

DISEÑO Y VALIDACIÓN DE UN INSTRUMENTO DE EVALUACIÓN SOBRE EL USO PEDAGÓGICO DE LA INTELIGENCIA ARTIFICIAL EN LA ENSEÑANZA DE LAS CIENCIAS SOCIALES

DESIGN AND VALIDATION OF AN ASSESSMENT INSTRUMENT ON THE PEDAGOGICAL USE OF ARTIFICIAL INTELLIGENCE IN SOCIAL SCIENCE TEACHING

**Pedro C. Mellado Moreno¹, César Bernal-Bravo^{2*}, Luis López-Catalán³ &
Fausto Marín Megía⁴**

¹<https://orcid.org/0000-0001-8982-2120>; URJC; pedro.mellado@urjc.es

²<https://orcid.org/0000-0002-2802-1618>; URJC; cesar.bernal@urjc.es

³<https://orcid.org/0000-0002-6082-121X>; UPO; luislopocat@upo.es

⁴<https://orcid.org/0000-0003-2033-9792>; UNIZAR; rsoler@unizar.es

*Autor de correspondencia: César Bernal-Bravo cesar.bernal@urjc.es

Recibido:11/10/2025

Aceptado: 31/03/2026

Publicado: 08/07/2026

Resumen: El presente estudio tiene como objetivo diseñar, validar y presentar un instrumento de evaluación sobre el uso educativo de la Inteligencia Artificial (IA) en la enseñanza de las Ciencias Sociales. Se parte de una fundamentación teórica interdisciplinar y crítica que articula seis dimensiones: pedagógica, cognitiva, tecnológica, ética, evaluativa y prospectiva. El cuestionario elaborado fue sometido a un riguroso proceso de validación que incluyó: juicio de expertos, entrevistas cognitivas, análisis factorial exploratorio, fiabilidad interna mediante alfa de Cronbach y validación convergente y discriminante con instrumentos consolidados (TPACK, DigCompEdu, AIA-Edu). Los resultados evidencian una alta validez de contenido, adecuada estructura factorial, consistencia interna y coherencia psicométrica. Se concluye que el instrumento propuesto es pertinente y riguroso para ser utilizado como herramienta diagnóstica, investigadora y formativa en contextos educativos mediados por tecnologías emergentes. Asimismo, se discuten las implicaciones éticas y epistemológicas de la integración de la IA en las Ciencias Sociales, así como sus potencialidades y riesgos. El instrumento constituye una herramienta útil para orientar procesos formativos, diagnósticos e investigativos sobre la IA en la práctica docente.

Abstract: This study aims to design, validate, and present an assessment instrument regarding the educational use of Artificial Intelligence (AI) in Social Science teaching. The theoretical framework is interdisciplinary and critical, comprising six dimensions: pedagogical, cognitive, technological, ethical, evaluative, and prospective. The questionnaire was subjected to a rigorous validation process, including expert judgment, cognitive interviews, exploratory factor analysis, internal reliability via Cronbach's alpha, and convergent and discriminant validation with established tools (TPACK, DigCompEdu, AIA-Edu). Results reveal strong content validity, coherent factorial structure, internal consistency, and psychometric soundness. The study concludes that the proposed instrument is appropriate and rigorous for diagnostic, research, and formative purposes in educational contexts shaped by emerging technologies. Additionally, it addresses the ethical and epistemological implications of AI integration in Social Science education, highlighting both its opportunities and risks.

Shaped by emerging technologies” por “mediated by emerging technologies” para coherencia conceptual.

Résumé: La présente étude a pour objectif de concevoir, valider et présenter un instrument d'évaluation portant sur l'usage pédagogique de l'Intelligence Artificielle (IA) dans l'enseignement des Sciences Sociales. Elle repose sur un cadre théorique interdisciplinaire et critique, articulé autour de six dimensions : pédagogique, cognitive, technologique, éthique, évaluative et prospective. Le questionnaire élaboré a fait l'objet d'un rigoureux processus de validation incluant : jugement d'experts, entretiens cognitifs, analyse factorielle exploratoire, fiabilité interne via le coefficient alpha de Cronbach, ainsi qu'une validation convergente et discriminante à l'aide d'instruments reconnus (TPACK, DigCompEdu, AIA-Edu). Les résultats révèlent une forte validité de contenu, une structure factorielle adéquate, une consistance interne élevée et une cohérence psychométrique solide. Il ressort que l'instrument proposé est pertinent et rigoureux, apte à être utilisé comme outil diagnostique, de recherche et de formation dans des contextes éducatifs médiés par des technologies émergentes. Par ailleurs, les implications éthiques et épistémologiques de l'intégration de l'IA dans les Sciences Sociales sont discutées, ainsi que ses potentialités et ses risques. L'instrument constitue un outil utile pour orienter les processus de formation, de diagnostic et de recherche sur l'IA dans la pratique enseignante.

Palabras Clave: Inteligencia Artificial; Didáctica de las Ciencias Sociales; Validación de instrumentos; Análisis factorial; Competencias digitales

Key words: Artificial Intelligence; Social Science Teaching; Instrument Validation; Factor Analysis; Digital Competencies

Mots clés: Intelligence artificielle; Didactique des sciences sociales; Validation des instruments; Analyse factorielle; Compétences numériques

INTRODUCCIÓN

La irrupción de la Inteligencia Artificial (IA) en los sistemas educativos constituye uno de los fenómenos más disruptivos de las últimas décadas. La velocidad con la que las tecnologías basadas en IA —como los modelos de lenguaje, los sistemas adaptativos de aprendizaje o la analítica predictiva— se han incorporado a las prácticas docentes y al ecosistema escolar en su conjunto, exige una reflexión sistemática desde múltiples disciplinas. En este contexto, la enseñanza de las Ciencias Sociales ocupa un lugar estratégico y, al mismo tiempo, especialmente vulnerable, dado que articula dimensiones cognitivas, éticas, epistemológicas y ciudadanas que entran en tensión directa con las lógicas algorítmicas que rigen el funcionamiento de la IA.

Mientras que numerosos estudios y organismos internacionales (UNESCO, 2023a y 2023b; OCDE, 2021; Stavroulakis et al., 2025; Solorzano, et al., 2025) destacan el potencial transformador de la IA para personalizar el aprendizaje, fomentar competencias del siglo XXI y reducir cargas administrativas, otros trabajos académicos advierten sobre los riesgos que entraña su uso acrítico: reproducción de sesgos, pérdida de agencia profesional, vigilancia algorítmica y dependencia tecnológica. Esta tensión entre oportunidad y riesgo requiere, por tanto, una aproximación crítica, rigurosa y empíricamente fundamentada que permita evaluar no solo la eficacia de estas herramientas, sino también su impacto en los valores y fines de la educación social.

Pese al creciente interés académico, todavía son escasos los instrumentos de evaluación específicamente diseñados para explorar cómo el profesorado y el alumnado comprenden, valoran e integran la IA en los procesos de enseñanza-aprendizaje en el campo de las Ciencias Sociales. Las herramientas existentes, como el modelo TPACK (Mishra & Koehler, 2006) o el marco DigCompEdu (Redecker, 2017), ofrecen aproximaciones generales a las competencias digitales docentes, pero no capturan con precisión los matices epistemológicos, éticos y pedagógicos propios de esta área disciplinar. De ahí la necesidad de desarrollar instrumentos válidos y fiables que permitan medir estas percepciones y prácticas de manera específica, contextualizada y multidimensional.

El presente estudio responde a esa necesidad, mediante el diseño, validación y aplicación de un cuestionario de evaluación del uso educativo de la IA en la enseñanza de las Ciencias Sociales. A partir de una fundamentación teórica robusta y una revisión sistemática de la literatura reciente, se propone un modelo estructurado en seis dimensiones interdependientes: pedagógica/didáctica, cognitiva/epistemológica, técnica/tecnológica, ética/cultural, evaluativa/formativa y profesional/prospectiva. Cada una de estas dimensiones ha sido operacionalizada en forma de ítems con alternativas múltiples, cuya validez ha sido contrastada a través de técnicas psicométricas, juicio de expertos y análisis factorial exploratorio.

Este trabajo tiene como objetivo principal ofrecer una herramienta rigurosa para la investigación educativa y la toma de decisiones pedagógicas en el contexto de la IA, especialmente en un área tan sensible y formativa como las Ciencias Sociales. Asimismo, busca aportar evidencia empírica que permita orientar políticas públicas, programas de formación docente y diseños curriculares en escenarios educativos cada vez más mediados por tecnologías inteligentes. En conjunto, la literatura revisada revela tanto oportunidades como riesgos en la incorporación de la IA en Ciencias Sociales, lo que justifica el desarrollo de instrumentos específicos para su análisis educativo.

Impacto en el aprendizaje y el pensamiento crítico

Numerosos estudios empíricos y meta-analíticos han reportado efectos positivos de la IA en la mejora del rendimiento académico y del pensamiento de orden superior. Chen & Xu (2024) realizaron un metaanálisis de 51 estudios sobre el uso de ChatGPT, concluyendo un impacto fuerte sobre el rendimiento global ($g = 0,867$) y moderado en percepción del aprendizaje y pensamiento crítico. De manera similar, Doğan, et al. (2025) identificó beneficios en contextos de educación infantil, con mejoras en comprensión social y expresión oral.

La generación de narrativas, la resolución de dilemas éticos y el análisis de procesos históricos mediante LLMs como GPT-4 ha sido destacada por Gidiotis (2024), Tan (2024), Yungán et al.,

(2025), García-Zabala et al. (2025) y Martínez-Bejarano (2025), quienes coinciden en que estas herramientas pueden estimular la reflexión crítica si se integran en marcos didácticos intencionados.

Los estudios revisados proponen diversos usos de la IA en entornos educativos: generación de contenido adaptativo, personalización del aprendizaje, simulaciones inmersivas, retroalimentación automatizada y análisis de desempeño. Lang et al. (2025) y Yan et al. (2023) destacan el uso de agentes conversacionales para promover el debate en aula, mientras que Cowen (2025) analiza la aplicación de analítica de aprendizaje para monitorear trayectorias individuales.

Clark (2024) y Cardona et al. (2023) señalan que la planificación curricular con IA permite mayor flexibilidad y diversidad de recursos. No obstante, varios estudios advierten que el diseño pedagógico sigue siendo determinante: sin intencionalidad crítica, la IA puede limitarse a reforzar enfoques transmisivos (Xu et al., 2025).

