369901>
- Escalante, J., Pack, A. y Barrett, A. (2023). AI-generated feedback on writing: Insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education*, 20(1), 40. <https://doi.org/10.1186/s41239-023-00425-2>
- European Commission. (2024). *Regulation (EU) 2024/1689 of the European Parliament and of the Council*. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32024R1689>
- Evans, C. (2013). Making sense of assessment feedback in higher education. *Review of Educational Research*, 83(1), 70–120. <https://doi.org/10.3102/0034654312474350>
- Gaddipati, S. K., Nair, D. y Plöger, P. G. (2020). Comparative evaluation of pretrained transfer learning models on automatic short answer grading. *arXiv*. <https://arxiv.org/abs/2009.01303>
- Gaddipati, S. K., Plöger, P., Hochgeschwender, N. y Metzler, M. (2021, April 5). *Automatic formative assessment for students' short text answers through feature extraction* [Doctoral dissertation, Hochschule Bonn-Rhein-Sieg].
- Gaona, J., Reguant, M., Valdivia, I., Vázquez, M. y Sancho-Vinuesa, T. (2018). Feedback by automatic assessment systems used in mathematics homework in the engineering field. *Computer Applications in Engineering Education*, 26(4), 921–934. <https://doi.org/10.1002/cae.21950>
- García-Peñalvo, F. J., Alier, M., Pereira, J. y Casany, M. J. (2024). Safe, transparent, and ethical artificial intelligence: Keys to quality sustainable education (SDG4). *International Journal of Educational Research and Innovation*, 22, 1–21. <https://doi.org/10.46661/ijeri.11036>
- González Fernández, M. O., Romero-López, M. A., Sgreccia, N. F. y Latorre Medina, M. J. (2025). Marcos normativos para una IA ética y confiable en la educación superior: Estado de la cuestión. *RIED-Revista Iberoamericana de Educación a Distancia*, 28(2), 181–208. <https://doi.org/10.5944/ried.28.2.43511>
- Grévisse, C. (2024). LLM-based automatic short answer grading in undergraduate medical education. *BMC Medical Education*, 24(1), 1060. <https://doi.org/10.1186/s12909-024-06026-5>
- György, A. y Vajda, I. (2007). Intelligent mathematics assessment in eMax. En *IEEE AFRICON Conference* (pp. 1-7). <https://doi.org/10.1109/AFRCON.2007.4401512>
- Hattie, J. y Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>
- Huisman, B., Saab, N., van Driel, J. y van den Broek, P. (2017). Peer feedback on college students' writing: Exploring the relation between students' ability match, feedback quality and essay performance. *Higher Education Research & Development*, 36(7), 1433–1446. <https://doi.org/10.1080/07294360.2017.1325854>
- Husein, R. A., Aburajouh, H. y Catal, C. (2025). Large language models for code completion: A systematic literature review. *Computer Standards & Interfaces*, 92, 103917. <https://doi.org/10.1016/j.csi.2024.103917>
- Hustad, E. y Arntzen, A. A. B. (2013). Facilitating teaching and learning capabilities in social learning management systems: Challenges, issues, and implications for design. *Journal of Integrated Design and Process Science*, 17(1), 33–46. <https://doi.org/10.3233/JID-2013-0003>
- Jia, Q., Cui, J., Xi, R., Liu, C., Rashid, P., Li, R. y Gehringer, E. (2024). On assessing the faithfulness of LLM-generated feedback on student assignments. En B. Paaßen y

- C. D. Epp (Eds.), *Proceedings of the 17th International Conference on Educational Data Mining* (pp. 491-499). International Educational Data Mining Society. <https://doi.org/10.5281/zenodo.12729868>
- Kim, T. W. (2023). Application of artificial intelligence chatbots, including ChatGPT, in education, scholarly work, programming, and content generation and its prospects: A narrative review. *Journal of Educational Evaluation for Health Professions*, 20, 38. <https://doi.org/10.3352/jeehp.2023.20.38>
- Klein, R., Kyrilov, A. y Tokman, M. (2011). Automated assessment of short free-text responses in computer science using latent semantic analysis. En *Proceedings of ITiCSE '11: The 16th Annual Conference on Innovation and Technology in Computer Science Education* (pp. 158-162). <https://doi.org/10.1145/1999747.1999793>
- Kuechler, W. y Vaishnavi, V. (2012). A framework for theory development in design science research: Multiple perspectives. *Journal of the Association for Information Systems*, 13(6), 395-423. <https://doi.org/10.17705/1JAIS.00300>
- Li, T. W., Hsu, S., Fowler, M., Zhang, Z., Zilles, C. y Karahalios, K. (2023). Am I wrong, or is the autograder wrong? Effects of AI grading mistakes on learning. En *Proceedings of the 2023 ACM Conference on International Computing Education Research (ICER '23)* (pp. 85-97). <https://doi.org/10.1145/3568813.3600124>
- Liu, T., Ding, W., Wang, Z., Tang, J., Huang, G. Y. y Liu, Z. (2019). Automatic short answer grading via multiway attention networks. En *Lecture Notes in Computer Science* (Vol. 11626, pp. 376-388). https://doi.org/10.1007/978-3-030-23207-8_32
- Lun, J., Zhu, J., Tang, Y. y Yang, M. (2020). Multiple data augmentation strategies for improving performance on automatic short answer scoring. En *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 34, pp. 13381-13388). <https://doi.org/10.1609/aaai.v34i09.7062>
- Mehrafarin, H., Rajaei, S. y Pilehvar, M. T. (2022). On the importance of data size in probing fine-tuned models. En *Findings of the Association for Computational Linguistics* (pp. 239-248). <https://doi.org/10.18653/v1/2022.findings-acl.20>
- Messer, M., Brown, N. C. C., Kölling, M. y Shi, M. (2024). Automated grading and feedback tools for programming education: A systematic review. *ACM Transactions on Computing Education*, 24(1), Article 1. <https://doi.org/10.1145/3636515>
- Metzler, T., Plöger, P. G. y Hees, J. (2024). Computer-assisted short answer grading using large language models and rubrics. En *INFORMATIK 2024: AI@WORK* (pp. 1383-1393). Gesellschaft für Informatik e.V. https://doi.org/10.18420/inf2024_121
- Nguyen, H., Bhat, S., Moore, S., Bier, N. y Stamper, J. (2022). Towards generalized methods for automatic question generation in educational domains. En S. Isotani, E. Millán, A. Ogan, P. Hastings, B. McLaren y R. Luckin (Eds.), *Intelligent Tutoring Systems. ITS 2022. Lecture Notes in Computer Science* (Vol. 13450, pp. 272-284). Springer. https://doi.org/10.1007/978-3-031-16290-9_20
- Nicol, D. y Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199-218. <https://doi.org/10.1080/03075070600572090>
- Novak, G. M. (2012). Just-in-time teaching. *New Directions for Teaching and Learning*, 2012(128), 63-73. <https://doi.org/10.1002/tl.469>
- Oates, B. J. (2006). *Researching information systems and computing*. SAGE Publications Ltd.

- OpenAI. (2024). *GPT-4o system card*. <https://openai.com/index/gpt-4o-system-card/>
- Padó, U., Eryilmaz, Y. y Kirschner, L. (2024). Short-answer grading for German: Addressing the challenges. *International Journal of Artificial Intelligence in Education*, 34(4), 1488-1510. <https://doi.org/10.1007/s40593-023-00383-w>
- Pang, J., Ye, F., Wong, D. F., Yu, D., Shi, S., Tu, Z. y Wang, L. (2025). Salute the classic: Revisiting challenges of machine translation in the age of large language models. *Transactions of the Association for Computational Linguistics*, 13, 73-95. https://doi.org/10.1162/tacl_a_00730
- Petridou, E. y Lao, L. (2024). Identifying challenges and best practices for implementing AI additional qualifications in vocational and continuing education: A mixed methods analysis. *International Journal of Lifelong Education*, 43(4), 385-400. <https://doi.org/10.1080/02601370.2024.2351076>
- Qiu, Y. y Jin, Y. (2024). ChatGPT and finetuned BERT: A comparative study for developing intelligent design support systems. *Intelligent Systems with Applications*, 21, 200308. <https://doi.org/10.1016/j.iswa.2023.200308>
- Rezaei, A. R. (2015). Frequent collaborative quiz taking and conceptual learning. *Active Learning in Higher Education*, 16(3), 189-204. <https://doi.org/10.1177/1469787415589627>
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18(2), 119-144. <https://doi.org/10.1007/BF00117714>
- Saha, S., Dhamecha, T. I., Marvaniya, S., Sindhgatta, R. y Sengupta, B. (2018). Sentence-level or token-level features for automatic short answer grading? Use both. En *Lecture Notes in Computer Science* (Vol. 10947, pp. 475-486). https://doi.org/10.1007/978-3-319-93843-1_37
- Sangapu, I. (2018). Artificial intelligence in education: From a teacher and a student perspective. *SSRN*. <https://doi.org/10.2139/ssrn.3372914>
- Schneider, J., Richner, R. y Riser, M. (2023). Towards trustworthy autograding of short, multi-lingual, multi-type answers. *International Journal of Artificial Intelligence in Education*, 33(1), 1-29. <https://doi.org/10.1007/s40593-022-00289-z>
- Senthilnathan, V., Sakthi Vaibhav, M. y Alexander, R. (2025). *Semantic refined prompting based automated essay scoring system*. En *Proceedings of the 2025 International Conference on Electronics and Renewable Systems (ICEARS)* (pp. 1344-1348). IEEE. <https://doi.org/10.1109/ICEARS64219.2025.10940227>
- Siddiqi, R. y Harrison, C. (2008). A systematic approach to the automated marking of short-answer questions. En *Proceedings of IEEE INMIC 2008: 12th International Multitopic Conference* (pp. 281-286). <https://doi.org/10.1109/INMIC.2008.4777758>
- Soulimani, Y. A., El Achaak, L. y Bouhorma, M. (2024). Deep learning-based Arabic short answer grading in serious games. *International Journal of Electrical and Computer Engineering*, 14(1), 841-853. <https://doi.org/10.11591/ijece.v14i1.pp841-853>
- Souza, F., Nogueira, R. y Lotufo, R. A. (2019). *Portuguese named entity recognition using BERT-CRF* [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.1909.10649>
- Sung, C., Ma, T., Dhamecha, T. I., Reddy, V., Saha, S. y Arora, R. (2019). Pre-training BERT on domain resources for short answer grading. En *Proceedings of EMNLP-IJCNLP 2019* (pp. 6076-6086). <https://doi.org/10.18653/v1/D19-1628>
- van Wynsberghe, A. (2021). Sustainable AI: AI for sustainability and the sustainability

- of AI. *AI and Ethics*, 1(3), 213-218. <https://doi.org/10.1007/s43681-021-00043-6>
- Villarroel, V., Bloxham, S., Bruna, D., Bruna, C. y Herrera-Seda, C. (2018). Authentic assessment: Creating a blueprint for course design. *Assessment & Evaluation in Higher Education*, 43(5), 840-854. <https://doi.org/10.1080/02602938.2017.1412396>
- Wang, A., Singh, A., Michael, J., Hill, F., Levy, O. y Bowman, S. R. (2018). GLUE: A multi-task benchmark and analysis platform for natural language understanding. En *Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP* (pp. 353-355). <https://doi.org/10.18653/v1/W18-5446>
- Wang, H. y Lehman, J. D. (2021). Using achievement goal-based personalized motivational feedback to enhance online learning. *Educational Technology Research and Development*, 69(2), 807-836. <https://doi.org/10.1007/s11423-021-09940-3>
- Wang, Y., Wang, C., Li, R. y Lin, H. (2022). On the use of BERT for automated essay scoring: Joint learning of multi-scale essay representation. En *Proceedings of NAACL 2022* (pp. 3432-3444). <https://doi.org/10.18653/v1/2022.naacl-main.249>
- Wang, Z., Lan, A. S., Waters, A. E., Grimaldi, P. y Baraniuk, R. G. (2019). A meta-learning augmented bidirectional transformer model for automatic short answer grading. En *Proceedings of the 12th International Conference on Educational Data Mining (EDM 2019)*.
- Winstone, N. E., Nash, R. A., Parker, M. y Rowntree, J. (2017). Supporting learners' agentic engagement with feedback: A systematic review and a taxonomy of recipience processes. *Educational Psychologist*, 52(1), 17-37. <https://doi.org/10.1080/00461520.2016.1207538>
- Xavier, C., Rodrigues, L., Costa, N., Neto, R., Alves, G., Falcão, T. P., Gašević, D. y Mello, R. F. (2025). Empowering instructors with AI: Evaluating the impact of an AI-driven feedback tool in learning analytics. *IEEE Transactions on Learning Technologies*, 18, 498-512. <https://doi.org/10.1109/TLT.2025.3562379>
- Xie, X. y Li, X. (2018). Research on personalized exercises and teaching feedback based on big data. En *Proceedings of the ACM International Conference* (pp. 217-221). <https://doi.org/10.1145/3232116.3232143>
- Xu, Z. y Zhu, P. (2023). Using BERT-based textual analysis to design a smarter classroom mode for computer teaching in higher education institutions. *International Journal of Emerging Technologies in Learning*, 18(19), 120-133. <https://doi.org/10.3991/ijet.v18i19.42483>
- Zhang, H., Cai, J., Xu, J. y Wang, J. (2019). Pretraining-based natural language generation for text summarization. En *Proceedings of CoNLL 2019* (pp. 789-798). <https://doi.org/10.18653/v1/K19-1074>
- Zhang, H., Yu, P. S. y Zhang, J. (2025). A systematic survey of text summarization: From statistical methods to large language models. *ACM Computing Surveys*. <https://doi.org/10.1145/3731445>
- Zhang, Z., Zhang, Z., Chen, H. y Zhang, Z. (2019). A joint learning framework with BERT for spoken language understanding. *IEEE Access*, 7, 168849-168858. <https://doi.org/10.1109/ACCESS.2019.2954766>
- Zhao, H., Chen, H., Yang, F., Liu, N., Deng, H., Cai, H., Wang, S., Yin, D. y Du, M. (2024). Explainability for large language models: A survey. *ACM Transactions on Intelligent Systems and Technology*, 15(2), Article 26. <https://doi.org/10.1145/3639372>
- 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.

International Journal of Educational Technology in Higher Education, 20(1), 12. <https://doi.org/10.1186/s41239-023-00415-4>

Fecha de recepción del artículo: 1 de junio de 2025

Fecha de aceptación del artículo: 6 de agosto de 2025

Fecha de aprobación para maquetación: 16 de septiembre de 2025

Fecha de publicación en OnlineFirst: 14 de octubre de 2025

Fecha de publicación: 1 de enero de 2026

Medición de la habilidad escrita en español como lengua extranjera con inteligencia artificial generativa

Measuring writing skills in Spanish as a foreign language with generative artificial intelligence



- ID María-Victoria Cantero Romero - *Universidad de Jaén, UJA (España)*
- ID María-Teresa Martín-Valdivia - *Universidad de Jaén, UJA (España)*
- ID Ana María Ortiz-Colón - *Universidad de Jaén, UJA (España)*
- ID Salud María Jiménez-Zafra - *Universidad de Jaén, UJA (España)*

RESUMEN

La irrupción de la Inteligencia Artificial Generativa (IAG), y en particular de los Modelos de Lenguaje de gran tamaño (LLMs) como ChatGPT, está transformando el ámbito educativo, especialmente en la enseñanza de lenguas extranjeras. Este artículo analiza el potencial de estas tecnologías para automatizar la evaluación de la competencia escrita en español como lengua extranjera (ELE), una tarea especialmente laboriosa al inicio de los cursos universitarios dirigidos a estudiantes Erasmus. La metodología se basa en tres experimentos con el Corpus de Aprendices de Español del Instituto Cervantes. En el primero, se utilizó la técnica de *zero-shot learning*, proporcionando al modelo un *prompt* con los descriptores del Plan Curricular del Instituto Cervantes. En el segundo y tercer experimentos, se ajustó el modelo mediante *fine-tuning* con el 90 % y el 80 % del corpus, respectivamente, reservando el resto para validación y prueba. Los resultados muestran que los modelos ajustados son capaces de identificar el nivel de competencia escrita con una precisión significativamente superior al enfoque sin entrenamiento previo. Estos hallazgos evidencian que los LLMs pueden emplearse para agilizar procesos de evaluación inicial en cursos de ELE, reduciendo la carga docente y mejorando la eficiencia. Se concluye que la IAG representa una herramienta complementaria valiosa en contextos educativos multiculturales y multilingües, siempre que su uso esté guiado por criterios pedagógicos sólidos.

Palabras clave: enseñanza de lenguas; expresión escrita; inteligencia artificial; innovación pedagógica relevante.

ABSTRACT

The emergence of Generative Artificial Intelligence (GAI)—particularly Large Language Models (LLMs) such as ChatGPT—is transforming the educational landscape, especially in the field of foreign language instruction. This article explores the potential of these technologies to automate the assessment of writing proficiency in Spanish as a Foreign Language (SFL), a task that is especially time-consuming at the beginning of university-level courses for Erasmus students. The study is based on three experiments conducted using the Spanish Learner Corpus compiled by the Instituto Cervantes. The first experiment applied a zero-shot learning approach by prompting the model with level descriptors from the Instituto Cervantes's Curriculum Plan. In the second and third experiments, the model was adjusted through fine-tuning using 90% and 80% of the corpus, respectively, with the remaining data reserved for testing and validation. The results indicate that the fine-tuned models significantly outperform the zero-shot configuration in identifying the correct proficiency levels of learner texts. These findings demonstrate that LLMs can be effectively employed to streamline the initial placement process in SFL courses, thus reducing the workload of instructors and improving efficiency. The study concludes that GAI can serve as a valuable complementary tool in multilingual and multicultural educational settings, provided its use is guided by sound pedagogical principles.

Keywords: language instruction; writing; artificial intelligence; teaching method innovations.

INTRODUCCIÓN

La evolución de la Inteligencia Artificial ha supuesto un punto de inflexión en todos los ámbitos y, en especial, en el educativo (Aparicio Gómez, 2023; Hernández-León y Rodríguez-Conde, 2024; Zambrano Campozano, 2025). El avance de esta disciplina abre un campo de nuevas investigaciones y trabajos con los que poder implantar la Inteligencia Artificial, en concreto, la generativa, en las aulas. Como ya nos señalan Bolaño-García y Duarte-Acosta (2024), la Inteligencia Artificial generativa ha ganado atención en educación, puesto que mejora la personalización del aprendizaje y la retroalimentación en tiempo real. Zapata Ros (2024) apoya esta idea de personalización además de la disponibilidad de información. Asimismo, Fajardo et al. (2023) remarcan la utilización de estas herramientas para la personalización del aprendizaje en la educación universitaria, adaptándolo a las preferencias y necesidades de cada estudiante mediante tutorías guiadas y virtuales. García-Peñalvo et al. (2024) señalan la importancia de preparar tanto a docentes como a discentes en el uso de la Inteligencia Artificial generativa, puesto que estará presente en todos los aspectos de la vida. Además, como nos indica Moreno (2019) hay que destacar el potencial de la Inteligencia Artificial generativa para transformar la educación al crear entornos de aprendizaje adaptativos, ajustados al rendimiento de los estudiantes, como pueden ser para estudiantes con necesidades educativas especiales. En su estudio sobre herramientas de Inteligencia Artificial generativa, Area-Moreira et al. (2024), además de las funciones ya mencionadas, indican que estas herramientas pueden ser utilizadas para la automatización de tareas como apoyo al profesorado en la labor docente e incluso como sistemas antiplagios. En consonancia con lo anterior, Chan y Tsi (2023) señalan el uso de la Inteligencia Artificial generativa como herramienta auxiliar para los profesores y no como una sustitución de estos. Moreno (2019) también señala tres enfoques con los que trabajar en educación: con la Inteligencia Artificial generativa, con robótica educativa y con las plataformas de autoaprendizaje. Es en el primero de ellos en el que nos vamos a centrar en este estudio. Tanto Barroso-Osuna y Cabero-Almenara (2025) como Owan et al. (2023) identifican una serie de beneficios del uso de la Inteligencia Artificial, en concreto, la generativa, en educación, entre los que destacan la optimización de tiempo docente y la evaluación automatizada y precisa. En este sentido, Crespo Mendoza et al. (2024) señalan que puede mejorar la precisión como la fiabilidad de las evaluaciones.

Como se ha mencionado anteriormente, el enfoque principal de este trabajo es la Inteligencia Artificial generativa, concretamente los Modelos de Lenguaje (LLMs, por sus siglas en inglés), que han supuesto una nueva herramienta con la que poder llevar a cabo estudios enfocados a mejorar las tareas docentes. Los LLMs son herramientas de generación de lenguaje natural entrenadas con una gran cantidad de textos (Wang, 2024). García-Peñalvo et al. (2024) señalan distintas funciones que pueden llevar a cabo los LLMs, como el apoyo a la investigación, la creación de contenido educativo, la generación de chatbots para interactuar con el alumnado ofreciendo

una retroalimentación autodirigida, el uso como complemento de los motores de búsqueda, el parafraseado de texto, la enseñanza de idiomas o la generación de exámenes y cuestionarios. Relacionado con estas dos últimas funciones es donde se enmarca el presente estudio, la nivelación en la enseñanza de idiomas y, en concreto, en el español como lengua extranjera.

La Inteligencia Artificial generativa presenta la oportunidad de poder realizar evaluaciones adaptativas a cada estudiante con una retroalimentación inmediata y específica, sugiriendo posibles soluciones (Barroso-Osuna y Cabero-Almenara, 2025). Asimismo, los Modelos de Lenguaje pueden ser usados para una automatización (García-Peñalvo, 2024) de la corrección tanto en pruebas de opción múltiple como en respuestas de texto abierto (Moreno, 2019). Como ventaja de usar los LLMs para evaluación, se pueden destacar la eficiencia (Area-Moreira et al., 2024) y la detección del plagio (García-Peñalvo et al., 2024). Asimismo, García-Peñalvo et al. (2024) señalan que los LLMs mejoran la productividad del profesorado.

Con respecto al uso de los Modelos de Lenguaje y la enseñanza de idiomas, diversos estudios (Baskara y Mukarto, 2023; Salguero Romero, 2023; Wang, 2024) han demostrado su utilidad. Baskara y Mukarto (2023) señalan cómo ChatGPT es capaz de generar textos realistas que acerquen a los estudiantes a la realidad de la lengua. Por otro lado, Hong (2023) señala como ventaja poder utilizar estas herramientas para agilizar la corrección de exámenes, liberando a los docentes de carga, dándoles oportunidad de centrarse en la preparación de las clases.

Uno de los retos de la enseñanza de idiomas y, en este caso, de la enseñanza de español como lengua extranjera es la adecuada nivelación de los estudiantes al iniciar los cursos de español. Que un estudiante se encuentre en un nivel adecuado de aprendizaje de español es esencial para una correcta progresión del aprendizaje, puesto que, si se encuentra en un nivel superior, el alumno puede frustrarse, y a la inversa, si se encuentra en un nivel inferior, el alumno puede no encontrar motivación. Es por ello que, la nivelación, es clave en la iniciación de los cursos de idiomas.

Hasta ahora, esta nivelación se ha realizado a través de pruebas de opción múltiple o entrevista (Biedma Torrecillas et al., 2012). Estas pruebas, principalmente, se centran en obtener el nivel que el alumno tiene con respecto a la expresión y comprensión tanto oral como escrita. Las pruebas de nivel de español existentes, como la prueba de nivel del Instituto Cervantes, consisten en contestar una serie de ítems escritos de respuesta cerrada (verdadero-falso; emparejar u ordenamiento) de dificultad ascendente (Centro Virtual Cervantes, s. f.). Esta misma tipología se observa actualmente en contextos internacionales, como en las pruebas de nivel de español de Columbia University y de University of Wisconsin–Madison, que también se basan en ejercicios de opción múltiple sin incluir producción escrita libre (Columbia University, s. f.; University of Wisconsin–Madison, s. f.). No obstante, las pruebas de nivel de español no se han centrado en las pruebas de expresión escrita de manera directa con escritura de textos completos, sino de manera indirecta, a

través de la respuesta de ítems, debido a las limitaciones que conlleva, como pueden ser la falta de inmediatez o la compleja nivelación cuando se trata de grandes grupos.

Para solucionar este problema, el presente estudio se centrará en la nivelación de la prueba de expresión escrita. Como se ha mencionado anteriormente, los Modelos de Lenguaje son capaces de automatizar la calificación de trabajos escritos (García-Peñalvo et al., 2024), además de manera rápida, ahorrando tiempo a los docentes (Area-Moreira et al., 2024), siendo este el principal motivo por el que no se ha incluido la prueba de expresión escrita en las pruebas de nivel actuales.

En este contexto, resulta necesario establecer una base conceptual que permita comprender los avances y limitaciones de la evaluación automatizada de la escritura y, de forma particular, los fundamentos lingüísticos que sustentan la clasificación de niveles A1–C1. En el siguiente apartado se desarrolla este marco teórico, que servirá de apoyo para la propuesta metodológica del presente estudio.

El presente artículo contiene la siguiente estructura. Tras la introducción y el marco teórico, se presenta la fundamentación tecnológica y elección del modelo, los objetivos del estudio y la metodología. A continuación, se exponen los resultados, seguidos de un apartado sobre la pertinencia pedagógica del modelo. Finalmente, el artículo concluye con la discusión y conclusiones, donde se ha añadido un subapartado sobre aspectos éticos y licencias de uso.

MARCO TEÓRICO

El presente apartado expone, en primer lugar, una revisión del estado del arte en torno a la evaluación automatizada de la escritura y los fundamentos lingüísticos que sustentan la clasificación por niveles del Marco Común Europeo de Referencia (MCER) y el Plan Curricular del Instituto Cervantes (PCIC). Esta base conceptual servirá para contextualizar la propuesta del estudio y, finalmente, para presentar los objetivos que guían la investigación.

Evaluación automatizada de la escritura: evolución y enfoques actuales

La clasificación automática de textos de aprendientes mediante modelos de lenguaje se inscribe en una tradición más amplia de evaluación automática de la escritura (AWE), cuyo desarrollo ha dado lugar a múltiples herramientas y sistemas que conviene tener en cuenta para contextualizar el presente trabajo. La AWE ha evolucionado considerablemente en las últimas décadas, convirtiéndose en una herramienta cada vez más común en contextos educativos. Sistemas pioneros como *e-rater* (Burstein et al., 2003), desarrollado por ETS, han sido empleados ampliamente en pruebas estandarizadas, valiéndose de métricas lingüísticas, gramaticales y discursivas para estimar la calidad textual. Por su parte, *Coh-Metrix* (McNamara et al., 2014) permite realizar análisis detallados de la cohesión, la complejidad sintáctica y la legibilidad, proporcionando una aproximación multifactorial al discurso escrito.

Estudios recientes, como el de Zhang (2021), ofrecen revisiones sistemáticas de estos sistemas, destacando su transición desde enfoques basados en reglas hacia modelos impulsados por el aprendizaje automático y el procesamiento del lenguaje natural. En esta misma línea, Wang et al. (2022) analizan enfoques actuales orientados a la evaluación de textos argumentativos, centrados en componentes discursivos como la estructura del razonamiento, la evidencia o la organización. Otras propuestas, como *Writing Mentor* (Burstein et al., 2018), integran la evaluación automática con retroalimentación formativa, favoreciendo procesos de autorregulación en la escritura académica. Asimismo, herramientas como *Write & Improve*, desarrollada por la Universidad de Cambridge, ejemplifican cómo es posible proporcionar feedback automático inmediato sobre textos producidos por estudiantes de lenguas extranjeras, facilitando el aprendizaje autónomo y guiado (Cambridge English, s. f.).

