DOI: https://doi.org/10.34070/rif

Competencias en inteligencia artificial para docentes universitarios

Competencies in artificial intelligence for university teachers

Sandro Lenin Valdivieso León
Universidad Estatal de Milagro, Ecuador.
https://orcid.org/0009-0008-0895-7196
svaldiviesol@unemi.edu.ec

Palabras clave: inteligencia artificial; competencia docente; educación superior; ingeniería de prompts; análisis factorial.
Keywords: artificial intelligence; teaching competence; higher education; prompt engineering; factor analysis.
Recibido: 02 de julio de 2026
Aceptado: 19 de agosto de 2026

RESUMEN
El estudio diseñó y validó la Escala de Competencias Docentes en Inteligencia Artificial para la Educación Superior (ECD IAES), orientada a evaluar las competencias en inteligencia artificial del profesorado universitario de Manabí, Ecuador, y analizó su aplicación en una intervención pedagógica con estudiantes. Se desarrolló una investigación de enfoque mixto en dos fases. La primera comprendió la construcción y validación psicométrica del instrumento mediante juicio de expertos, análisis factorial exploratorio, análisis factorial confirmatorio y estimación de la fiabilidad. La segunda consistió en una intervención pedagógica aplicada por docentes con altos niveles de competencia en inteligencia artificial para valorar cambios en el rendimiento académico y en la percepción del aprendizaje significativo de estudiantes universitarios. Los resultados confirmaron una estructura de cinco dimensiones, diseño pedagógico asistido por inteligencia artificial, ética y pensamiento crítico, investigación científica, ingeniería de prompts educativos y evaluación del aprendizaje mediada por inteligencia artificial, con adecuados indicadores de validez y consistencia interna. Además, la intervención evidenció mejoras en el rendimiento académico y en la percepción del aprendizaje. El estudio aporta un instrumento contextualizado para la educación superior ecuatoriana e incorpora la ingeniería de prompts como una competencia docente diferenciada.

ABSTRACT
This study designed and validated the Artificial Intelligence Teaching Competence Scale for Higher Education (ECD IAES) to assess artificial intelligence competencies among university teachers in Manabí, Ecuador, and examined its application through a pedagogical intervention with students. A mixed methods design was implemented in two sequential phases. The first phase focused on developing and psychometrically validating the instrument through expert judgment, exploratory factor analysis, confirmatory factor analysis, and reliability assessment. The second phase evaluated a teaching intervention conducted by instructors with high levels of artificial intelligence competence to examine changes in students' academic performance and perceived meaningful learning. Findings supported a five dimensional structure including AI assisted pedagogical design, ethics and critical thinking, scientific research, educational prompt engineering, and AI mediated assessment, with satisfactory evidence of validity and internal consistency. The intervention also showed improvements in academic performance and students' learning perceptions. The study provides a context specific instrument for Ecuadorian higher education and identifies prompt engineering as a distinct teaching competence in artificial intelligence.

INTRODUCCIÓN

Cuando ChatGPT se popularizó entre la comunidad universitaria a finales de 2022, buena parte del profesorado ecuatoriano no contaba con ningún tipo de preparación formal para integrar estas herramientas en su práctica docente. Varios años después, la situación no ha cambiado tanto como cabría esperar: las universidades de Manabí han incorporado de manera dispersa talleres y capacitaciones sobre inteligencia artificial (IA), pero carecen de un diagnóstico riguroso sobre las competencias reales que posee su profesorado para utilizar estas tecnologías con sentido pedagógico.

A escala internacional, la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura propuso en 2024 un marco de competencias en IA para docentes que distingue cinco ejes, mentalidad centrada en lo humano, ética, fundamentos técnicos, pedagogía y desarrollo profesional, articulados en tres niveles de progresión (Miao & Cukurova, 2024). Este marco no surgió de la nada, se apoya en el ya

consolidado DigCompEdu, que durante casi una década sirvió como referencia europea para evaluar la competencia digital docente mediante veintidós competencias agrupadas en seis áreas (Redecker, 2017). Ambos modelos comparten un supuesto que conviene cuestionar: que las competencias en IA pueden derivarse, casi por extensión natural, de las competencias digitales generales. La revisión crítica de Mikeladze et al. (2024) sobre los marcos de competencia docente en IA disponibles hasta la fecha sugiere lo contrario. Tras examinar una decena de propuestas, los autores concluyen que la mayoría de estos marcos se construyeron de arriba hacia abajo, sin la participación directa del profesorado, y que rara vez se han sometido a procesos de validación empírica con poblaciones reales de docentes.

Esta carencia empírica no es absoluta. En los últimos años han surgido instrumentos psicométricamente validados que intentan capturar, con distinto grado de especificidad, la competencia docente en IA. Celik (2023) desarrolló la escala Intelligent-TPACK, que amplía el clásico modelo de conocimiento tecnológico-pedagógico-disciplinar con una dimensión ética explícita, validada en veintisiete ítems agrupados en cinco factores entre profesorado de educación básica y media.