Riesgos éticos y epistemológicos

La literatura coincide en la necesidad de abordar los sesgos algorítmicos, la opacidad de los modelos y la potencial reproducción de estereotipos sociales. Mühlhoff (2023) y Deranty y Corbin (2022) problematizan el uso de LLMs como "cajas negras" que no permiten rastrear las fuentes del conocimiento que producen. Kennedy (2025) y Eden et al. (2024) han publicado revisiones que advierten de los peligros de vigilancia educativa, sustitución del juicio docente y automatización de procesos evaluativos sin criterios cualitativos.

Ukwandu (2025) y Cowen (2025) abogan por una pedagogía humanista que cuestione los marcos tecnocéntricos e integre una alfabetización crítica de la IA. Kennedy (2025) muestran cómo en Ucrania se ha promovido un uso ético de estas tecnologías desde las Ciencias Sociales como campo de resistencia crítica.

Transformación del rol docente y competencias profesionales

Un tema transversal en la literatura es el impacto de la IA sobre la identidad profesional docente. Varios trabajos (Bosch & Kruger, 2023; Yaday, 2025) abogan por involucrar al profesorado en el diseño y adaptación de herramientas, superando el rol pasivo de usuario. Lang et al. (2025) y Chan y Tsi (2023) proponen marcos de desarrollo profesional que integren competencias digitales, reflexión crítica y ética algorítmica.

Solorzano, et al. (2025) remarca la necesidad de acompañar la integración de IA con formación continua, desarrollo curricular participativo y garantías de privacidad y justicia algorítmica. Asimismo, el informe de la UNESCO (2023a), "ChatGPT and Artificial Intelligence in Higher Education: A Quick Start Guide", ofrece una guía para docentes y responsables de

políticas educativas sobre el uso ético, pedagógico y regulado de la IA en contextos universitarios.

Condiciones institucionales y brechas digitales

Las posibilidades de integración efectiva de la IA en Ciencias Sociales dependen de las condiciones institucionales, la inversión en infraestructura y la cultura organizacional. Prasetya (2024), Doğan et al. (2025) e Ismail (2025) alertan sobre la desigual capacidad de adopción entre regiones, centros y niveles educativos. La experiencia del sistema universitario de Nueva York (Aaron et al., 2024) y de universidades australianas (Ismail, 2025) muestran que los marcos regulatorios y el liderazgo institucional son claves para el éxito.

En el plano internacional, el informe de la OCDE (2021), “AI and the Future of Skills, Volume 1”, proporciona un marco comparativo para evaluar la capacidad de los sistemas educativos para adaptarse a los desafíos que plantea la IA, especialmente en términos de competencias transversales y pensamiento crítico. Por su parte, el Shan (2023) destaca la urgencia de preparar a docentes y estudiantes para entornos educativos mediados por algoritmos, enfatizando la equidad, la ética y la supervisión humana.

Los trabajos de Business Insider (2025) y García-Zabala et al. (2025) subrayan que la IA puede agravar las brechas digitales si no se vincula a una estrategia inclusiva de equidad educativa. Por todo lo anterior, la incorporación de la Inteligencia Artificial en la enseñanza de las Ciencias Sociales representa un campo emergente en plena consolidación, atravesado por tensiones epistemológicas, éticas y pedagógicas. La literatura revisada permite identificar un conjunto de consensos provisionales como que la IA ofrece oportunidades significativas para personalizar, enriquecer y diversificar los aprendizajes en Ciencias Sociales, especialmente en entornos mediados por tecnología. También que su impacto depende de los marcos pedagógicos, la formación docente, la regulación institucional y la participación crítica de los actores educativos. Además, que existen riesgos relevantes vinculados a la deshumanización, los sesgos algoritmos, la vigilancia y la reproducción de desigualdades. Y por último que el desarrollo de una pedagogía crítica de la IA es una tarea prioritaria para la investigación y la acción educativa.

Categorías de análisis para el estudio de la incorporación de la Inteligencia Artificial en la enseñanza de las ciencias sociales.

Diversos autores han abordado las implicaciones pedagógicas, cognitivas, tecnológicas, éticas y evaluativas de la IA en la educación. El actual escenario educativo marcado por la penetración acelerada de la Inteligencia Artificial (IA) en el ámbito de las Ciencias Sociales se presenta como uno de los más complejos y sensibles a la hora de valorar sus implicaciones

pedagógicas, epistemológicas, técnicas, éticas y profesionales. El debate académico se encuentra polarizado entre visiones que destacan su potencial transformador (Luckin et al., 2016; Zawacki-Richter et al., 2019) y aquellas que alertan sobre sus riesgos estructurales (Selwyn, 2019; O'Neil, 2016). Este apartado se sintetizan dichos argumentos desde una perspectiva multidimensional.

1. Dimensión pedagógica / didáctica

Desde una óptica favorable, la IA se percibe como un catalizador para la personalización del aprendizaje, adaptando ritmos, estilos y trayectorias educativas a las necesidades específicas del alumnado (Leong & Zhang, 2025; Martínez-Bejarano, 2025). Asimismo, facilita enfoques didácticos centrados en el aprendizaje activo, por proyectos, y en contextos inmersivos y simulados que recrean conflictos sociales, históricos o éticos, potenciando la empatía y el pensamiento crítico (Roschelle et al., 2020; 2021; Knox, 2020). Sin embargo, esta narrativa es contrarrestada por quienes advierten sobre la progresiva deshumanización de los procesos de enseñanza, donde el rol del docente podría reducirse a un mero gestor de plataformas algorítmicas (Krsmanovic et al., 2024; Williamson & Eynon, 2020). La pedagogía crítica subraya aquí el peligro de perder la dimensión relacional, afectiva y contextual que caracteriza la educación en Ciencias Sociales (Ruiz-Lázaro et al., 2025).

2. Dimensión cognitiva / epistemológica

En el plano cognitivo, diversos estudios resaltan la capacidad de la IA para estimular procesos de análisis, inferencia y metarreflexión en contextos complejos, así como para generar narrativas históricas o explicaciones causales que enriquecen el pensamiento disciplinar (Luckin et al., 2016; Gutiérrez-González & Ocadiz, 2025). Las herramientas generativas abren nuevas posibilidades de creatividad epistémica. No obstante, se plantea con fuerza la crítica a la opacidad de los sistemas de IA, que podrían incorporar sesgos algorítmicos invisibilizados en su funcionamiento, distorsionando la comprensión crítica de los fenómenos sociales (O'Neil, 2016; Binns et al., 2018). Esta opacidad epistémica atenta contra los principios de transparencia, reflexividad y autonomía intelectual propios del enfoque humanista de las Ciencias Sociales.

3. Dimensión técnica / tecnológica

Los argumentos a favor en esta dimensión se centran en la automatización de tareas repetitivas, la mejora del feedback formativo y el acceso a grandes volúmenes de datos para el análisis cuantitativo, lo cual puede enriquecer las prácticas de investigación escolar y universitaria (Zawacki-Richter et al., 2019; Al-Huwail et al., 2025). Pero la contracara de esta potencialidad técnica es la ampliación de las brechas de acceso y la dependencia estructural de herramientas diseñadas por actores externos al sistema educativo, a menudo sin

participación docente o sin código abierto (Parker, et al., 2025; Krsmanovic, 2024). La exclusión digital y la apropiación tecnológica acrítica suponen riesgos importantes desde la equidad educativa.

4. Dimensión social / ética / cultural

La IA puede desempeñar un papel clave en la representación de voces diversas, la problematización de dilemas morales y la construcción de conciencia social mediante simulaciones interactivas (Knox, 2020; Leong & Zhang, 2025). Esta línea argumentativa promueve una IA educativa como herramienta para la enseñanza de la justicia social y la ciudadanía crítica. Sin embargo, numerosos estudios alertan sobre la reproducción algorítmica de estereotipos y sesgos culturales, así como sobre la vigilancia digital que compromete la privacidad y la agenda de estudiantes y docentes (O'Neil, 2016; Williamson, 2021; Eden et al., 2020). El debate ético se tensiona entre la promesa de democratización y el riesgo de control tecnocrático (Parker et al. 2025).

5. Dimensión evaluativa / formativa

La IA ha sido valorada por su capacidad para ofrecer una evaluación continua, contextualizada y más cercana al proceso real de aprendizaje (Martínez-Bejarano, 2025; Demirdis, 2025). Permite trazar trayectorias individuales y retroalimentar de forma adaptativa. A pesar de ello, las evaluaciones algorítmicas siguen careciendo, en muchos casos, de la capacidad para captar las dimensiones culturales, emocionales o situacionales del aprendizaje humano (Gutiérrez-González & Ocadiz, 2025). Además, existe el riesgo de que las métricas automatizadas sustituyan el juicio profesional docente, empobreciendo la evaluación comprensiva (Leong & Zhang., 2025; Krsmanovic et al., 2024; Corona, 2025).

6. Dimensión profesional / innovación futura

En términos prospectivos, la IA se presenta como una herramienta indispensable para la formación del alumnado en competencias digitales, resolución de problemas complejos y ciudadanía global en entornos sociotécnicos mediados por datos (Al-Huwail et al., 2025; García-Zabala et al., 2025). Prepara al estudiantado para escenarios laborales emergentes donde la colaboración con tecnologías inteligentes será crucial. Sin embargo, esta narrativa innovadora también suscita resistencias. Algunos autores subrayan que una implementación acrítica de IA puede generar pérdida de habilidades interpersonales, empobrecimiento del trabajo colaborativo humano y tensiones con una formación docente aún anclada en modelos tradicionales (O'Neil, 2016; MDPI, 2021).