Fundamentos lingüísticos de la clasificación por niveles A1–C1

La clasificación de los textos producidos por aprendices de español como lengua extranjera en niveles A1–C1 se fundamenta en los descriptores establecidos por el Marco Común Europeo de Referencia para las Lenguas (Consejo de Europa, 2002) y el Plan Curricular del Instituto Cervantes (Instituto Cervantes, 2006). Estos documentos definen de manera detallada las competencias lingüísticas, pragmáticas y sociolingüísticas asociadas a cada nivel, proporcionando una base sólida para la evaluación.

En el nivel A1, el repertorio léxico es muy limitado y se restringe a transacciones básicas y expresiones cotidianas. Los textos son muy breves y sencillos, con oraciones simples y un promedio reducido de palabras por enunciado. Predomina el uso de las formas regulares del presente de indicativo y se emplea un repertorio gramatical elemental (Consejo de Europa, 2002; Instituto Cervantes, 2006).

En el nivel A2, el aprendiz puede producir textos breves que transmiten información sencilla sobre temas familiares. Se observa un mayor repertorio léxico y estructuras ligeramente más complejas, incorporando tiempos verbales del pasado de indicativo (pretérito perfecto, imperfecto e indefinido) y algunas formas irregulares de presente. También aparece el uso del imperativo afirmativo (Consejo de Europa, 2002; Instituto Cervantes, 2006).

En el nivel B1, el repertorio léxico es más amplio y permite elaborar textos que cumplen una tarea comunicativa concreta, manteniendo una estructura coherente. Gramaticalmente, se manejan con cierta soltura tiempos como el futuro simple, el condicional simple y el pretérito pluscuamperfecto, además de introducir el presente de subjuntivo y el imperativo negativo. El discurso muestra una mayor cohesión y variedad de recursos conectores (Consejo de Europa, 2002; Instituto Cervantes, 2006).

En el nivel B2, el usuario dispone de un repertorio lingüístico amplio y preciso, capaz de sostener argumentaciones complejas y descripciones detalladas. Se

manejan con fluidez oraciones compuestas y subordinadas, así como un uso seguro de los tiempos verbales de indicativo (presente, pasados, futuros y condicionales) y de subjuntivo (presente, imperfecto, perfecto y pluscuamperfecto). La cohesión textual es consistente y se emplean matices léxicos adecuados a diferentes registros (Consejo de Europa, 2002; Instituto Cervantes, 2006).

En el nivel C1, el repertorio lingüístico y no lingüístico es lo suficientemente amplio y flexible para abordar cualquier tipo de transacción o interacción comunicativa, incluso en contextos académicos o profesionales exigentes. El aprendiz es capaz de producir textos extensos y complejos con una estructura clara y bien organizada, utilizando con precisión todos los tiempos verbales y un amplio rango de recursos sintácticos y léxicos (Consejo de Europa, 2002; Instituto Cervantes, 2006).

FUNDAMENTACIÓN TECNOLÓGICA Y ELECCIÓN DEL MODELO

En el presente estudio se ha utilizado el modelo de lenguaje ChatGPT, desarrollado por OpenAI, en su versión 3.5, lanzada en noviembre de 2022 (OpenAI, 2022). A pesar de no ser un modelo de acceso en abierto, esta versión permite realizar un *fine-tuning* a través de la API de OpenAI. La selección de la versión 3.5 frente a la versión posterior se justifica por la posibilidad de realizar este proceso, lo cual no es posible en la última versión de ChatGPT.

Para esta investigación se optó por emplear GPT-3.5 en lugar de modelos más recientes o de código abierto debido a una combinación de factores técnicos, económicos y metodológicos. Al tratarse de una tarea novedosa —la nivelación automática de la competencia escrita en español como lengua extranjera—, se consideró pertinente evaluar el rendimiento de un modelo generalista ampliamente probado, como GPT-3.5, tanto en su configuración base como mediante *fine-tuning*. Este modelo, accesible a través de la API de OpenAI, no requiere infraestructuras computacionales avanzadas y ofrece una arquitectura sólida con buena cobertura del español (Li, 2023; Pourpanah et al., 2023). Además, presenta una relación adecuada entre rendimiento, coste y velocidad de respuesta, factores clave para validar la viabilidad del enfoque en esta fase exploratoria (Roumeliotis et al., 2024). Por contraste, en el momento de los experimentos, modelos como GPT-4 suponían un coste económico considerablemente mayor y tiempos de inferencia más prolongados, lo que reforzó la decisión de utilizar GPT-3.5 como referencia inicial.

No obstante, conviene señalar que GPT-3.5 presenta limitaciones frente a modelos más recientes, como GPT-4, que ofrecen una comprensión contextual más precisa, un entrenamiento con conjuntos de datos más amplios y heterogéneos y una mayor capacidad de razonamiento. Estas características los convierten en candidatos especialmente adecuados para investigaciones futuras, tanto en la evaluación automatizada de la competencia escrita alineada con los niveles del MCER como en la generación de corpus sintéticos de alta calidad que permitan entrenar sistemas especializados.

OBJETIVO DEL ESTUDIO

Para llevar a cabo el presente estudio se han adaptado los descriptores establecidos por el Marco Común Europeo de Referencia para las Lenguas (Consejo de Europa, 2002) y el Plan Curricular del Instituto Cervantes con el fin de establecer criterios claros y operativos que permitan a un modelo de lenguaje, en este caso ChatGPT, clasificar de forma automática las producciones escritas de aprendices de español según su nivel de competencia.

A pesar de los avances mencionados en la evaluación automática de la escritura, pocos estudios se han centrado en la nivelación de textos escritos producidos por aprendices de lenguas extranjeras y, en concreto, en el caso específico de la enseñanza de español como lengua extranjera. Las investigaciones existentes suelen abordar la retroalimentación o corrección automática, pero no la clasificación de producciones escritas según niveles del MCER o del Plan Curricular del Instituto Cervantes, especialmente en tareas de evaluación inicial. Esta laguna es especialmente relevante en contextos como el universitario, donde la heterogeneidad del alumnado internacional, como en los programas Erasmus, exige herramientas eficientes para la nivelación. Este estudio propone una solución innovadora basada en Modelos de Lenguaje generativos (ChatGPT), que permite clasificar automáticamente textos escritos de aprendices de enseñanza de español como lengua extranjera (ELE) según su nivel de competencia, reduciendo la carga docente y mejorando la gestión del proceso de inicio de curso. Así, el trabajo no solo complementa investigaciones previas centradas en la mejora textual, sino que amplía el alcance de la evaluación automatizada hacia tareas de diagnóstico inicial en contextos de enseñanza de segundas lenguas.

Es por ello que, como objetivo general del estudio, se ha planteado evaluar la eficacia del modelo de lenguaje ChatGPT como herramienta innovadora para nivelar la expresión escrita de los estudiantes de español. Asimismo, se espera conseguir una serie de objetivos específicos: i) conocer si ChatGPT con su entrenamiento previo es capaz de nivelar de manera adecuada; ii) comprobar si ChatGPT es capaz de nivelar si es ajustado con un corpus nivelado con el Plan Curricular del Instituto Cervantes y el Marco Común Europeo de Referencia para las lenguas; y iii) determinar el impacto de ChatGPT en la eficiencia del proceso de nivelación del español como lengua extranjera.

METODOLOGÍA

Para llevar a cabo la presente investigación se ha utilizado el Corpus de Aprendices de Español (en adelante CAES) (Palacios Martínez et al., 2019), elaborado por el Instituto Cervantes en colaboración con la Universidad de Santiago de Compostela. Asimismo, se ha empleado el modelo de lenguaje ChatGPT en su versión 3.5, desarrollado por OpenAI (OpenAI, 2022).

En este estudio se han realizado tres tipos de pruebas experimentales: una de ellas utilizando la técnica de *zero-shot learning*, y las otras dos mediante *fine-tuning* del modelo.

El procedimiento de *fine-tuning* se realizó con una única época de entrenamiento, manteniendo el resto de parámetros por defecto según la API de OpenAI. Para garantizar la reproducibilidad, se empleó una semilla aleatoria fija (valor 42), es decir, un valor de referencia que permite que los experimentos puedan reproducirse siempre con los mismos resultados. No se aplicaron técnicas de balanceo de clases, puesto que se optó por conservar la distribución real del corpus, reflejando así condiciones auténticas de nivelación en el aula. De este modo, los resultados del sistema se ajustan mejor a los desafíos propios de escenarios educativos reales, sin introducir modificaciones artificiales en la representación de los niveles. No obstante, se reconoce que en investigaciones futuras podrían aplicarse estrategias de compensación para comparar su efecto en la equidad y la robustez del modelo. Tampoco se empleó validación cruzada, en coherencia con el carácter inicial y experimental del estudio.

En relación con los parámetros técnicos del entrenamiento, se mantuvieron los valores predeterminados de la API de OpenAI en aspectos como el tamaño de lote (*batch size*, es decir, el número de ejemplos procesados a la vez), la tasa de aprendizaje (*learning rate*, que indica la velocidad con la que el modelo ajusta sus parámetros durante el entrenamiento) y las funciones de pérdida (métricas que miden la diferencia entre la predicción del modelo y el resultado esperado). No se aplicaron técnicas adicionales de regularización ni estrategias de *early stopping* (interrupción anticipada del entrenamiento para evitar sobreajuste), puesto que el objetivo principal era validar la viabilidad del enfoque más que optimizar el modelo al máximo rendimiento. Esta decisión se enmarca en el carácter exploratorio de la investigación, orientado a comprobar la aplicabilidad del modelo a la tarea de nivelación automática de textos.

Corpus CAES

El corpus CAES fue recogido por la Universidad de Santiago de Compostela y financiado por el Instituto Cervantes. Se recopilieron datos informatizados desde octubre de 2011 hasta diciembre de 2020 de centros, en su mayoría universitarios, de distintos países como España, Estados Unidos, Brasil, Egipto, Irlanda o Portugal. Los estudiantes que formaron parte del proyecto procedían de once lenguas de origen distintas (inglés, chino mandarín, portugués, árabe, ruso, alemán, francés, griego, italiano, japonés y polaco).

En este estudio se ha trabajado con la versión 2.1 del corpus CAES de marzo de 2022, con la que actualizaron la primera recogida de datos, que contaba con un número menor de datos, con un total de 1423 participantes, frente a los 2544 participantes de la actualización de 2022.

En el corpus se encuentran ejemplos de distintos niveles de español según el Marco Común Europeo de Referencia para las Lenguas, desde A1 hasta C1. En los niveles A1, A2 y B1 se recopilan textos pertenecientes a tres tareas de distinta tipología que los estudiantes debían escribir con una extensión de entre 30 y 200 palabras. Los niveles B2 y C1 cuentan con una muestra de dos tareas por nivel y con una extensión de entre 275 y 500 palabras.

Los temas que se han identificado en cada nivel son los siguientes: En A1, la primera tarea consiste en un correo electrónico sobre cambiar de trabajo, con un total de 728 ejemplos; la segunda tarea consiste en un correo electrónico sobre su familia, con 703 ejemplos; la última tarea consiste en una nota acerca de llegar tarde, con una muestra de 705. En el nivel A2, la primera tarea que encontramos es una biografía, con 673 ejemplos; la segunda tarea consiste en la reserva de una habitación de hotel, con una muestra de 603 textos; y la última consiste en escribir una postal sobre sus vacaciones, que cuenta con una muestra de 701 ejemplos. Se puede comprobar que las muestras recogidas de los niveles iniciales rondan los 2 000 ejemplos por nivel, dando una muestra significativa de dichos niveles. Asimismo, la temática de los textos se encuentra relacionada con las funciones y los productos textuales correspondientes de su nivel según el Plan Curricular del Instituto Cervantes (Instituto Cervantes, 2006).

En el nivel B1 se encuentran, igualmente, tres tareas distintas. La primera, escribir una carta a un amigo, con una muestra de 528 textos; la segunda tarea, escribir un correo electrónico sobre una reclamación a una compañía aérea, con una muestra de 454; y la última tarea, una narración de una historia, con una muestra de 382. Hay que señalar que estas tareas, al igual que las de nivel A1 y A2, se encuentran relacionadas con las funciones que se describen en el Plan Curricular.

En los niveles B2 y C1, las tareas se reducen a dos. En el nivel B2, en la primera tarea deben redactar una solicitud de admisión, con un total de 375 muestras; mientras que la segunda tarea del nivel B2 trata sobre un texto donde deben argumentar el tema de fumar en lugares públicos, con una muestra de 356.

Con respecto al nivel C1, la primera tarea consiste en escribir una reclamación a una compañía de gas, con una muestra de 169; y en la segunda tarea, deben escribir una reseña de una película, con una muestra de 184 textos. Se puede ver cómo la muestra se reduce significativamente en estos niveles al disminuir el número de tareas. Asimismo, es notable, cómo en el nivel C1 las muestras son menores, no llegando a 400 muestras en este nivel.

Al igual que en los niveles anteriores, se puede observar que las tareas de los niveles B2 y C1 corresponden igualmente a las funciones descritas en el Plan Curricular.

La anotación del corpus fue realizada por especialistas en enseñanza de español como lengua extranjera de la Universidad de Santiago de Compostela. Cada texto fue clasificado en uno de los niveles del MCER siguiendo los criterios establecidos en el Plan Curricular del Instituto Cervantes y las pautas definidas en el propio

proyecto (Palacios Martínez et al., 2019; Universidad de Santiago de Compostela, s. f.). Para garantizar la fiabilidad de la clasificación, los textos fueron evaluados de forma independiente por varios anotadores y, posteriormente, revisados de manera conjunta hasta alcanzar un consenso. Este procedimiento asegura que la asignación de niveles sea coherente y consistente, lo que convierte al corpus en un recurso sólido para la investigación y la evaluación automática de la producción escrita.

En la Tabla 1, se presenta un resumen de los niveles con respecto a las tareas abordadas y al número total de muestras recogidas.

Asimismo, Cantero (2024) realizó un estudio de este corpus en el que se pueden sacar los siguientes resultados. En el nivel A1, el promedio de palabra por oración se encuentra entre 10.9 y 11.7 y las palabras más frecuentes son conectores simples y un léxico limitado. En el nivel A2, el promedio de palabras por oración es mayor, tratándose entre 12.5 a 14.6, con un léxico frecuente simple. En el siguiente nivel, B1, se puede observar que aumenta a 12.0 - 16.5 el promedio de palabras por oración, asimismo, el vocabulario es más amplio y complejo que el de niveles anteriores. Con respecto al nivel B2, el promedio de palabras por oración es aún más complejo, llegando a encontrarse entre 17.7 y 21.5, y en relación con el léxico se utilizan conectores más complejos y un vocabulario más especializado. Por último, en el nivel C1, el promedio de palabras por oración se encuentra entre 20.7 y 23.3 con un vocabulario complejo y variado.

Tabla 1
Resumen de muestras

Nivel	Tareas	Muestras
A1	Correo electrónico cambio de trabajo	728
	Correo electrónico familia	703
	Nota llegar tarde	705
A2	Biografía persona que admiras	673
	Reserva habitación de hotel	603
	Postal de vacaciones	701
B1	Carta a un amigo	528
	Historia graciosa	382
	Reclamación compañía aérea	454
B2	Solicitud admisión	375
	Fumar en lugares públicos	356
C1	Reseña película	184
	Reclamación compañía de gas	169

Prompt utilizado

Una vez analizado el corpus, se ha elaborado un *prompt* específico para realizar las pruebas con el modelo. Como señala Morales-Chan (2023), un buen *prompt* puede garantizar el éxito de la tarea. Por ello, es importante definir el objetivo y proporcionar un contexto suficiente.

El siguiente *prompt* fue utilizado en las tres pruebas realizadas en este estudio — *zero-shot learning* (Prueba 1) y *fine-tuning* (Pruebas 2 y 3)— con el fin de mantener una coherencia metodológica en los criterios de evaluación. El diseño del *prompt* se basó en los descriptores lingüísticos del Plan Curricular del Instituto Cervantes y del Marco Común Europeo de Referencia para las Lenguas.

Tú eres un experto lingüista especializado en enseñanza de español como lengua extranjera. Tu tarea es indicar el nivel de español como lengua extranjera de los textos siguiendo el Plan Curricular del Instituto Cervantes.

Aquí tienes una descripción de los distintos niveles.

Niveles A1 y A2 Transacciones básicas relacionadas con su entorno.

A1: Repertorio limitado de léxico, textos muy breves y sencillos, un promedio de 10 palabras por oración. Formas regulares del presente de indicativo.

A2: Textos breves con información sencilla, un promedio de 12 palabras por oración. Tiempos verbales del pasado de indicativo: Pretérito perfecto, imperfecto e indefinido. Formas irregulares de presente de indicativo. Imperativo afirmativo.

Niveles B1 y B2 desenvolverse con textos sobre temas de su interés gustos y preferencias.

B1: Vocabulario amplio pero sencillo, realizar textos con una tarea concreta. Presente de indicativo, pretérito perfecto, imperfecto e indefinido de indicativo, futuro simple, condicional simple, pretérito pluscuamperfecto de indicativo, presente de subjuntivo. Imperativo negativo.

B2: Repertorio lingüístico amplio, oraciones subordinadas. Tiempos verbales de indicativo: presente, pretérito perfecto, imperfecto, indefinido, futuro simple y compuesto, condicional simple y compuesto, pretérito pluscuamperfecto. Tiempos verbales de subjuntivo: presente, pretérito imperfecto, pretérito perfecto y pluscuamperfecto.

C1 transacciones de todo tipo. Disponen de un repertorio de recursos lingüísticos y no lingüísticos lo suficientemente amplio y rico. Pueden enfrentarse a una amplia serie de textos extensos y complejos. Todos los tiempos verbales de indicativo y de subjuntivo el presente, pretérito perfecto, imperfecto y pluscuamperfecto.

Ahora vas a recibir un TEXTO y teniendo en cuenta lo explicado anteriormente y los errores gramaticales indica al final de tu respuesta con la etiqueta 'NIVEL:' el nivel del TEXTO (A1, A2, B1, B2 o C1).

TEXTO: "..."

Con este *prompt* se ha añadido información descriptiva breve de cada nivel, siguiendo el análisis del corpus comentado anteriormente. Asimismo, siguiendo el Plan Curricular del Instituto Cervantes (Instituto Cervantes, 2006), se ha incluido una descripción de los tiempos verbales utilizados en cada uno de los niveles y los tipos de textos siguiendo las indicaciones de los productos textuales. De esta forma, se le proporciona al modelo un contexto más amplio para que su respuesta sea más precisa y ajustada a las necesidades que se le solicitan.

Pruebas con el Modelo de Lenguaje ChatGPT

En el estudio se han realizado un total de tres pruebas distintas para evaluar la capacidad del modelo al nivelar los textos del corpus.

La primera de las pruebas fue un *zero-shot learning*. En este proceso, el modelo no recibe ejemplos específicos, sino que se basa en el conocimiento previo del mismo. Para realizar esta prueba, se ha utilizado únicamente el *prompt* que se ha mencionado en el apartado anterior.

En las pruebas dos y tres se ha realizado *fine-tuning*. Este proceso consiste en realizar una especialización de un modelo preentrenado para que realice una tarea específica, adaptándolo a un conjunto específico de datos que se le proporciona. Este conjunto de datos lo obtuvimos del corpus CAES, considerándolo un conjunto de ejemplos claros para el modelo. Para realizar el *fine-tuning*, se le proporciona al modelo una entrada-salida especificando la entrada y el tipo de salida que queremos que nos proporcione. En este caso, el *prompt* es el que se ha mencionado en el apartado anterior. Por otro lado, la salida que se le ha solicitado ha sido el nivel de español. Asimismo, para realizar el *fine-tuning*, se ha reservado una parte del conjunto de datos para verificar la respuesta.

Asimismo, las pruebas 2 y 3 se diferencian entre sí por la división que se ha realizado del corpus. Para la segunda prueba, se realizó una división del corpus en 90 %, 5 % y 5 %. El 90 % del corpus se utilizó para entrenar el modelo, un 5 % para validarlo y el 5 % restante para testarlo. En la tercera prueba, se realizó una división de 80 % - 20 %, utilizando el 80 % del corpus para el entrenamiento y el 20 % para el testeo.

RESULTADOS

Para analizar los resultados, se han empleado tres medidas de evaluación ampliamente utilizadas en tareas de clasificación:

- **Precisión** (*precision*): indica el porcentaje de ejemplos que el modelo clasificó en un determinado nivel y que realmente pertenecen a ese nivel.
- **Cobertura** (*recall*): señala la proporción de ejemplos de un nivel específico que el modelo identificó correctamente. Por ejemplo, el porcentaje de textos de nivel A1 detectados como A1 respecto al total de textos A1 existentes en el corpus.
- **F1-score**: valor único que combina precisión y cobertura mediante su media armónica, proporcionando una medida equilibrada del rendimiento. Esta métrica es especialmente útil cuando es importante que el modelo no solo acierte, sino que también detecte todos los casos posibles de cada categoría.

A continuación, se presentan los resultados de las tres experimentaciones:

Experimentación *zero-shot learning*

Como se ha mencionado anteriormente, en esta técnica, al modelo no se le proporciona ningún ejemplo, se realiza mediante el *prompt* elaborado. En este caso los resultados que se han obtenido son los siguientes:

Tabla 2
Resultados experimentación ZSL

Etiquetas	Precisión	Cobertura	F1-score
A1	0.9375	0.1402	0.2439
A2	0.3333	0.7677	0.4648
B1	0.2400	0.2609	0.2500
B2	0.4286	0.8110	0.1364
C1	0.0000	0.0000	0.0000

La Tabla 2 muestra que, en el nivel A1, la precisión es alta, es decir, los textos que clasifica como nivel A1 tienen una probabilidad alta de ser de este nivel. Sin embargo, la cobertura es bastante baja, el modelo ha detectado como A1 un porcentaje bajo de textos del corpus en general. Por tanto, el F1-score da un resultado bajo.

En el nivel A2, ocurre el fenómeno opuesto al descrito en el nivel A1. La precisión es baja, por lo que acierta en menor medida, sin embargo, la cobertura es alta. Por lo tanto, podemos decir que, en este nivel, aunque detecta en mayor medida los textos de A2, tiene una baja precisión a la hora de detectar el nivel correcto.

En relación con el nivel B1, ambos factores, precisión y cobertura, son bajos. En este nivel, el modelo tiene problemas tanto para detectar los textos de nivel B1 como para identificarlos en su nivel correcto.

Con respecto al nivel B2, se observa que la cobertura es alta, detecta de la mayoría de los textos de nivel B2, no obstante, la precisión es moderada, no acierta en la mayoría de las clasificaciones.

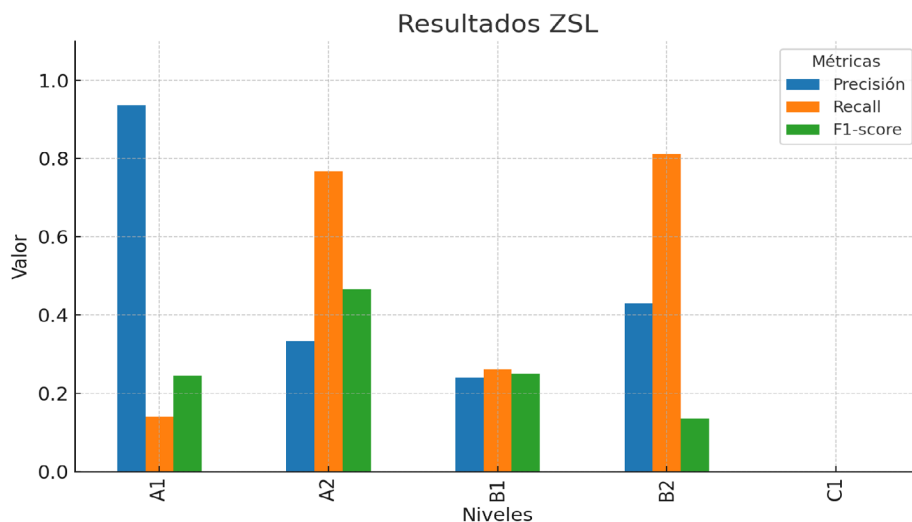
Por último, es llamativo el caso del nivel C1, ya que tanto precisión como cobertura son 0, no detecta ningún texto de dicho nivel.

Asimismo, al tratarse de una experimentación a través de un *prompt*, el modelo no solo dio en sus respuestas el nivel del texto, sino que también añadió comentarios en cada una sobre los errores encontrados. Ejemplos completos de estas respuestas, que incluyen los textos originales y las correcciones propuestas por el modelo, se presentan en el Anexo 1.

La Figura 1 complementa esta información mostrando de forma comparativa los valores de precisión, cobertura y F1-score para cada nivel evaluado en la configuración *zero-Shot Learning*.

Figura 1

Resultados por nivel en la experimentación zero-shot learning



Experimentación *fine-tuning* 90-5-5

En esta segunda experimentación, como ya se ha mencionado anteriormente, se ha realizado un *fine-tuning* del corpus y se ha dividido en 90 %, 5 % y 5 % para el entrenamiento, validación y prueba del modelo. A continuación, se pueden observar los resultados de esta experimentación:

Tabla 3
Experimentación fine-tuning 90-5-5

Etiquetas	Precisión	Cobertura	F1-score
A1	0.9905	0.9720	0.9811
A2	0.9519	1.0000	0.9754
B1	1.0000	0.9565	0.9778
B2	1.0000	1.0000	1.0000
C1	1.0000	1.0000	1.0000

La Tabla 3 evidencia que los resultados presentan niveles de precisión y cobertura más altos que los realizados con el *zero-shot learning*. En el nivel A1, se puede ver que el modelo casi no comete errores y que detecta casi todos los textos del nivel A1.

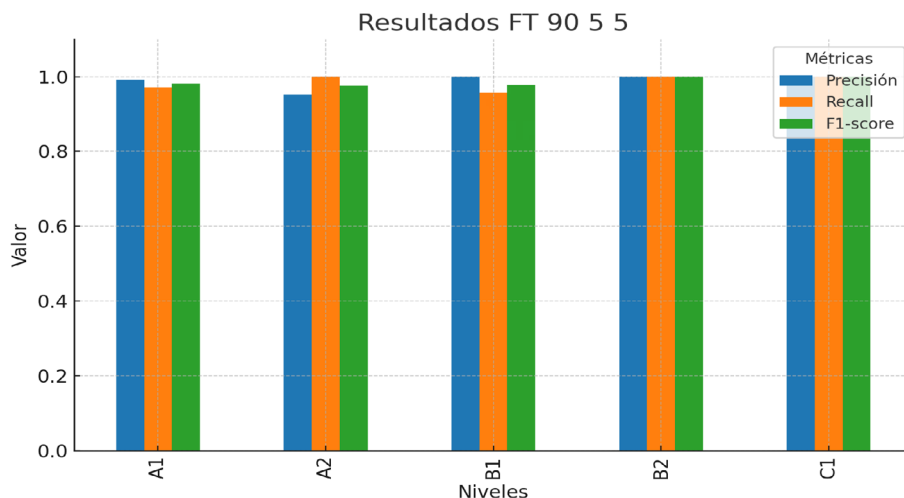
En el nivel A2, como podemos apreciar, el modelo detecta todos los textos del nivel A2, con una precisión muy alta.