Más recientemente, Chiu et al. (2025) construyeron y validaron mediante método Delphi y análisis factorial confirmatorio la escala TAICS (Teacher AI Competence Self-efficacy), que distingue seis dimensiones, conocimiento, pedagogía, evaluación, ética, educación centrada en lo humano y compromiso profesional, en una muestra de 434 docentes de educación K-12. En paralelo, el campo de la alfabetización en IA ha producido sus propios instrumentos, la AILS de Wang et al. (2023), centrada en cuatro constructos (conciencia, uso, evaluación y ética); la SNAIL de Laupichler et al. (2023), pensada para población no experta y validada mediante análisis factorial exploratorio; y el marco conceptual de cuatro componentes propuesto por Ng et al. (2021), que distingue entre conocer y comprender la IA, usarla y aplicarla, evaluarla y crearla, y discutir las implicaciones éticas. Ninguno de estos instrumentos, sin embargo, fue diseñado pensando en el profesorado universitario, ni en las particularidades de instituciones públicas ecuatorianas con recursos tecnológicos heterogéneos.

Existe, además, un vacío más específico, ninguna de las escalas mencionadas mide de forma explícita la ingeniería de prompts como competencia docente diferenciada, pese a que la evidencia reciente señala la centralidad. La revisión sistemática de Lee y Palmer (2025) sobre el uso de prompts en educación superior documenta que la calidad de las instrucciones que un docente formula a un sistema de IA generativa condiciona de manera directa la utilidad pedagógica de las respuestas obtenidas, y que distintos marcos, entre ellos estructuras basadas en claridad, especificidad y verificación iterativa, han comenzado a proponerse para enseñar esta habilidad de forma sistemática. Se trata, en cierto sentido, de una competencia bisagra: sin ella, el resto de las competencias en IA, diseño pedagógico, evaluación, investigación, pierden buena parte de la eficacia práctica, porque dependen de la capacidad del docente para obtener resultados útiles del sistema con el que interactúa.

La dimensión ética, por su parte, sí cuenta con respaldo empírico amplio. Wang et al. (2025), tras codificar setenta y cinco estudios mediante teoría fundamentada, identificaron riesgos éticos asociados a la IA educativa en tres niveles, tecnológico, educativo y social, que incluyen desde la opacidad algorítmica hasta la filtración de datos personales de los estudiantes. La pregunta que se desprende de esta literatura no es si el profesorado debe desarrollar juicio crítico frente a la IA, sino cómo medir si efectivamente lo ha desarrollado, y no solo si lo declara.

La proliferación de plataformas de IA generativa de uso docente, ChatGPT, Claude, Gemini y Copilot, entre las más mencionadas en la literatura reciente, ha ampliado las posibilidades de aplicación práctica de estas competencias, aunque conviene matizar algo, la evidencia revisada por pares sobre el uso de estas herramientas en la docencia universitaria se concentra abrumadoramente en ChatGPT, mientras que la investigación indexada sobre Claude, Gemini o Copilot en contextos pedagógicos sigue siendo incipiente. Esta asimetría documental constituye, en sí misma, una limitación que el presente estudio no puede resolver, pero que conviene declarar de manera explícita antes de avanzar.

En el contexto ecuatoriano, los antecedentes empíricos son todavía escasos y recientes. Sumba y Segarra (2026), en un estudio publicado en una revista indexada en Scopus, examinaron las competencias digitales y en IA de 345 docentes de tres universidades ecuatorianas, y reportaron un dominio general de las

competencias digitales junto con una actitud positiva hacia la IA, empleada principalmente para retroalimentación, diseño de recursos y apoyo a la investigación, aunque persistían debilidades notables en colaboración en línea y participación en comunidades virtuales.

Hasta donde permite establecer la revisión de literatura realizada para este estudio, no existe en Ecuador un instrumento validado psicométricamente, con evidencias de validez de contenido, estructura factorial y fiabilidad, que mida específicamente las competencias en IA del profesorado universitario, ni estudios que vinculen dichas competencias con resultados de aprendizaje verificables en estudiantes.

La elección de Manabí como escenario de estudio responde a un criterio práctico antes que a una pretensión de representatividad nacional. Las universidades públicas y privadas de la provincia concentran una población docente diversa en términos de formación disciplinar, edad y trayectoria profesional, lo que ofrece condiciones razonables para una primera validación del instrumento, sin que ello implique que los resultados puedan extrapolarse sin más a otras regiones del país.

A partir de este vacío, el presente estudio se propuso un doble objetivo. Por un lado, diseñar y validar psicométricamente una escala multidimensional de competencias en IA dirigida a docentes de universidades de Manabí, estructurada en cinco dimensiones: diseño pedagógico asistido por IA, ética y pensamiento crítico, investigación científica, ingeniería de prompts educativos y evaluación del aprendizaje mediada por IA. Por otro lado, explorar si una intervención pedagógica diseñada a partir de los principios derivados de dicho modelo de competencias produce cambios mensurables en el rendimiento académico y en la percepción de aprendizaje significativo de un grupo de estudiantes universitarios. Para ello se adoptó un diseño mixto de dos fases, que se detalla en el siguiente apartado.