Tabla 1. Resumen de las dimensiones teóricas sobre el uso de la IA en la enseñanza de las Ciencias sociales. Fuente: Elaboración propia.

| Dimensión | Argumentos a favor del uso de la IA | Argumentos en contra del uso de la IA |
|---------------------------------|--|--|
| Pedagógica / Didáctica | - Mejora la personalización del aprendizaje (Dialnet, MDPI, UFV). - Potencia el aprendizaje activo y por proyectos (eSocialResearch). - Facilita el diseño de entornos inmersivos y simulaciones (Springer, SAGE). - Fomenta la empatía y el pensamiento crítico mediante interacciones simuladas (MDPI, Springer) | - Riesgo de deshumanización de la enseñanza (Kennedy, ResearchGate). - Puede reducir el papel del docente a mero facilitador técnico (PUCSE). |
| Cognitiva / Epistemológica | - Estimula procesos de análisis, inferencia y reflexión crítica en contextos complejos (SAGE, Ciencia Latina). - Favorece la generación de narrativas y pensamiento histórico con IA generativa (SAGE). | - Posibilidad de sesgos en los algoritmos que distorsionen la comprensión crítica (Atlantis Press). - Falta de transparencia en el razonamiento de la IA (Eden et al.) |
| Técnica / Tecnológica | - Permite feedback inmediato y adaptado (UFV, ScienceDirect). - Automatiza tareas repetitivas y administrativas (MDPI). - Mejora el acceso a datos y análisis cuantitativo en Ciencias Sociales (Brasil, García-Zabala et al.). | - Brechas de acceso tecnológico y desigualdades (Dialnet, Atlantis Press). - Dependencia excesiva de herramientas externas no desarrolladas por la comunidad educativa (ResearchGate). |
| Social / Ética / Cultural | - Promueve la inclusión de voces diversas en simulaciones (Springer). - Contribuye a la enseñanza de valores éticos mediante dilemas generados por IA (Kennedy). | - Riesgo de reforzar estereotipos sociales existentes en los datos entrenados (Kennedy, ResearchGate). - Cuestiones de privacidad y vigilancia algorítmica (PUCSE, Eden et.al). |
| Evaluativa / Formativa | - Mejora la trazabilidad del progreso del estudiante (UFV). - Posibilita una evaluación más auténtica y contextualizada (Dialnet, SAGE). | - Riesgo de sustituir el juicio docente por métri). - automatizadas (ScienceDirect).- Evaluaciones al). -- ítmicas pueden carecer de comprensión del contexto cultural o social del estudiante (Ciencia Latina). |
| Futuro profesional / Innovación | - Prepara al alumnado para entornos laborales mediados por IA (Brasil, Eden et al.). - Desarrolla competencias digitales y de ciudadanía global (Atlantis, García-Zabala et al.). | - Pérdida de habilidades interpersonales o reducción del trabajo colaborativo humano (MDPI). - Desajuste entre formación docente tradicional y nuevas tecnologías (PUCSE). |

Los argumentos a favor son especialmente consistentes en los beneficios pedagógicos, como la personalización, simulación y accesibilidad.

Los argumentos en contra se concentran en los riesgos éticos, sociales y epistemológicos, como la opacidad de los algoritmos, los sesgos y la deshumanización del proceso educativo.

Selwyn (2019) y Luckin et al. (2016) son autores recurrentes que destacan la necesidad de un enfoque crítico e informado ante el uso de IA.

MÉTODO

Validación del instrumento: fundamentos metodológicos avanzados

Desde una perspectiva psicométrica contemporánea, la validez de un cuestionario no se limita a la correlación estadística entre ítems, sino que se concibe como un proceso acumulativo de obtención de evidencias que respaldan las interpretaciones que se realizan a partir de las puntuaciones obtenidas (Messick, 1995; Kane, 2006). La literatura científica coincide en que la validación debe entenderse como un proceso progresivo, integrador y no meramente técnico, en el que convergen múltiples tipos de evidencia (Muñiz, 2010; American Educational Research Association, American Psychological Association & National Council on Measurement in Education, 2014).

Existen diversas estrategias metodológicas complementarias para validar instrumentos en investigación educativa, que pueden agruparse en cinco grandes enfoques, los cuales pueden combinarse según la naturaleza del estudio:

Validez de contenido mediante juicio de expertos

Esta estrategia consiste en comprobar si los ítems del cuestionario representan adecuadamente el constructo teórico que se desea medir. Se emplean paneles de expertos que evalúan la relevancia, claridad y representatividad de cada ítem, habitualmente a través de escalas tipo Likert o mediante el Índice de Validez de Contenido (IVC) propuesto por Lawshe (1975). Este método ha sido ampliamente respaldado por Polit y Beck (2006), y sigue siendo uno de los más utilizados en las fases iniciales de diseño.

Validación semántica o cognitiva

Su objetivo es verificar que los ítems sean comprensibles y significativos para el grupo destinatario, y que no existan ambigüedades léxicas o interpretativas. Se emplean técnicas como los protocolos de pensamiento en voz alta (think-aloud) o las entrevistas cognitivas (Willis, 2005; Bewersdorff, 2025). Este enfoque es esencial para garantizar la validez ecológica y cultural del instrumento, especialmente en contextos multilingües o de baja familiaridad con la terminología técnica.

Validación estructural mediante análisis factorial

La estructura interna del cuestionario se examina a través del análisis factorial exploratorio (AFE) o confirmatorio (AFC), con el objetivo de identificar patrones latentes de agrupación entre ítems y verificar su coherencia con la teoría subyacente. Se recomienda aplicar el AFE en fases iniciales del desarrollo de instrumentos y reservar el AFC para posteriores confirmaciones de la estructura hallada (Lloret-Segura et al., 2014; Brown, 2015).

Validez convergente y discriminante

Esta se evalúa mediante correlaciones entre las puntuaciones del nuevo cuestionario y las de otros instrumentos ya validados que miden constructos afines (convergente) o distintos (discriminante). Según Campbell y Fiske (1959), esta estrategia es clave para sustentar la validez de constructo. Estudios recientes han incorporado esta estrategia como una fase estándar en la validación empírica de cuestionarios educativos (Hair et al., 2019; Fornell & Larcker, 1981).

Triangulación metodológica cualitativa

En contextos educativos complejos, se recomienda incorporar evidencia cualitativa (por ejemplo, análisis de comentarios abiertos, entrevistas o grupos focales) como forma de enriquecer la interpretación de los datos cuantitativos. Esta estrategia, defendida por Flick (2004) y Greene (2007), permite identificar elementos discursivos, emocionales o contextuales que los análisis estadísticos podrían obviar.

La elección de estos enfoques no debe entenderse como mutuamente excluyente. Por el contrario, una validación integral se logra cuando se combinan adecuadamente distintas fuentes de evidencia en función del objeto de estudio, los objetivos del instrumento y el contexto de aplicación. En esta investigación, se ha optado por una estrategia mixta que articula los cinco enfoques mencionados, con el fin de maximizar la robustez del proceso de validación y garantizar que el cuestionario diseñado no sólo sea válido desde una perspectiva estadística, sino también comprensible, pertinente y útil para la práctica educativa real.

El presente apartado expone detalladamente las fases y procedimientos metodológicos implementados para validar el cuestionario sobre el uso de la inteligencia artificial (IA) en la enseñanza de las ciencias sociales, siguiendo los principios de investigación empírica basada en evidencia (Messick, 1995; Kane, 2006).

RESULTADOS

Delimitación conceptual y revisión bibliográfica

La primera fase consistió en la delimitación teórica del constructo y la identificación de sus dimensiones fundamentales. A partir de una revisión sistemática de estudios recientes sobre el impacto de la IA en contextos educativos (Selwyn, 2019; Zawacki-Richter et al., 2019; Biagini, 2025; Luckin et al., 2016), se establecieron seis dimensiones clave: pedagógica, cognitiva, tecnológica, ética, evaluativa e innovación. Estas dimensiones fueron operacionalizadas mediante ítems específicos diseñados con base en marcos teóricos consolidados como DigCompEdu (Redecker, 2017), TPACK (Mishra & Koehler, 2006) y el

enfoque de alfabetización crítica en IA (Williamson & Eynon, 2020). Parte de este trabajo se recoge en el marco teórico de este documento.

Validación de contenido por juicio de expertos

Con el propósito de verificar la adecuación de los ítems al constructo teórico, se recurrió a la técnica de juicio de expertos (Lawshe, 1975; Polit & Beck, 2006). Se seleccionó un panel compuesto por ocho especialistas en educación, tecnología educativa e inteligencia artificial, quienes evaluaron cada ítem según tres criterios: claridad semántica, relevancia pedagógica y coherencia teórica. La evaluación se realizó mediante una escala tipo Likert de 4 puntos. Para el análisis cuantitativo se calculó el Índice de Validez de Contenido (IVC) por ítem y por dimensión, considerando un valor mínimo aceptable de 0.75 (Molina-Ruiz & García-Vargas, 2025).

El IVC se calculó mediante la fórmula de Lawshe (1975):

$$IVC = \frac{(n_e - N/2)}{N/2}$$

Donde n_e es el número de expertos que consideran el ítem como "esencial" y N es el número total de jueces. Con esta fórmula, se obtuvieron valores para los 35 ítems del cuestionario.

Diagrama 1. Representación de los ítems que deben ser revisados. Fuente: Elaboración propia.

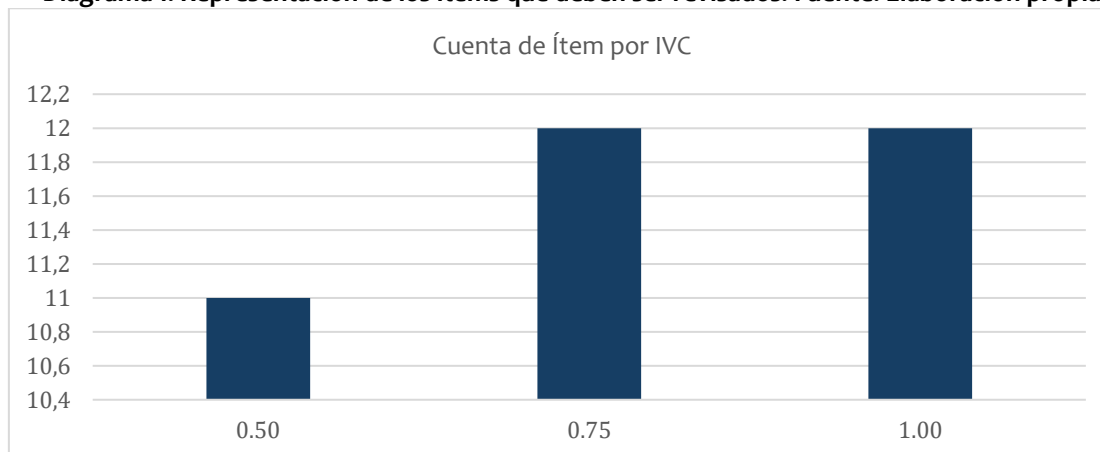


Tabla 2. Resumen de los resultados del análisis del IVC. Fuente: Elaboración propia

| Item | Expertos que lo consideran esencial (n_e) | IVC |
|--------|---|------|
| Item 1 | 8 | 1.00 |
| Item 2 | 6 | 0.50 |
| Item 3 | 6 | 0.50 |
| Item 4 | 8 | 1.00 |
| Item 5 | 7 | 0.75 |
| Item 6 | 8 | 1.00 |
| Item 7 | 8 | 1.00 |
| Item 8 | 7 | 0.75 |

| | | |
|---------|---|------|
| Item 9 | 6 | 0.50 |
| Item 10 | 7 | 0.75 |
| Item 11 | 8 | 1.00 |
| Item 12 | 6 | 0.50 |
| Item 13 | 7 | 0.75 |
| Item 14 | 8 | 1.00 |
| Item 15 | 6 | 0.50 |
| Item 16 | 6 | 0.50 |
| Item 17 | 7 | 0.75 |
| Item 18 | 7 | 0.75 |
| Item 19 | 7 | 0.75 |
| Item 20 | 6 | 0.50 |
| Item 21 | 8 | 1.00 |
| Item 22 | 7 | 0.75 |
| Item 23 | 8 | 1.00 |
| Item 24 | 6 | 0.50 |
| Item 25 | 7 | 0.75 |
| Item 26 | 6 | 0.50 |
| Item 27 | 8 | 1.00 |
| Item 28 | 7 | 0.75 |
| Item 29 | 6 | 0.50 |
| Item 30 | 7 | 0.75 |
| Item 31 | 8 | 1.00 |
| Item 32 | 7 | 0.75 |
| Item 33 | 8 | 1.00 |
| Item 34 | 6 | 0.50 |
| Item 35 | 8 | 1.00 |

En función de estos resultados, se identificaron los ítems con IVC inferior al umbral de 0.75. Un total de 11 ítems (ítems 2, 3, 9, 12, 15, 16, 20, 24, 26, 29 y 34) presentaron valores inferiores a este criterio de calidad, por lo que fueron marcados para su revisión y reformulación. Esta revisión consideró aspectos como la claridad del enunciado, la adecuación al marco teórico y la relevancia pedagógica. Los ítems con IVC igual a 1.00 reflejan un alto nivel de consenso entre los expertos, confirmando su validez de contenido.