Con respecto al nivel B1, el modelo predice de manera correcta todos los textos, asimismo, cuenta con una cobertura alta, con un pequeño porcentaje no detectado.

Por último, con los niveles B2 y C1, los resultados muestran que el modelo detecta todos los textos de estos niveles y acierta en todas las ocasiones. No obstante, los resultados de este experimento demuestran que el modelo puede predecir correctamente todos los niveles con valores cercanos o iguales a 1. La Figura 2 complementa esta información mostrando de forma comparativa los valores de precisión, cobertura y F1-score para cada nivel evaluado en la configuración *fine-tuning* 90-5-5.

Figura 2

Resultados por nivel en la experimentación fine-tuning 90-5-5



Experimentación fine-tuning 80-20

En la tercera experimentación, se realizó un *fine-tuning* dividiendo el corpus en dos porcentajes, un 80 % para el entrenamiento y un 20 % para la validación del modelo. A continuación, se pueden observar los resultados obtenidos.

En esta experimentación, como refleja la Tabla 4, en el nivel A1, tanto la precisión como la cobertura muestran un nivel elevado de aciertos y de detección de textos.

Los niveles A2 y B1 muestran un resultado similar al de A1, aun teniendo un resultado ligeramente inferior en la precisión en el nivel A2 y en la cobertura en el nivel B1.

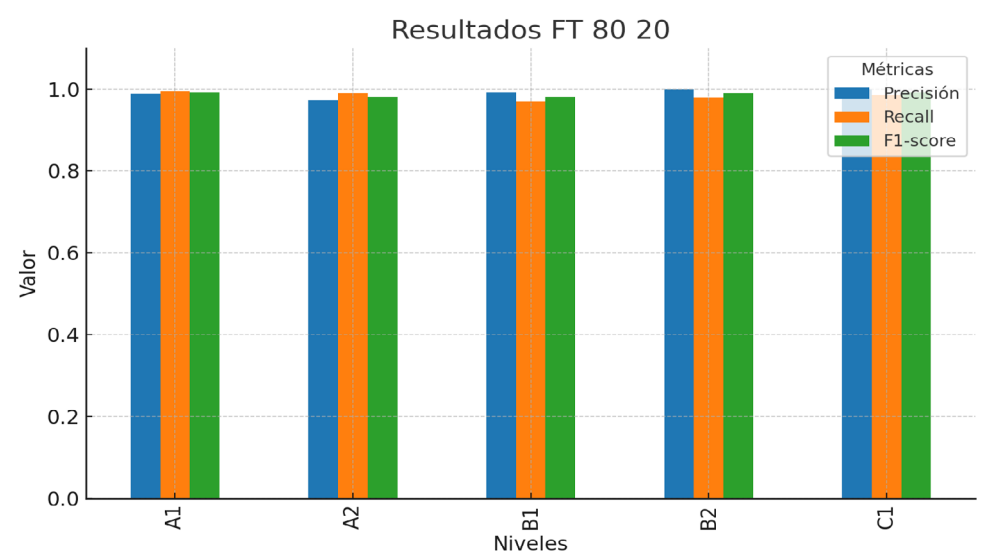
Con respecto a los niveles B2 y C1, observamos que la precisión del modelo es excelente, puesto que predice de manera correcta estos niveles, aunque la cobertura sea moderadamente menor.

Tabla 4
Experimentación fine-tuning 80-20

Etiquetas	Precisión	Cobertura	F1-score
A1	0.9884	0.9953	0.9918
A2	0.9727	0.9899	0.9812
B1	0.9925	0.9707	0.9815
B2	1.0000	0.9795	0.9896
C1	1.0000	0.9859	0.9929

La Figura 3 complementa esta información mostrando de forma comparativa los valores de precisión, cobertura y F1-score para cada nivel evaluado en la configuración *fine-tuning 80-20*.

Figura 3
Resultados por nivel en la experimentación fine-tuning 80-20



Análisis estadístico

Para evaluar la significación estadística de los resultados, se estimaron intervalos de confianza al 95 % (IC95 %) de los valores Macro-F1 mediante **bootstrap**, una técnica de remuestreo con reemplazo que permite calcular márgenes de confianza sin necesidad de asumir una distribución estadística específica. Se realizaron 1 000 réplicas para obtener estos intervalos, los cuales indican el rango dentro del cual se espera que se encuentre el valor verdadero del Macro-F1 con un 95 % de probabilidad. Adicionalmente, se llevaron a cabo pruebas de hipótesis para contrastar los resultados. El experimento *zero-shot learning* obtuvo un valor Macro-F1 de 0,2190 con un intervalo de confianza IC95 % = [0,1785, 0,2617], mientras que los experimentos con *fine-tuning* alcanzaron valores cercanos a la perfección: un valor Macro-F1 de 0,9869 con un intervalo de confianza IC95 % = [0,9748, 0,9959] en la partición 90-5-5 y un valor Macro-F1 de 0,9874 (IC95 % = [0,9812, 0,9933]) en la partición 80-20. Las comparaciones se realizaron mediante pruebas de hipótesis no paramétricas basadas en bootstrap —es decir, contrastes estadísticos que no requieren asumir una distribución concreta de los datos y que se apoyan en múltiples remuestreos aleatorios—. Estos análisis mostraron que ambos experimentos con *fine-tuning* superan de forma estadísticamente significativa al experimento *zero-shot* ($\Delta^1 \approx 0,77$; $p < 0,001$ en ambos casos). Por el contrario, no se observaron diferencias significativas entre las dos experimentaciones de *fine-tuning* ($\Delta = -0,0005$; IC95 % = [-0,0133, 0,0113]; $p = 0,964$). Estos resultados confirman que el *fine-tuning* mejora sustancialmente la capacidad del modelo para evaluar la competencia escrita en ELE.

PERTINENCIA PEDAGÓGICA DEL MODELO Y APLICACIÓN TANTO EN CONTEXTOS PRESENCIALES COMO A DISTANCIA

Los resultados obtenidos muestran que un modelo de lenguaje ajustado con un corpus especializado puede alcanzar un rendimiento muy alto en la clasificación automática de textos escritos según los niveles del MCER. Esta capacidad tiene un claro potencial para facilitar la labor docente, especialmente en la etapa inicial de colocación del estudiantado, constituyendo una innovación pedagógica en el uso de inteligencia artificial para la enseñanza de lenguas.

En un entorno de aprendizaje real, el sistema podría integrarse en un *Learning Management System* (LMS) como módulo de evaluación diagnóstica. El flujo de uso sería sencillo: el estudiante envía un texto, el modelo lo clasifica automáticamente por nivel y presenta el resultado al docente mediante una rúbrica alineada con los descriptores del Plan Curricular del Instituto Cervantes (PCIC). De esta forma, el profesorado podría asignar al estudiante el curso correspondiente a su nivel real, optimizando el ajuste de los grupos y evitando desfases que puedan afectar al progreso de aprendizaje.

En implementaciones futuras, y a partir de análisis más exhaustivos, el sistema podría generar informes más detallados basados en el PCIC, identificando áreas lingüísticas específicas a reforzar (por ejemplo, uso de tiempos verbales, cohesión textual o repertorio léxico). Esta información permitiría que el curso no solo se ajuste al nivel del alumno, sino también a sus principales carencias, orientando la programación didáctica hacia la mejora de dichos aspectos.

En contextos de educación a distancia y aprendizaje en línea, la integración en un LMS tendría la misma función diagnóstica, clasificando automáticamente al alumnado desde el primer acceso. Esto permitiría mantener la eficiencia en la organización de grupos incluso en entornos sin contacto presencial, algo especialmente relevante en cursos masivos o programas virtuales de matrícula continua.

En comparación con herramientas como *Write & Improve* (Cambridge English, s. f.), centradas principalmente en retroalimentación correctiva y evaluación formativa, la propuesta aquí presentada aporta un enfoque complementario: la clasificación automática por niveles MCER. Esta función, combinada con la posibilidad de integración directa en plataformas educativas, refuerza su valor como herramienta de apoyo docente para optimizar la colocación inicial y agilizar la labor de planificación de los cursos.

En todos los casos, el sistema se concibe como apoyo docente y no como sustituto, con el objetivo de agilizar las tareas de evaluación inicial y liberar tiempo para actividades de mayor valor añadido, como la retroalimentación individualizada o el seguimiento del progreso.

Entre las posibles mejoras futuras, se plantea la ampliación del corpus a más géneros y temáticas, la inclusión de descriptores léxicos y pragmáticos adicionales, la experimentación con otros modelos de lenguaje y la adaptación del sistema a otros idiomas y niveles educativos. También resultaría pertinente evaluar su impacto real en el rendimiento y la motivación del alumnado, así como su integración fluida en distintos LMS y contextos educativos.

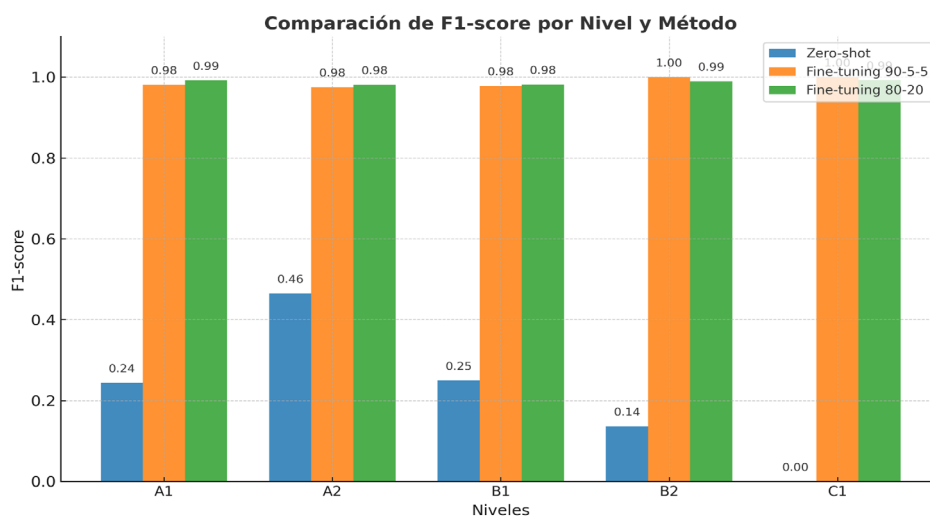
DISCUSIÓN Y CONCLUSIONES

Una vez realizadas las pruebas y obtenidos los resultados, se observa un contraste claro entre el desempeño del modelo en la configuración *zero-shot learning* y en las experimentaciones con *fine-tuning*. En la primera, el modelo no logra una precisión adecuada en ninguno de los niveles, destacando únicamente un valor relativamente alto en A1, aunque con una cobertura muy baja. En los niveles A2 y B2, la cobertura es medio-alta (0,76 y 0,81), lo que indica que el modelo detecta la mayoría de los textos de esos niveles, pero con baja precisión. El nivel C1 es especialmente problemático, con valores nulos tanto en precisión como en cobertura. Estos datos confirman que, con su preentrenamiento original, el modelo no es capaz de nivelar textos de ELE de forma fiable siguiendo el Plan Curricular del Instituto Cervantes.

En contraste, las dos pruebas de *fine-tuning* (90-5-5 y 80-20) proporcionaron valores de precisión y cobertura muy próximos a 1 en todos los niveles, con especial solidez en B2 y C1. La comparación entre ambos experimentos sugiere que un mayor volumen de datos de entrenamiento puede favorecer el rendimiento, aunque incluso con menos datos el modelo mantiene una alta capacidad de predicción. Este comportamiento coincide con lo observado en trabajos previos sobre adaptación de modelos para tareas específicas de procesamiento del lenguaje natural (García-Peñalvo, 2024; García-Peñalvo et al., 2024), donde la personalización del sistema incrementa notablemente su eficacia.

Figura 4

Comparación resultados de los experimentos



Si se compara con herramientas como *Write & Improve* (Cambridge English, s. f.), centradas en retroalimentación formativa y detección de errores, el presente estudio aporta un enfoque complementario: la clasificación automática por niveles MCER, que podría integrarse en entornos de aprendizaje para optimizar la colocación inicial de estudiantes y agilizar la labor docente. Asimismo, estudios como los de Burstein et al. (2003, *e-rater*) o McNamara et al. (2014, *Coh-Metrix*) ya habían mostrado la utilidad de combinar características lingüísticas y métricas automáticas para evaluar la escritura; nuestros resultados refuerzan esta línea, evidenciando que un modelo ajustado con corpus específicos alcanza niveles de rendimiento muy altos.

Por otro lado, en la experimentación *zero-shot learning*, el modelo, además de indicar el nivel, generó comentarios sobre errores y sugerencias de corrección

de los textos. Esta funcionalidad, aunque no relevante para el objetivo central de este estudio, podría explorarse en futuras investigaciones como recurso de apoyo didáctico para que los estudiantes identifiquen y corrijan sus errores.

En cuanto a las limitaciones, el estudio se ha realizado sobre un único corpus debido a la escasez de corpus especializados y nivelados de ELE que cumplan criterios homogéneos y controlados por expertos. Asimismo, la clasificación se ha basado en descriptores gramaticales presentes en el *prompt* (por ejemplo, repertorio verbal, número promedio de palabras por oración, uso de ciertos tiempos y modos), sin integrar de forma sistemática estructuras sintácticas complejas, recursos léxicos específicos o aspectos pragmáticos como la adecuación o la coherencia discursiva, que resultan especialmente relevantes en niveles intermedios y avanzados.

Como líneas de trabajo futuro, se propone:

1. Diseñar y validar nuevos corpus de ELE nivelados por expertos, con mayor diversidad temática y representatividad de niveles.
2. Explorar la aplicación del método con otros modelos de lenguaje para contrastar su rendimiento.
3. Ampliar el conjunto de descriptores lingüísticos, incorporando indicadores sintácticos, léxicos y pragmáticos.
4. Experimentar con variaciones controladas del *prompt* para evaluar su impacto en la clasificación.
5. Validar el sistema con textos auténticos de estudiantes en contextos educativos reales e integrar la herramienta en plataformas de gestión del aprendizaje (LMS).

En conjunto, los resultados alcanzados confirman que la inteligencia artificial generativa y, en particular, los modelos de lenguaje ajustados con datos específicos pueden convertirse en un recurso eficaz para agilizar la nivelación inicial del estudiantado de ELE, apoyando la labor docente y optimizando los procesos de enseñanza-aprendizaje.

Dentro de esta reflexión resulta pertinente abordar también los aspectos éticos y legales implicados en el uso de corpus de aprendientes, que se desarrollan en el subapartado siguiente.

Aspectos éticos y de licencia de uso

En el presente estudio se ha utilizado el Corpus de Aprendices de Español (CAES), que se encuentra disponible en línea para fines académicos y de investigación (Universidad de Santiago de Compostela, s. f.). Sus textos han sido anonimizados previamente, de modo que no incluyen datos personales identificables, y fueron recopilados en contextos educativos regulados y con validación experta, lo que garantiza un tratamiento adecuado de la información.

Desde un punto de vista ético, resulta necesario considerar que la ampliación o combinación del corpus con otros recursos debe atender tanto a la claridad en las licencias de uso como a la prevención de sesgos asociados a la lengua materna o al contexto sociocultural. Estos factores son esenciales para asegurar una clasificación justa y reproducible de los textos y, en última instancia, para garantizar un uso responsable de la inteligencia artificial en entornos educativos.

Agradecimientos

Esta publicación del proyecto Desarrollo de Modelos ALIA está financiada por el Ministerio para la Transformación Digital y de la Función Pública y por el Plan de Recuperación, Transformación y Resiliencia –Financiado por la Unión Europea– NextGenerationEU. Este trabajo también ha sido parcialmente financiado por el Proyecto CONSENSO (PID2021-122263OB-C21), el Proyecto MODERATES (TED2021-130145B-I00) y el Proyecto SocialTox (PDC2022-133146-C21), financiados por MCIN/AEI/10.13039/501100011033 y por la Unión Europea NextGenerationEU/PRTR, el proyecto ROMANET (CERV-2024-CHAR-LITI-101215052), financiado por la Unión Europea en el marco del programa Ciudadanos, Igualdad, Derechos y Valores, y el proyecto HEART-NLP-UJA (PID2024-156263OB-C21), financiado por MICIU/AEI/10.13039/501100011033 y por FEDER/UE. El trabajo de investigación realizado por Salud María Jiménez-Zafra es parte de la ayuda RYC2023-044481-I, financiada por MICIU/AEI/10.13039/501100011033 y por el FSE+.

NOTAS

- ¹ Para cuantificar las diferencias entre modelos, se utilizó el símbolo Δ (delta), que en estadística se emplea convencionalmente para indicar la diferencia entre dos valores. Así, $\Delta = \text{Macro-F1}(\text{fine-tuning}) - \text{Macro-F1}(\text{zero-shot}) \approx 0,77$ significa que la diferencia de Macro-F1 entre los experimentos comparados es aproximadamente 0,77.

REFERENCIAS

- Aparicio Gómez, W. O. (2023). La inteligencia artificial y su incidencia en la educación: transformando el aprendizaje para el siglo XXI. *Revista Internacional de Pedagogía e Innovación Educativa*, 3(2), 217-229. <https://dialnet.unirioja.es/servlet/articulo?codigo=9624350>
- Area-Moreira, M., Del Prete, A., Sanabria-Mesa, A. L. y Sannicolás-Santos, M. B. (2024). No todas las herramientas de IA son iguales: análisis de aplicaciones inteligentes para la enseñanza universitaria. *Digital Education Review*, 45, 141-149. <https://doi.org/10.1344/der.2024.45.141-149>
- Barroso-Osuna, J. y Cabero-Almenara, J. (2025). Potencialidades de la inteligencia artificial en la personalización de la educación. En P. Román-Graván, J. Barroso-Osuna, J. Cabero-Almenara y C. Llorente-Cejudo (Eds.), *Visiones sobre la integración educativa de la inteligencia*

- artificial (1.^a ed.). Dykinson. <https://doi.org/10.14679/4177>
- Baskara, R. y Mukarto, M. (2023). Exploring the implications of ChatGPT for language learning in higher education. *Indonesian Journal of English Language Teaching and Applied Linguistics*, 7(2), 343-358. <https://doi.org/10.21093/ijeltal.v7i2.1387>
- Biedma Torrecillas, A., Chamorro Guerrero, M. D., Lozano, G. y Sánchez Cuadrado, A. (2012). Diseño y validación de las pruebas de nivel del CLM de la Universidad de Granada. En *Actas del VII Congreso ACLES: Multilingüismo en los centros de lengua universitarios: evaluación, acreditación, calidad y política lingüística* (pp. 26-37). ACLES. <https://dialnet.unirioja.es/servlet/libro?codigo=501925>
- Bolaño-García, M. y Duarte-Acosta, N. (2024). Una revisión sistemática del uso de la inteligencia artificial en la educación. *Revista Colombiana de Cirugía*, 39(1), 51-63. <https://doi.org/10.30944/20117582.2365>
- Burstein, J., Elliot, N., Beigman Klebanov, B., Madnani, N., Napolitano, D., Schwartz, M., Houghton, P. y Molloy, H. (2018). Writing mentor: Writing progress using self-regulated writing support. *Journal of Writing Analytics*, 2, 285-313. <https://doi.org/10.37514/JWA-J.2018.2.1.12>
- Burstein, J., Kukich, K., Wolff, S., Lu, C., Chodorow, M., Braden-Harder, L. y Harris, M. D. (2003). E-rater as a diagnostic tool for writing instruction. En *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology: Demonstrations* (pp. 79-81). Association for Computational Linguistics.
- Cambridge English. (s. f.). *Write & Improve*. <https://writeandimprove.com>
- Cantero, M. V. (2024). Aproximación a un posible uso de ChatGPT para nivelar la expresión escrita en ELE. En F. M. Sirignano, R. Martínez Roig y A. López Padrón (Eds.), *Enseñanza y aprendizaje en la era digital desde la investigación y la innovación* (pp. 55-64). Octaedro.
- Centro Virtual Cervantes. (s. f.). *Ítem de respuesta cerrada*. https://cvc.cervantes.es/ensenanza/biblioteca_ele/diccio_ele/diccionario/itemrespuestacerrada.htm
- Chan, C. K. Y. y Tsi, L. H. Y. (2023). *The AI revolution in education: Will AI replace or assist teachers in higher education* [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2305.01185>
- Columbia University, Department of Latin American and Iberian Cultures. (s. f.). *Spanish placement exam*. Recuperado el 25 de julio de 2025, de <https://laic.columbia.edu/content/spanish-second-language-placement-exam>
- Consejo de Europa. (2002). *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación*. Instituto Cervantes; Ministerio de Educación, Cultura y Deporte. https://cvc.cervantes.es/ensenanza/biblioteca_ele/marco/cvc_mer.pdf
- Crespo Mendoza, R., Rodríguez López, W., Montenegro Patrel, M. y Tomalá Tomalá, G. (2024). IA: una herramienta para asistir a los docentes en la evaluación de los estudiantes. *Conocimiento Global*, 9(2), 305-323. <https://doi.org/10.70165/cglobal.v9i2.423>
- Fajardo, G. M., Ayala, D. C., Arroba, E. M. y López, M. (2023). Inteligencia artificial y la educación universitaria: una revisión sistemática. *Magazine de las Ciencias: Revista de Investigación e Innovación*, 8(1), 109-131. <https://doi.org/10.33262/rmc.v8i1.2935>
- García-Peñalvo, F. J. (2024). Cómo afecta la inteligencia artificial generativa a los procesos de evaluación. *Cuadernos de Pedagogía*, (549).
- García-Peñalvo, F. J., Llorens-Largo, F. y Vidal, J. (2024). La nueva realidad de la educación ante los avances de la

- inteligencia artificial generativa. *RIED-Revista Iberoamericana de Educación a Distancia*, 27(1), 9-39. <https://doi.org/10.5944/ried.27.1.37716>
- Hernández-León, N. y Rodríguez-Conde, M. J. (2024). Inteligencia artificial aplicada a la educación y la evaluación educativa en la universidad: Introducción de sistemas de tutorización inteligentes, sistemas de reconocimiento y otras tendencias futuras. *Revista de Educación a Distancia (RED)*, 24(78), Artículo 6. <https://doi.org/10.6018/red.594651>
- Hong, W. C. H. (2023). *The impact of ChatGPT on foreign language teaching and learning: Opportunities in education and research*. *Journal of Educational Technology and Innovation*, 5(1), 38-53. <https://doi.org/10.61414/jeti.v5i1.103>
- Instituto Cervantes. (2006). *Plan curricular del Instituto Cervantes: Niveles de referencia para el español* (3 vols.). Biblioteca Nueva. https://cvc.cervantes.es/ensenanza/biblioteca_ele/plan_curricular/
- Li, Y. (2023). *A practical survey on zero-shot prompt design for in-context learning* [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2309.13205>
- McNamara, D. S., Graesser, A. C., McCarthy, P. M. y Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511894664>
- Morales-Chan, M. A. (2023). *Explorando el potencial de ChatGPT: Una clasificación de prompts efectivos para la enseñanza*. Universidad Galileo. <https://biblioteca.galileo.edu/tesario/handle/123456789/1348>
- Moreno, R. D. (2019). La llegada de la inteligencia artificial a la educación. *Revista de Investigación en Tecnologías de la Información*, 7(14), 260-270. <https://doi.org/10.36825/riti.07.14.022>
- OpenAI. (2022). *ChatGPT (versión 3.5) [Modelo de lenguaje de inteligencia artificial]*. <https://openai.com>
- Owan, V. J., Abang, K. B., Idika, D. O., Etta, E. O. y Bassey, B. A. (2023). Exploring the potential of artificial intelligence tools in educational measurement and assessment. *EURASIA Journal of Mathematics, Science and Technology Education*, 19(8), em2307. <https://doi.org/10.29333/ejmste/13428>
- Palacios Martínez, I., Barcala Rodríguez, F. M. y Rojo, G. (2019). El corpus de aprendices de español (CAES) y sus aplicaciones para la enseñanza y aprendizaje del español como lengua extranjera. En M. Blanco, H. Olbertz y V. Vázquez Rozas (Eds.), *Corpus y construcciones: Perspectivas hispánicas* (pp. 273-301). Universidade de Santiago de Compostela (Verba, Anexo 79). <https://doi.org/10.15304/9788417595876>
- Pourpanah, F., Abdar, M., Luo, Y., Zhou, X., Wang, R. y Lim, C. P. (2023). A review of generalized zero-shot learning methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4), 4051-4070. <https://doi.org/10.1109/TPAMI.2022.3182926>
- Roumeliotis, K. I., Tselikas, N. D. y Nasiopoulos, D. K. (2024). Next-generation spam filtering: Comparative fine-tuning of LLMs, NLPs, and CNN models for email spam classification. *Electronics*, 13(11), 2034. <https://doi.org/10.3390/electronics13112034>
- Salguero Romero, P. (2023). *La traducció pedagògica i l'ús de ChatGPT-3 a classes d'anglès com a segona llengua per a nens i nenes* [Trabajo de fin de grado, Universitat Autònoma de Barcelona]. Repositorio UAB. <https://ddd.uab.cat/record/279383>
- Universidad de Santiago de Compostela. (s. f.). *Corpus de aprendices de español (CAES)*. <https://galvan.usc.es/caes>
- University of Wisconsin–Madison, Testing and Evaluation Services. (s. f.). *Spanish*

- placement test. University of Wisconsin–Madison. <https://testing.wisc.edu/centerpages/spanishtest.html>
- Wang, L. (2024). *El uso de ChatGPT y Gemini para la preparación de actividades de comprensión oral de CLE: Un estudio comparativo* [Trabajo de fin de máster, Universitat Autònoma de Barcelona]. Repositorio UAB. <https://ddd.uab.cat/record/300353>
- Wang, X., Lee, Y. y Park, J. (2022). *Automated evaluation for student argumentative writing: A survey* [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2205.04083>
- Zambrano Campozano, R. F. (2025). Impacto de la inteligencia artificial en la transformación digital de la educación. *Serie Científica de la Universidad de las Ciencias Informáticas*, 18(1), 177-192. https://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2306-24952025000100177
- Zhang, S. (2021). Review of automated writing evaluation systems. *Journal of China Computer-Assisted Language Learning*, 1(1), 170-176. <https://doi.org/10.1515/jccall-2021-2007>
- Zapata Ros, M. (2024). IA generativa y ChatGPT en educación: Un reto para la evaluación y ¿una nueva pedagogía? *Revista Paraguaya de Educación a Distancia (REPED)*, 5(1), 12-44. <https://doi.org/10.56152/reped2024-vol5num1-art2>

ANEXOS

Anexo 1. Ejemplos completos de respuestas del modelo

Texto del corpus:

“Hola Eva.

Que hace dos meses, hemos estudiado espanol en madrid.

He vuela a Japon la semana pasada.

Gracias por tu ayuda.

Yo escribo mi familia, mi mujer, hija y hijo.

Mi mujer trabaja en Idioma.

Hija trabaja en una tienda de palo, hijo en un enterprise electrico.

Vivimos en yokohama cerca de tokyo.

Vamos a ir a madrid que mes viene

Por favor, vernos en madrid.

Hasta pronto.” (CAES, 2022)

Respuesta del modelo:

Errores encontrados en el texto:

- “Que hace dos meses, hemos estudiado espanol en madrid.” La forma correcta sería “Hace dos meses, estudiamos español en Madrid.”