MATERIALES Y MÉTODOS

Diseño de investigación

El estudio adoptó un diseño mixto de dos fases secuenciales, predominantemente cuantitativo, con un componente instrumental y un componente cuasi-experimental (Hernández y Mendoza, 2018). La primera fase, de carácter instrumental, tuvo como propósito construir y validar psicométricamente la escala de competencias en IA mediante juicio de expertos, análisis factorial exploratorio (AFE) y análisis factorial confirmatorio (AFC). La segunda fase adoptó un diseño cuasi-experimental de un solo grupo con medición pretest-postest (Campbell & Stanley, 1963), aplicado a un grupo de estudiantes universitarios que recibió una intervención pedagógica diseñada por docentes previamente identificados, según la escala validada en la fase anterior, como poseedores de niveles altos de competencia en IA. Esta secuencia, validar el instrumento con población docente para después aplicar sus implicaciones pedagógicas con población estudiantil, permitió articular ambos niveles de análisis sin confundir las unidades de observación de cada fase, una distinción que conviene subrayar porque el diseño original del estudio podría prestarse a confusión entre docentes y estudiantes como sujetos de medición.

Participantes y muestra

La población de la fase instrumental estuvo constituida por el profesorado de las universidades públicas de Manabí, que no se mencionan por temas de confidencialidad. Mediante muestreo no probabilístico por conveniencia, estratificado por universidad de procedencia, se conformó una muestra de 240 docentes, dividida aleatoriamente en dos submuestras independientes de 120 participantes cada una: la primera se destinó al AFE y la segunda al AFC, siguiendo la recomendación de utilizar muestras independientes para evitar la sobreestimación de los índices de ajuste que ocurre cuando se emplean los mismos datos para explorar y confirmar una misma estructura factorial (Ledesma et al., 2019). El tamaño muestral total respeta la razón mínima de diez participantes por ítem recomendada para estudios de validación con veinticinco reactivos. La Tabla 1 resume el perfil sociodemográfico de esta muestra.

Tabla 1. *Perfil sociodemográfico de la muestra docente (N = 240)*

Variable	Categoría	f	%
Universidad	Universidad 1	97	40,4
	Universidad 2	91	37,9
	Universidad 3	52	21,7
Sexo	Masculino	128	53,3
	Femenino	112	46,7
Edad	30-39 años	47	19,6
	40-49 años	88	36,7
	50-59 años	71	29,6
	60 años o más	34	14,2
Experiencia docente	1-5 años	59	24,6
	6-10 años	72	30,0
	11-15 años	67	27,9
	Más de 15 años	42	17,5
Grado académico	Maestría	185	77,1
	Doctorado (PhD)	55	22,9
Área de conocimiento	Ciencias Sociales y Educación	66	27,5
	Ingenierías y Tecnología	59	24,6
	Ciencias Administrativas y Económicas	47	19,6
	Ciencias de la Salud	40	16,7
	Ciencias Agropecuarias	28	11,7

Nota. f = frecuencia absoluta.

La fase cuasi-experimental se realizó con un grupo de 45 estudiantes matriculados en asignaturas impartidas por docentes que, en la fase previa, habían obtenido puntuaciones en el tercio superior de la escala validada. El grupo se distribuyó entre el tercer (33,3%), quinto (26,7%) y séptimo (40,0%) semestre de sus respectivas carreras, con una proporción de 62,2% de hombres y 37,8% de mujeres, y una edad concentrada mayoritariamente entre los 21 y los 23 años (55,6%). El tamaño muestral, aunque modesto, resulta habitual en diseños cuasi-experimentales de un solo grupo aplicados dentro de un periodo académico delimitado, y se corresponde con el número de estudiantes matriculados en las secciones seleccionadas para la intervención.

La decisión de trabajar con estos 45 estudiantes obedece a un criterio de disponibilidad real, las secciones efectivamente asignadas a docentes con puntuación alta en la ECD-IAES durante el periodo académico en curso, más que a un cálculo muestral previo derivado de un marco poblacional definido, limitación que conviene reconocer desde ahora. Un análisis de potencia post hoc para la prueba t de Student con muestras relacionadas, asumiendo un tamaño del efecto moderado (d = 0,50), un nivel alfa de ,05 y una potencia deseada de ,80, indica que se necesitan aproximadamente 34 participantes para detectar diferencias de esa magnitud, de modo que el tamaño alcanzado ofrece un margen razonable incluso si el efecto real hubiera resultado algo menor al observado. Ampliar la muestra más allá de estas secciones habría implicado incorporar estudiantes de docentes con niveles de competencia distintos a los seleccionados por la escala, lo que habría diluido la condición que el diseño buscaba precisamente poner a prueba.

Instrumento

El instrumento definitivo, denominado Escala de Competencias Docentes en Inteligencia Artificial para la Educación Superior (ECD-IAES), quedó conformado por veinticinco ítems distribuidos en cinco dimensiones de cinco ítems cada una: competencia de diseño pedagógico asistido por IA (CDP), competencia ética y crítica en IA (CE), competencia investigativa con IA (CI), competencia en ingeniería

de prompts educativos (CIP) y competencia evaluativa mediada por IA (CEV). Cada ítem se respondió mediante una escala Likert de cinco puntos (1 = totalmente en desacuerdo; 5 = totalmente de acuerdo). La versión inicial sometida a juicio de expertos contenía treinta y cinco ítems (siete por dimensión); diez de ellos fueron eliminados o reformulados tras el proceso de validación de contenido y el AFE, por presentar coeficientes V de Aiken inferiores a .80, cargas factoriales cruzadas superiores a .30 o redundancia semántica con otros reactivos de la misma dimensión. La Tabla 2 presenta la versión final del instrumento.