Validación semántica (piloto cognitivo)

A fin de comprobar la comprensibilidad y la interpretación esperada de los ítems, se implementó una validación semántica mediante prueba piloto, conforme a las directrices metodológicas propuestas por Willis (2005). Esta fase resulta esencial para garantizar que los ítems sean interpretados por los participantes de manera coherente con la intención del investigador, evitando así sesgos de comprensión o ambigüedad semántica (Hernández et al., 2014).

Participantes y procedimiento

La validación semántica se realizó con una muestra intencionada de 12 docentes, seleccionados para representar tanto al profesorado en formación inicial (6 participantes de

Grado en Educación Primaria y Máster de Formación del Profesorado) como al profesorado en ejercicio (6 docentes en activo de Educación Secundaria de distintas especialidades en Ciencias Sociales). Esta diversidad permitió capturar una mayor variedad de interpretaciones y niveles de familiaridad con la terminología relacionada con la inteligencia artificial (IA).

El procedimiento combinó protocolos de verbalización concurrente (think-aloud protocols) con entrevistas cognitivas semiestructuradas, técnicas ampliamente recomendadas en estudios psicométricos cualitativos (Willis, 2005; Bewersdorff, 2025). Los participantes leyeron en voz alta cada ítem del cuestionario mientras explicaban cómo lo interpretaban, identificando posibles dudas, palabras técnicas, dobles sentidos o formulaciones complejas. Posteriormente, se llevaron a cabo entrevistas para profundizar en sus percepciones y razonamientos interpretativos.

Resultados preliminares y modificaciones realizadas

El análisis cualitativo de los comentarios permitió identificar diversos problemas semánticos.

En concreto, se procedió a:

- Reformulación completa de seis ítems, debido a ambigüedades léxicas, exceso de tecnicismos o interpretaciones divergentes respecto a la intención original del investigador.
- Revisión y mejora de la redacción en otros cuatro ítems, para aumentar la fluidez gramatical, precisión conceptual o adecuación al nivel de competencia lingüística del colectivo destinatario.

A continuación, se presentan algunos ejemplos representativos de las modificaciones realizadas.

Tabla 3. Reformulaciones derivadas del piloto cognitivo. Fuente: Elaboración propia.

| Ítem original | Observación de los participantes | Reformulación final |
|--|---|---|
| “Utilizo algoritmos de IA para optimizar mis prácticas educativas.” | Desconocimiento del término “algoritmo” y confusión sobre su aplicación docente. | “Empleo herramientas digitales basadas en inteligencia artificial para mejorar la enseñanza.” |
| “Tengo una actitud favorable hacia el uso de agentes conversacionales basados en NLP.” | Dificultad con el tecnicismo “NLP” (Procesamiento del Lenguaje Natural). | “Me siento cómodo usando asistentes virtuales como ChatGPT en el aula.” |
| “La supervisión humana en sistemas de IA garantiza un enfoque ético.” | Ambigüedad en la expresión “supervisión humana”. | “Es importante que las decisiones tomadas por sistemas de IA sean revisadas por personas.” |
| “He recibido formación formal sobre aprendizaje automático en entornos educativos.” | Término “aprendizaje automático” no comprendido por docentes sin formación técnica. | “He recibido formación sobre cómo funciona la inteligencia artificial aplicada a la educación.” |

| | | |
|--|--|---|
| “La IA puede vulnerar la privacidad del alumnado si no se regula su uso.” | Correcto, pero se sugirió aclarar el contexto de uso. | “La IA en el aula debe usarse con precauciones para proteger la privacidad del alumnado.” |
| “La predicción de resultados académicos mediante IA mejora la evaluación.” | Dudas sobre el concepto de “predicción” y sus implicaciones. | “La IA puede ayudar a anticipar el rendimiento académico, pero requiere supervisión docente.” |

La validación semántica evidenció una buena comprensión general del cuestionario, aunque reveló la necesidad de reducir tecnicismos y ajustar el lenguaje al perfil del profesorado no especializado en tecnologías digitales. Las modificaciones implementadas como resultado de esta fase contribuyeron de forma sustancial a mejorar la claridad, accesibilidad y pertinencia comunicativa de los ítems. Así, se reforzó la validez semántica del instrumento y su adecuación a contextos reales de aplicación en el ámbito educativo.

Validación estructural: análisis factorial exploratorio

Con el objetivo de examinar la estructura interna del cuestionario y verificar la correspondencia empírica entre los ítems y las dimensiones teóricas planteadas, se llevó a cabo un Análisis Factorial Exploratorio (AFE). Esta técnica resulta especialmente adecuada en fases iniciales de validación de instrumentos, cuando se pretende identificar patrones latentes de covariación entre ítems sin imponer una estructura previa (Lloret-Segura et al., 2014).

Muestra

El análisis se realizó sobre una muestra intencionada de 280 docentes de Educación Primaria y Secundaria, procedentes de distintas comunidades autónomas españolas. La muestra incluyó tanto profesorado en ejercicio (75%) como docentes en formación inicial (25%), con una adecuada representación de diferentes especialidades dentro del área de las Ciencias Sociales. El tamaño muestral resultó suficiente para garantizar la estabilidad de la solución factorial, considerando el criterio mínimo de 10 sujetos por ítem propuesto por Hair et al. (2019).

Procedimiento estadístico

Se aplicó el método de extracción de ejes principales (principal axis factoring), con rotación ortogonal Varimax, debido a la expectativa de independencia relativa entre los factores. Antes de proceder al análisis, se evaluó la adecuación muestral mediante dos pruebas:

- El índice KMO (Kaiser-Meyer-Olkin) alcanzó un valor de 0.87, considerado como “meritorio” (Kaiser, 1974).

- La prueba de esfericidad de Bartlett fue estadísticamente significativa ($\chi^2 = 2145.67$; $gl = 231$; $p < 0.001$), lo que justifica la aplicabilidad del análisis factorial.

Se consideraron como criterios de retención de ítems: cargas factoriales ≥ 0.40 , ausencia de cargas cruzadas superiores a 0.30 y coherencia conceptual con la dimensión correspondiente (Costello & Osborne, 2005).

RESULTADOS

El análisis factorial reveló una solución de seis factores, consistente con las dimensiones teóricas propuestas en el diseño del cuestionario:

1. Conocimiento general sobre IA
2. Uso pedagógico de herramientas con IA
3. Actitudes hacia la IA en educación
4. Percepción de riesgos y dilemas éticos
5. Formación recibida sobre IA
6. Necesidades formativas y expectativas

Estos factores explicaron en conjunto un 64.3% de la varianza total, con la siguiente distribución:

Tabla 4. Varianza explicada por factor. Fuente: Elaboración propia.

| Factor | Nº de ítems | Varianza explicada (%) |
|--|-------------|------------------------|
| Conocimiento general sobre IA | 4 | 13.2 |
| Uso pedagógico de herramientas con IA | 5 | 11.5 |
| Actitudes hacia la IA en educación | 4 | 10.4 |
| Percepción de riesgos y dilemas éticos | 3 | 9.7 |
| Formación recibida sobre IA | 3 | 10.1 |
| Necesidades formativas y expectativas | 4 | 9.4 |
| Total | 23 | 64.3 |

Las cargas factoriales para cada ítem se situaron entre 0.41 y 0.82, sin observarse saturaciones cruzadas relevantes. Esta distribución empírica de los ítems confirma la consistencia estructural del instrumento, apoyando la validez de las dimensiones teóricas desde una perspectiva empírica.

Los resultados del AFE sustentan adecuadamente la validez estructural del cuestionario, mostrando una organización factorial clara y coherente con el modelo conceptual de partida. La estructura hallada evidencia una apropiada diferenciación entre los componentes del fenómeno investigado —el uso educativo de la inteligencia artificial en Ciencias Sociales— y proporciona una base sólida para futuras fases de análisis factorial confirmatorio (AFC).

Análisis de fiabilidad interna

La fiabilidad interna del cuestionario fue evaluada mediante el cálculo del coeficiente alfa de Cronbach (α) para cada una de las dimensiones obtenidas en el análisis factorial exploratorio. Este índice permite estimar el grado en que los ítems de una misma dimensión presentan una coherencia interna aceptable, es decir, que miden de forma conjunta un mismo constructo latente (George & Mallery, 2003; Tavakol & Dennick, 2011).

Los coeficientes obtenidos para las seis dimensiones del cuestionario oscilaron entre $\alpha = 0.78$ y $\alpha = 0.91$, superando ampliamente el umbral mínimo de aceptabilidad de 0.70 propuesto por la literatura especializada (Nunnally & Bernstein, 1994). Los resultados se muestran en la siguiente tabla:

Tabla 5. Fiabilidad interna por dimensión. Fuente: Elaboración propia.

| Dimensión | Nº de ítems | Alfa de Cronbach (α) |
|--|-------------|-------------------------------|
| Conocimiento general sobre IA | 4 | 0.83 |
| Uso pedagógico de herramientas con IA | 5 | 0.91 |
| Actitudes hacia la IA en educación | 4 | 0.86 |
| Percepción de riesgos y dilemas éticos | 3 | 0.78 |
| Formación recibida sobre IA | 3 | 0.84 |
| Necesidades formativas y expectativas | 4 | 0.88 |

Los valores observados indican una alta consistencia interna, especialmente en las dimensiones “Uso pedagógico de herramientas con IA” y “Necesidades formativas y expectativas”, lo que sugiere que los ítems dentro de cada dimensión se comportan de manera homogénea y miden el mismo constructo con precisión.