- “He vuela a Japon la semana pasada.” La forma correcta sería “He vuelto a Japón la semana pasada.”

- “Gracias por tu ayuda.” Debería ser “Gracias por tu ayuda.”

- “Yo escribo mi familia, mi mujer, hija y hijo.” Se debería decir “Te escribo sobre mi familia: mi mujer, mi hija y mi hijo.”

- “Mi mujer trabaja en Idioma.” La forma correcta sería “Mi mujer trabaja en Idiomas.”

- “Hija trabaja en una tienda de palo, hijo en un enterprise electrolico.” Debería ser “Mi hija trabaja en una tienda de ropa, y mi hijo en una empresa electrónica.”

- “Vamos a ir a madrid que mes viene.” La forma correcta sería “Vamos a ir a Madrid el mes que viene.”

- “Hasta plonto.” Debería ser “Hasta pronto.”

NIVEL: A2 (OpenAI, 2022)

—
Texto del corpus:

“El 24 de Noviembre

Estimados señores

El 22 de noviembre pasado, viajé con Tonta Airlines de Badaros a Madrid con el vuelo TA 222.

Estoy aquí en Madrid hasta el 27 para algunas citas profesionales.

Cuando embarqué en Badaros, la azafata me dijo que mi maleta estaba demasiado grande y que tenía que facturarla.

La maleta esta de marca Samsonite, de modelo XYZ1234, color negra, con mis nombre y dirección muy visibles.

Cuando llegamos a Madrid, la maleta no fue encontrada.

Hace a dos días que estoy esperando pero nadie de su compañía me ha llamado con información.

Evidentemente, no puedo quedar sin cambiarme o lavarme me.

Entonces compré algunas cosas y me parece que Tonta Airlines tenga que indemnizarme.

Yo incluido la lista de mis compras y les pide que se pongan en contacto para decirme si sean de acuerdo de pagar el 50% des estas compras.

Se pueden contactarme por correo electronico o llamarme en el movil 660 12 34 45 y evidentemente, en cualquier caso que Tonta Airlines encuentre mi maleta, me daría mucha alegría.

Un saludo” (CAES, 2022)

Respuesta del modelo:

El texto presenta varios errores gramaticales y algunos elementos propios de un nivel de español como lengua extranjera B1:

- Uso incorrecto de tiempos verbales: “tenia” en lugar de “tenía”.

- Errores en el uso de los artículos: “la maleta esta” en lugar de “la maleta está”.

- Uso incorrecto de pronombres: “Hace a dos días” en lugar de “Hace dos días”.

- Errores de concordancia de género y número: “me parece que Tonta Airlines tenga que indemnizarme” debería ser “me parece que Tonta Airlines debe indemnizarme”.

NIVEL: B1 (OpenAI, 2022)

Fecha de recepción del artículo: 1 de junio de 2025

Fecha de aceptación del artículo: 9 de septiembre de 2025

Fecha de aprobación para maquetación: 8 de octubre de 2025

Fecha de publicación en OnlineFirst: 28 de octubre de 2025

Fecha de publicación: 1 de enero de 2026

Metodología participativa para la creación de recursos digitales de microaprendizaje con inteligencia artificial generativa

Participatory methodology for creating microlearning digital resources with generative AI



- Bell Manrique-Losada - *Universidad de Medellín, UdeMedellín (Colombia)*
 Sandra Isabel Arango-Vásquez - *Universidad de Medellín, UdeMedellín (Colombia)*
 Mariutsi Alexandra Osorio-Sanabria - *Instituto Tecnológico Metropolitano, ITM (Colombia)*

RESUMEN

En los últimos años, los avances en las Tecnologías de la Información (TI) e Internet han transformado el ámbito educativo. Estos avances han creado nuevas oportunidades de aprendizaje e interacción colaborativa en la producción de recursos educativos digitales. Tanto estudiantes como docentes reconocen esta oportunidad y están aprovechando el potencial de los modelos, métodos, TI y enfoques derivados de la inteligencia artificial generativa (IAGen). Este estudio analiza y describe el proceso de implementación de una metodología basada en principios de microaprendizaje para guiar la interacción entre actores y lograr la cocreación de recursos de microaprendizaje (microRED) con IAGen. La investigación adopta una perspectiva cualitativa y se estructura metodológicamente desde la articulación de la Investigación Acción Participativa (IAP) y la Metodología de Solución Creativa de Problemas Sociales (SCPS). Este estudio se desarrolla en tres fases: 1) Reconocimiento del contexto y delimitación temática participativa exploratoria; 2) Diseño colaborativo de microRED con estudiantes y docentes de secundaria y educación superior, integrando retos y evaluación de herramientas de IAGen; y, 3) Validación de la metodología propuesta a través de talleres de cocreación, *bootcamp* y análisis cualitativo. La recopilación y el análisis cualitativo de los datos se realiza a través de talleres de cocreación con estudiantes y docentes de básica secundaria y educación superior. Como resultado del análisis de las contribuciones de los actores en cada fase, se concluye que los enfoques de IAP y SCPS mejoran la interacción y la generación de conocimiento asociado al uso de la IAGen para la cocreación de RED.

Palabras clave: educación; aprendizaje; inteligencia artificial; tecnología de la información.

ABSTRACT

In recent years, advances in information technology and the Internet have transformed education by creating new opportunities for learning and collaboration in the production of Digital Educational Resources (DERs). Both students and teachers are leveraging the potential of models, methods, and approaches derived from Generative Artificial Intelligence (GenAI). The study analyzes and describes a microlearning-based methodology to guide interaction between actors and support the co-creation of Microlearning Resources (abbreviated microDERs) using GenAI. Adopting a qualitative approach, the study integrates Participatory Action Research (PAR) and the Creative Social Problem Solving (CSPS) methodology. It unfolds in three phases: (1) recognizing the context and conducting exploratory, participatory, thematic delimitation; (2) collaboratively designing microDERs with high school and higher education students and teachers, including the evaluation of GenAI tools; and (3) validating the methodology through co-creation workshops, bootcamps, and qualitative analysis. Data are collected and analyzed through co-creation workshops with students and teachers from elementary, high school, and higher education. Findings on the contributions of the actors in each phase indicate that combining PAR and CSPS improves interaction and knowledge building in the co-creation of DERs using GenAI.

Keywords: education; learning; artificial intelligence; information technology.

INTRODUCCIÓN

La revisión de estrategias educativas para mejorar los Resultados de Aprendizaje de los estudiantes es clave para fortalecer su creatividad y el pensamiento crítico. En este escenario, el microaprendizaje emerge como una alternativa, al ofrecer contenidos breves y accesibles a través de Recursos Educativos Digitales (RED) (Denojean-Mairet et al., 2024; Torgerson e Iannone, 2019). Su impacto positivo se potencia con la integración de la Inteligencia Artificial Generativa (IAGen), que facilita la personalización y creación de materiales adaptados a contextos específicos (Miao y Holmes, 2024; Piantari et al., 2024).

La cocreación de RED entre estudiantes y docentes promueve la autonomía y la Apropiación Social del Conocimiento, entendida como el proceso mediante el cual las comunidades producen, usan y aplican el conocimiento, fortaleciendo su democratización e impacto en la transformación social (Gonnet, 2019).

Este estudio analiza la implementación de una metodología participativa en contextos educativos para facilitar la cocreación de recursos de microaprendizaje (microRED) mediante herramientas de IAGen. La investigación se desarrolló en tres fases: 1) Reconocimiento del contexto y mapeo de actores, necesidades, desafío y conocimiento; 2) Diseño colaborativo de microRED con estudiantes y docentes de secundaria (10° y 11° grado) y educación superior (9° semestre), integrando retos y evaluación de herramientas de IAGen; y 3) Validación de la metodología propuesta a través de talleres de cocreación, *bootcamp* y análisis cualitativo. La investigación adopta el enfoque de Investigación-Acción Participativa (IAP) y el método de Solución Creativa de Problemas Sociales (SCPS), dado que involucra activamente a estudiantes y docentes en la construcción colaborativa de conocimiento mediante ciclos reflexivos y participativos (Kemmis y McTaggart, 2005). Estos enfoques favorecen el diálogo de saberes y la participación comunitaria (Ritzer, 2008) asegurando que la cocreación de microRED responde a las necesidades y perspectivas de sus beneficiarios.

El artículo se organiza en cuatro apartados: marco conceptual y antecedentes orientado a los fundamentos de IAP y SCPS, microaprendizaje e IAGen. Luego se presenta el proceso metodológico, donde se describen las tres fases definidas en el proceso de cocreación de microRED. En el siguiente apartado se encuentran los resultados, donde se presentan las actividades desarrolladas, los actores involucrados y los artefactos generados en cada etapa de la metodología propuesta. Finalmente, se describe la discusión, conclusiones y trabajo futuro.

MARCO CONCEPTUAL

Investigación-acción participativa

La IAP es una metodología que permite la comprensión de los fenómenos sociales, combinando teoría, acción y participación, adaptándose a contextos regionales para promover la construcción colectiva de conocimiento (Fals Borda, 2001; Sirvent y Rigal, 2012). Esta metodología, reconoce a los sujetos no como objetos de estudio, sino como agentes capaces de transformar sus realidades mediante procesos colaborativos de investigación (Fals Borda, 2001). Además, fortalece la apropiación social y las transiciones de ecosistemas sociales al integrar prácticas democráticas, saberes y una mirada sistémica de las problemáticas (Arnanz et al., 2023).

Solución Creativa de Problemas Sociales

La SCPS es una metodología participativa basada en *design thinking*, orientada al diseño colaborativo de soluciones educativas y sociales contextualizadas (Mirón, 2022). Esta metodología, promueve a través de herramientas de creatividad, análisis crítico y codiseño espacios de cocreación, pensamiento crítico y empoderamiento de distintos actores para la identificación de problemas, diseño de estrategias y evaluación del impacto. Su integración con la IAP permite ampliar la capacidad diagnóstica y fortalecer el diseño de soluciones en escenarios educativos, comunitarios y organizacionales.

Inteligencia artificial generativa (IAGen)

La IAGen es una rama de la Inteligencia Artificial (IA) que, mediante redes neuronales profundas y procesamiento del lenguaje natural (Cain, 2024), automatiza el procesamiento de información y la presentación de resultados en todas las representaciones simbólicas clave del pensamiento humano, para ofrecer productos de conocimiento semiacabados en diferentes formatos (Cabeza-Rodríguez, 2025; Samoilil et al., 2021). En educación, esta tecnología potencia la creación de materiales didácticos personalizados, mejora la productividad docente para diseñar recursos educativos y entornos de aprendizaje y facilita experiencias de aprendizaje adaptadas a las necesidades e intereses de los estudiantes (Abdous, 2023; Cain, 2024; Dickey y Bejarano, 2023).

Microaprendizaje

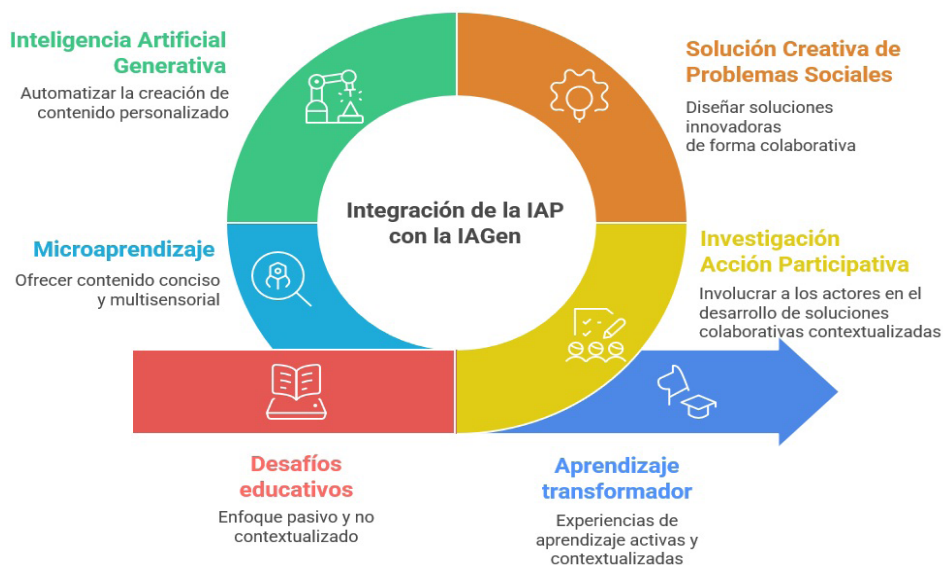
El microaprendizaje es una estrategia educativa que proporciona contenido digital breve y focalizado, incorporando elementos multisensoriales y multimodales para

facilitar la comprensión y adaptación del aprendizaje (Denojean-Mairet et al., 2024; Dolasinski y Reynolds, 2020; Durán y Escudero, 2023; Torgerson e Iannone, 2019). Esta estrategia plantea que los RED se estructuran en siete dimensiones (tiempo, contenido, currículo, formato, proceso, medio y modalidad de aprendizaje) (Hug et al., 2006) y se presentan en diferentes formatos, como vídeos, textos, micropodcasts y mensajes cortos en redes sociales (Díaz Redondo et al., 2021; Mateus-Nieves y Moreno, 2021).

En la Figura 1 se presenta la interacción funcional de los cuatro pilares conceptuales (IAP, SCPS, IAGen y el Microaprendizaje) que soportan el desarrollo de la metodología planteada con esta investigación. Como se puede inferir de la figura, esta interacción funcional inicia desde la definición del *enfoque del proceso* a partir de desafíos educativos en el aula. Luego se inicia el proceso de interacción con los actores y diálogo de saberes en el marco de las metodologías de Investigación acción participativa y la Solución Creativa de Problemas Sociales. El proceso fluye hacia la etapa central del proceso asociada hacia la creación de contenido personalizado, apoyado con la IAGen, para finalmente materializar dichos recursos en contenidos de microaprendizaje.

Figura 1

Interacción funcional de los pilares conceptuales del estudio



Fuente: Elaboración propia. Software: [napkin.ai](https://www.napkin.ai)

Antecedentes de referencia y estudios relacionados

El avance acelerado de la IA ha impulsado modelos, procesos y estrategias educativas más personalizadas y automatizadas, generando oportunidades para transformar prácticas pedagógicas, creación de recursos de aprendizaje y sistemas de evaluación. Sin embargo, persisten desafíos relacionados con la adaptación a características individuales (Yang, 2024), procesos de aprendizaje dinámicos (Tu et al., 2025), empatía (Zhang et al., 2024), y metas de alfabetización de los estudiantes (Adeshola y Adepoju, 2023).

Desde el enfoque del microaprendizaje, la IAGen potencia la colaboración entre docentes y tecnologías, promoviendo el desarrollo de habilidades cognitivas, psicomotoras y afectivas de los estudiantes quienes se posicionan como eje del proceso formativo (Memon y Kwan, 2025; Bygstad et al., 2022). De otra parte, se han logrado avances en la generación de microcontenidos adaptativos mediante el uso de IAGen (Boumalek et al., 2024).

Este contexto plantea dos líneas de intervención: la primera, desde la reconfiguración del rol docente, dado que cede su rol central y ubica al estudiante en el centro del proceso de enseñanza-aprendizaje, y reconoce a las herramientas de IAGen como actores importantes en dicho proceso (Bozkurt, 2023). La segunda línea, desde el aprovechamiento de las tecnologías emergentes para diseñar y crear contenidos atractivos y pertinentes (Holmes et al., 2023; Karras et al., 2020; Luckin et al., 2016), buscando alinearlos con las necesidades e intereses de los jóvenes de hoy (Holmes et al., 2023; Sengar et al., 2025). La metodología presentada en este estudio responde a ambas perspectivas, proponiendo un modelo colaborativo para la creación de microRED a partir de la integración y uso de herramientas de IAGen dirigido a estudiantes jóvenes.

METODOLOGÍA

Este estudio se desarrolla desde una perspectiva cualitativa y participativa, articulando los enfoques de IAP y SCPS. Esta convergencia permite abordar problemáticas complejas de forma situada, colaborativa y transformadora desde cuatro dimensiones: 1) Herramienta orientada a la apropiación social de conocimiento, al promover la colaboración entre los participantes que experimentan la problemática y quienes serán los usuarios finales de los microRED; 2) Orientación a la acción, ya que el conocimiento generado se traduce en artefactos y acciones concretas en la comunidad (microRED en construcción/uso); 3) Fomento de creatividad e innovación buscando soluciones a los retos identificados; y 4) Comunicación y diálogo, al propiciar comunicación efectiva y abierta entre los participantes, propiciando relaciones colaborativas y procesos de apropiación.

La investigación parte de la hipótesis: “la implementación de una metodología de cocreación de microRED, mediante la participación de profesores y jóvenes

estudiantes, apoyada en IAGen, favorece el desarrollo del pensamiento computacional y la interacción social”. Esta hipótesis responde a la pregunta de investigación ¿Cómo diseñar una metodología de cocreación que, mediante el uso de herramientas de IAGen, permita a profesores y jóvenes estudiantes crear microRED?

Los instrumentos de recolección de información incluyeron: (i) guías de grupo focal aplicadas en talleres de cocreación con estudiantes y profesores, validadas mediante procesos de pilotaje; (ii) cuestionarios semiestructurados para recoger percepciones sobre el proceso de la metodología propuesta y alrededor de los productos generados; y (iii) matrices de análisis y bitácoras de investigación que documentaron las interacciones, decisiones y ajustes realizados. La validez de contenido de estos instrumentos se realizó mediante retroalimentación de los actores participantes y ejercicios de revisión cruzada por parte del equipo investigador.

El análisis de los datos se fue construyendo de manera progresiva a partir de un ejercicio de codificación abierta y axial que permitió dar forma a categorías emergentes y reconocer los vínculos que se establecían entre ellas. Para ordenar la información y facilitar su interpretación, se elaboraron matrices comparativas que ayudaron a visualizar patrones y a conectar los hallazgos con las preguntas de investigación. Este proceso no se desarrolló de manera aislada: se contrastaron las interpretaciones con los participantes y se aplicaron estrategias de triangulación que reforzaron la coherencia y la transparencia de los resultados. La saturación teórica se alcanzó cuando, en los talleres finales, ya no aparecieron nuevas categorías, lo que confirmó que el análisis había llegado a un punto de consistencia y solidez suficiente para sustentar las conclusiones.

La saturación teórica se realizó a través de un proceso de triangulación y contraste. Se analizaron los microRED cocreados, las matrices de sistematización y las bitácoras de investigación, lo que permitió verificar la coherencia de las categorías emergentes en diferentes fuentes. Asimismo, las interpretaciones fueron contrastadas con los marcos conceptuales de la IAP, la SCPS, la IAGen y el microaprendizaje, garantizando que los hallazgos se sostuvieran teóricamente y no dependieran únicamente de la dinámica de los grupos. Esto permitió confirmar que las categorías alcanzaron un nivel de consistencia y solidez suficiente para sustentar las conclusiones.

Participantes

Siguiendo la visión de Manzini (2015) sobre el diseño colaborativo en innovación social, se involucraron actores clave del contexto educativo: equipo investigador, docentes y estudiantes de secundaria de grados 10° y 11°, así como docentes y estudiantes de pregrado (9° semestre) vinculados a un semillero de investigación de la Universidad en Medellín, Colombia. La investigación se basó en una muestra intencional, diversa y representativa. En la Tabla 1 se presenta el nivel educativo, actor, edad y número de los participantes en el diseño y validación de la metodología.

Tabla 1
Datos de los participantes

Nivel educativo	Actor	Edad	Participantes 2024-2	Participantes 2025-1
Básico secundario	Estudiantes grados 10º y 11º	13 y 16 años	33	32
	Profesores	40 – 50 años	2	2
Universitario	Estudiantes noveno semestre	18 - 22 años	13	19
	Estudiantes del semillero de investigación	18 - 22 años	8	5

Fuente: Datos del estudio.

Procedimiento

La investigación se estructuró en tres fases: 1) Reconocimiento del contexto y delimitación temática a partir de la IAP, 2) Diseño y cocreación de microRED, y 3) validación. En la Figura 2 se presentan estas fases y su articulación metodológica.

Figura 2
Proceso metodológico



Fuente: Elaboración propia.

Fase 1: Reconocimiento del contexto y delimitación temática participativa

Esta fase se desarrolló bajo los principios de la IAP, priorizando la comprensión del contexto educativo desde la experiencia y la voz de los actores. Durante el semestre 2024-2 se realizaron talleres de cocreación con los estudiantes y docentes de las instituciones participantes.

Fase 2: Diseño y cocreación de microRED

Esta fase se iteró en dos semestres (2024-2 y 2025-1) y se articuló a los principios de la IAP y la SCPS, orientando el trabajo hacia soluciones colaborativas y creativas. A partir del mapa de conocimiento, los estudiantes universitarios formularon y elaboraron bocetos de microRED, que fueron socializados y enriquecidos por los estudiantes de secundaria en sesiones participativas. Esta interacción permitió la cocreación progresiva de prototipos. En paralelo, los estudiantes del semillero de investigación exploraron un conjunto de herramientas de IAGen, evaluándolas según criterios como alcance, tipo de licencia y recurso generable para determinar aquellas más adecuadas para la creación de microRED.

Fase 3: Validación

Esta fase tuvo como propósito verificar la coherencia estructural (etapas y actividades) y funcional (tareas) de la metodología de cocreación de microRED en un contexto real de aplicación. Para ello, se empleó un caso práctico con un grupo piloto de estudiantes vinculados a un semillero de investigación universitario. La validación se desarrolló en dos momentos complementarios. El primero, evaluó la organización y secuencia de etapas y actividades propuestas; y, el segundo, aplicó la metodología en el caso práctico, con el fin de analizar su eficacia y pertinencia en condiciones reales.

Procedimiento ético para la aplicación

La aplicación de la metodología siguió un procedimiento ético orientado a proteger la integridad de los participantes, especialmente estudiantes menores de edad. Este incluyó un consentimiento informado y un protocolo de confidencialidad y anonimato. El consentimiento se presentó en un formato claro y comprensible para estudiantes de básica secundaria, con firma de padres/acudientes y asentimiento de los menores, asegurando la comprensión del propósito de la investigación y condiciones de su participación.

El protocolo de confidencialidad incluyó el uso de códigos/seudónimos en registros, informes y publicaciones; almacenamiento seguro de datos en repositorios

de acceso restringido al equipo investigador; y eliminación de información personal al finalizar la investigación y plazos de resguardo. El procedimiento se alineó con la Ley 1581 de 2012 (Congreso de la República de Colombia, 2012) y el Decreto 1377 de 2013 (Presidente de la República de Colombia, 2013) sobre protección de datos personales en Colombia, asegurando la recolección mínima necesaria y el uso ético de imágenes, videos o audios.

RESULTADOS

A continuación, se describen los resultados por cada fase metodológica.

Fase 1: Reconocimiento del contexto y delimitación temática participativa

Los resultados de esta fase se estructuran en dos momentos, lo que permitió avanzar en un proceso colaborativo de diagnóstico y construcción colectiva de conocimiento:

Momento 1. Talleres con enfoque participativo.

Durante los semestres 2024-2 y 2025-1 entre el equipo investigador y con apoyo de los profesores de secundaria y estudiantes universitarios se realizaron cinco talleres con 32 estudiantes de básica secundaria (dos con un grupo de 10° y tres con 11°). Las sesiones se diseñaron como espacios de cocreación para explorar las realidades educativas, sociales y emocionales del estudiantado. A partir de las dinámicas participativas, se identificaron problemáticas prioritarias: falta de orientación vocacional, acceso limitado a oportunidades educativas (becas y emprendimiento) y gestión emocional como eje para la convivencia escolar.

Momento 2. Mapa de conocimiento del contexto.

Los hallazgos fueron sistematizados en un mapa de conocimiento que visibiliza las necesidades formativas, las potencialidades estudiantiles y las dinámicas del entorno escolar. Este insumo orientó las siguientes fases metodológicas, definiendo retos y dos ejes temáticos que guiaron el desarrollo de los microRED (orientación vocacional y gestión de emociones).

Fase 2: Diseño y cocreación de microRED.

Esta fase tradujo necesidades del mapa de conocimiento en productos que integraron microRED con el fin de abordar los retos identificados en la Fase 1. El desarrollo de estos productos se articuló a procesos colaborativos entre estudiantes

universitarios y de secundaria, y se basó en el uso pedagógico de herramientas de IAGen. Los resultados de esta fase se desglosan en dos momentos.

Momento 1. Cocreación de los microRED.

Durante los semestres 2024-2 y 2025-1, los estudiantes universitarios diseñaron bocetos en respuesta a los retos identificados en la Fase 1. Dichos bocetos fueron socializados con los estudiantes de secundaria, quienes desde su experiencia hicieron realimentación a través de preguntas orientadoras. Como resultado de la interacción entre los participantes del estudio se desarrollaron cuatro productos que muestran procesos de sensibilización y formación adaptados al contexto escolar. En la Tabla 2 se presenta la caracterización del trabajo realizado en el Momento 1 organizado por cada semestre académico, y en las Figuras 3 y 4 se muestran algunas evidencias de los productos cocreados en cada semestre.

Tabla 2
Momento 1. Cocreación de los MicroRED

Semestre	Participantes	Acciones	Preguntas orientadoras	Producto
2024-2	13 estudiantes universitarios 33 estudiantes de secundaria grado 10	• Revisar retos. • Diseñar bocetos de productos. • Socializar bocetos. • Revisar y proponer ajustes.	• ¿Cómo reducir la incertidumbre vocacional para evitar la deserción y fortalecer el proyecto profesional? • ¿Cómo implementar estrategias de gestión de emociones que mejoren la comunicación asertiva en estudiantes de secundaria?	• <i>Vía Pro</i>. Sitio web de orientación vocacional con cápsulas temáticas, simuladores de elección vocacional y videos con testimonios estudiantiles. • <i>Moodies</i>. Aplicación móvil para la gestión de emociones, con ejercicios de autorregulación, contenidos interactivos de reflexión emocional y guías de comunicación asertiva.

Semestre	Participantes	Acciones	Preguntas orientadoras	Producto
2025-1	19 estudiantes universitarios 32 estudiantes de secundaria grado 11	• Revisar retos y bocetos del semestre anterior. • Proponer nuevo reto. • Identificar mejoras en retos previos.	• ¿Cómo mejorar el relacionamiento interpersonal entre estudiantes mediante estrategias de comunicación que favorezcan la convivencia? • ¿Cómo mejorar la herramienta Vía Pro para que los estudiantes encuentren información relevante para su continuidad académica? • ¿Cómo mejorar la plataforma Moodies para motivar la participación de los estudiantes?	• <i>La vida no es un juego</i>. Prototipo de juego de mesa que promueve la convivencia escolar, con microRED análogos como tarjetas temáticas y guías impresas de reflexión colectiva. • <i>EnRuta</i>. Plataforma web que optimiza la usabilidad del producto <i>Vía Pro</i>, integrando cápsulas interactivas con contenidos vocacionales. • <i>Moodies +</i>. Plataforma web gamificada con microRED de desafíos, autoevaluaciones y contenidos multimedia para fomentar la participación estudiantil.

Fuente: Datos del estudio.

Figura 3

Evidencia de prototipos cocreados 2024-2



a) Prototipo Vía Pro



b) Aplicación Moodies

Fuente: Datos del estudio.