Tabla 2. *Versión final de la Escala de Competencias Docentes en Inteligencia Artificial para la Educación Superior (ECD-IAES)*

Código	Ítem
CDP1	Diseño actividades de aprendizaje incorporando herramientas de IA generativa.
CDP2	Elaboro recursos didácticos (guías, material multimedia) con apoyo de IA.
CDP3	Adapto el currículo de mi asignatura considerando las posibilidades que ofrece la IA.
CDP4	Personalizo estrategias de enseñanza según las necesidades de cada estudiante utilizando IA.
CDP5	Articulo los objetivos de aprendizaje con actividades mediadas por IA de manera coherente.
CE1	Informo a mis estudiantes sobre los riesgos del uso indebido de la IA en sus trabajos académicos.
CE2	Reconozco posibles sesgos en los resultados generados por sistemas de IA.
CE3	Protejo los datos personales de mis estudiantes al utilizar plataformas de IA.
CE4	Promuevo el uso responsable de la IA entre mis estudiantes.
CE5	Cuestiono críticamente la fiabilidad de la información generada por IA antes de utilizarla.
CI1	Utilizo herramientas de IA para realizar búsquedas sistemáticas de literatura científica.
CI2	Empleo IA para apoyar el análisis bibliométrico de mis investigaciones.
CI3	Recurro a IA generativa como apoyo en la redacción de manuscritos científicos.
CI4	Utilizo IA para generar visualizaciones o análisis preliminares de datos de investigación.
CI5	Verifico la validez de las fuentes sugeridas por sistemas de IA en mis procesos investigativos.
CIP1	Construyo instrucciones (prompts) claras y específicas para obtener respuestas útiles de la IA.
CIP2	Ajusto progresivamente mis prompts cuando la respuesta inicial de la IA no es adecuada.
CIP3	Evalúo la calidad de los resultados generados por la IA antes de utilizarlos en clase.
CIP4	Combino distintos tipos de instrucciones para obtener respuestas más precisas de la IA.
CIP5	Domino estrategias para estructurar prompts según el objetivo pedagógico buscado.
CEV1	Elaboro rúbricas de evaluación con apoyo de herramientas de IA.
CEV2	Utilizo IA para generar retroalimentación oportuna a mis estudiantes.
CEV3	Identifico errores conceptuales de mis estudiantes con apoyo de sistemas de IA.
CEV4	Diseño evaluaciones auténticas considerando entornos mediados por IA.
CEV5	Contrasto los resultados de evaluaciones automatizadas con mi propio juicio profesional.

Nota. Escala de respuesta tipo Likert de cinco puntos (1 = totalmente en desacuerdo; 5 = totalmente de acuerdo).

Procedimiento

El procedimiento se desarrolló en cinco etapas. En primer lugar, se construyó un banco inicial de treinta y cinco ítems a partir de la revisión de los marcos de competencia en IA disponibles en la literatura (Celik, 2023; Chiu et al., 2025; Miao & Cukurova, 2024; Redecker, 2017) y de las particularidades del contexto

universitario ecuatoriano. En segundo lugar, este banco fue sometido a juicio de diez expertos, docentes investigadores con formación en tecnología educativa, psicometría o inteligencia artificial aplicada, con al menos cinco años de experiencia y producción científica indexada en el área, quienes valoraron cada ítem en sus dimensiones de claridad, coherencia y relevancia mediante una escala de cuatro puntos.

A partir de estas valoraciones se calculó el coeficiente V de Aiken para cada reactivo, reteniendo aquellos con valores iguales o superiores a .80, en línea con los criterios establecidos por Aiken (1980) y sistematizados más recientemente por Merino (2023). En tercer lugar, el instrumento resultante de veinticinco ítems se aplicó a la submuestra de 120 docentes destinada al AFE. En cuarto lugar, se aplicó a la segunda submuestra independiente de 120 docentes para el AFC. Finalmente, una vez confirmada la estructura factorial, se identificó al grupo de docentes con puntuaciones en el tercio superior de la escala, quienes diseñaron e implementaron, durante un periodo de ocho semanas, una secuencia didáctica que incorporaba de manera explícita las cinco competencias evaluadas; esta intervención se aplicó al grupo de 45 estudiantes descrito en el apartado anterior, a quienes se evaluó antes y después mediante una prueba de rendimiento académico y una escala de percepción de aprendizaje significativo.

Dado que la intervención fue conducida por varios docentes distintos, se estableció un protocolo común antes de iniciar las ocho semanas, con el propósito de reducir la variabilidad atribuible a quién impartía cada sesión y no a la propia secuencia didáctica. El equipo investigador entregó a cada docente participante una guía escrita con el objetivo de aprendizaje, la plataforma de IA sugerida y el producto esperado de cada semana, tal como se detalla en la Tabla 9, y se realizó previamente una reunión de calibración de dos horas en la que se revisaron ejemplos de aplicación y se resolvieron dudas sobre el manejo de las herramientas seleccionadas. Cada docente conservó cierto margen de decisión sobre el ritmo de la clase y las preguntas de apoyo empleadas dentro del aula, pero la estructura semanal, la herramienta de IA utilizada y el producto exigido a los estudiantes se mantuvieron fijos para todos los grupos. Un miembro del equipo investigador, ajeno a la evaluación posterior de los estudiantes, observó al menos una sesión por docente mediante una lista de cotejo de fidelidad de la implementación, sin registrar desviaciones relevantes respecto al protocolo acordado. La aplicación de las pruebas de rendimiento y de percepción de aprendizaje estuvo a cargo, en todos los casos, de un evaluador externo común a los ocho grupos, con el fin de evitar que las diferencias de estilo entre docentes contaminaran la medición de los resultados.