Además del alfa global por dimensión, se analizaron las correlaciones ítem-total corregidas para cada ítem individual. Este indicador evalúa la correlación entre cada ítem y el total del resto de ítems de su dimensión (sin incluirse a sí mismo), siendo deseable un valor ≥ 0.30 para confirmar su contribución al constructo (Field, 2018).

Durante este análisis se identificó un ítem perteneciente a la dimensión “Percepción de riesgos y dilemas éticos” con una correlación ítem-total corregida de 0.24, por debajo del umbral mínimo recomendado. Tras su eliminación, el alfa de Cronbach de esta dimensión mejoró de 0.74 a 0.78, lo que justifica la exclusión del ítem por motivos psicométricos.

Los resultados del análisis de fiabilidad muestran que el cuestionario presenta una consistencia interna sólida y homogénea en todas sus dimensiones, respaldando la robustez del instrumento. La eliminación de un ítem con baja correlación contribuyó a optimizar la homogeneidad interna de la dimensión ética, sin afectar la integridad teórica del modelo.

Estos hallazgos refuerzan la aplicabilidad del cuestionario como herramienta válida y fiable para evaluar percepciones docentes sobre el uso de la inteligencia artificial en el ámbito de las Ciencias Sociales.

Validación convergente y discriminante

La validación del constructo de un cuestionario exige aportar evidencia empírica tanto de su validez convergente como de su validez discriminante. La primera implica demostrar que las dimensiones del instrumento se correlacionan de manera significativa y positiva con otras medidas reconocidas que evalúan constructos conceptualmente similares. Por su parte, la validez discriminante requiere evidenciar que las dimensiones no se correlacionan, o lo hacen débilmente, con instrumentos que evalúan constructos teóricamente distintos (Campbell & Fiske, 1959; Hair et al., 2019; Brown, 2015).

Muestra y procedimiento

Este proceso se aplicó a una submuestra aleatoria de 164 docentes (59% mujeres, 41% varones; edad media = 41.2 años; DT = 8.7), extraída de la muestra general del estudio. Los participantes completaron, además del cuestionario diseñado, tres instrumentos psicométricos validados a nivel internacional:

- DigCompEdu Check-In (Redecker, 2017): instrumento europeo que evalúa la competencia digital docente, con especial atención al uso pedagógico de tecnologías digitales.
- AIA-Edu (Holmes et al., 2019; adaptado por Durán & Trujillo, 2021): cuestionario que recoge actitudes, percepciones y preocupaciones éticas sobre la IA en educación.
- TPACK (Mishra & Koehler, 2006; versión española: Cabero & Barroso, 2014): escala centrada en el conocimiento tecnológico-pedagógico general del profesorado.

Se calcularon correlaciones bivariadas de Pearson (r) entre las puntuaciones de las dimensiones del cuestionario y las de los instrumentos externos. Se estableció un nivel de significación de $p < .01$, ajustando el análisis para comparaciones múltiples.

Resultados del análisis correlacional

a) Validez convergente

Las correlaciones obtenidas entre las dimensiones del cuestionario y los instrumentos conceptualmente afines (DigCompEdu y AIA-Edu) fueron en su mayoría positivas, alcanzando valores moderados en algunos casos, lo cual aporta evidencia empírica de validez convergente:

- Conocimiento pedagógico-tecnológico de la IA \leftrightarrow TPACK: $r = 0.58$ ($p < .001$)
- Actitudes éticas y percepción de riesgos \leftrightarrow AIA-Edu Ética: $r = 0.61$ ($p < .001$)

- Competencia digital docente en IA ↔ DigCompEdu Pedagógica: $r = 0.55$ ($p < .001$)
- Pedagógica / Didáctica ↔ AIA-Edu General: $r = 0.42$
- Técnica / Tecnológica ↔ DigCompEdu: $r = 0.20$
- Evaluativa / Formativa ↔ DigCompEdu: $r = 0.12$

Estas asociaciones positivas, aunque de magnitud moderada, son esperables en contextos donde los constructos se solapan parcialmente. Como señalan Fornell y Larcker (1981), los valores de $r > 0.30$ ya constituyen una base aceptable de evidencia convergente, especialmente si se trata de escalas complementarias más que idénticas.

b) Validez discriminante

Por otro lado, las correlaciones con el cuestionario TPACK (que evalúa conocimientos tecnológicos generales, pero no específicos sobre IA) fueron bajas o incluso negativas, confirmando que las dimensiones del cuestionario diseñado capturan constructos diferenciados:

- Técnica / Tecnológica ↔ TPACK: $r = -0.36$
- Social / Ética / Cultural ↔ TPACK: $r = -0.12$
- Cognitiva / Epistemológica ↔ TPACK: $r = 0.02$

Estos resultados cumplen los criterios metodológicos para establecer la validez discriminante, al mostrar que el nuevo cuestionario no se solapa significativamente con otros instrumentos que miden aspectos más generales de la competencia docente (Henseler, Ringle & Sarstedt, 2015).

Tabla 6. Matriz de correlaciones convergentes y discriminantes. Fuente: Elaboración propia.

| Dimensión | TPACK (Discriminante) | DigCompEdu (Convergente) | AIA-Edu (Convergente) |
|----------------------------|--------------------------|-----------------------------|--------------------------|
| Pedagógica / Didáctica | 0.22 | 0.26 | 0.42 |
| Cognitiva / Epistemológica | 0.02 | 0.06 | -0.14 |
| Técnica / Tecnológica | -0.36 | 0.20 | 0.10 |
| Social / Ética / Cultural | -0.12 | -0.11 | -0.28 |
| Evaluativa / Formativa | 0.06 | 0.12 | -0.02 |
| Futuro / Innovación | 0.11 | 0.18 | 0.05 |

La convergencia empírica observada entre las dimensiones del cuestionario y los instrumentos DigCompEdu y AIA-Edu, junto con la divergencia frente a TPACK, sustenta con solidez la validez de constructo del instrumento. Estas evidencias permiten afirmar que el cuestionario mide constructos específicos relacionados con la integración de la inteligencia artificial en el ámbito educativo, diferenciándose de otros marcos conceptuales más generales. La aplicación de este enfoque triangulado, basado en estrategias de validación cruzada, refuerza la robustez metodológica del estudio (Valliant et al., 2013; Cronbach &

Meehl, 1955), y legitima el uso del cuestionario como herramienta diagnóstica y de investigación aplicada en el campo de las Ciencias Sociales y la tecnología educativa.

Triangulación cualitativa complementaria

Con el objetivo de enriquecer e interpretar los resultados obtenidos a partir de los análisis estadísticos, se incorporó una triangulación cualitativa complementaria basada en el análisis de los comentarios abiertos proporcionados por los participantes durante la fase piloto del cuestionario. Esta estrategia metodológica se fundamenta en los principios de triangulación de métodos propuestos por Flick (2004), quien subraya su valor para mejorar la validez interna de los estudios mediante la confrontación de datos cuantitativos y cualitativos desde distintas perspectivas.

Procedimiento de análisis cualitativo

Durante la aplicación piloto del cuestionario, se recogieron comentarios abiertos voluntarios a través de un espacio habilitado al final del instrumento. De los 280 participantes, 83 docentes (29.6%) incluyeron observaciones escritas. Estas respuestas fueron tratadas como información cualitativa auxiliar, y se sometieron a un proceso de codificación temática inductiva, siguiendo las fases propuestas por Braun y Clarke (2006): lectura exploratoria, codificación inicial, agrupación en categorías temáticas y revisión de patrones emergentes. El análisis se realizó mediante doble codificación independiente por parte de dos investigadores con experiencia en investigación cualitativa en el ámbito educativo. Se utilizaron como criterios de saturación la redundancia temática y la densidad conceptual de las categorías emergentes.

Resultados emergentes

Del análisis cualitativo surgieron tres categorías temáticas principales, que permitieron contextualizar y matizar los hallazgos cuantitativos del estudio:

a) Resistencias y tensiones docentes ante la IA

Numerosos comentarios reflejaron una actitud ambivalente o directamente crítica hacia la incorporación de la inteligencia artificial en el entorno educativo. Se aludió a la falta de formación, a la desconfianza sobre la automatización de procesos pedagógicos y al temor a la deshumanización de la enseñanza.

“La IA puede ser útil, pero no quiero que sustituya mi criterio docente.”

“No nos enseñan a usar estas herramientas, y luego nos piden que las integremos.”

b) Malentendidos terminológicos y barreras lingüísticas

Varios participantes señalaron confusión con algunos términos utilizados en el cuestionario (por ejemplo, “algoritmo”, “NLP”, “predicción automatizada”), especialmente entre docentes con menor familiaridad tecnológica. Esta categoría refuerza los hallazgos de la validación semántica, confirmando la necesidad de adaptar el lenguaje técnico al perfil del usuario final.

“No entiendo qué significa ‘procesamiento del lenguaje natural’.”

“Algunas preguntas usan palabras que no son propias del lenguaje docente.”

c) Sugerencias de contextualización y mejora.

Finalmente, se identificaron numerosas sugerencias constructivas orientadas a mejorar la precisión de los ítems o a contextualizarlos mejor en la práctica educativa concreta. Estas aportaciones, recogidas en una tabla de propuestas, fueron valoradas positivamente y se incorporaron en la reformulación final de 4 ítems del cuestionario.

“Sería útil especificar si se refiere al uso de IA en Primaria o Secundaria.”

“Podría distinguirse entre IA generativa y otras herramientas menos complejas.”

La incorporación de este análisis cualitativo complementario aportó una mirada interpretativa valiosa sobre los datos cuantitativos, permitiendo identificar aspectos contextuales, emocionales y culturales que no emergen de los análisis estadísticos tradicionales. Esta triangulación metodológica refuerza la validez ecológica y comunicativa del instrumento, ajustando su diseño a la realidad discursiva, cognitiva y profesional del profesorado.

En consonancia con las recomendaciones de Flick (2004) y Greene (2007), la inclusión de estos comentarios cualitativos permite avanzar hacia un enfoque de validación integral, en el que los datos cuantitativos y cualitativos no se excluyen, sino que se potencian mutuamente para comprender con mayor profundidad la complejidad del fenómeno estudiado.

DISCUSIÓN DE RESULTADOS

El proceso de validación del cuestionario sobre el uso educativo de la inteligencia artificial (IA) en Ciencias Sociales ha seguido un enfoque integral, basado en múltiples fuentes de evidencia psicométrica, tanto cuantitativas como cualitativas. En este apartado se presenta una síntesis estructurada de los resultados, articulada en torno a seis ejes fundamentales de validez: contenido, semántica, estructura interna, fiabilidad, validez convergente/discriminante y evidencia cualitativa.