Figura 4

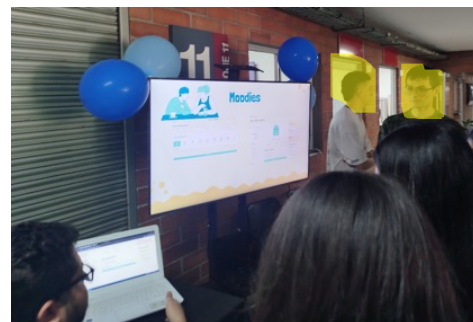
Evidencia de prototipos cocreados 2025-1



a) Juego análogo de la vida no es un juego



b) Sitio web de EnRuta



b) Socialización plataforma de Moodies +

Fuente: Datos del estudio.

Momento 2. Exploración del uso de herramientas IAGen.

Los estudiantes del semillero de investigación exploraron 24 herramientas de IAGen, clasificadas por tipo de recurso (texto, imagen, audiovisual y código), alcance y licencia de uso (libre o propietaria).

Para los recursos textuales, se analizaron seis herramientas centradas en generación, traducción y modelado del lenguaje (*ChatGPT*, *Claude AI*, *Google Gemini*, *DeepL*, *Hugging Face* y *Notion AI*). Esta categoría fue la más explorada, destacando *Hugging Face* como única opción de código abierto. Las demás, aunque son útiles para redacción pedagógica, traducción multilingüe y creación de guías formativas, requieren licencias propietarias. También se exploraron asistentes de escritura como Grammarly y de comprensión documental como Acrobat IA y Humata, que facilitan la revisión y síntesis de textos académicos.

En cuanto a recursos visuales, se exploraron cinco herramientas para crear imágenes digitales a partir de texto (*Adobe Firefly*, *DALL-E*, *Magic Design*, *Midjourney* y *Stable Diffusion*). De esta revisión se identificó que solo *Stable Diffusion* opera bajo licencia libre. Estas aplicaciones permiten crear microRED visuales en distintos formatos, tales como imágenes 2D, imágenes estilizadas, efectos visuales, diseños automatizados e ilustraciones de alta resolución. La exploración fue limitada, dado su acceso restringido, lo cual exige una evaluación avanzada posterior.

Respecto a recursos audiovisuales, se exploraron cuatro herramientas para creación y edición de video mediante avatares, conversión de texto a video y efectos visuales (*Lumen5*, *Pictory*, *RunwayML* y *Synthesia*) con licencia propietaria lo que puede limitar su uso institucional. Se analizaron además dos herramientas útiles para personalizar la construcción de recursos educativos, *Whisper*, para la transcripción de audio a texto con licencia libre y *ElevenLabs*, para la síntesis de voz realista.

En la categoría de herramientas para generar código de aplicativos software y pruebas automatizadas, se evaluaron *GitHub*, *Copilot*, *Tabnine*, *Testim* y *Applitools*, por su capacidad para generar, asistir o validar fragmentos de código. Aunque su aplicación directa en educación escolar es limitada, se reconoció su potencial en actividades y procesos de formación mayormente técnica.

Alineado con la hipótesis de la investigación que orienta este estudio, desde la ejecución de esta fase se logró comprobar que:

- Es viable involucrar a profesores y jóvenes estudiantes en procesos colaborativos de creación de microRED.
- Los estudiantes pueden asumir un rol protagónico en su aprendizaje, guiados por una metodología que integre su participación, la mediación docente y las herramientas de IAGen.
- Una metodología de cocreación apoyada por herramientas de IAGen facilita el diseño e implementación de microRED.
- El proceso de diseño y cocreación de microRED fomenta y/o fortalece el pensamiento computacional y la interacción social entre los estudiantes.

Fase 3: Validación

Las acciones y hallazgos del proceso de cocreación de microRED se documentaron mediante ilustraciones, extractos de los documentos de análisis, fotografías del trabajo con estudiantes y con los artefactos generados. En la Tabla 3 se detallan los actores involucrados de esta fase, según el momento, la sesión, la modalidad, los participantes y/o actores y las actividades.

Tabla 3
Momentos y actores involucrados en la Fase Validación

Momento	Sesión	Modalidad	Participantes/ Actores	Actividad
1. Evaluación de la estructura de la metodología	S1	Presencial	• Equipo investigador • Profesor Coordinador del semillero de investigación. • Estudiantes universitarios integrantes del semillero	Intercambio creativo de ideas
2. Validación de la implementación de la metodología	S2	Presencial	• Equipo investigador • Estudiantes universitarios integrantes del semillero	Reestructuración y ajustes
	S3	Presencial	• Equipo investigador	Organización visual
2. Validación de la implementación de la metodología	S1	Presencial	• Equipo investigador • Profesor Coordinador del semillero de investigación. • Estudiantes universitarios integrantes del semillero	Implementación y evaluación de la metodología

Fuente: Elaboración propia.

Momento 1. Evaluación de la estructura de la metodología

Este momento incluyó tres sesiones diseñadas para involucrar a los participantes del estudio en la construcción de la propuesta de la metodología, en coherencia con sus necesidades, contextos y perspectivas.

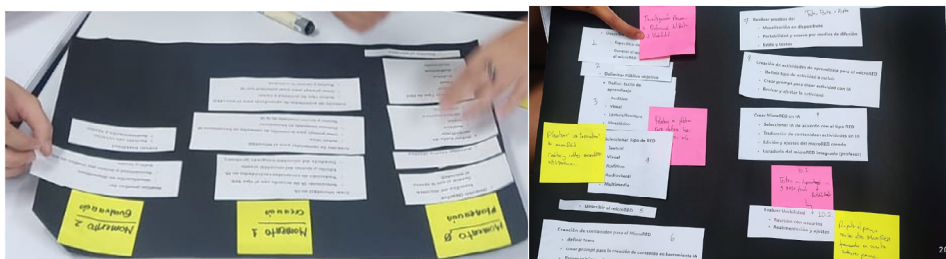
En la Sesión 1 se propició un espacio colaborativo para el intercambio creativo de ideas entre los participantes (equipo investigador, estudiantes y profesor coordinador del Semillero). Los estudiantes se dividieron en tres grupos, el profesor ejerció como apoyo orientador, y el equipo investigador actuó como facilitador del proceso, dirigiendo las actividades clave que partieron de la reflexión individual, puesta en común, análisis y construcción conjunta. Estas actividades incluyeron

acciones de ideación, análisis del proceso, priorización de elementos clave y análisis de experiencias previas en creación de recursos.

Durante la sesión, los estudiantes definieron rutas de trabajo y flujos de acción para diversos actores, mediante un conjunto de tareas orientadas a la cocreación de microRED. Se incluyeron elementos a la metodología, promoviendo una actitud reflexiva frente a los aportes realizados. Como resultado, se construyó la primera versión de la metodología, integrando tanto los elementos iniciales de referencia propuestos por los investigadores, como las contribuciones de los estudiantes. En la Figura 5 se presentan las evidencias de la sesión.

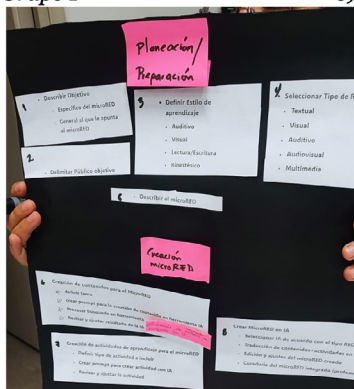
Figura 5

Primera versión de la Metodología de cocreación de MicroRED -Sesión 1



a) Avance metodología Grupo 1

b) Avance metodología Grupo 2



c) Avance metodología Grupo 3

Fuente: Datos del estudio.

En la sesión 2 se reunieron los estudiantes del semillero de investigación y el equipo investigador para revisar los aportes de la Sesión 1, comparar propuestas y construir de forma conjunta una versión unificada de la metodología para la creación de microRED. La sesión siguió tres actividades: socialización de propuestas, análisis comparativo y construcción colaborativa de la versión unificada (Tabla 4). En la

Tabla 5 se presentan los cambios realizados sobre las versiones de la metodología, evidenciando la trazabilidad de aportes y la gestión de cambios.

Tabla 4
Actividades de la Sesión 2

N.º	Actividad	Descripción
1	Socialización de propuestas	Cada grupo presentó su versión preliminar de la metodología, destacando los pasos, herramientas de IAGen, roles y observaciones generales. La presentación se soportó en el artefacto tipo esquema/mapa lo que facilitó la comprensión colectiva.
2	Análisis comparativo	Se identificaron similitudes, diferencias, fortalezas y aspectos no relevantes entre las propuestas mediante una matriz de análisis que registró los elementos comunes y divergentes, así como argumentos a favor o en contra de algunas decisiones. El equipo investigador moderó la discusión, promoviendo la argumentación y reflexión.
3	Construcción colaborativa	A partir del análisis comparativo se generó una versión unificada de la metodología (V1.0). Esta versión se trabajó en un documento compartido que fue intervenido por los participantes, mientras se discutían y consensuaban los componentes (fases, actividades, recursos, roles y observaciones). Esta versión integró los aportes relevantes de cada grupo fortaleciendo la coherencia general de la metodología.

Fuente: Datos del estudio.

Tabla 5
Trazabilidad de cambios de la metodología

Elementos	Preliminar Vo.1	V1.0
Fecha	Septiembre de 2024 a marzo de 2025	Abril a mayo de 2025
Etapas	Se definen tres etapas: E1 Contexto. E2 Construcción MicroRED E3 Validación.	• Se mantienen las tres etapas. • Se precisa la denominación de las E1 y E3: E1 Definición del contexto (Planeación / preparación). E3 Validación / Evaluación.
Subetapas	Se definen subetapas para las etapas E2 y E3. E2 Construcción MicroRED E2.SE1 Creación de contenidos. E2.SE2 Creación de actividades de aprendizaje. E2.SE3 Crear microRED en IA. E3 Validación E3.SE1 Pruebas. E3.SE2 Usabilidad.	• Se mantienen las subetapas para las etapas E2 y E3. • Se ajusta la denominación de las subetapas E2.SE3, E3.SE1 y E3.SE2. E2. Construcción MicroRED E2.SE3 Creación microRED en IA. E3. Validación / Evaluación E3.SE1 Desarrollo de pruebas. E3.SE2 Evaluación de usabilidad del microRED.

Elementos	Preliminar Vo.1	V1.0
	Se definen actividades para las etapas E1, E2 y E3.	• Se agrega una actividad a la E1. • Se ajusta la denominación de 15 actividades de todas las etapas.
Actividades	E1 Contexto E1.A1 Público objetivo. E1.A2 Descripción. E1.A3 Objetivo (Específico del microRED y General al que le apunta el microRED). E1.A4 Estilo de aprendizaje (Auditivo, Visual, Lectura/Escritura, Kinestésico). E1.A5 Tipo de RED (Textual, Visual, Auditivo, Audiovisual y Multimedia).	E1 Contexto E1.A1 Identificar un problema de aprendizaje. E1.A2 Describir el Objetivo General de aprendizaje del RED. E1.A3 Describir el Objetivo Específico del microRED. E1.A4 Definir / Delimitar el público objetivo E1.A5 Definir el Estilo de aprendizaje (Auditivo, Visual, Lectura/Escritura, Kinestésico). E1.A6 Seleccionar el Tipo de microRED (Textual, Visual, Auditivo, Audiovisual y Multimedia).
	E2 Construcción MicroRED E2.SE1 Creación de contenidos E2.SE1.A1 Definir tema. E2.SE1.A2 Crear prompt para la creación de contenido con IA. E2.SE1.A3 Procesar búsqueda en herramienta. E2.SE1.A4 Revisar y ajustar el resultado de la IA. E2.SE2 Creación de actividades de aprendizaje E2.SE2.A1 Definir tipo de actividad a incluir. E2.SE2.A2 Crear prompt para actividad con IA. E2.SE2.A3 Revisar y ajustar la actividad. E2.SE3 Creación microRED en IA. E2.SE3.A1 Seleccionar IA de acuerdo con el tipo RED. E2.SE3.A2 Traducir contenidos + actividades en IA. E2.SE3.A3 Editar y ajustar el microRED creado. E2.SE3.A4 Curaduría del microRED integrado (profesor).	E2 Construcción MicroRED E2.SE1 Creación de contenidos E2.SE1.A3 Procesar búsqueda en herramienta IA (ajustar y/o refinar prompt). E2.SE3 Creación microRED en IA. E2.SE3.A1 Probar herramientas de IA de acuerdo con el tipo de RED para seleccionar la más adecuada. E2.SE3.A2 Realizar edición y ajustes del microRED generado por IA. E2.SE3.A4 Curaduría del microRED integrado por parte del profesor.
	E3 Validación. E3.SE1 Pruebas. E3.A1 Pruebas E3.SE2 Usabilidad. E3.A2 Validación	E3 Validación. E3.SE1 Desarrollo de pruebas. E3.SE1.A1 Realizar pruebas de visualización en dispositivos. E3.SE1.A2 Realizar pruebas de portabilidad y acceso por medios de difusión. E3.SE1.A3 Realizar pruebas de estilo y textos. E3.SE2 Evaluación de usabilidad del microRED. E3.SE2.A1 Revisar usabilidad con usuarios. E3.SE2.A2 Realizar feedback y ajustes sobre el microRED.

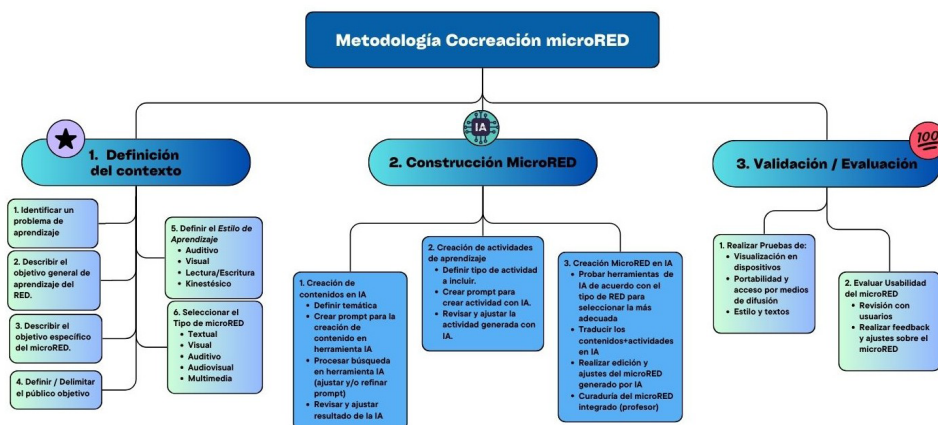
Fuente: Datos del estudio.

En la Sesión 3 el equipo investigador estructuró y representó gráficamente la versión unificada de la metodología para la cocreación de microRED, con el propósito de traducir los acuerdos de las Sesiones 1 y 2 en un artefacto visual claro, funcional y accesible que facilitara la implementación y apropiación por parte de diferentes públicos.

Partiendo de la versión integrada de la metodología construida en la Sesión 2 – que incluía fases, actividades, roles y condiciones de implementación– se exploraron modelos de visualización gráfica para su representación final. Posteriormente, se realizó una validación interna sobre la claridad, coherencia y fidelidad del diseño en la metodología. Como resultado, se generó una versión gráfica de la metodología, lista para su validación en escenarios de aplicación (ver Figura 6).

Figura 6

Representación gráfica de la metodología propuesta



Fuente: Elaboración propia.

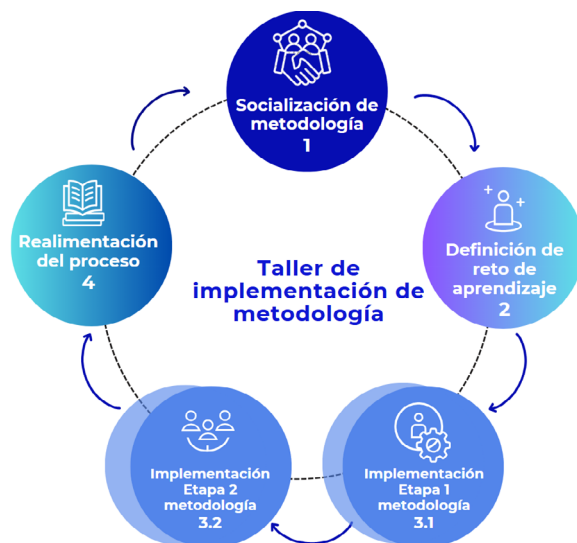
Momento 2. Implementación de la metodología

Este momento consistió en la aplicación de la metodología diseñada en un ejercicio con un grupo piloto conformado por: el equipo investigador, un profesor y un grupo de estudiantes del Semillero de Investigación participante. El objetivo de este momento fue evaluar su relevancia, aplicabilidad y coherencia en la creación de microRED mediante IAGen. Se desarrolló un taller tipo *bootcamp* (ver Figura 7), cuyos productos creados fueron evaluados por los participantes, permitiendo ajustar el proceso y establecer criterios de mejora. La sesión siguió cuatro actividades: socialización de la metodología, definición del contexto de aprendizaje,

implementación de la metodología y realimentación del proceso (ver detalle en Tabla 6). Finalmente, en la Figura 8 se muestra la evidencia de la planeación de los microRED piloto, mientras que en la Figura 9 y Figura 10 se presentan las evidencias de los microRED contruidos.

Figura 7

Estructura del taller de implementación de la metodología



Fuente: Elaboración propia.

Tabla 6

Actividades de la Sesión 1

N.º	Actividad	Descripción
1.	Socialización de la metodología	El equipo investigador presentó el proyecto que sustenta la metodología propuesta y explicó las actividades de las tres fases definidas.
2.	Definición del contexto de aprendizaje	Se eligió el ‘diseño de videojuegos’ como reto, aprovechando los conocimientos previos de los participantes del semillero de investigación para facilitar la definición temática y la revisión de contenidos y actividades asociadas.

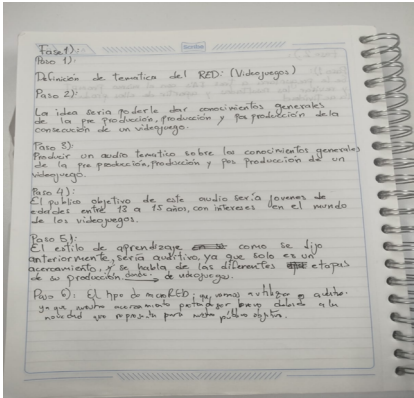
N.º	Actividad	Descripción
3.	Implementación de la metodología	Etapas 1. Se organizaron grupos de trabajo entre los estudiantes, quienes recibieron una guía impresa y la representación gráfica de la metodología (Figura 8.a), como ruta de trabajo para la implementación de cada actividad. Cada grupo ejecutó las actividades de manera autónoma (algunas evidencias de estas actividades se presentan en la Figura 8.b). Etapas 2. Cada grupo de trabajo construyó su microRED utilizando herramientas de IAGen (ver evidencias de esta etapa en las Figuras 9 y 10).
4.	Realimentación del proceso	Los grupos de trabajo presentaron la versión inicial de sus microRED, describiendo el proceso de creación y las herramientas de IAGen utilizadas. La audiencia aportó observaciones y sugerencias tanto al proceso y a la aplicación, como al producto diseñado, generando reflexiones y buenas prácticas para próximas validaciones de la metodología.

Fuente: Datos del estudio.

Figura 8
Evidencias planeación de un microRED – Etapa1

PLANEACIÓN

1. Problema de aprendizaje: diseño de videojuegos.
2. Objetivo general: Identificar qué género de videojuego es apto para hacer realidad una idea
3. Objetivo específico:
 - Diseñar un test que ayude a las personas que planean el desarrollo de un videojuego a identificar cuáles son el género y tipo más acertados para su proyecto
 - Facilitar a las personas en la toma de decisiones sobre el género y tipo de videojuego que se quiere desarrollar
 - Dar a conocer los tipos de videojuegos que existen en la actualidad y cuáles son sus diferencias
4. Público objetivo: desarrolladores y diseñadores de videojuegos principiantes de 16 a 22 años de edad
5. Estilo de aprendizaje: Escritura/lectura
6. Tipo de microRED: Multimedia (texto e imagen)



a) Evidencia Fase planeación Grupo1

b) Evidencia Fase planeación Grupo2

Fuente: Datos del estudio.

Figura 9

Evidencias construcción de microRED – Etapa2

Fase 2: Construcción de MicroRED

1. Creación de Contenidos

*Definir el tema: Conocimientos generales de la preproducción, producción y posproducción de los videojuegos.

El prompt: “*Para la generación de una actividad MicroRED quiero explicar de manera breve e introductoria conocimientos básicos de la preproducción, producción y post producción de los videojuegos, para adolescentes de entre 13 y 15 años. Podrías generar un texto sobre esto*”.

*Procesar Búsqueda en Herramientas de IA: Para el ejercicio pusimos el mismo prompt en tres IA diferentes para comparar sus resultados.

*Generar resultados:

Chat GPT|

¿Cómo se crea un videojuego?

Conoce las etapas básicas: Preproducción, Producción y Postproducción

Crear un videojuego es como hacer una película o montar una obra de teatro: necesita planificación, trabajo en equipo y mucha creatividad. Aquí te contamos las tres etapas principales que hacen posible que un juego llegue a nuestras manos:

a) Evidencia tareas 1, 2 y 3 de actividad 1 –Etapa2

ChatGPT ▾

Para la generación de una actividad MicroRED quiero explicar de manera breve e introductoria conocimientos básicos de la preproducción, producción y post producción de los videojuegos, para adolescentes de entre 13 y 15 años. Podrías generar un texto sobre esto

Claro, aquí tienes un texto introductorio y breve sobre las etapas de **preproducción, producción y postproducción de videojuegos**, pensado para adolescentes entre 13 y 15 años, en el contexto de una actividad MicroRED:

¿Cómo se crea un videojuego?

Descubre las etapas de su desarrollo

Crear un videojuego no es solo sentarse a programar. Es un proceso creativo y técnico que se divide en tres grandes etapas: **preproducción, producción y postproducción**. Aquí te contamos en qué consiste cada una:

1. Preproducción: ¡Las ideas cobran forma!

Esta es la etapa en la que se **planifica** todo. El equipo decide de qué tratará el juego, quiénes serán los personajes, cómo será el mundo y qué tipo de mecánicas se usarán (por ejemplo, ¿el personaje salta? ¿dispara? ¿explora?).

También se crean **dibujos, guiones y prototipos** para imaginar cómo se verá y funcionará el juego. Es como cuando haces un boceto antes de una gran obra. ↓

b) Evidencia y resultados tareas 3 y 4 de actividad 1 -Etapa2

Fuente: Datos del estudio.

Figura 10

Ejemplo microRED creado – Etapa2



Fuente: Datos del estudio.

Para finalizar esta etapa, se discutieron los roles de los actores, las rutas de acción, los formatos de curaduría y el grado de intervención humana necesario en la revisión de los resultados de cada IAGen en cada fase propuesta en la metodología. Este análisis permitió ajustar y generar acuerdos sobre la operatividad del proceso. Los aportes fueron sistematizados en una nueva versión que se planea evaluar con grupos piloto más amplios.

DISCUSIÓN Y CONCLUSIONES

Este estudio presenta una metodología para la cocreación de microRED apoyada por herramientas de IAGen. El proceso se sustenta en las metodologías de IAP y SCPS, mediante estrategias colaborativas como sesiones de cocreación, *bootcamp* y diálogo de saberes con los actores de las comunidades involucradas y orientada hacia la solución de problemas y el logro de resultados de aprendizaje. Estas dinámicas favorecen la exploración temática, la interacción entre pares y el desarrollo de prácticas distintas a las propuestas curriculares tradicionales (Ortiz Jaramillo, 2022).

En estos espacios, los estudiantes experimentan y descubren nuevas posibilidades de lo que pueden llegar a hacer a partir de sus intereses, despertando curiosidad por temáticas y habilidades no conocidas.

El enfoque de IAP es el núcleo de la metodología propuesta, sirviendo como base de la apropiación social, en contextos de aprendizaje participativo. Esta metodología promueve el uso pedagógico de IAGen como herramienta para rediseñar los entornos de enseñanza-aprendizaje y apoyar los procesos complejos de producción de RED. En línea con propuestas como la de Gimenez y Siqueira (2024), se plantea un framework para fragmentar, organizar, secuenciar y adaptar contenidos educativos, con el objetivo de enriquecer la experiencia de aprendizaje de los estudiantes.

La metodología presentada se estructura en tres fases interrelacionadas. En la Fase 1, el trabajo colaborativo entre estudiantes permitió caracterizar el contexto escolar y recopilar información significativa desde la perspectiva de los actores educativos. Este proceso aportó insumos para delimitar temáticamente la investigación y fortaleció la interacción de los participantes, en consonancia con Fals Borda (2001) y Arnanz et al. (2023). La Fase 2 integró las capacidades técnicas y reflexivas de los estudiantes universitarios y de educación secundaria, quienes exploraron herramientas y recursos en el aula. Los productos desarrollados por los estudiantes se articularon como espacios de aplicación, donde se diseñaron y evaluaron propuestas de solución concretas a partir de microRED. Finalmente, en la Fase 3 se logró una validación inicial de la propuesta metodológica, incorporando criterios de relevancia, aplicabilidad y coherencia. Esta validación constituye un insumo clave para la siguiente etapa del proyecto, en la que se enmarca el trabajo presentado en este estudio. Esta etapa es orientada hacia el diseño y la publicación de los microRED en entornos virtuales de acceso abierto, lo que permitirá evaluar su impacto en el desarrollo del pensamiento computacional en jóvenes estudiantes. Se comprobó, además, que la metodología es operativa en contextos reales de aula y genera aprendizajes significativos en educación secundaria y superior.

La metodología propuesta, acorde con lo planteado por Leal-Urueña (2020), contempla actividades, actores, procedimientos de evaluación y realimentación que garantizan el refinamiento del resultado final, es decir, los microRED. Esta propuesta se distingue por su estructura formal e integradora, especialmente en sus fases inicial y final, frente a otras como la de Zhang y West (2020), que definen un flujo de eventos instruccionales organizados en un tópico clave para crear experiencias de aprendizaje, o como el proceso de curaduría y diseño de elementos de *microlearning* basado en conjuntos de conceptos conectados intencionalmente (Kohler et al., 2021).

Asimismo, la metodología se enriquece respecto a modelos previos como el marco de diseño instruccional propuesto por Monib et al. (2025), al incorporar una fase central de creación del microRED con apoyo de IAGen, además de los tres componentes asociados con las características individuales, situacionales y explícitas de los beneficiarios del recurso propuesta por Monib.

La sistematización de las contribuciones generadas en cada fase mencionada sustenta la viabilidad de la propuesta y guía su proyección hacia nuevas acciones, como la implementación con grupos piloto y la publicación de microRED en plataformas de acceso abierto. Se toma como referencia plataformas como 7taps.com y SafetyCulture.com para crear cursos de *microlearning online* pensados para su consumo en pantallas conectadas de *smartphones* y distribución por redes sociales. En este marco, se consolida una metodología con potencial de transferencia y escalabilidad, articulada con una estrategia para la apropiación social mediada por IAGen. La metodología propuesta permite a los estudiantes asumir un rol protagónico en su aprendizaje y fortalecer competencias vinculadas al pensamiento computacional y a la interacción social.