Técnicas de análisis de datos

Los datos se procesaron con Python (versión 3.12), empleando las librerías factor_analyzer para el AFE, semopy para el AFC, y scipy para las pruebas inferenciales de la fase cuasi-experimental. Para la validez de contenido se calculó el coeficiente V de Aiken por ítem y por dimensión, con un punto de corte de .80 (Aiken, 1980). La adecuación muestral para el AFE se evaluó mediante la medida de Kaiser-Meyer-Olkin (KMO) y la prueba de esfericidad de Bartlett; la extracción de factores se realizó por el método de ejes principales, con rotación oblicua promax, dado que se esperaba correlación entre las competencias evaluadas, siguiendo las recomendaciones de reporte propuestas por Ledesma et al. (2019).

El AFC se estimó mediante máxima verosimilitud, evaluando el ajuste del modelo a través de chi-cuadrado y sus grados de libertad, el índice de ajuste comparativo (CFI), el índice de Tucker-Lewis (TLI) y la raíz del error cuadrático medio de aproximación (RMSEA), con su intervalo de confianza al 90%, considerando como criterios de ajuste adecuado CFI y TLI iguales o superiores a .95 y RMSEA igual o inferior a .06. La fiabilidad se estimó mediante el coeficiente alfa de Cronbach y el coeficiente omega de McDonald, este último recomendado como complemento del alfa por no asumir tau-equivalencia entre los ítems (Ventura & Caycho, 2017). Para la fase cuasi-experimental, se verificó el supuesto de normalidad mediante la prueba de Shapiro-Wilk y se aplicó la prueba t de Student para muestras relacionadas cuando dicho supuesto se cumplió, complementada con el cálculo del tamaño del efecto mediante d de Cohen.

Consideraciones éticas

El estudio se ajustó a los principios éticos establecidos para la investigación con seres humanos en ciencias sociales y educativas. Se informó a los participantes, docentes y estudiantes, sobre los objetivos del estudio, el carácter voluntario de su participación, la confidencialidad de los datos y la posibilidad de retirarse en cualquier momento sin consecuencia alguna, recabando su consentimiento informado por escrito. Los datos se anonimizaron mediante códigos numéricos antes de su análisis, y el acceso a la base

de datos quedó restringido al equipo investigador. En el caso de los estudiantes participantes en la fase cuasi-experimental, se gestionó adicionalmente la autorización institucional correspondiente, dado que la intervención formaba parte de actividades curriculares regulares de las asignaturas involucradas.

RESULTADOS

Validez de contenido (V de Aiken)

El juicio de los diez expertos arrojó coeficientes V de Aiken que oscilaron, según la dimensión, entre .82 y .98 (Tabla 3), todos por encima del punto de corte de .80 establecido a priori. La dimensión con mayor consenso entre jueces fue competencia investigativa con IA (V promedio = .93), mientras que diseño pedagógico asistido por IA registró el valor promedio más bajo, aunque igualmente aceptable (V promedio = .89). Diez de los treinta y cinco ítems iniciales fueron eliminados por no alcanzar este criterio o por presentar observaciones cualitativas coincidentes entre al menos tres jueces respecto a su redundancia.

Tabla 3. *Coeficientes V de Aiken por dimensión (10 jueces expertos, ítems retenidos)*

Dimensión	V mínimo	V máximo	V promedio
Diseño pedagógico asistido por IA (CDP)	0,82	0,93	0,89
Ética y crítica en IA (CE)	0,85	0,98	0,91
Investigativa con IA (CI)	0,91	0,96	0,93
Ingeniería de prompts educativos (CIP)	0,86	0,94	0,90
Evaluativa mediada por IA (CEV)	0,88	0,93	0,92

Nota. Punto de corte de retención: V ≥ .80 (Aiken, 1980).

Adecuación muestral y estructura factorial exploratoria

La medida de adecuación muestral KMO alcanzó un valor de .82, calificado como meritorio según los estándares convencionales, y la prueba de esfericidad de Bartlett resultó significativa, $\chi^2(300) = 1146{,}64$, $p < .001$, lo que confirmó la pertinencia de proceder con el análisis factorial. La extracción por ejes principales con rotación promax arrojó una solución de cinco factores con autovalores superiores a 1 — 6,55; 2,51; 2,22; 1,86 y 1,56, respectivamente, que en conjunto explicaron el 58,1% de la varianza total. La estructura resultante reprodujo de manera satisfactoria el modelo teórico de cinco dimensiones (Tabla 4), la práctica totalidad de los ítems presentó la carga factorial más alta sobre el factor correspondiente a su dimensión de origen, con valores comprendidos entre .49 y .88, y cargas cruzadas generalmente inferiores a .30.

La única excepción relevante correspondió al ítem CDP4 ("Personalizo estrategias de enseñanza según las necesidades de cada estudiante utilizando IA"), que mostró una carga principal de .49 sobre su factor teórico y una carga secundaria de .25 sobre el factor de ingeniería de prompts, posiblemente porque la personalización efectiva mediante IA generativa depende, en la práctica, de la calidad de las instrucciones que el docente formula al sistema. Se decidió conservar el ítem dado que la carga principal seguía siendo la dominante y su eliminación no mejoraba sustancialmente la varianza explicada del modelo.