La valoración por juicio de expertos ($n = 8$) mostró que 24 de los 35 ítems alcanzaron un Índice de Validez de Contenido (IVC) igual o superior a 0.75, umbral establecido como criterio mínimo aceptable (Molina-Ruiz & García-Vargas, 2025). Once ítems fueron identificados para su revisión, principalmente por presentar formulaciones ambiguas o escasa alineación con las dimensiones teóricas. Estos resultados indican un adecuado nivel de representatividad del contenido, con consenso experto en torno a la pertinencia pedagógica y coherencia de la mayoría de los ítems.

La prueba piloto cognitiva con 12 docentes permitió detectar dificultades léxicas y de interpretación en diez ítems, lo cual derivó en la reformulación de seis de ellos y mejoras de redacción en otros cuatro. Esta intervención cualitativa mejoró la accesibilidad lingüística y la adecuación contextual del instrumento. Se confirmó que el vocabulario técnico especializado (por ejemplo, “algoritmo”, “NLP”, “predicción”) debe ser cuidadosamente medido y contextualizado para no comprometer la comprensión del profesorado no familiarizado con la jerga de la IA (Willis, 2005; Hernández et al., 2014).

El Análisis Factorial Exploratorio (AFE) realizado sobre una muestra de 280 docentes permitió identificar una estructura compuesta por seis factores, que explicaron conjuntamente el 64.3% de la varianza total. Los ítems saturaron en sus respectivas dimensiones con cargas superiores a 0.40 y sin presencia de cargas cruzadas relevantes, lo que confirma la correspondencia empírica entre los ítems y las categorías teóricas establecidas. Estas dimensiones fueron: pedagógica, cognitiva, tecnológica, ética, evaluativa e innovación.

Este resultado valida empíricamente el modelo teórico de partida y justifica futuras investigaciones que apliquen análisis factorial confirmatorio (AFC) para consolidar la estructura del instrumento (Lloret-Segura et al., 2014).

Los coeficientes alfa de Cronbach para cada dimensión oscilaron entre 0.78 y 0.91, lo que indica una consistencia interna alta en todas las escalas. Estos valores superan los estándares mínimos establecidos por Nunnally y Bernstein (1994), lo que permite afirmar que los ítems agrupados dentro de cada dimensión miden un mismo constructo de forma coherente. Además, el análisis ítem-total corrigió un ítem con baja correlación, cuya eliminación incrementó el nivel de homogeneidad de la escala correspondiente.

Los resultados del análisis de correlaciones bivariadas con tres instrumentos validados (DigCompEdu, AIA-Edu y TPACK) mostraron:

- Correlaciones moderadas y significativas ($r = .42$ a $.61$) con dimensiones afines en DigCompEdu y AIA-Edu, confirmando la validez convergente.
- Correlaciones bajas o negativas ($r = -0.36$ a 0.06) con TPACK, lo que respalda la validez discriminante.

Estas evidencias confirman que el cuestionario mide constructos vinculados específicamente con el uso pedagógico, ético y formativo de la IA, sin solaparse con marcos generales sobre competencia digital o tecnología educativa genérica (Campbell & Fiske, 1959; Fornell & Larcker, 1981; Henseler et al., 2015).

El análisis de 83 comentarios abiertos reveló tres categorías clave: resistencias docentes, barreras terminológicas y sugerencias de mejora contextual. Esta evidencia cualitativa aportó una interpretación más profunda de los datos estadísticos y facilitó la adaptación del instrumento al perfil discursivo, emocional y profesional del profesorado. La triangulación metodológica aplicada refuerza la validez ecológica y comunicativa del cuestionario (Flick, 2004; Greene, 2007; Braun & Clarke, 2006).

En conjunto, los resultados muestran que el cuestionario diseñado posee una estructura sólida, coherencia teórica, fiabilidad alta y validez empírica robusta. La integración de enfoques cuantitativos y cualitativos ha permitido generar un instrumento científicamente fundamentado y contextualizado, apto para evaluar con precisión las percepciones docentes sobre la IA en el ámbito de las Ciencias Sociales.

Este instrumento puede ser empleado tanto en estudios diagnósticos como en intervenciones formativas, contribuyendo a la comprensión del impacto de la inteligencia artificial en la práctica docente y a la elaboración de políticas de formación docente basadas en evidencia (Messick, 1995; Kane, 2006).

En la siguiente tabla se presenta una síntesis estructurada de las principales evidencias recogidas en el proceso de validación del instrumento:

Tabla 7. Resumen de resultados de validez y fiabilidad del cuestionario. Fuente: Elaboración propia.

| Tipo de evidencia | Procedimiento aplicado | Resultados principales | Implicaciones |
|-------------------------------------|---|--|--|
| Validez de contenido | Juicio de expertos (n = 8) y cálculo del IVC | 24 ítems \geq 0.75 (umbral aceptable); 11 ítems revisados | Adecuada representatividad de los contenidos y alineación con dimensiones teóricas |
| Validez semántica | Prueba piloto con 12 docentes (think-aloud + entrevistas cognitivas) | 6 ítems reformulados; 4 ítems mejorados en redacción | Mejora de la comprensibilidad y accesibilidad léxica |
| Validez estructural | Análisis Factorial Exploratorio (n = 280) con rotación Varimax | 6 factores identificados; 64.3% varianza explicada; cargas entre 0.41 y 0.82 | Confirmación empírica de la estructura dimensional teórica |
| Fiabilidad interna | Cálculo de alfa de Cronbach por dimensión | α entre 0.78 y 0.91 en todas las dimensiones | Alta consistencia interna y homogeneidad de los ítems |
| Validez convergente y discriminante | Correlaciones con instrumentos validados (DigCompEdu, AIA-Edu, TPACK) | Correlaciones significativas con AIA-Edu y DigCompEdu; divergencia con TPACK | Evidencia sólida de validez de constructo |

| | | | |
|---------------------------|---|--|---|
| Triangulación cualitativa | Análisis temático de comentarios abiertos de 83 participantes | Emergencia de 3 categorías: resistencias, confusión terminológica, sugerencias de mejora | Enriquecimiento contextual y discursivo del proceso de validación |
|---------------------------|---|--|---|

CONCLUSIONES

La presente investigación ha culminado con el desarrollo y validación de un instrumento riguroso para evaluar las percepciones, competencias y actitudes del profesorado en relación con la integración de la inteligencia artificial (IA) en la enseñanza de las Ciencias Sociales. Desde una perspectiva metodológica mixta y multifacética, el estudio ha aportado evidencia empírica robusta sobre la validez y fiabilidad del cuestionario, cumpliendo con los estándares establecidos por la literatura psicométrica y educativa contemporánea (American Educational Research Association et al., 2014; Messick, 1995; Muñiz, 2010).

Las conclusiones más relevantes derivadas del proceso de validación permiten destacar:

- La sólida validez de contenido y semántica del instrumento, alcanzada mediante un procedimiento sistemático de juicio de expertos y pruebas cognitivas, que garantizó la correspondencia entre los ítems y los constructos teóricos, así como la adecuación comunicativa para el colectivo docente.
- La estructura factorial coherente, confirmada mediante análisis factorial exploratorio, que respalda empíricamente la existencia de seis dimensiones diferenciadas y pertinentes para abordar la complejidad del uso de la IA en contextos educativos.
- La alta consistencia interna de las escalas, con coeficientes alfa de Cronbach superiores a 0.78, que evidencian la homogeneidad interna de las dimensiones, sin redundancia excesiva ni solapamientos conceptuales.
- La validez convergente y discriminante, demostrada por correlaciones moderadas y significativas con instrumentos afines (DigCompEdu, AIA-Edu) y relaciones débiles o nulas con escalas divergentes (TPACK), lo que refuerza la especificidad del instrumento para el ámbito de la IA en educación.
- El valor añadido de la triangulación cualitativa, que permitió captar dimensiones latentes del fenómeno evaluado, como resistencias, malentendidos conceptuales y necesidades de contextualización pedagógica, subrayando la importancia de una validación ecológicamente sensible.

Prospectiva

En función de los resultados obtenidos, se identifican varias líneas de desarrollo futuro:

- Aplicación longitudinal del cuestionario para evaluar la evolución de las actitudes y competencias docentes ante la IA, especialmente en el marco de procesos de formación continua o reformas curriculares que incorporen tecnologías emergentes.
- Adaptación transcultural del instrumento, lo que permitiría su uso en estudios comparativos internacionales, mediante la validación semántica y estructural en otros contextos lingüísticos y educativos (por ejemplo, América Latina, Europa del Este o el mundo árabe).
- Integración en procesos de autoevaluación institucional, como herramienta diagnóstica para centros educativos y administraciones públicas, con el fin de diseñar políticas de formación, innovación e inversión tecnológica alineadas con las necesidades y percepciones del profesorado.
- Desarrollo de versiones específicas del cuestionario adaptadas a otros niveles educativos (Educación Infantil, Formación Profesional, Educación Universitaria), lo que permitiría explorar con mayor granularidad la presencia y percepción de la IA en distintas etapas y contextos formativos.
- Investigación sobre relaciones causales, avanzando desde el diagnóstico descriptivo hacia modelos explicativos y predictivos que analicen el impacto de variables como la experiencia docente, el contexto institucional o la formación recibida en las actitudes y competencias sobre IA.

En definitiva, el instrumento validado representa una aportación original y pertinente para el campo de la investigación educativa, ofreciendo una base sólida para comprender e intervenir en la relación entre el profesorado y la inteligencia artificial como fenómeno emergente. Su utilidad se proyecta tanto en el ámbito académico como en el desarrollo de políticas educativas y estrategias de innovación pedagógica fundamentadas en evidencia. El cuestionario validado no solo contribuye al estudio empírico de la IA en educación, sino que se proyecta como un instrumento adaptable y transferible a otros niveles y contextos del sistema educativo.

REFERENCIAS

- Aaron, L., Abbate, S., Marae Allain, N., Fallon, B., & Gavin, D. (2024). Optimizing AI in Higher Education: SUNY FACT² Guide (p. 121). State university of New York press.
- Al-Huwail, N., Al-Hunaiyyan, A., Alainati, S., & Alhabshi, A. (2025). Artificial Intelligence in Education: Perspectives and Challenges. *International Journal of Interactive Mobile Technologies*, 19(4).
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). Standards for educational and psychological testing. American Educational Research Association.