La implementación de talleres de cocreación con la comunidad educativa, la participación del semillero de investigación y el desarrollo de proyectos de aula demuestran tanto la viabilidad operativa como la relevancia formativa de los enfoques utilizados para proponer la metodología de cocreación de microRED en diversos entornos educativos.

El incremento de la investigación sobre microaprendizaje y el uso de IAGen en educación evidencia su potencial transformador e innovador en los procesos y prácticas pedagógicas (Gimenez y Siqueira, 2024). Sin embargo, la adaptación o rediseño de dichas prácticas requiere abordar desafíos asociados con la interacción social, la colaboración entre actores y los aspectos intrínsecos al diseño de RED, como la parametrización, fragmentación y fiabilidad. Incorporar componentes innovadores en las prácticas pedagógicas implica asegurar la integración de nuevas herramientas, metodologías y enfoques que faciliten el logro de resultados de aprendizaje mediante la resolución de problemas de contextos educativos (Villarreal et al., 2022). En este sentido, y como afirma Stanford et al. (2016), una práctica innovadora debe comprender un conjunto de acciones que contribuyan al logro de objetivos, integrar elementos novedosos que la diferencien de prácticas tradicionales, y promover la interacción de los actores y el aprovechamiento de recursos existentes.

Los hallazgos de esta investigación tienen implicaciones prácticas y teóricas que aportan reflexiones sobre el valor del microaprendizaje, el rol de la IAGen en la educación y la transformación de los roles de profesores y estudiantes en los nuevos entornos de enseñanza-aprendizaje. El estudio refuerza la utilidad del microaprendizaje como un enfoque instruccional centrado en ofrecer contenido breve y específico a estudiantes diversos en tiempos cortos, alineada con la teoría de carga cognitiva (CLT, por sus siglas en inglés) que sustenta que, minimizando dicha carga, el aprendizaje se potencia (Sweller, 2011). De este modo, la metodología se proyecta como una herramienta que no solo organiza el proceso de cocreación de microRED, sino que también contribuye a repensar los entornos de aprendizaje desde la colaboración y el uso responsable de la IAGen.

Para superar los desafíos identificados, se plantea que la metodología de cocreación de microRED sea un insumo para orientar a educadores, estudiantes, diseñadores instruccionales y curriculares y creadores de contenidos en la generación

de microRED y otros recursos relacionados, aprovechando las oportunidades y potencialidades que ofrecen los avances de la Inteligencia Artificial en la educación. En cuanto a su implementación en entornos virtuales como plataformas *Learning Management Systems* (LMS) y sistemas de *Massive Open Online Courses* (MOOC), se recomienda integrar el uso de herramientas colaborativas que faciliten la interacción, observación y sistematización de aportes de los actores involucrados en cada etapa del proceso.

Los hallazgos evidencian que la integración de la IAGen no sustituye los procesos pedagógicos, sino que los complementa al potenciar la creatividad, la exploración temática y la construcción colaborativa de conocimiento, como aspectos claves para el fortalecimiento del pensamiento computacional. Asimismo, su implementación impulsa escenarios de aprendizaje dinámicos y adaptativos, en los que los estudiantes desarrollan competencias transversales como la resolución de problemas, el análisis crítico y la innovación en diversos contextos educativos. De esta manera, la IAGen se convierte en un recurso estratégico que favorece la inclusión, la participación y la generación de experiencias formativas más significativas y sostenibles en el tiempo.

La metodología que se plantea muestra que el microaprendizaje, cuando se articula con enfoques participativos como la IAP y la SCPS, adquiere un carácter innovador y transformador que facilita tanto la apropiación social como la creación de recursos educativos transferibles a distintos contextos escolares; esta articulación fomenta procesos de cocreación que enriquecen la interacción entre actores, fortalecen la capacidad de adaptación pedagógica y promueven una visión interdisciplinaria de la enseñanza. En este sentido, el microaprendizaje no solo amplía las posibilidades de mediación tecnológica, sino que también contribuye a consolidar prácticas educativas más inclusivas, flexibles y orientadas al desarrollo de competencias para afrontar los retos contemporáneos en educación.

Como trabajo futuro se propone validar la metodología en otras instituciones de educación secundaria y superior, tanto en entornos presenciales como virtuales, con el fin de identificar ajustes y mejoras que fortalezcan su replicabilidad. En razón a que en esta investigación la metodología fue aplicada únicamente en dos instituciones de educación básica secundaria y superior, se limita la generalización de los resultados a otros niveles (primaria o técnica) o modalidades educativas (virtual o a distancia). Además, el estudio se centró en validar el diseño y construcción de los microRED, sin evaluar longitudinalmente su efectividad pedagógica. Se recomienda ampliar la aplicación de la metodología en otros contextos y niveles educativos, así como analizar las implicaciones éticas del uso de la IAGen en entornos vulnerables o con acceso limitado a tecnología.

Conflicto de intereses

No se declara ningún potencial conflicto de intereses con respecto a la investigación, autoría y/o publicación de este artículo.

Agradecimientos

Este artículo es parte de los resultados preliminares del Proyecto de Investigación 1375-2024 “Metodología de cocreación de recursos digitales de microaprendizaje activo para la enseñanza de pensamiento computacional y la interacción social por jóvenes estudiantes de informática” financiado por la Universidad de Medellín y el Instituto Tecnológico Metropolitano (ITM) de Medellín. Nos gustaría agradecer a la Agencia de Traducción ITM (traducciones@itm.edu.co) por la versión del manuscrito en inglés.

REFERENCIAS

- Abdous, M. (2023, 22 de marzo). How AI is shaping the future of higher ed. *Inside Higher Ed*. <https://www.insidehighered.com/views/2023/03/22/how-ai-shaping-future-higher-ed-opinion>
- Adeshola, I. y Adepoju, A. P. (2023). The opportunities and challenges of ChatGPT in education. *Interactive Learning Environments*, 32(10), 6159-6172. <https://doi.org/10.1080/10494820.2023.2253858>
- Arnanz Monreal, L., García Montes, N. y Villasante Prieto, T. R. (2023). La investigación y acción participativa y transformadora basada en los grupos motores. *Prisma Social*, (43), 34-56. <https://revistaprismasocial.es/article/view/5176>
- Boumalek, K., El Mezouary, A., Hmedna, B. y Bakki, A. (2024). Transforming microlearning with generative AI: Current advances and future challenges. En M. Lahby, Y. Maleh, A. Bucchiarone y S. E. Schaeffer (Eds.), *General aspects of applying generative AI in higher education*. Springer Nature. https://doi.org/10.1007/978-3-031-65691-0_13
- Bozkurt, A. (2023). Generative artificial intelligence (AI) powered conversational educational agents: The inevitable paradigm shift. *Asian Journal of Distance Education*, 18(1). <https://www.asianjde.com/ojs/index.php/AsianJDE/article/view/718>
- Bygstad, B., Øvrelid, E., Ludvigsen, S. y Dæhlen, M. (2022). From dual digitalization to digital learning space: Exploring the digital transformation of higher education. *Computers & Education*, 182, 104463. <https://doi.org/10.1016/j.compedu.2022.104463>
- Cabeza-Rodríguez, M.-Á. (2025). Asistentes ChatGPT en educación superior en línea y satisfacción del alumnado: Un caso de estudio. *RIED-Revista Iberoamericana de Educación a Distancia*, 28(2). <https://doi.org/10.5944/RIED.28.2.43552>
- Cain, W. (2024). Prompting change: Exploring prompt engineering in large language model AI and its potential to transform education. *TechTrends*, 68(1), 47-57. <https://doi.org/10.1007/s11528-023-00896-0>
- Congreso de la República de Colombia. (2012, 17 de octubre). *Ley 1581 de 2012, por la cual se dictan disposiciones generales para la protección de datos personales* [Ley estatutaria; Diario Oficial 48.587, 18 de octubre de 2012]. <https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=49981>
- Denojean-Mairet, M., López-Pernas, S., Agbo, F. J. y Tedre, M. (2024). A literature review on the integration of microlearning and social media. *Smart*

- Learning Environments*, 11, 46. <https://doi.org/10.1186/s40561-024-00334-5>
- Díaz Redondo, R. P., Caeiro Rodríguez, M., López Escobar, J. J. y Fernández Vilas, A. (2021). Integrating microlearning content in traditional e-learning platforms. *Multimedia Tools and Applications*, 80(2), 3121-3151. <https://doi.org/10.1007/s11042-020-09523-z>
- Dickey, E. y Bejarano, A. (2023). *A model for integrating generative AI into course content development* [Preprint]. arXiv. <https://arxiv.org/abs/2308.12276>
- Dolasinski, M. y Reynolds, J. (2020). Microlearning: A new learning model. *Journal of Hospitality and Tourism Research*, 44(3), 551-561. <https://doi.org/10.1177/10963480209015>
- Durán Alcalá, M. y Escudero Nahón, A. (2023). Microlearning en el entorno educativo. *Revista de Investigación Educativa de la REDIECH*, 14, e1763. https://doi.org/10.33010/ie_rie_rediech.v14i0.1763
- Fals Borda, O. (2001). Transformaciones del conocimiento social aplicado: Lo que va de Cartagena a Ballarat. *Análisis Político*, 42, 93-101. <https://revistas.unal.edu.co/index.php/anpol/article/view/75309>
- Gimenez, P. y Siqueira, S. (2024). *Microlearning e IA generativa na educação online: Fragmentação, sequenciamento e adaptação de conteúdos para aprendizagem centrada no aluno*. En *Anais Estendidos do XIII Congresso Brasileiro de Informática na Educação* (pp. 314-319). SBC – Sociedade Brasileira de Computação. https://sol.sbc.org.br/index.php/cbie_estendido/article/view/31804
- Gonnet, J. P. (2019). ¿Por qué la interacción? Una reconstrucción de los escritos tempranos de Erving Goffman. *Revista Reflexiones*, 99(1), 0-15. <https://doi.org/10.15517/rr.v99i1.35308>
- Holmes, W., Bialik, M. y Fadel, C. (2023). *Artificial intelligence in education*. Globethics Publications. <https://doi.org/10.58863/20.500.12424/4276068>
- Hug, T., Lindner, M. y Bruck, P. A. (Eds.). (2006). *Micromedia & e-learning 2.0: Gaining the big picture* [Actas de la Microlearning Conference 2006]. Innsbruck University Press. <http://library.oapen.org/handle/20.500.12657/39646>
- Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J. y Aila, T. (2020). Analyzing and improving the image quality of StyleGAN. En *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 8107-8116). Computer Vision Foundation/IEEE. <https://doi.org/10.1109/CVPR42600.2020.00813>
- Kemmis, S. y McTaggart, R. (2005). Participatory action research: Communicative action and the public sphere. En N. K. Denzin y Y. S. Lincoln (Eds.), *The SAGE handbook of qualitative research* (3ª ed., pp. 559-603). SAGE.
- Kohler, M., Gamrat, C., Raish, V. y Gross, E. (2021). Microlearning and micro-credentials in higher education. En J. R. Corbeil, B. H. Khan y M. E. Corbeil (Eds.), *Microlearning in the digital age: The design and delivery of learning in snippets* (pp. 109-128). Routledge. <https://doi.org/10.4324/9780367821623-10>
- Leal-Urueña, L. A. (2020). *Definición de las especificaciones de diseño*. CIDET Moodle (Universidad Pedagógica Nacional). <https://cidetmoodle.pedagogica.edu.co/mod/page/view.php?id=94698>
- Luckin, R., Holmes, W., Griffiths, M. y Forcier, L. B. (2016). *Intelligence unleashed: An argument for AI in education*. Pearson. <https://edu.google.com/pdfs/Intelligence-Unleashed-Publication.pdf>
- Manzini, E. (2015). *Design, when everybody designs: An introduction to design for social innovation*. MIT Press. <https://doi.org/10.7551/mitpress/9873.001.0001>

- Mateus-Nieves, E. y Moreno Moreno, E. (2021). Microlearning como estrategia para una educación asincrónica. *Épsilon: Revista de la Sociedad Andaluza de Educación Matemática "THALES"*, (109), 41-58. https://thales.cica.es/epsilon/sites/default/files/2023-02/epsilon109_3.pdf
- Memon, T. D. y Kwan, P. (2025). A collaborative model for integrating teacher and GenAI into future education. *TechTrends*. <https://doi.org/10.1007/s11528-025-01105-w>
- Miao, F. y Holmes, W. (2024, 4 de abril). *Guía para el uso de IA generativa en educación e investigación*. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000389227>
- Mirón, L. (2022). La Escuela Creativa de Innovación Social. En *Jornadas Hacia una Nueva Cultura Científica* (pp. 25-34). Editorial Universitat Politècnica de València. <https://doi.org/10.4995/NCC2022.2022.15773>
- Monib, W. K., Qazi, A. y Apong, R. A. (2025). Microlearning beyond boundaries: A systematic review and a novel framework for improving learning outcomes. *Heliyon*, 11(2), e41413. <https://doi.org/10.1016/j.heliyon.2024.e41413>
- Ortiz Jaramillo, I. (2022). *Diseño de bootcamp de programación y pensamiento computacional con enfoque transdisciplinar* [Tesis de maestría, Universidad de los Andes]. Repositorio Universidad de los Andes. <https://hdl.handle.net/1992/64050>
- Piantari, E., Al Husaeni, D. N., Anisyah, A. y Erlangga, E. (2024). Analysis of publication trends on artificial intelligence in education: A systematic literature review. En *Proceedings of the 9th Mathematics, Science, and Computer Science Education International Seminar (MSCEIS 2023)* (pp. 430-443). Atlantis Press. <https://www.atlantis-press.com/proceedings/msceis-23/126003203>
- Presidente de la República de Colombia. (2013, 27 de junio). *Decreto 1377 de 2013, por el cual se reglamenta parcialmente la Ley 1581 de 2012* [Diario Oficial 48.834, 27 de junio de 2013]. <https://www.alcaldiabogota.gov.co/sisjur/normas/Norma1.jsp?i=53646>
- Ritzer, G. (2008). *Teoría sociológica moderna* (Cuarta edición). McGraw-Hill.
- Samoili, S., López Cobo, M., Delipetrev, B., Martínez-Plumed, F., Gómez Gutiérrez, E. y De Prato, G. (2021). *AI Watch. Defining artificial intelligence 2.0*. Publications Office of the European Union. <https://doi.org/10.2760/019901>
- Sengar, S. S., Hasan, A. B., Kumar, S. y Carroll, F. (2025). Generative artificial intelligence: A systematic review and applications. *Multimedia Tools and Applications*, 84, 23661-23700. <https://doi.org/10.1007/s11042-024-20016-1>
- Sirvent, M. T. y Rigal, L. A. (2012). *Investigación acción participativa: Un desafío de nuestros tiempos para la construcción de una sociedad democrática*. Proyecto Páramo Andino. <http://hdl.handle.net/11336/187628>
- Stanford, C., Cole, R., Froyd, J., Friedrichsen, D., Khatri, R. y Henderson, C. (2016). Supporting sustained adoption of education innovations: The Designing for Sustained Adoption Assessment Instrument. *International Journal of STEM Education*, 3, Artículo 1. <https://doi.org/10.1186/s40594-016-0034-3>
- Sweller, J. (2011). Cognitive load theory. En J. P. Mestre y B. H. Ross (Eds.), *Psychology of Learning and Motivation* (Vol. 55, pp. 37-76). Academic Press. <https://doi.org/10.1016/B978-0-12-387691-1.00002-8>
- Torgerson, C. y Iannone, S. (2019). *Designing microlearning* (What Works in Talent Development). ATD Press (Association for Talent Development). <https://www.td.org/product/book--designing-microlearning/111919>

- Tu, Y., Chen, J. y Huang, C. (2025). Empowering personalized learning with generative artificial intelligence: Mechanisms, challenges and pathways. *Frontiers of Digital Education*, 2, 19. <https://doi.org/10.1007/s44366-025-0056-9>
- Villarreal Contreras, R., Salas Álvarez, D. y Alemán, A. (2022). *Prácticas educativas innovadoras desde un enfoque STEAM+A*. Fondo Editorial Universidad de Córdoba. <https://repositorio.unicordoba.edu.co/entities/publication/d29d8a9e-1990-4682-b97d-adb53844525b>
- Yang, J. (2024). Deep integration of generative AI with higher education: Scenarios, risks and recommendations. *China Higher Education*, 5, 52-56.
- Zhang, J. y West, R. E. (2020). Designing microlearning instruction for professional development through a competency-based approach. *TechTrends*, 64(2), 310-318. <https://doi.org/10.1007/s11528-019-00449-4>
- Zhang, Y., Kong, F., Wang, P., Sun, S., Wang, L., Feng, S., Wang, D., Zhang, Y. y Song, K. (2024). *STICKERCONV: Generating multimodal empathetic responses from scratch* [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2402.01679>

Fecha de recepción del artículo: 1 de junio de 2025

Fecha de aceptación del artículo: 2 de septiembre de 2025

Fecha de aprobación para maquetación: 3 de octubre de 2025

Fecha de publicación en OnlineFirst: 24 de octubre de 2025

Fecha de publicación: 1 de enero de 2026

Aprendizaje activo en la educación superior de ingeniería: un decálogo para diseñar quizzes online post-clase

Active learning in higher engineering education: a decalogue for the design of post-lecture online quizzes

-----  -----

 Dubravka Novkovic - *La Salle-Universidad Ramon Llull, URL (España)*
 Ricardo Torres Kompen - *La Salle-Universidad Ramon Llull, URL (España)*
 Josep Petchamé - *La Salle-Universidad Ramon Llull, URL (España)*

RESUMEN

Los *quizzes* son una valiosa herramienta entre las diversas estrategias de aprendizaje activo en la educación superior. En contextos de aprendizaje a distancia se utilizan con frecuencia para evaluar el nivel de conocimientos de los estudiantes, fomentar su participación y apoyar la retención de conocimientos o el rendimiento académico. Existen numerosos estudios sobre los *quizzes* en general, y los *quizzes pre-clase*, existiendo escasa investigación empírica sobre los *quizzes post-clase*. Este artículo analiza las experiencias de estudiantes de Ingeniería en Tecnologías de Información y Comunicación (TIC) con *quizzes online post-clase* semanales en una asignatura de gestión. Los participantes realizaron un *quiz* de un solo intento, calificado automáticamente, en el Sistema de Gestión del Aprendizaje (en inglés, LMS) Moodle de la institución después de cada unidad temática. Completada la experiencia de usuario del alumnado, se llevó a cabo una investigación cualitativa mediante la herramienta *Bipolar Laddering (BLA)*, diseñada para extraer las percepciones positivas y negativas del alumnado sobre el modelo de *quizzes* implementado. En general, nuestro análisis muestra que los estudiantes consideran los *quizzes online post-clase* una valiosa estrategia de aprendizaje activo, especialmente útil para la preparación de exámenes. Los encuestados percibieron que los *quizzes* semanales fomentan tanto el *engagement* como el aprendizaje autónomo. Basándose en las observaciones de los estudiantes, los investigadores formularon diez recomendaciones de diseño, que podrían orientar el desarrollo de los *quizzes online post-clase* eficaces que se pueden implementar fácilmente tanto en contextos de aprendizaje remoto como presencial.

Palabras clave: aprendizaje activo; estrategia de aprendizaje; *quiz* en línea; examen posterior a la clase; ingeniería; *Bipolar Laddering (BLA)*.

ABSTRACT

Quizzes are a valuable addition to the ever-growing toolkit of active learning strategies in higher education. In distance learning contexts, quizzes are frequently used to assess students' level of knowledge, to encourage their participation and to support knowledge retention or academic performance. Despite a wealth of studies about quizzes in general, and pre-lecture quizzes in particular, there is relatively little empirical work examining post-lecture online quizzes. This paper discusses Information and Communication Technology (ICT) Engineering students' experiences with weekly post-lecture online quizzes in a management subject. Participants took a single attempt, automatically graded, quiz on Moodle Learning Management System (LMS) after each subject unit. Once the learning experience was completed, qualitative research was carried out through a *Bipolar Laddering (BLA)* tool, designed to extract positive and negative perceptions of the students on the implemented online quiz model. Overall, our analysis shows that students view post-lecture online quizzes as a valuable active learning strategy, especially useful as a practice for high-stakes exams. Regular low-stakes quizzing was perceived to foster ongoing engagement with the subject material, to support continuous review and self-directed learning. Building on the student-reported insights, the researchers formulated ten design considerations that may inform the development of effective post-lecture online quizzes, which can be easily implemented in both remote and face-to-face learning contexts.

Keywords: active learning; learning strategy; online quiz; post-lecture test; engineering; *Bipolar Laddering (BLA)*.

INTRODUCCIÓN

La educación actual ha experimentado una transformación fundamental gracias a la integración de las Tecnologías de la Información y la Comunicación (TIC), cambiando radicalmente la forma en que se accede, se imparte y se construye el conocimiento. En este contexto, los entornos de aprendizaje en línea, que pueden implementarse en modelos educativos en línea, híbridos y presenciales, han ampliado y fortalecido elementos clave del aprendizaje flexible, como el acceso sin restricciones en términos de tiempo, ubicación y ritmo (Veletsianos et al., 2021). El presente estudio se basa en la teoría del aprendizaje activo, un marco teórico que enfatiza la participación activa del alumnado en la construcción del conocimiento, y tiene por objetivo entender las percepciones de los estudiantes de Ingenierías TIC sobre una de las herramientas de aprendizaje activo, los *quizzes online*, en el contexto de una asignatura de gestión e implementados a través de un Sistema de Gestión del Aprendizaje (*Learning Management System*, LMS). Como primer paso, se presenta una investigación sobre el aprendizaje activo y los *quizzes*. A partir de esta, se desarrolló un trabajo de investigación dirigido a identificar las percepciones de los estudiantes sobre los *quizzes* para comprender la experiencia real del estudiante. Los investigadores utilizaron el método de escalamiento bipolar (*Bipolar Laddering*, BLA), un método cualitativo con formato de cuestionario abierto, para identificar los elementos positivos y negativos que facilitan y dificultan el aprendizaje. A partir del análisis de las respuestas de los estudiantes, este estudio ofrece una guía práctica para la implementación eficaz de *quizzes online post-clase*.

El término «aprendizaje activo» surgió formalmente en la literatura de educación superior hace más de tres décadas como un enfoque pedagógico transformador (Bonwell y Eison, 1991), que transformó el enfoque tradicional centrado en el contenido, pasando de la transmisión pasiva de conocimientos a una instrucción centrada en el estudiante (Hartikainen et al., 2019). En cuanto a la definición de aprendizaje activo, un artículo reciente revisado por pares sobre la literatura de educación en biología y la comunidad docente (Driessen et al., 2020) reveló que la mayor parte de la literatura sobre aprendizaje activo revisada (148 artículos) no conseguía aclarar el verdadero significado del término. En el estudio que nos ocupa, el aprendizaje activo se define como “un enfoque centrado en el estudiante para la construcción de conocimiento, enfocado en actividades y estrategias que fomentan el pensamiento de orden superior” (Doolittle et al., 2023, p. 1).

En cuanto a los beneficios del aprendizaje activo, un metaanálisis fundamental de 225 estudios en las disciplinas de Ciencia, Tecnología, Ingeniería y Matemáticas (STEM) (Freeman et al., 2014) reveló que la sustitución de las clases magistrales tradicionales por aprendizaje activo condujo a puntuaciones significativamente más altas en los exámenes de los participantes y a una menor tasa de suspensos. Si bien el aumento de la participación estudiantil y el rendimiento académico son beneficios consolidados y reconocidos del aprendizaje activo (Prince, 2004), se confirmó que

también promueve la comunicación y la interactividad, fomenta un mayor sentido de comunidad y conexión, mejora la satisfacción del alumnado con la experiencia de aprendizaje y proporciona mayor flexibilidad en el proceso educativo (Allsop et al., 2020).

Por otro lado, en el ámbito de la ingeniería se han reportado respuestas mixtas al aprendizaje activo, según Tharayil et al. (2018). A partir de la observación de aulas y entrevistas a 17 profesores de ingeniería en Estados Unidos, los autores encontraron que los estudiantes eran más propensos a percibir positivamente esas metodologías cuando los instructores explicaban sus beneficios y proporcionaban una explicación experta de las actividades, lo que aumentaba la participación y mejoraba la evaluación del curso. Implementado como una experiencia de laboratorio basada en la indagación en el contexto de prácticas científicas esenciales (Owens et al., 2020), el uso de aprendizaje activo causó una cierta resistencia estudiantil, debido al mayor esfuerzo requerido por todos los participantes. Según los autores, los instructores pueden mitigar el problema mediante prácticas de andamiaje y brindando retroalimentación clara y consistente, lo que ayuda a disminuir la frustración de los estudiantes frente a tareas científicas desconocidas o desafiantes. Curiosamente, en otro estudio STEM en Texas, Andrews et al. (2020) compararon las actitudes y comportamientos de instructores y estudiantes con respecto al aprendizaje activo. La encuesta aplicada a los instructores indagó en sus actitudes hacia el aprendizaje activo, las estrategias utilizadas para reducir la resistencia de los estudiantes y sus percepciones sobre el comportamiento estudiantil. Paralelamente, se aplicó una encuesta estudiantil que pedía valorar los métodos de enseñanza y reflexionar sobre las propias actitudes y conducta en clase. Ambos instrumentos se reforzaron con observaciones directas en el aula. El hallazgo clave fue la tendencia de los instructores a sobrestimar la resistencia de los estudiantes al aprendizaje activo. En ocasiones, los instructores temían reacciones negativas, especialmente cuando su titularidad, promoción y calificaciones de mérito dependían de las valoraciones de los estudiantes al final del semestre. Sin embargo, los datos recogidos mostraron una realidad diferente: los estudiantes apreciaban las actividades de aprendizaje activo, participaban voluntariamente, mantenían la concentración y expresaban la intención de evaluar favorablemente tanto el curso como al instructor.

Muchas estrategias de enseñanza se engloban dentro del concepto de aprendizaje activo. Por ejemplo, la investigación previamente citada de Doolittle et al. (2023) enumera quince enfoques diferentes que aparecieron al menos diez veces en 547 artículos revisados, siendo los más mencionados el aprendizaje basado en problemas, la clase invertida y los debates en grupos pequeños. Otras revisiones bibliográficas también destacan una amplia gama de actividades de enseñanza categorizadas según diferentes aspectos de interés (Arthurs y Kreager, 2017; Kozanitis y Nenciovici, 2023; Mitchell et al., 2017). En la formación en ingeniería, Villalonga Pons et al. (2023) citan estrategias como el aprendizaje basado en problemas, el aprendizaje basado en proyectos, el aprendizaje cooperativo y el aprendizaje práctico, considerados

indicadores clave de calidad para crear experiencias de clase eficaces, especialmente en entornos de aprendizaje en línea. De igual manera, Gosavi y Arora (2022) observaron que las estrategias de aprendizaje activo suelen operar en conjunto, y la implementación de una suele desencadenar el uso de otras.