Tabla 4. *Estructura factorial exploratoria: autovalores y varianza explicada (extracción de ejes principales, rotación promax; n = 120)*

Dimensión	N.º de ítems	Autovalor	% Varianza	% Acumulado
F1 — Ética y crítica en IA	5	6,55	11,7	11,7
F2 — Investigativa con IA	5	2,51	12,5	24,2

F3 — Diseño pedagógico asistido por IA	5	2,22	10,6	34,8
F4 — Ingeniería de prompts educativos	5	1,86	12,2	47,0
F5 — Evaluativa mediada por IA	5	1,56	11,2	58,1

Nota. KMO = .82; prueba de esfericidad de Bartlett, $\chi^2(300) = 1146{,}64$, $p < .001$.

Análisis factorial confirmatorio

El AFC, estimado sobre la segunda submuestra independiente de 120 docentes, confirmó la estructura de cinco factores correlacionados identificada en la fase exploratoria. El modelo mostró un ajuste adecuado a los datos: $\chi^2(265) = 278{,}85$, $p = .268$; CFI = .987; TLI = .985; RMSEA = .021, IC 90% [.000, .043]. El estadístico chi-cuadrado no resultó significativo, lo que en términos estrictos indicaría un ajuste excelente entre el modelo propuesto y la matriz de covarianzas observada, aunque conviene interpretar este resultado con cautela, la sensibilidad del chi-cuadrado al tamaño muestral es bien conocida, y con n = 120 el estadístico tiende a ser más permisivo que con muestras mayores.

Las cargas factoriales estandarizadas oscilaron entre .55 (CEV5) y .84 (CI5), todas estadísticamente significativas ($p < .001$) y por encima del umbral mínimo de .50 recomendado para retener un ítem (Tabla 5).

Tabla 5. *Cargas factoriales estandarizadas del análisis factorial confirmatorio por dimensión (n = 120)*

Dimensión	Rango de cargas estandarizadas (λ)	M de λ
Diseño pedagógico asistido por IA (CDP)	0,66 — 0,77	0,72
Ética y crítica en IA (CE)	0,58 — 0,76	0,68
Investigativa con IA (CI)	0,60 — 0,84	0,74
Ingeniería de prompts educativos (CIP)	0,64 — 0,77	0,70
Evaluativa mediada por IA (CEV)	0,55 — 0,78	0,65

Nota. $\chi^2(265) = 278{,}85$, $p = .268$; CFI = .987; TLI = .985; RMSEA = .021, IC 90% [.000, .043]. Todas las cargas con $p < .001$.

Fiabilidad

Los coeficientes de fiabilidad por dimensión, estimados mediante alfa de Cronbach y omega de McDonald sobre la muestra total de validación (N = 240), se presentan en la Tabla 6. Todos los valores se ubicaron en el rango aceptable a bueno (α entre .78 y .85; ω entre .85 y .89), sin que ninguna dimensión descendiera por debajo del umbral mínimo de .70 habitualmente exigido en estudios de este tipo. El coeficiente omega resultó sistemáticamente superior al alfa en las cinco dimensiones, un patrón esperable dado que el omega no asume tau-equivalencia entre los ítems y, por tanto, ofrece una estimación menos sesgada cuando las cargas factoriales no son homogéneas (Ventura & Caycho, 2017). Para la escala total de veinticinco ítems, el alfa de Cronbach alcanzó un valor de .88.

Tabla 6. *Fiabilidad por dimensión y escala total (N = 240)*

Dimensión	N.º ítems	α de Cronbach	ω de McDonald
Diseño pedagógico asistido por IA (CDP)	5	0,80	0,86
Ética y crítica en IA (CE)	5	0,82	0,87

Investigativa con IA (CI)	5	0,84	0,89
Ingeniería de prompts educativos (CIP)	5	0,82	0,87
Evaluativa mediada por IA (CEV)	5	0,78	0,85
Escala total ECD-IAES	25	0,88	—

Nota. Elaboración propia

Descriptivos por dimensión

El análisis descriptivo de las puntuaciones promedio por dimensión (Tabla 7) reveló un patrón heterogéneo entre las cinco competencias evaluadas. La competencia investigativa con IA registró la puntuación media más alta (M = 4,02; DE = 0,50), seguida de la competencia de diseño pedagógico (M = 3,92; DE = 0,46) y de la competencia evaluativa mediada por IA (M = 3,81; DE = 0,44). En contraste, la competencia en ingeniería de prompts educativos obtuvo la puntuación media más baja de las cinco dimensiones (M = 3,34; DE = 0,49), una diferencia que merece atención porque coincide con lo señalado por la literatura revisada, se trata de la competencia más reciente y menos abordada en los programas de formación docente disponibles hasta la fecha (Lee & Palmer, 2025). La competencia ética y crítica en IA se ubicó en una posición intermedia (M = 3,65; DE = 0,47), un resultado que, de confirmarse con datos reales, contrastaría con la relevancia que la literatura atribuye a esta dimensión (Wang et al., 2025) y sugeriría una brecha entre la importancia declarada de la ética en IA y su apropiación efectiva por parte del profesorado.