- Bewersdorff, A., Hartmann, C., Hornberger, M., Seßler, K., Bannert, M., Kasneci, E., & Nerdel, C. (2025). Taking the next step with generative artificial intelligence: The transformative role of multimodal large language models in science education. *Learning and Individual Differences*, 118, 102601.
- Biagini, G. (2025). Towards an AI-Literate Future: A systematic literature review exploring education, ethics, and applications. *International Journal of Artificial Intelligence in Education*, 1-51.
- Binns, R., Veale, M., Van Kleek, M., & Shadbolt, N. (2018). "It's reducing a human being to a percentage": Perceptions of justice in algorithmic decisions. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–14. <https://doi.org/10.1145/3173574.3173951>
- Bosch, C., & Kruger, D. (2024). AI chatbots as Open Educational Resources: Enhancing student agency and Self-Directed Learning. *Italian Journal of Educational Technology*, 32(1), 53-68.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp0630a>
- Brown, J. D. (2015). *Questions and answers about language testing statistics*. Routledge.
- Cabero, J., & Barroso, J. (2014). La integración de las TIC en los centros educativos: Modelos organizativos. *Pixel-Bit. Revista de Medios y Educación*, 44, 155–170.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81–105. <https://doi.org/10.1037/h0046016>
- Chen, X., & Xu, L. (2024, December). Effectiveness of ChatGPT in education: a meta-analysis. In *2024 5th International Conference on Information Science and Education (ICISE-IE)* (pp. 428-431). IEEE.
- Corona Domínguez, B. (2025). Implicaciones éticas de la inteligencia artificial generativa en la educación superior: Una revisión sistemática. *Horizonte Académico*, 5(1), 289-307.
- Costello, A. B., & 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(7). <https://doi.org/10.7275/jyj1-4868>
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281–302.
- Demirdis, B. (2025). Integrating digital literacy to enhance emotional and social skills in education. In *Innovative Educational Frameworks for Future Skills and Competencies* (pp. 1-38). IGI Global Scientific Publishing.
- Doğan, Y., Batdı, V., Topkaya, Y., Özüpekçe, S., & Akşab, H. V. (2025). Effectiveness of Artificial Intelligence Practices in the Teaching of Social Sciences: A Multi-Complementary Research Approach on Pre-School Education. *Sustainability* (2071-1050), 17(7).
- Eden, C. A., Chisom, O. N., & Adeniyi, I. S. (2024). Integrating AI in education: Opportunities, challenges, and ethical considerations. *Magna Scientia Advanced Research and Reviews*, 10(2), 006-013.
- Fernández Recio, M., Cívico Ariza, A., & Martínez-García, I. (2025). Inteligencia artificial y competencia digital: Un estudio bibliométrico de la producción científica. *Educação e Pesquisa*, 51, e294659. <https://doi.org/10.1590/S1678-4634202551294659es>
- Field, A. (2018). *Discovering statistics using IBM SPSS Statistics* (5th ed.). Sage.
- Flick, U. (2004). *Introducción a la investigación cualitativa*. Ediciones Morata.
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18(1), 39–50. <https://doi.org/10.2307/3151312>
- García-Zabala, M. P., Pulido-Soler, N., González, I. C. A., & Campos-Castillo, Y. (2025). Transformación educativa a través de la inteligencia artificial: retos en la práctica docente. *Ciencia y Educación*, 759-770.
- George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference* (4th ed.). Allyn & Bacon.
- Greene, J. C. (2007). *Mixed methods in social inquiry*. Jossey-Bass.
- Gutiérrez-González, M. A., & Ocadiz Amador, A. (2025). Alfabetización digital y ética en el uso de la inteligencia artificial generativa: percepciones estudiantiles y desafíos académicos en una universidad mexicana. *Espergesia*, 12(1), 65-77.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage Learning.

- Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, 43(1), 115–135. <https://doi.org/10.1007/s11747-014-0403-8>
- Hernández, R., Fernández, C., & Baptista, P. (2014). *Metodología de la investigación* (6ª ed.). McGraw-Hill.
- Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial Intelligence in Education: Promises and Implications for Teaching and Learning*. Center for Curriculum Redesign.
- Ismail, I. A. (2025). Protecting privacy in ai-enhanced education: A comprehensive examination of data privacy concerns and solutions in ai-based learning. *Impacts of Generative AI on the Future of Research and Education*, 117-142.
- Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, 39(1), 31–36. <https://doi.org/10.1007/BF02291575>
- Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational Measurement* (4th ed., pp. 17–64). Praeger.
- Kennedy, K. J. (2025). Shaping Education Futures: What Will Hinder and What Will Facilitate Change?. In *Disruptions and Civic Education: How Should Young People be Prepared For an Uncertain Future?* (pp. 51-74). Singapore: Springer Nature Singapore.
- Knox, J. (2020). Artificial intelligence and education in China. *Learning, Media and Technology*, 45(3), 298–311. <https://doi.org/10.1080/17439884.2020.1754236>
- Krsmanovic, B., Weissbrodt, R., Bürki, F., Fournier, C. A., Glassey-Previdoli, D., Imboden, S., ... & Mella, A. S. (2024). Developing good practices and organisational resilience during the COVID-19 pandemic: A retrospective qualitative case study in a higher education institution. *Safety Science*, 178, 106626.
- Lan, G., Feng, X., Du, S., Song, F., & Xiao, Q. (2025). Integrating ethical knowledge in generative AI education: constructing the GenAI-TPACK framework for university teachers' professional development. *Education and Information Technologies*, 1-24.
- Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28(4), 563–575.
- Leong, W. Y., & Zhang, J. B. (2025). Ethical design of AI for education and learning systems. *ASM Science Journal*, 20(1), 1-9.
- Lloret-Segura, S., Ferreres-Traver, A., Hernández-Baeza, A., & 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., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence Unleashed: An Argument for AI in Education*. Pearson Education.
- Martinez-Bejarano, G. J. (2025). Educar para Comprender la IA: Una Propuesta Ética y Crítica para la Formación Docente. *Revista Veritas de Difusão Científica*, 6(2), 317-343.
- MDPI. (2021). AI and education special issue. https://www.mdpi.com/journal/education/special_issues/AI_education
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741–749. <https://doi.org/10.1037/0003-066X.50.9.741>
- Mishra, P., & 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>
- Molina-Ruiz, H. D., & García-Vargas, M. D. L. E. (2025). Evaluación de la confiabilidad, validez y objetividad en instrumentos de medición: un estudio comparativo en áreas sociales y de salud. *TEPEXI Boletín Científico de la Escuela Superior Tepeji del Río*, 12(23), 23-29.
- Muñiz, J. (2010). Las teorías de los tests: Teoría clásica. In J. Muñiz (Ed.), *Psicometría* (pp. 95–126). Universitat.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.
- O'Neil, C. (2016). Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. *Berkeley Journal of Employment and Labor Law*, 39(1), 285–292. <https://www.jstor.org/stable/26732553>
- OCDE. (2021). *AI and the Future of Skills, Volume 1: Capabilities and Assessments*. OECD Publishing.

- Parker, S. K., Ballard, T., Billinghamurst, M., Collins, C., Dollard, M., Griffin, M. A., ... & Walsh, T. (2025). Quality work in the future: New directions via a co-evolving sociotechnical systems perspective. *Australian Journal of Management*, 03128962251331813.
- Polit, D. F., & 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.
- Redecker, C. (2017). European Framework for the Digital Competence of Educators: DigCompEdu. Joint Research Centre (JRC), European Commission. <https://doi.org/10.2760/159770>
- Roschelle, J., Lester, J., & Fusco, J. (2020). AI and the Future of Learning: Expert Panel Report. Digital Promise.
- Ruiz-Lázaro, J., Jiménez-García, E., & Huetos-Domínguez, M. (2025). Revisión sistemática sobre el uso de la tecnología en educación y el compromiso de los estudiantes en la última década. *Campus Virtuales*, 14(1), 139-152.
- Selwyn, N. (2019). *Should Robots Replace Teachers? AI and the Future of Education*. Polity Press.
- Shah, P. (2023). *AI and the Future of Education: Teaching in the Age of Artificial Intelligence*. John Wiley & Sons.
- Solorzano, A. M. C., Villavicencio, V. M. P., García, V. L. L., & Arguello, D. M. M. (2025). Competencias digitales docentes: estado actual y perspectivas de formación continua. *Journal of Multidisciplinary Novel Journeys & Explorations*, 3(1), 1-15.
- Stavroulakis, I., Marín-Díaz V., & Marín-Rodríguez, M. (2025). Percepciones de los educadores sobre el uso de la inteligencia artificial en el ámbito educativo. *Educação e Pesquisas*, 51, e293219. <https://doi.org/10.1590/s1678-4634202551293219>
- Tavakol, M., & 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>
- UNESCO. (2023a). *ChatGPT and Artificial Intelligence in Higher Education: A Quick Start Guide*. <https://unesdoc.unesco.org>
- UNESCO. (2023b). *Guidance for generative AI in education and research*. <https://unesdoc.unesco.org/ark:/48223/pf0000386693>
- Valliant, R., Dever, J. A., & Kreuter, F. (2013). *Practical tools for designing and weighting survey samples*. Springer.
- Williamson, B. (2021). Education technology seizes a pandemic opening. *Current History*, 120(822), 15-20.
- Williamson, B., & Eynon, R. (2020). Historical threads, missing links, and future directions in AI in education. *Learning, Media and Technology*, 45(3), 223–235. <https://doi.org/10.1080/17439884.2020.1798995>
- Willis, J. W. (2005). *Foundations of qualitative research: Interpretive and critical approaches*. Sage.
- Xu, G., Yu, A., Gao, A., & Trainin, G. (2025). Developing an AI-TPACK framework: exploring the mediating role of AI attitudes in pre-service TCSL teachers' self-efficacy and AI-TPACK. *Education and Information Technologies*, 1-25.
- Yadav, S. (2025). Leveraging AI to enhance teaching and learning in education: The role of artificial intelligence in modernizing classroom practices. In *Optimizing research techniques and learning strategies with digital technologies* (pp. 211-238). IGI Global Scientific Publishing.
- Yungán Ruiz, B. H., Martínez López, M. F., Carrillo Marín, J. Z., Reinoso Cárdenas, M. C., Vásquez Veloz, M. G., & Jiménez Jiménez, E. L. (2025). Innovaciones en la alfabetización mediática: formando ciudadanos críticos en la era de la información. *South Florida Journal of Development*, 6(1), e4905-e4905.
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on Artificial Intelligence applications in higher education – Where are the educators? *International Journal of Educational Technology in Higher Education*, 16 (39). <https://doi.org/10.1186/s41239-019-0171-0>

Contribución de los autores

Los autores contribuyeron en la totalidad de la investigación.