En relación con los *quizzes*, Mitchell et al. (2017) analizaron su uso como estrategia instruccional en cursos de Sistemas de Información. Una revisión sistemática reciente de evaluaciones en línea en educación superior (Heil e Ifenthaler, 2023) identificó los *quizzes* como un tipo de evaluación automatizada. El término «quiz» surgió alrededor de 1780 y, se consolidó junto con otros como «prueba» o «examen» (Dengri et al., 2021). Estos autores revisaron once estudios que utilizaron *quizzes* como herramienta de enseñanza y aprendizaje, reportando incrementos en las puntuaciones de pruebas posteriores, así como comentarios estudiantiles positivos. Estudios anteriores habían explorado el uso de *quizzes pre-clase* como estrategia para incentivar y, en ocasiones, recompensar la preparación del estudiantado en clase (Brink, 2013; Cook y Babon, 2017; Howell, 2021). Más recientemente, un estudio comparativo en seis facultades de una universidad pública estadounidense destacó que los estudiantes valoraban tanto los *quizzes pre-clase* como *post-clase*, considerándolos útiles para mantener la concentración y favorecer la autoevaluación (Klein et al., 2023). En un programa *online* de educación superior en Australia, Ross et al. (2018) hallaron que el 88,6 % de los participantes coincidió en que los *quizzes* les permitieron identificar áreas de desconocimiento. De manera similar, Raes et al. (2020) mostraron que, al digitalizarse e incorporar mejoras tecnológicas, los *quizzes* influyeron positivamente en la motivación intrínseca del alumnado, siendo percibidos como útiles y atractivos tanto en entornos presenciales como remotos. La dimensión digital también fue central en el estudio de Rossi et al. (2021) en Brasil, donde se implementaron *quizzes* con corrección inmediata en un curso experimental. Aunque no se evaluó su efecto aislado, los autores subrayaron que la retroalimentación inmediata favoreció la consolidación de conceptos clave. En línea con ello, Gowele (2024) recomendó integrar los *quizzes* en los planes de estudio de matemáticas, resaltando el potencial de las plataformas *online* para mejorar participación, rendimiento y seguimiento del progreso estudiantil. De hecho, las plataformas basadas en juegos como Kahoot! han mostrado efectos notables en el aprendizaje derivados de su uso constante (Nuci et al., 2021). Además, Kirzner et al. (2021) administraron *quizzes* de Kahoot! al final de cada clase para repasar contenidos diarios. Sus hallazgos, tanto cuantitativos como cualitativos, confirmaron que el alumnado reconoció la relación entre la gamificación, mejora del aprendizaje y retención de contenidos.

Si bien la evidencia existente respalda ampliamente los beneficios educativos de los *quizzes*, algunos estudios, por el contrario, ofrecen perspectivas o factores críticos que pueden afectar la efectividad de las pruebas. Los críticos señalan, por ejemplo, la ansiedad y el estrés que los *quizzes* generan en algunos estudiantes (Stoyanova y Giannouli, 2022). Esto sugiere que una implementación efectiva puede exigir atención

tanto a la frecuencia de los *quizzes* como a mantener su carácter de actividad de bajo riesgo. No obstante, los autores descubrieron que la ansiedad de los estudiantes disminuyó y que estaban más motivados cuando se les permitió realizar múltiples intentos usando la retroalimentación correctiva. Estas y otras preocupaciones similares se investigaron en un metaanálisis del «efecto de las pruebas» realizado por un equipo de académicos asiáticos (Yang et al., 2021). Sus hallazgos revelan que los beneficios de las pruebas abarcaron 18 disciplinas académicas, apoyando no solo la retención de información, sino también una comprensión más profunda y la aplicación práctica del conocimiento.

Para enmarcar el presente estudio, cabe mencionar que La Salle-Universidad Ramon Llull (URL) ha implementado un modelo educativo denominado Nuevo Contexto de Aprendizaje (NCA) con el objetivo de satisfacer las necesidades del alumnado contemporáneo, priorizando la flexibilidad, la participación activa y la integración de la tecnología (Petchamé et al., 2021). Introducido durante el curso académico 2018-2019, el modelo NCA se fundamenta en cinco principios pedagógicos (Interioridad; Mente, Cuerpo y Movimiento; Construcción del Pensamiento; Conducta Autorregulada; Dimensión Social del Aprendizaje) que se materializan en cinco entornos de aprendizaje (Acogida, Seminario, Taller, Proyecto y Cierre), cada uno diseñado para apoyar las diferentes etapas del proceso de aprendizaje. Por lo tanto, el modelo NCA está plenamente alineado con el uso de metodologías de aprendizaje activo.

El NCA se apoya en una serie de infraestructuras tecnológicas y herramientas de aprendizaje activo implementadas en el Campus La Salle-URL para ofrecer a los estudiantes opciones flexibles de asistencia, permitiéndoles asistir a clases de forma remota (mediante un sistema de videoconferencia) o presencial en el Campus (Petchamé et al., 2023). La piedra angular de este sistema de aula inteligente, conocido como *Smart Classroom*, es un LMS Moodle, recomendado por López-Tocón (2021) como un método eficaz para el aprendizaje de contenido científico.

A diferencia de estudios previos relacionados principalmente con el uso de *quizzes pre-clases* (Evans et al., 2021; Gyllen et al., 2021), el presente trabajo analiza los *quizzes post-clases* como herramienta para la evaluación continua y formativa de estudiantes de ingeniería. Este artículo examina la participación activa de estudiantes de grado de un programa de Ingeniería TIC matriculados en un curso de “Economía Financiera”, distribuido en siete unidades e impartido durante un semestre. Para promover una participación constante y reforzar el contenido de las clases, se asigna a los estudiantes un *quiz* posterior a cada unidad temática, aproximadamente una vez por semana. Los *quizzes* constan de 15 a 20 preguntas de respuesta múltiple que se administran y califican automáticamente a través de un LMS Moodle. Se espera que los estudiantes completen cada *quiz* individualmente (un solo intento por *quiz*) dentro de los cinco días posteriores a la clase, lo que proporciona al instructor tiempo suficiente para revisar los resultados agregados antes de la siguiente sesión de clase. Si bien las calificaciones individuales de los *quizzes* se revisan para fines

de retroalimentación, no se incluyen directamente en la calificación del curso. En cambio, la participación en el *quiz*, definida como la entrega oportuna del *quiz*, sí que cuenta dentro del 10 % de la calificación final. Este enfoque enfatiza el aprendizaje formativo sobre la presión del rendimiento, fomentando hábitos de estudio regulares y verificaciones de comprensión. Finalmente, para apoyar la preparación del examen, todos los *quizzes* se vuelven a abrir cuatro semanas antes del examen final. Moodle reorganiza automáticamente las preguntas en cada nuevo intento de *quiz*. En esa etapa, los estudiantes tienen acceso sin restricciones e intentos ilimitados del *quiz*, lo que les resulta útil como un recurso de autoaprendizaje. El examen final puede presentar hasta dos preguntas adaptadas de los *quizzes* de la unidad dentro de la estructura del examen de 10 preguntas. La evaluación sumativa del curso se lleva a cabo con base en la fórmula ponderada, calculada de la siguiente manera: tres exámenes en papel durante el semestre (50 %); un examen final escrito (40 %); y entrega oportuna del *quiz* (10 %).

El objetivo de este trabajo de investigación es doble: (1) profundizar la comprensión de las percepciones de los estudiantes sobre los *quizzes online post-clases* como un instrumento de aprendizaje activo implementado en una materia relacionada con los negocios de un grado en Ingeniería TIC, y (2) ofrecer orientación práctica a los instructores que buscan integrar *quizzes online post-clases* en su enseñanza en cualquier tipo de entorno de aprendizaje.

METODOLOGÍA

El presente estudio se centró en analizar la experiencia de estudiantes de Ingeniería TIC con *quizzes post-clase*, utilizando la metodología BLA (*Bipolar Laddering*) desarrollada por Pifarré y Tomico (2007), que se ha utilizado en diversas investigaciones a nivel de educación superior (Horna-Saldaña y Canaleta, 2024; Petchamé et al., 2021, 2023, 2024). Según la definición de Pifarré y Tomico (2007), se trata de un enfoque de investigación cualitativo que combina preguntas abiertas con técnicas tradicionales de entrevista en escalera. Siguiendo el principio socrático de la tabla rasa, sin ninguna indicación previa, se les pide a los estudiantes que identifiquen libremente los aspectos positivos y negativos de su experiencia con el *quiz*. Estos elementos bipolares, que representan fortalezas y debilidades, son calificados por los estudiantes en una escala de importancia de 10 puntos (1 = mínimo; 10 = máximo) y justificados mediante preguntas de seguimiento con el “¿Por qué?”. En la fase final, se anima a los estudiantes a proponer mejoras, sugiriendo cómo potenciar los aspectos positivos o abordar los negativos.

Este método permite analizar la experiencia del usuario y minimiza posibles sesgos, ya que las respuestas surgen de las experiencias directas de los estudiantes y no de preguntas sugerentes. Por lo tanto, este enfoque es adecuado para diversos entornos, incluidos los contextos digitales examinados en esta investigación. Cuando se utiliza BLA no se requieren muestras grandes, ya que el objetivo es capturar la

variedad y la jerarquía de las percepciones del usuario, y la saturación suele alcanzarse tras un número limitado de respuestas (Hennink y Kaiser, 2022; Petchamé et al., 2021). El propósito de la muestra es revelar dimensiones cognitivas distintivas para su posterior agrupamiento y análisis, en lugar de construir modelos generalizables. Por lo tanto, más que la representatividad estadística, el enfoque se centra en lograr la saturación cualitativa, el punto en el que las respuestas adicionales ya no introducen nuevos elementos o temas a los constructos emergentes. Por otro lado, dado que las herramientas de aprendizaje activo que analizamos en esta investigación (*quizzes online*) se basan en la tecnología y no están vinculadas a ninguna materia o disciplina específica, las recomendaciones presentadas pueden aplicarse ampliamente en diferentes contextos de educación superior.

Para el presente estudio, los puntos de datos se etiquetaron como se describe en el artículo de Petchamé et al. (2024). Los elementos positivos de la retroalimentación de los estudiantes se etiquetaron como Elementos (E) Positivos (P) que son Comunes (C) —si son repetidos espontáneamente por más de un estudiante— o Únicos (U) —si son mencionados por un solo estudiante—. El esquema de etiquetado correspondiente se ha aplicado a las afirmaciones negativas, en este caso con la letra N (Negativo) utilizada en lugar de la letra P (Positivo). Tanto los elementos positivos como los negativos se escalonaron según el número de Menciones, que representa el número de veces que los estudiantes citaron cada elemento como su percepción de la experiencia del *quiz*. Luego calculamos el promedio para cada elemento, que representa la media simple de las calificaciones de importancia individuales asignadas por los encuestados, donde las calificaciones variaron de 1 (mínima importancia) a 10 (máxima importancia). Finalmente, se calculó la varianza para indicar el nivel de heterogeneidad entre las evaluaciones individuales del mismo elemento.

Dada la brevedad de las respuestas de los estudiantes, no se requirió una herramienta especializada de análisis cualitativo. Por lo tanto, tres investigadores analizaron individualmente todas las respuestas breves recopiladas y posteriormente se reunieron para consensuar cada elemento (Petchamé et al., 2021). Este enfoque incluyó un proceso de triangulación para mejorar la validez de los resultados (Vaismoradi y Snelgrove, 2019).

RESULTADOS

Nuestra encuesta BLA se realizó de forma anónima al final del primer trimestre académico 2024-2025 entre cuatro grupos de estudiantes de segundo año de Ingeniería TIC en La Salle-URL. Antes de la encuesta, todos los participantes completaron su experiencia formativa en la asignatura “Economía Financiera”. El número total de estudiantes matriculados en el curso fue de 209, y 51 de ellos respondieron a la encuesta BLA. Solo dos hombres encuestados tenían entre 23 y 25 años. Además de la edad, la cohorte de estudiantes encuestados compartía otras similitudes, como el nivel de alfabetización digital, el nivel educativo y el nivel

socioeconómico. Los datos demográficos de los encuestados se muestran en la Tabla 1 a continuación.

Tenga en cuenta que la encuesta BLA debe administrarse durante la última sesión de clase para garantizar que los estudiantes hayan completado su aprendizaje antes de brindar retroalimentación. Realizar la encuesta en ese momento permite una evaluación más completa de todo el proceso de aprendizaje y se ajusta a los patrones habituales de asistencia a clase observados en la mayoría de las asignaturas en esa etapa. Si bien el número de respuestas no es una limitación en la metodología BLA, la encuesta podría distribuirse antes del final de la última sesión de clase o podría extenderse el plazo para completarla a fin de obtener más respuestas.

Tabla 1

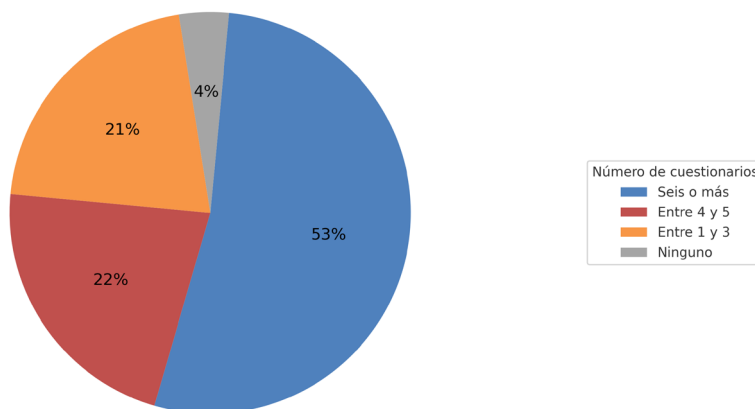
Encuestados de la encuesta de BLA por género y grupo de edad

	Número	%	Edad M	Edad DE
Hombres	38	75	20,2	0,9
Mujeres	11	21	20	0
Prefiere no contestar	2	4	20	0
Total	51	100	20,2	0,8

Como se muestra en la Figura 1 a continuación, más de la mitad de los encuestados (n=27; 53 %) tenían una experiencia considerable en *quizzes post-clase*.

Figura 1

Número de quizzes contestados por los encuestados



A través de la encuesta, los investigadores identificaron ocho Elementos Comunes Positivos (PCE-1 a PCE-8) y cinco Elementos Únicos Positivos (PUE-1 a PUE-5), según se muestra a continuación en la Tabla 2. La tabla muestra los elementos positivos de la experiencia del *quiz* indicados por los encuestados en forma de comentarios. La columna ‘Menciones’ muestra el número de veces que se citó cada elemento y los porcentajes asociados basados en el número total de menciones ($n = 105$). Debido a que cada participante podía indicar hasta tres elementos positivos, los porcentajes de ‘Menciones’ no suman el 100 % de los encuestados. Estos resultados indican que, para los estudiantes, los aspectos más impactantes de la experiencia del *quiz online post-clase* fueron la capacidad de profundizar su comprensión del material del curso y ayudarlos en la preparación del examen (por ejemplo, “porque me parece una forma simple y rápida de prepararme para el examen”). Los participantes también destacaron la retroalimentación inmediata que brindan los *quizzes online*, lo que les permite reconocer rápidamente sus lagunas de conocimiento. PCE-4 y PCE-5 recibieron menos menciones, lo que sugiere que estos elementos podrían jugar un papel de soporte complementario: la realización de exámenes regulares para mantenerse al día con los temas del curso (“Es positivo porque es una forma obligatoria de estar al día con la materia y contribuye a la calificación”), así como la accesibilidad, disponibilidad y variedad de formatos de preguntas para los *quizzes online*. Algunos estudiantes destacaron la naturaleza dinámica y atractiva de los *quizzes*, que hace que el aprendizaje sea más ameno (“es divertido y motivador”). Las menciones de elementos de PUE fueron escasas; cada una representó poco menos del 1 % del total. Se animó a los encuestados a calificar en una escala del 1 al 10 (1=mínimo; 10=máximo) la importancia de cada afirmación, que se muestra a continuación en la columna “Media”, mientras que la “Var” refleja el grado de variación entre esas calificaciones individuales.

Tabla 2

Percepción de los estudiantes sobre elementos que mejoran su aprendizaje

Comentario	Elemento	Mencio- nes n	Mencio- nes %	Media	Var
Los <i>quizzes</i> me ayudan a comprender mejor el material del curso, repasar y reforzar los conceptos básicos de la materia.	PCE-1	28	27	8,6	1,17
Los <i>quizzes</i> son una valiosa herramienta de práctica para la preparación de exámenes.	PCE-2	25	24	8,2	1,92
Los <i>quizzes</i> me ayudan a identificar lagunas en mi comprensión mediante retroalimentación inmediata.	PCE-3	14	13	8,0	2,43

Comentario	Elemento	Mencio- nes n	Mencio- nes %	Media	Var
Los quizzes semanales me ayudan a mantenerme al día con el curso.	PCE-4	9	9	8,4	1,36
Los quizzes son breves y flexibles, con diferentes formatos de preguntas y tiempo suficiente para completarlos.	PCE-5	9	9	8,2	1,95
Los quizzes están organizados por unidad, son de fácil acceso y siempre están abiertos.	PCE-6	7	7	9,0	0,86
Los quizzes tienen un impacto positivo en mi calificación final.	PCE-7	5	4	8,2	1,36
Los quizzes son dinámicos y atractivos, lo que hace que mi aprendizaje sea más ameno y motivador.	PCE-8	3	2	8,0	0,67
Los quizzes incluyen preguntas más allá de las diapositivas de la clase, lo que me ayuda a consolidar los temas recién introducidos.	PUE-1	1	1	10,0	
Los quizzes me animaron a repasar las diapositivas de la clase, centrarme en los puntos clave y recordar el contenido después.	PUE-2	1	1	10,0	
Los quizzes fomentan el aprendizaje activo al aplicar la teoría a situaciones de la vida real.	PUE-3	1	1	10,0	
Los quizzes crean un elemento competitivo que me motiva a sobresalir.	PUE-4	1	1	7,0	
Aleatorizar el orden de las preguntas del quiz ayuda a evitar la memorización de patrones y promueve una comprensión real.	PUE-5	1	1	6,0	

Nota: Los elementos (E) se identifican como Positivos (P), Comunes (C) o Únicos (U)

Por otro lado, 51 estudiantes compartieron un total de 63 elementos que consideraron como retos dentro de su experiencia de aprendizaje. Nuevamente, debido a que cada participante podía reportar hasta tres elementos negativos, los porcentajes de 'Menciones' no suman el número total de encuestados sino el número total de menciones ($n = 63$). Todas las percepciones negativas se agrupan en nueve temas comúnmente reportados (NCE-1 a NCE-9) y cinco temas únicos

(NUE-1 a NUE-5), como se muestra en la Tabla 3. Los estudiantes expresaron varias inquietudes sobre los *quizzes post-clase* y ofrecieron sugerencias que podrían potencialmente aumentar el valor educativo de esta actividad. El punto planteado con mayor frecuencia (n=17; 27 %) se refería a las limitaciones de tiempo para completar el *quiz*, particularmente cuando las preguntas parecían largas, repetitivas o poco claras (n=10; 16 %). Otras declaraciones fueron acompañadas de ideas para mejorar la experiencia de aprendizaje, tales como: permitir múltiples intentos de contestar el *quiz*, y continuar teniendo acceso al mismo después de haberlo finalizado; alinear el contenido del *quiz* más estrechamente con el material del examen; incluir más ejemplos prácticos y preguntas basadas en el razonamiento; proporcionar retroalimentación más detallada, especialmente explicaciones para las respuestas incorrectas (como comentó un estudiante: “No me gusta que solo te diga si tus respuestas son correctas o incorrectas sin darte argumentos que las respalden”). Si bien estas observaciones reflejan las experiencias de algunos estudiantes, pueden ofrecer puntos de partida útiles para la implementación de esta tarea de aprendizaje activo. Otros puntos se mencionaron con menos frecuencia.

Tabla 3

Percepciones de los estudiantes sobre los elementos que obstaculizan su aprendizaje

Comentario	Elemento	Mencio- nes n	Mencio- nes %	Media	Var
Algunos <i>quizzes</i> requieren demasiado tiempo para completarse dentro del margen de tiempo disponible.	NCE-1	17	27	3,5	1,31
Algunas preguntas del <i>quiz</i> son demasiado largas, poco claras o repetidas.	NCE-2	10	16	4,0	0,60
Los <i>quizzes</i> no permiten múltiples intentos ni acceso después de completarlos.	NCE-3	6	9	4,3	0,56
Algunas preguntas del <i>quiz</i> no se ajustan bien a los exámenes.	NCE-4	6	9	4,2	2,14
Los <i>quizzes</i> carecen de ejemplos prácticos y preguntas basadas en el razonamiento.	NCE-5	5	8	3,4	2,64
El número de preguntas del <i>quiz</i> es insuficiente.	NCE-6	4	6	5,0	0,00
No realizar un <i>quiz</i> obligatorio o tener un rendimiento inferior al esperado puede afectar negativamente la calificación final.	NCE-7	4	6	4,3	0,19

Comentario	Elemento	Mencio- nes n	Mencio- nes %	Media	Var
Los <i>quizzes</i> carecen de retroalimentación, como explicaciones de errores.	NCE-8	4	6	3,3	1,69
Algunos gráficos en los <i>quizzes</i> no son visibles.	NCE-9	2	3	1,5	0,25
Los <i>quizzes</i> no incluyen conjuntos de preguntas aleatorias.	NUE-1	1	2	5,0	
La falta de recordatorios de fechas límite por correo electrónico provoca la pérdida de intentos de contestar el <i>quiz</i> .	NUE-2	1	2	5,0	
Los <i>quizzes</i> tienen poco peso en la calificación final.	NUE-3	1	2	4,0	
Algunas preguntas del <i>quiz</i> cubren material que no se abordó durante las sesiones de clase.	NUE-4	1	2	2,0	
Los <i>quizzes</i> fueron solo parcialmente útiles para la retención a largo plazo de los conceptos teóricos.	NUE-5	1	2	1,0	

Nota: Los elementos (E) se identifican como Negativos (N), Comunes (C) o Únicos (U)

Además, los estudiantes propusieron mejoras para cada elemento percibido de su experiencia con los *quizzes*, ofreciendo sugerencias para potenciar los beneficios o mitigar las limitaciones. Por ejemplo, aumentar el número de *quizzes* (p. ej., “Solo hay un *quiz* por unidad, y quizás sería beneficioso añadir uno o dos más para practicar más” o “Añadir más preguntas y que cambien entre un *quiz* y otro”); notificar a los estudiantes sobre el lanzamiento de los *quizzes* (p. ej., “Se debería notificar a los estudiantes por correo electrónico cuando se publique un nuevo *quiz*”); adaptar los *quizzes* para usarlos como práctica (p. ej., “Se podría conceder acceso ilimitado a la revisión del *quiz* inmediatamente después de la fecha límite de entrega” o “Añadir ejercicios a los *quizzes*”).

DISCUSIÓN

Este estudio se ha focalizado en las percepciones de un grupo de estudiantes de ingeniería sobre *quizzes online post-clase* como herramienta de aprendizaje activo, destacando los beneficios percibidos, así como las posibles áreas de mejora. Si bien los hallazgos podrían orientar a los instructores que buscan implementar y

optimizar este enfoque pedagógico, es importante interpretar la retroalimentación de los estudiantes de forma reflexiva y en contexto.

La mayoría de los estudiantes valoraron los *quizzes online post-clase* como una herramienta de estudio valiosa, especialmente para la preparación de exámenes. Esto coincide con investigaciones previas que demuestran que las pruebas regulares pueden promover la retención de conocimientos mediante el “efecto de prueba” (Roediger y Karpicke, 2006). En ese sentido, los *quizzes* pueden servir como herramienta de andamiaje y construcción de conocimiento, proporcionando un soporte estrechamente alineado con los resultados de aprendizaje previstos. La coherencia general del uso de la herramienta podría mejorarse al asegurar una mejor alineación entre los objetivos de las evaluaciones formativas y los criterios de evaluación de las evaluaciones sumativas. Varios estudiantes destacaron la ventaja de la retroalimentación inmediata que ofrecen los *quizzes online*, lo que, en su opinión, les permitió identificar y abordar rápidamente las lagunas de conocimiento, una visión coherente con el modelo de retroalimentación efectiva en entornos de aprendizaje de Hattie y Timperley (2007). Si bien los estudiantes tienden a enfatizar las calificaciones, la literatura sugiere que la retroalimentación en sí misma, y no la calificación, constituye el verdadero mecanismo acelerador del aprendizaje. Las respuestas a la herramienta BLA sugieren además que los *quizzes* semanales podrían animar a los estudiantes a interactuar con mayor regularidad con el contenido del curso, reduciendo potencialmente la tendencia a preparar el contenido de la asignatura solo en fechas próximas al examen. Aunque estas percepciones se alinean con la evidencia que el uso de mecanismos de prácticas, distribuidas de forma regular durante el curso, pueda resultar más efectiva a largo plazo para el proceso de retención y asimilación (Dunlosky et al., 2013), se necesitaría evidencia adicional para confirmar si este beneficio se logró de forma consistente en la práctica. Los comentarios de los estudiantes también destacaron los posibles beneficios motivacionales de la evaluación regular, mientras que algunos participantes señalaron que los *quizzes* eran entretenidos, lo que sugiere que los elementos de gamificación de los *quizzes* podrían contribuir de forma positiva a la mejora de la experiencia del estudiante (Dicheva et al., 2015; Majuri et al., 2018).

Los valores medios asociados a la importancia de los Elementos Comunes Positivos oscilaron entre 8,0 y 9,0, lo que indica una retroalimentación generalmente positiva en los elementos principales. Los valores de varianza oscilaron entre 0,67 para el ítem “Los *quizzes* son dinámicos y atractivos” y 2,43 para “Los *quizzes* me ayudan a identificar lagunas en mi comprensión mediante retroalimentación inmediata”, lo que sugiere que las valoraciones de diversos elementos no se mantuvieron en la misma línea. El ítem “Los *quizzes* están organizados por unidad, son de fácil acceso y siempre están abiertos” recibió la puntuación media más alta, 9,0 (DE = 0,86), lo que indica un fuerte consenso entre los siete estudiantes que lo mencionaron.

Los comentarios de los participantes también revelaron diversas oportunidades de mejora. Las preocupaciones más frecuentes se relacionaron con las limitaciones de

tiempo y el diseño de las preguntas. Los estudiantes consideraron que las preguntas largas, repetitivas o formuladas de forma ambigua obstaculizaban el aprendizaje efectivo. Esto coincide con la investigación de Gillmor et al. (2015), donde se apunta que una construcción pobre de las preguntas puede generar una carga cognitiva innecesaria que interfiere con los resultados de aprendizaje previstos. Los encuestados también expresaron su preferencia por disponer de intentos ilimitados de respuesta a los *quizzes* y acceso inmediato a los resultados, lo que puede reflejar su preferencia por el uso de los *quizzes* como herramientas de aprendizaje y no tanto como instrumentos de evaluación. Esta perspectiva se alinea con el paradigma de evaluación formativa descrito por Black y Wiliam (1998), que enfatizó la evaluación como una forma de guiar el aprendizaje en lugar de como una medida de rendimiento. Otra mejora que comentaron los estudiantes fue la oportunidad de vincular los conceptos teóricos con su aplicación práctica. Los encuestados manifestaron su preferencia por más preguntas basadas en el razonamiento, lo que apunta a objetivos de aprendizaje de orden superior en la taxonomía del aprendizaje de Bloom (Tofade et al., 2013). Esta consideración iría en la línea de minimizar las preguntas de memorización de conocimientos y, en su lugar, priorizar las preguntas de aplicación, análisis y evaluación para facilitar una comprensión más profunda y la retención del conocimiento a largo plazo. Por último, pero no menos importante, los estudiantes comentaron sobre el mecanismo de retroalimentación, destacando en particular la ausencia de explicaciones para las respuestas incorrectas. La importancia de la retroalimentación explicativa sobre los errores fue estudiada por Shute (2008), quien demostró que la retroalimentación elaborada era más eficaz que la retroalimentación de verificación por sí sola. Implementar explicaciones adecuadas tanto para las respuestas correctas como para las incorrectas podría constituir un elemento que podría incrementar el aprendizaje del alumnado a partir de la simple añadidura de explicaciones.