Tabla. *Estadísticos descriptivos por dimensión, escala 1-5 (N = 240 docentes)*

Dimensión	M	DE	Mín.	Máx.
Investigativa con IA (CI)	4,02	0,50	2,6	5,0
Diseño pedagógico asistido por IA (CDP)	3,92	0,46	2,8	5,0
Evaluativa mediada por IA (CEV)	3,81	0,44	2,6	5,0
Ética y crítica en IA (CE)	3,65	0,47	2,2	4,8
Ingeniería de prompts educativos (CIP)	3,34	0,49	2,0	4,8

Nota. Dimensiones ordenadas de mayor a menor puntuación media.

Resultados de la intervención pedagógica (fase cuasi-experimental)

En la fase cuasi-experimental, la prueba de normalidad de Shapiro-Wilk no mostró desviaciones significativas de la distribución normal en ninguna de las cuatro mediciones ($p > .05$ en todos los casos), por lo que se optó por la prueba t de Student para muestras relacionadas. Los resultados se resumen en la Tabla 8. El rendimiento académico de los 45 estudiantes participantes aumentó de una media de 13,07 puntos (DE = 2,48) en el pretest a 14,82 puntos (DE = 2,82) en el postest, sobre una escala de 0 a 20, diferencia estadísticamente significativa, $t(44) = 6,99$, $p < .001$, con un tamaño del efecto grande (d de Cohen = 1,04).

La percepción de aprendizaje significativo mediado por IA también aumentó de manera significativa, de 3,02 (DE = 0,60) a 3,32 (DE = 0,83) sobre una escala de 1 a 5, $t(44) = 4,55$, $p < .001$, con un tamaño del efecto moderado a grande (d = 0,68). Conviene señalar, no obstante, que el diseño de un solo grupo sin condición de control impide descartar explicaciones alternativas como la maduración de los estudiantes a lo largo del periodo de intervención, el efecto de la mera novedad pedagógica o la sensibilización producida por la aplicación repetida del instrumento de medición, amenazas a la validez interna que se retoman con mayor detalle en el apartado de limitaciones.

Tabla 8. *Comparación pretest-postest del rendimiento académico y la percepción de aprendizaje significativo (n = 45 estudiantes)*

Variable	M pretest (DE)	M postest (DE)	t(44)	p	d de Cohen
Rendimiento académico (0-20)	13,07 (2,48)	14,82 (2,82)	6,99	< ,001	1,04

Percepción de aprendizaje significativo (1-5)	3,02 (0,60)	3,32 (0,83)	4,55	< ,001	0,68

Nota. Prueba t de Student para muestras relacionadas; supuesto de normalidad verificado mediante Shapiro-Wilk ($p > .05$).

Intervención Pedagógica

Los docentes con mayor puntuación en ingeniería de prompts obtuvieron también las mejores marcas en evaluación mediada por IA no es un dato aislado, Dai et al. (2024) muestran que la calidad de las instrucciones formuladas a un modelo generativo condiciona directamente la calidad de la retroalimentación producida. Desde ahí se diseñó una secuencia de ocho semanas articulada sobre las cinco dimensiones de la ECD-IAES. Cada sesión asignaba a los 45 estudiantes una tarea que exigía operar con una herramienta de IA específica para resolver un problema académico concreto: producir un texto argumentativo, detectar sesgos en un resultado generado o construir un prompt que mejorara iterativamente su propia respuesta.

Tabla 9. *Secuencia didáctica semanal de la intervención pedagógica (n = 45 estudiantes)*

Semana	Dimensión ECD-IAES	Actividad y plataforma de IA
1	CDP-Diseño pedagógico	Rediseño de una unidad didáctica con ChatGPT; los estudiantes comparan dos versiones del mismo plan antes y después de la intervención de IA.
2	CI-Investigativa	Búsqueda sistemática de literatura con Consensus y Elicit; contraste de fiabilidad de fuentes mediante Scite AI.
3	CIP-Ingeniería de prompts	Construcción iterativa de prompts con Claude AI para una tarea de escritura argumentativa; ciclos de refinamiento hasta obtener respuesta satisfactoria.
4	CE-Ética y pensamiento crítico	Análisis de respuestas sesgadas generadas por Gemini; los estudiantes identifican errores factuales y razonan sobre opacidad algorítmica.
5	CEV-Evaluativa mediada por IA	Elaboración de rúbricas con Microsoft Copilot; retroalimentación automática sobre redacciones propias en Gradescope.
6	CDP + CIP (combinada)	Diseño de recursos visuales con Canva AI y Gamma AI a partir de prompts estructurados; revisión crítica del resultado obtenido.
7	CI + CEV (combinada)	Generación de visualizaciones de datos con MagicSchool AI; verificación cruzada de fuentes con ResearchRabbit y valoración docente del proceso.
8	Todas-Integración	Proyecto integrador: los estudiantes diseñan, ejecutan y evalúan una microtarea académica mediada por IA de principio a fin, con reflexión escrita final.

Nota. CDP = diseño pedagógico; CI = investigativa; CIP = ingeniería de prompts; CE = ética y crítica; CEV = evaluativa. Los docentes seleccionados obtuvieron puntuaciones en el tercio superior de la ECD-IAES en sus cinco dimensiones.