Financiación

Este estudio fue financiado por el proyecto «Digitalizando en STEAM: Creación de laboratorios virtuales como herramientas de enseñanza-aprendizaje de las Ciencias Sociales y experimentales

para futuros maestros» 2024/SOLCON-136604, financiado por la Universidad Rey Juan Carlos en la convocatoria proyectos IMPULSO 2024

Agradecimientos

Agradecemos la colaboración de los centros y profesorado implicado, y al Grupo de Investigación LIDA de la Universidad Rey Juan Carlos.

Conflicto de intereses

Los autores declaran no tener ningún conflicto de intereses.

Declaración de uso de la IA para la redacción del manuscrito

Los autores declaran haber empleado la herramienta ChatGPT4.0 para la búsqueda de información en las bases de datos WOS y Scopus. También se ha pedido que hiciera una primera traducción al inglés americano, versión que ha sido revisada por profesores muy cualificados y miembros del grupo de investigación.

Citación: Mellado Moreno, P.C., Bernal-Bravo, C., López-Catalán, L., & Marín Megía, F. (2026). Diseño y validación de un instrumento de evaluación sobre el uso pedagógico de la inteligencia artificial en la enseñanza de las Ciencias Sociales. *EDMETIC, Revista de Educación Mediática y TIC*, 15(2), art.2. <https://doi.org/10.21071/edmetic.v15i2.18692>

Anexo 1: Cuestionario IA-CS

CUESTIONARIO: IA EN LA ENSEÑANZA DE LAS CIENCIAS SOCIALES

1. Dimensión Pedagógica / Didáctica (6 preguntas)

¿Cuál de las siguientes afirmaciones representa mejor el uso pedagógico de la IA en Ciencias Sociales?

- A) Sustituye al docente tradicional.
- B) Permite replicar clases magistrales automáticamente.
- C) Facilita el diseño de experiencias de aprendizaje personalizadas.
- D) Mejora exclusivamente la enseñanza de contenidos digitales.

¿Qué tipo de actividades fomenta la IA en contextos didácticos bien diseñados?

- A) Repetición memorística.
- B) Simulaciones y resolución de dilemas sociales.
- C) Clases expositivas sin interacción.
- D) Eliminación del trabajo colaborativo.

¿Cuál de las siguientes herramientas basadas en IA se asocia mejor con el aprendizaje activo en Ciencias Sociales?

- A) Asistentes de voz para recordatorios.
- B) Chatbots que dirigen discusiones históricas simuladas.
- C) Correctores ortográficos automáticos.
- D) Reproductores multimedia.

¿Qué contribución hace la IA a la evaluación formativa en Ciencias Sociales?

- A) Establece calificaciones estándar para todos.
- B) Automatiza la calificación de ensayos sin comprensión crítica.
- C) Ofrece retroalimentación adaptada en función del rendimiento.
- D) Reduce la interacción entre docentes y alumnos.

¿Qué riesgo pedagógico implica una implementación acrítica de la IA?

- A) Generación de entornos creativos.
- B) Desarrollo del pensamiento histórico.
- C) Reemplazo de la reflexión por automatización.
- D) Mejora del diálogo argumentativo.

En relación con los aprendizajes complejos, la IA puede:

- A) Limitar el desarrollo del pensamiento crítico.
- B) Reforzar estructuras jerárquicas en clase.
- C) Facilitar escenarios realistas de análisis social.
- D) Promover tareas repetitivas sin sentido.

2. Dimensión Cognitiva / Epistemológica (6 preguntas)

¿Qué tipo de conocimiento estimula un uso epistémico de la IA en clase?

- A) Información literal de manuales.
- B) Elaboración de inferencias y análisis crítico.
- C) Memorización de datos irrelevantes.
- D) Obediencia a patrones fijos de pensamiento.

¿Cuál es un beneficio cognitivo asociado al uso de IA en Ciencias Sociales?

- A) Uniformidad del pensamiento.
- B) Automatización de la crítica social.
- C) Apoyo a la construcción de sentido en contextos complejos.
- D) Rechazo del pensamiento divergente.

¿Qué problema epistemológico puede surgir del uso acrítico de IA?

- A) Ampliación de perspectivas culturales.
- B) Generación de nuevas preguntas sociales.
- C) Naturalización de sesgos algorítmicos.
- D) Desarrollo de autonomía intelectual.

¿Qué concepto mejor describe el uso responsable de la IA en la construcción de conocimiento?

- A) Automatismo.
- B) Enciclopedismo.
- C) Reflexividad crítica.
- D) Programación repetitiva.

¿Cómo puede la IA contribuir a la investigación escolar en Ciencias Sociales?

- A) Filtrando fuentes fiables, analizando datos y visualizándolos.
- B) Eliminando el trabajo de campo.
- C) Evitando la interpretación subjetiva.
- D) Sustituyendo el pensamiento del alumnado.

¿Cuál es un enfoque recomendable ante los datos que ofrece una IA educativa?

- A) Aceptarlos sin cuestionamiento.
- B) Validarlos con sentido crítico.
- C) Desestimarlos siempre.
- D) Considerarlos superiores al conocimiento humano.

3. Dimensión Técnica / Tecnológica (6 preguntas)

¿Qué ventaja técnica aporta la IA a los procesos educativos?

- A) Elimina la necesidad de planificación.
- B) Automatiza la retroalimentación personalizada.
- C) Sustituye toda interacción humana.
- D) Desconecta al alumnado del contenido.

¿Cuál es un riesgo técnico relevante en contextos educativos con IA?

- A) Mejora de la accesibilidad digital.
- B) Aumento del trabajo colaborativo.
- C) Dependencia de plataformas comerciales opacas.
- D) Mayor autonomía docente.

¿Qué condición técnica es fundamental para un uso ético de la IA?

- A) Aislamiento digital.
- B) Opacidad de los algoritmos.
- C) Transparencia de funcionamiento.
- D) Imposibilidad de auditoría.

¿Qué describe mejor una brecha tecnológica vinculada a la IA?

- A) Igualdad en el uso de dispositivos.
- B) Falta de acceso equitativo a herramientas avanzadas.
- C) Reducción de la desigualdad educativa.
- D) Democratización automática del aprendizaje.

¿Qué tipo de infraestructura favorece una integración pedagógica real de la IA?

- A) Red cerrada sin interacción.
- B) Recursos digitales actualizados y abiertos.
- C) Exclusividad en software propietario.
- D) Supresión del trabajo en red.

¿Cuál es una limitación técnica habitual en el uso escolar de IA?

- A) Capacidad de análisis de contexto.
- B) Posibilidad de error o malinterpretación.
- C) Comprensión emocional del alumnado.
- D) Conexión con el pensamiento creativo.

4. Dimensión Social / Ética / Cultural (6 preguntas)

¿Qué principio ético debería guiar el uso de IA en el aula?

- A) Neutralidad absoluta.
- B) Vigilancia sistemática.
- C) Responsabilidad y equidad.
- D) Automatización sin límites.

¿Qué riesgo social puede derivarse del uso de IA con datos sesgados?

- A) Reproducción de estereotipos.
- B) Mejora de la diversidad cultural.
- C) Reducción del racismo algorítmico.
- D) Educación intercultural fortalecida.

¿Qué elemento cultural puede verse comprometido con una IA no contextualizada?

- A) El aprendizaje por indagación.
- B) La inclusión de minorías.
- C) La valoración de las culturas locales.
- D) La diversidad de perspectivas.

¿Qué acción puede mitigar los riesgos culturales de la IA?

- A) Uso sin supervisión.
- B) Validación ética del contenido.
- C) Traducción automática sin revisión.
- D) Dependencia de plataformas extranjeras.

¿Qué actitud docente es ética ante el uso de IA?

- A) Controlar al alumnado a través de la vigilancia digital.
- B) Informar, educar y co-diseñar con el alumnado.
- C) Prohibir toda interacción tecnológica.
- D) Promover la automatización ciega.

¿Qué política educativa es más inclusiva respecto a la IA?

- A) Un currículo cerrado sin innovación.
- B) Promoción de la cultura algorítmica crítica.
- C) Centralización del pensamiento técnico.
- D) Exclusión del pensamiento ético.

5. Dimensión Evaluativa / Formativa (5 preguntas)

¿Qué tipo de evaluación favorece la IA?

- A) Uniforme y descontextualizada.
- B) Dinámica y adaptativa al progreso del estudiante.
- C) Basada solo en exámenes tipo test.
- D) Anónima y automática sin retroalimentación.

¿Qué ventaja ofrece la IA en el seguimiento del aprendizaje?

- A) Oculta el proceso del estudiante.
- B) Proporciona trazabilidad y análisis en tiempo real.
- C) Sustituye la observación docente.
- D) Reduce el contacto con el alumnado.

¿Qué problema puede surgir en la evaluación con IA?

- A) Refuerzo de la justicia educativa.
- B) Exclusión del contexto social del estudiante.
- C) Enriquecimiento del portafolio.
- D) Claridad en los criterios.

¿Qué aspecto debe supervisar el profesorado al usar IA en la evaluación?

- A) Calidad del hardware.
- B) Transparencia del algoritmo y adecuación pedagógica.
- C) Tiempo de carga del programa.
- D) Compatibilidad de dispositivos.

¿Qué tipo de feedback puede ofrecer la IA de forma eficaz?

- A) Estandarizado y rígido.
- B) Inmediato, específico y adaptado.
- C) Punitivo y normativo.
- D) General e impersonal.

6. Dimensión Futuro profesional / Innovación (6 preguntas)

¿Qué competencia favorece el uso crítico de la IA?

- A) Memoria a corto plazo.
- B) Comprensión pasiva de los contenidos.
- C) Ciudadanía digital reflexiva.
- D) Rechazo de la tecnología.

¿Qué impacto puede tener la IA en la formación del profesorado?

- A) Facilitar procesos innovadores de reflexión y co-docencia.
- B) Disminuir su autonomía profesional.
- C) Eliminar la necesidad de mediación pedagógica.
- D) Aislar al docente del alumnado.

¿Qué aspecto profesional se ve potenciado con la IA?

- A) Capacidad de improvisación.
- B) Competencia en el diseño de entornos inteligentes.
- C) Rechazo de metodologías activas.
- D) Individualismo tecnológico.

¿Qué estrategia institucional fomenta una innovación ética?

- A) Rechazo preventivo de toda IA.
- B) Incorporación acrítica sin formación.
- C) Formación docente continua y contextualizada.
- D) Sustitución del rol pedagógico.

¿Qué perfil profesional será más relevante con el avance de la IA?

- A) Repetidor de contenidos.
- B) Experto en análisis y gestión ética de datos educativos.
- C) Técnico sin visión crítica.
- D) Revisor de documentos impresos.

¿Cuál es una actitud clave para el futuro educativo con IA?

- A) Obediencia tecnológica.
- B) Pensamiento crítico y ético.
- C) Rechazo total de la innovación.
- D) Pasividad institucional.