Al observar las puntuaciones medias y las varianzas de los Elementos Comunes Negativos, la afirmación más frecuentemente mencionada sobre el tiempo para completar los *quizzes* ($n=17$; 27 %) recibió un promedio moderado de 3,5 de importancia, lo que sugiere que, si bien a menudo se reconoce, las opiniones al respecto fueron algo mixtas. Por otro lado, la afirmación “El número de preguntas del *quiz* es insuficiente” recibió un promedio máximo de 5,0 ($DE = 0,00$), lo que indica un completo acuerdo entre los cuatro estudiantes que la mencionaron. Otros dos elementos atrajeron nuestra atención debido a la variedad de opiniones que suscitaron: la afirmación sobre la falta de preguntas basadas en el razonamiento ($DE = 2,64$) y el bajo nivel de alineación con los exámenes ($DE = 2,14$); algunos encuestados las consideraron muy importantes, mientras que otros les dieron calificaciones bajas.

En cuanto a las limitaciones de este estudio, su propia naturaleza, centrada en la identificación de elementos positivos y negativos (obtenidos mediante la herramienta BLA) asociados con los *quizzes online post-clase*, condujo a la exclusión

deliberada de datos asociados al rendimiento académico de los participantes. Futuras investigaciones podrían incorporar medidas objetivas de los resultados académicos, ampliar el tamaño de la muestra y emplear mecanismos de triangulación metodológica para reforzar la validez y la aplicabilidad de los hallazgos. Puede resultar también de interés el investigar la correlación entre los patrones de participación en los *quizzes* y el rendimiento en los exámenes para proporcionar evidencia más sólida que permita optimizar el diseño de los *quizzes*. Asimismo, estudios experimentales que comparen diferentes mecanismos de retroalimentación podrían ofrecer una guía más precisa para maximizar la eficacia pedagógica de los *quizzes online post-clase*. Además, este estudio se realizó con alumnado de segundo año de Ingeniería TIC. Investigaciones futuras podrían incluir participantes de diferentes niveles académicos, por ejemplo, estudiantes de primero, tercero o último año, para evaluar si los hallazgos se aplican a diferentes niveles de experiencia y conocimiento. Los investigadores podrían replicar este estudio en otras disciplinas que utilizan *quizzes online*, para evaluar si los resultados son específicos de la ingeniería o generalizables a otros ámbitos. Para incrementar la diversidad de la muestra, futuras investigaciones podrían también incluir estudiantes internacionales o participantes de otras facultades dentro o fuera de la institución.

CONCLUSIONES Y RECOMENDACIONES

Este estudio exploró el uso de *quizzes online post-clase* semanales como una herramienta de evaluación formativa para el alumnado de ingeniería del ámbito TIC en una asignatura de gestión. Mientras que los *quizzes pre-clase* han sido bien estudiados, los *quizzes post-clase* han recibido comparativamente una menor atención. En esta intervención, los estudiantes completaron los *quizzes* evaluados de forma automática en un sistema Moodle LMS después de cada unidad temática de una asignatura, cuya participación impactó en el 10 % de su calificación final. Esta investigación integra un marco teórico sobre el aprendizaje activo con una metodología cualitativa basada en la Psicología Constructivista, con el objetivo de examinar las experiencias de usuario, clasificar las percepciones de los participantes en fortalezas y debilidades asociándoles un valor para ponderar su importancia, y recoger ideas para potenciar las ventajas y mitigar las limitaciones del modelo implementado.

Los resultados de la encuesta anónima realizada con la herramienta *Bipolar Laddering* (BLA) sugieren que, desde la perspectiva del estudiante, los *quizzes online post-clase* pueden ser una estrategia útil de aprendizaje activo, especialmente para apoyar la preparación de exámenes, la participación continua y el *engagement* con el material del curso. Los estudiantes manifestaron que los *quizzes* frecuentes y de bajo impacto en la evaluación pueden facilitar la memorización y comprensión del contenido y la retención de conocimientos, fomentar el aprendizaje autodirigido y promover la revisión regular del contenido. Además, los estudiantes indicaron

que el aprendizaje activo puede mejorarse cuando los *quizzes* van más allá de la mera recuperación de información e incluyen preguntas que fomentan una reflexión más profunda. Asimismo, las percepciones de los estudiantes sobre la efectividad de los *quizzes* estaban estrechamente relacionadas con la usabilidad del *quiz* (p. ej., acceso intuitivo, extensión manejable, diversos formatos) y el soporte a nivel de manejo del sistema (p. ej., notificaciones oportunas, claridad visual, etc.). Los encuestados señalaron que la efectividad de los *quizzes online post-clase* puede depender en parte de equilibrar la estructura (p. ej., organización por unidad, disponibilidad y retroalimentación constantes) y la flexibilidad (p. ej., múltiples intentos, retroalimentación explicativa, acceso posterior a la finalización). Por último, nuestros hallazgos sugieren que las características del diseño de los *quizzes* pueden fomentar la motivación de los estudiantes, en particular cuando se percibe que los objetivos de las evaluaciones formativas están alineados con los criterios de evaluación sumativa y las preguntas están bien alineadas con los exámenes del curso.

Basándose en las respuestas de los estudiantes, este estudio propone las siguientes diez recomendaciones para diseñar *quizzes post-clase* que puedan ser compatibles con los formatos educativos tanto en línea como presenciales:

- Diseño del *quiz*: que sea conciso, formulando preguntas de forma clara y que se encuentren alineadas con los objetivos de aprendizaje, sin que sean excesivamente largas y complejas.
- Peso en la evaluación: considerar un peso adecuado de los *quizzes* en relación con el total de la evaluación, para incentivar el *engagement* del alumnado en la asignatura, evitando en lo posible incrementar la ansiedad derivada del proceso de evaluación.
- Alineación con el contenido de los exámenes: en la medida de lo posible, alinear el contenido de los *quizzes* con los exámenes de la asignatura para reforzar el valor percibido del alumnado de cara a la preparación de los exámenes.
- Calidad de las preguntas: incluir preguntas prácticas y centradas en escenarios concretos que puedan facilitar la transferencia de aprendizaje a partir de contextos reales. Cuando sea factible, incorporar elementos basados en el razonamiento para potenciar una mayor comprensión.
- Formato de la pregunta: usar distintos formatos de pregunta (p. ej., múltiple elección, verdadero/falso, respuesta corta, rellenar espacios, etc.) para facilitar distintas aproximaciones de aprendizaje y distintas dimensiones de comprensión.
- *Feedback* de las respuestas: proporcionar breves explicaciones cuando sea posible y considerar la opción de permitir al menos un reintento de respuesta para transformar eventuales errores en oportunidades de aprendizaje.
- Acceso a los *quizzes*: volver a permitir el acceso a los *quizzes* a los que el alumnado ya ha dado respuesta en un ‘modo de estudio’, o sea, con un número ilimitado de intentos para enfatizar el carácter formativo a la vez que reducir en lo posible el carácter evaluativo de los mismos.

- Tiempo: balancear la comprensibilidad de los *quizzes* con unos requisitos de tiempo razonables para minimizar la fatiga del alumnado y mantener una carga cognitiva manejable. Cuando se considere apropiado, pueden generarse múltiples *quizzes* en distintas unidades formativas de la asignatura.
- Aleatoriedad: generar un banco de preguntas que permita aleatorizar preguntas con la finalidad de desincentivar la simple memorización, promoviendo el aprendizaje integral de la asignatura a partir de dar respuesta a diversos *quizzes*.
- Reflexión: incluir breves actividades de reflexión una vez realizados los *quizzes post-clase* para fomentar la metacognición del alumnado, así como la identificación de áreas para profundizar en su estudio.

Estas recomendaciones deben interpretarse en el contexto de las limitaciones del estudio, ya que se basan principalmente en las percepciones de los estudiantes. Futuras investigaciones podrían verificar y evaluar su impacto en los resultados de aprendizaje mediante medidas objetivas de rendimiento, así como muestras más amplias y diversas.

REFERENCIAS

- Allsop, J., Young, S. J., Nelson, E. J., Piatt, J. y Knapp, D. (2020). Examining the benefits associated with implementing an active learning classroom among undergraduate students. *International Journal of Teaching and Learning in Higher Education*, 32(3), 418-426.
- Andrews, M. E., Graham, M., Prince, M., Borrego, M., Finelli, C. J. y Husman, J. (2020). Student resistance to active learning: Do instructors (mostly) get it wrong? *Australasian Journal of Engineering Education*, 25(2), 142-154. <https://doi.org/10.1080/22054952.2020.1861771>
- Arthurs, L. A. y Kreager, B. Z. (2017). An integrative review of in-class activities that enable active learning in college science classroom settings. *International Journal of Science Education*, 39(15), 2073-2091. <https://doi.org/10.1080/09500693.2017.1363925>
- Black, P. y Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-74. <https://doi.org/10.1080/0969595980050102>
- Bonwell, C. C. y Eison, J. A. (1991). *Active learning: Creating excitement in the classroom* (1991 ASHE-ERIC Higher Education Reports). ERIC Clearinghouse on Higher Education, The George Washington University. <https://eric.ed.gov/?id=ED336049>
- Brink, A. G. (2013). The impact of pre- and post-lecture quizzes on performance in intermediate accounting II. *Issues in Accounting Education*, 28(3), 461-485. <https://doi.org/10.2308/iace-50445>
- Cook, B. R. y Babon, A. (2017). Active learning through online quizzes: Better learning and less (busy) work. *Journal of Geography in Higher Education*, 41(1), 1--14. <https://doi.org/10.1080/03098265.2016.1185772>
- Dengri, C., Gill, A., Chopra, J., Dengri, C., Koritala, T., Khedr, A., Korsapati, A. R., Adhikari, R., Jain, S., Zec, S., Chand, M., Kashyap, R., Pattan, V., Khan, S. A. y Jain, N. K. (2021). A review of the quiz, as a new

- dimension in medical education. *Cureus*, 13(10), e18854. <https://doi.org/10.7759/cureus.18854>
- Dicheva, D., Dichev, C., Agre, G. y Angelova, G. (2015). Gamification in education: A systematic mapping study. *Journal of Educational Technology & Society*, 18(3), 75-88.
- Doolittle, P., Wojdak, K. y Walters, A. (2023). Defining active learning: A restricted systematic review. *Teaching and Learning Inquiry*, 11, Article 25. <https://doi.org/10.20343/teachlearninqu.11.25>
- Driessen, E. P., Knight, J. K., Smith, M. K. y Ballen, C. J. (2020). Demystifying the meaning of active learning in postsecondary biology education. *CBE-Life Sciences Education*, 19(4), ar52. <https://doi.org/10.1187/cbe.20-04-0068>
- Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J. y Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, 14(1), 4-58. <https://doi.org/10.1177/1529100612453266>
- Evans, T., Kensington-Miller, B. y Novak, J. (2021). Effectiveness, efficiency, engagement: Mapping the impact of pre-lecture quizzes on educational exchange. *Australasian Journal of Educational Technology*, 37(1), Article 1. <https://doi.org/10.14742/ajet.6258>
- Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H. y Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, 111(23), 8410-8415. <https://doi.org/10.1073/pnas.1319030111>
- Gillmor, S., Poggio, J. y Embretson, S. (2015). Effects of reducing the cognitive load of mathematics test items on student performance. *Numeracy*, 8(1), Article 4. <https://doi.org/10.5038/1936-4660.8.1.4>
- Gosavi, C. S. y Arora, S. (2022). Active learning strategies for engaging students in higher education. *Journal of Engineering Education Transformations*, 36(S1), 1-7. <https://doi.org/10.16920/jeet/2022/v36is1/22167>
- Gowe, J. E. (2024). The impact of quizzes on improving mathematics performance among higher education students. *The Journal of Informatics*, 4(1), Article 1. <https://doi.org/10.59645/tji.v4i1.397>
- Gyllen, J. G., Stahovich, T. F., Mayer, R. E., Entezari, N. y Darvishzadeh, A. (2021). Priming productive study strategies with preparatory quizzes in an engineering course. *Applied Cognitive Psychology*, 35(1), 169-180. <https://doi.org/10.1002/acp.3750>
- Hartikainen, S., Rintala, H., Pylväs, L. y Nokelainen, P. (2019). The concept of active learning and the measurement of learning outcomes: A review of research in engineering higher education. *Education Sciences*, 9(4), Article 276. <https://doi.org/10.3390/educsci9040276>
- Hattie, J. y Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112. <https://doi.org/10.3102/003465430298487>
- Heil, J. y Ifenthaler, D. (2023). Online assessment in higher education: A systematic review. *Online Learning*, 27(1), 187-218. <https://doi.org/10.24059/olj.v27i1.3398>
- Hennink, M. y Kaiser, B. N. (2022). Sample sizes for saturation in qualitative research: A systematic review of empirical tests. *Social Science & Medicine*, 292, 114523. <https://doi.org/10.1016/j.socscimed.2021.114523>
- Horna-Saldaña, C. y Canaleta, X. (2024). Application of universal design for learning and digital fabrication in the creation of a tool for inclusive teaching of the ordering of chemical elements. *Journal of Chemical Education*, 101(12), 5261-5271. <https://doi.org/10.1021/acs.jchemed.4c00679>

- Howell, R. A. (2021). Engaging students in education for sustainable development: The benefits of active learning, reflective practices and flipped classroom pedagogies. *Journal of Cleaner Production*, 325, 129318. <https://doi.org/10.1016/j.jclepro.2021.129318>
- Kirzner, R. S., Alter, T. y Hughes, C. A. (2021). Online quiz as exit ticket: Using technology to reinforce learning in face-to-face classes. *Journal of Teaching in Social Work*, 41(2), 151-171. <https://doi.org/10.1080/08841233.2021.1898521>
- Klein, K., Calabrese, J., Aguiar, A., Mathew, S., Ajani, K., Almajid, R. y Aarons, J. (2023). Evaluating active lecture and traditional lecture in higher education. *Journal on Empowering Teaching Excellence*, 7(2), Article 2. <https://doi.org/10.26077/ba42-a5cc>
- Kozanitis, A. y Nenciovici, L. (2023). Effect of active learning versus traditional lecturing on the learning achievement of college students in humanities and social sciences: A meta-analysis. *Higher Education*, 86(6), 1377-1394. <https://doi.org/10.1007/s10734-022-00977-8>
- López-Tocón, I. (2021). Moodle quizzes as a continuous assessment in higher education: An exploratory approach in physical chemistry. *Education Sciences*, 11(9), Article 500. <https://doi.org/10.3390/educsci11090500>
- Majuri, J., Koivisto, J. y Hamari, J. (2018). Gamification of education and learning: A review of empirical literature. En *Proceedings of the 2nd International GamiFIN Conference* (pp. 11-19). CEUR Workshop Proceedings, 2186.
- Mitchell, A., Petter, S. y Harris, A. (2017). Learning by doing: Twenty successful active learning exercises for information systems courses. *Journal of Information Technology Education: Innovations in Practice*, 16, 21-46. <https://doi.org/10.28945/3643>
- Nuci, K. P., Tahir, R., Wang, A. I. y Imran, A. S. (2021). Game-based digital quiz as a tool for improving students' engagement and learning in online lectures. *IEEE Access*, 9, 91220-91234. <https://doi.org/10.1109/ACCESS.2021.3088583>
- Owens, D. C., Sadler, T. D., Barlow, A. T. y Smith-Walters, C. (2020). Student motivation from and resistance to active learning rooted in essential science practices. *Research in Science Education*, 50(1), 253-277. <https://doi.org/10.1007/s11165-017-9688-1>
- Petchamé, J., Iriondo, I., Korres, O. y Paños-Castro, J. (2023). Digital transformation in higher education: A qualitative evaluative study of a hybrid virtual format using a smart classroom system. *Heliyon*, 9(6), e16675. <https://doi.org/10.1016/j.heliyon.2023.e16675>
- Petchamé, J., Iriondo, I., Torné, O. y Solanellas, F. (2024). Student perceptions of project-based learning when studying information systems in an ICT engineering program: Taking advantage of active learning. *Education Sciences*, 14(11), Article 1148. <https://doi.org/10.3390/educsci14111148>
- Petchamé, J., Iriondo, I., Villegas, E., Fonseca, D., Romero Yesa, S. y Aláez, M. (2021). A qualitative approach to help adjust the design of management subjects in ICT engineering undergraduate programs through user experience in a smart classroom context. *Sensors*, 21(14), 4762. <https://doi.org/10.3390/s21144762>
- Pifarré, M. y Tomico, O. (2007). Bipolar laddering (BLA): A participatory subjective exploration method on user experience. En *Proceedings of the 2007 Conference on Designing for User eXperiences* (pp. 2-13). ACM. <https://doi.org/10.1145/1389908.1389911>
- Prince, M. (2004). Does active learning work? A review of the research. *Journal of Engineering Education*, 93(3), 223-231. <https://doi.org/10.1002/j.2168-9830.2004.tb00809.x>

- Raes, A., Vanneste, P., Pieters, M., Windey, I., Van Den Noortgate, W. y Depaepe, F. (2020). Learning and instruction in the hybrid virtual classroom: An investigation of students' engagement and the effect of quizzes. *Computers & Education*, 143, 103682. <https://doi.org/10.1016/j.compedu.2019.103682>
- Roediger, H. L., III, y Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17(3), 249-255. <https://doi.org/10.1111/j.1467-9280.2006.01693.x>
- Ross, B., Chase, A.-M., Robbie, D., Oates, G. y Absalom, Y. (2018). Adaptive quizzes to increase motivation, engagement and learning outcomes in a first year accounting unit. *International Journal of Educational Technology in Higher Education*, 15(1), Article 30. <https://doi.org/10.1186/s41239-018-0113-2>
- Rossi, I. V., de Lima, J. D., Sabatke, B., Nunes, M. A. F., Ramirez, G. E. y Ramirez, M. I. (2021). Active learning tools improve the learning outcomes, scientific attitude, and critical thinking in higher education: Experiences in an online course during the COVID-19 pandemic. *Biochemistry and Molecular Biology Education*, 49(6), 888-903. <https://doi.org/10.1002/bmb.21574>
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153-189. <https://doi.org/10.3102/0034654307313795>
- Stoyanova, S. y Giannouli, V. (2022). Online testing as a means of enhancing students' academic motivation during the coronavirus pandemic. *Education Sciences*, 13(1), Article 25. <https://doi.org/10.3390/educsci13010025>
- Tharayil, S., Borrego, M., Prince, M., Nguyen, K. A., Shekhar, P., Finelli, C. J. y Waters, C. (2018). Strategies to mitigate student resistance to active learning. *International Journal of STEM Education*, 5(1), Article 7. <https://doi.org/10.1186/s40594-018-0102-y>
- Tofade, T., Elsner, J. y Haines, S. T. (2013). Best practice strategies for effective use of questions as a teaching tool. *American Journal of Pharmaceutical Education*, 77(7), Article 155. <https://doi.org/10.5688/ajpe777155>
- Vaismoradi, M. y Snelgrove, S. (2019). Theme in qualitative content analysis and thematic analysis. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 20(3), Article 23. <https://doi.org/10.17169/fqs-20.3.3376>
- Veletsianos, G., Kimmons, R., Larsen, R. y Rogers, J. (2021). Temporal flexibility, gender, and online learning completion. *Distance Education*, 42(1), 22-36. <https://doi.org/10.1080/01587919.2020.1869523>
- Villalonga Pons, J., Besalú, M., Samà Camí, A. y Sancho-Vinuesa, T. (2023). Estrategias de aprendizaje de estudiantes de ingeniería en línea. *RIED-Revista Iberoamericana de Educación a Distancia*, 26(2), 237-256. <https://doi.org/10.5944/ried.26.2.36257>
- Yang, C., Luo, L., Vadillo, M. A., Yu, R. y Shanks, D. R. (2021). Testing (quizzing) boosts classroom learning: A systematic and meta-analytic review. *Psychological Bulletin*, 147(4), 399-435. <https://doi.org/10.1037/bul0000309>

Fecha de recepción del artículo: 1 de junio de 2025

Fecha de aceptación del artículo: 6 de agosto de 2025

Fecha de aprobación para maquetación: 16 de septiembre de 2025

Fecha de publicación en OnlineFirst: 13 de octubre de 2025

Fecha de publicación: 1 de enero de 2026

La Política Editorial de la *RIED. Revista Iberoamericana de Educación a Distancia*, se concreta en los siguientes criterios:

- **De la AIESAD.** La RIED. *Revista Iberoamericana de Educación a Distancia* se configura como el instrumento de la Asociación Iberoamericana de Educación Superior a Distancia (AIESAD) para la difusión de trabajos de carácter científico, experiencias, convocatorias e información bibliográfica, dentro del ámbito de la enseñanza/aprendizaje abierto y a distancia en sus diferentes formulaciones y presentaciones.
- **Arbitrada.** La RIED es una publicación arbitrada que utiliza el sistema de evaluación externa de revisión por pares (doble ciego), identificándose cada trabajo con un DOI (*Digital Object Identifier System*).
- **Periodicidad y formato.** La RIED. *Revista Iberoamericana de Educación a Distancia*, tiene una periodicidad semestral (un volumen anual con dos números). Se edita en doble versión: impresa (ISSN: 1138-2783) y electrónica (E-ISSN: 1390-33061).
- **Idioma de los trabajos.** Podrán presentarse trabajos en lengua española, portuguesa e inglesa.
- **Requisitos.** Toda propuesta de colaboración deberá reunir los siguientes requisitos:
 - hacer referencia al campo de especialización propio de la RIED;
 - estar científicamente fundada y gozar de unidad interna;
 - suponer una ayuda para la profundización en las diversas dimensiones y ámbitos de la educación abierta y a distancia y de las TIC aplicadas a la educación.
 - Se primarán los trabajos sujetos al modelo IMyD (*Introducción, Metodología, Resultados y Discusión*) y que puedan tener incidencia en la educación superior.
- **Trabajo original.** Los trabajos enviados a la RIED para su publicación deberán constituir una colaboración original no publicada previamente en soporte alguno, ni encontrarse en proceso de publicación o valoración en cualquiera otra revista o proyecto editorial.
- **Normas de redacción y presentación.** Los trabajos deberán atenerse a las normas de redacción y presentación de carácter formal de la RIED. Las colaboraciones enviadas a la RIED que no se ajusten a ellas serán desestimadas.
- **Recepción de originales.** La Secretaría de la RIED acusará la recepción del manuscrito enviado por el autor/es. El Consejo de Redacción revisará el artículo enviado informando al autor/es, en caso necesario, si se adecua al campo temático de la revista y al cumplimiento de las normas y requisitos formales de redacción y presentación. En el caso de que todos los aspectos sean favorables, se procederá a la revisión por pares del artículo.
- **Revisión externa.** Antes de la publicación, los manuscritos enviados serán valorados de forma anónima por dos miembros del Comité Científico o Evaluadores Externos (revisión por pares), por el sistema de doble ciego que, en su caso, realizarán sugerencias para la revisión y mejora en vistas a la elaboración de una nueva versión. Para la publicación definitiva se requiere la valoración positiva de ambos revisores. En caso de controversia evidente por parte de éstos, se requerirá de una tercera valoración para su aceptación, modificación o rechazo definitivos de la publicación.
- **Criterios de Evaluación del Comité Científico y Evaluadores Externos.** Los criterios de valoración de cada artículo que justifican la decisión de aceptación/modificación/rechazo se basan en los siguientes ejes:
 - interés del campo de estudio al ámbito de los formatos educativos no presenciales, prioritariamente con posible incidencia en la educación superior.
 - relevancia, originalidad e información valiosa de las aportaciones,
 - aplicabilidad de los resultados para la resolución de problemas.
 - actualidad y novedad,
 - avance del conocimiento científico,
 - fiabilidad y validez científica: calidad metodológica contrastada,
 - correcta organización, redacción y estilo de la presentación del material.
- **Información.** La Secretaría de la RIED informará a los autores de la decisión de aceptación, modificación y rechazo de cada uno de los artículos. La corrección de pruebas de imprenta la hará la RIED cotejando con el original.
- **Política de privacidad:** Se mantendrá y preservará en todos los casos y circunstancias el anonimato de los autores y el contenido de los artículos desde la recepción del manuscrito hasta su publicación. La información obtenida en el proceso de revisión y evaluación tendrá carácter confidencial.
- **Fuentes.** Los autores citarán debidamente las fuentes de extracción de datos, figuras e información de manera explícita y tangible tanto en la bibliografía, como en las referencias. Si el incumplimiento se detectase durante el proceso de revisión o evaluación se desestimará automáticamente la publicación del artículo.
- **Responsabilidad.** RIED no se hará responsable de las ideas y opiniones expresadas en los trabajos publicados. La responsabilidad plena será de los autores de los mismos.
- **Licencia.** Los textos publicados en esta revista están sujetos a una Licencia Creative Commons Atribución 4.0 Internacional". Puede copiarlos, distribuirlos, comunicarlos públicamente, hacer obras derivadas y usos comerciales siempre que reconozca los créditos de las obras (autoría, nombre de la revista, institución editora) de la manera especificada por los autores o por la propia RIED.

OTRAS INFORMACIONES DE INTERÉS

- Declaración ética sobre publicación y malas prácticas: <https://revistas.uned.es/index.php/ried/about#custom-o>
- Directrices para autores. Normas para publicar en RIED: <https://revistas.uned.es/index.php/ried/about/submissions#authorGuidelines>
- Lista de comprobación previa de los envíos: <https://revistas.uned.es/index.php/ried/about/submissions>

Revista Iberoamericana de Educación a Distancia

Inteligencia artificial en la educación superior: diseño, competencias, evaluación y desafíos

ESTUDIOS E INVESTIGACIONES

Inteligencia artificial generativa y autonomía educativa: metáforas históricas y principios éticos para la transformación pedagógica

Inteligencia artificial e inteligencia colectiva en la educación superior digital: estudio cuasi-experimental con la plataforma Kampal

Impacto de la IA generativa en competencias digitales universitarias: evidencia experimental basada en el marco DigComp

AI-ED-SAT: diseño y validación de un cuestionario para autoevaluar competencias docentes en IA educativa

Autopercepción y utilidad de la inteligencia artificial generativa en docentes en formación

Docentes y divulgación científica en redes sociales: desarrollo y validación inicial de instrumentos de evaluación

Tecnologías inmersivas en la universidad: perfiles del profesorado y obstáculos para su integración

Desarrollo y validación de la escala de Aprendizaje Autorregulado con IA (AI-SRL)

Influencia de la motivación hacia el aprendizaje y la procrastinación en la dependencia a ChatGPT

Retroalimentación de aprendizajes con inteligencia artificial generativa en estudiantes universitarios

Aplicación de PathRAG en aprendizaje adaptativo con IA generativa para una educación inclusiva y sostenible

Nivelación inteligente: estrategia de aprendizaje adaptativo con IA en educación superior

Recomendador de evaluación para preguntas cortas utilizando modelos de lenguaje en propiedad intelectual

Medición de la habilidad escrita en español como lengua extranjera con inteligencia artificial generativa

Metodología participativa para la creación de recursos digitales de microaprendizaje con inteligencia

Aprendizaje activo en la educación superior de ingeniería: un decálogo para diseñar quizzes online post-clase