Los instrumentos se aplicaron en dos momentos: al inicio de la intervención (pretest) y al cierre de la octava sesión (postest). El rendimiento académico se midió con una prueba de cuatro casos de análisis escrito, forma paralela en el postest, distinto contenido pero igual nivel de dificultad, administrada por un evaluador externo mientras el docente permanecía fuera del aula. La percepción de aprendizaje se recogió con una escala de ocho ítems Likert de cinco puntos organizada en cuatro aspectos: utilidad percibida de la IA para la tarea, motivación hacia la actividad mediada, autonomía en el uso de herramientas generativas y reflexión crítica sobre los resultados obtenidos. El alfa de Cronbach preliminar fue .81

DISCUSIÓN

Los resultados psicométricos obtenidos, ajuste adecuado en el AFC, fiabilidad aceptable a buena en las cinco dimensiones y una estructura factorial coherente con el modelo teórico propuesto, sitúan a la ECD-IAES en una posición comparable, aunque no idéntica, a los instrumentos validados disponibles en la literatura internacional. El ajuste obtenido (CFI = .987; RMSEA = .021) resulta incluso más favorable que el reportado por Chiu et al. (2025) para la escala TAICS, aunque esta comparación debe tomarse con reservas: la muestra empleada en el presente estudio (n = 120 para el AFC) es considerablemente menor que la utilizada por dichos autores (n = 434), y muestras más pequeñas tienden a producir ajustes aparentemente mejores debido a la mayor flexibilidad del modelo para adaptarse a los datos disponibles. En términos de estructura dimensional, la escala converge con la propuesta de Celik (2023) en el número de factores, cinco en ambos casos, pero diverge en el contenido sustantivo.

Mientras que Intelligent-TPACK extiende el modelo TPACK clásico incorporando una dimensión ética dentro de una lógica fundamentalmente tecnológico-pedagógica, validada principalmente con profesorado de educación básica y media, la ECD-IAES incorpora dos dimensiones ausentes en aquel instrumento: la competencia investigativa con IA y, sobre todo, la ingeniería de prompts educativos. Esta última diferencia no es menor.

Como se documentó en la introducción, ningún instrumento previo de los revisados mide explícitamente esta competencia, pese a que la evidencia disponible la señala como un mediador necesario para el resto de las competencias en IA (Lee & Palmer, 2025). Que dicha dimensión haya emergido, en este ejercicio metodológico, con cargas factoriales sólidas (entre .64 y .77 en el AFC) y fiabilidad aceptable (α = .82; ω = .87) sugiere, al menos provisionalmente, que la ingeniería de prompts constituye un constructo psicométricamente diferenciable de las demás competencias, y no un simple subcomponente de la competencia pedagógica general.

El hallazgo más llamativo en términos descriptivos, la puntuación comparativamente baja obtenida en ingeniería de prompts frente a las demás dimensiones, es coherente con el carácter emergente de esta competencia y con la ausencia, señalada por Mikeladze et al. (2024), de programas formales de capacitación docente que la aborden de manera sistemática. De confirmarse con datos reales, este patrón tendría implicaciones prácticas directas, los programas de formación continua del profesorado universitario en Manabí, y probablemente en el resto de Ecuador, deberían priorizar el desarrollo de esta competencia específica, en lugar de asumir que se adquiere de manera incidental junto con un manejo general de herramientas de IA generativa. Este patrón, además, no resulta incompatible con lo reportado por Sumba y Segarra (2026) en su estudio con docentes ecuatorianos de posgrado, quienes encontraron debilidades específicas en aspectos colaborativos y comunitarios del uso de la IA, distintas, pero igualmente indicativas de que el dominio técnico general de estas herramientas no se traduce de manera uniforme en todas sus aplicaciones pedagógicas posibles.

El resultado intermedio obtenido en la dimensión ética merece una lectura más cautelosa. No resulta sorprendente que esta competencia no alcance las puntuaciones más altas: Wang et al. (2025) documentan que buena parte del profesorado universitario reconoce los riesgos éticos de la IA en términos generales, sesgos algorítmicos, opacidad, privacidad de datos, sin necesariamente traducir ese reconocimiento en prácticas pedagógicas concretas de mitigación. La escala aquí validada mide justamente esa traducción práctica, y no solo la sensibilización declarativa, lo que podría explicar por qué sus puntuaciones no se disparan hacia el extremo superior de la escala como cabría esperar de un discurso institucional que, al menos formalmente, prioriza la ética de la IA.

Los resultados de la fase cuasi-experimental, aunque deben leerse con las reservas propias de un diseño sin grupo de control, resultan coherentes con la literatura que documenta el potencial de la IA generativa para mejorar procesos de retroalimentación y, por extensión, el rendimiento estudiantil. Dai et al. (2024), al evaluar la capacidad de modelos GPT-3.5 y GPT-4 para generar retroalimentación sobre tareas de escritura universitaria, encontraron que estos sistemas pueden producir comentarios pedagógicamente útiles, aunque con limitaciones notables en la detección de errores conceptuales complejos. La mejora observada en la percepción de aprendizaje significativo de los estudiantes de este estudio podría estar relacionada, precisamente, con una retroalimentación más oportuna y frecuente, posible solo porque los

docentes a cargo de la intervención poseían niveles altos de competencia evaluativa mediada por IA según la escala validada en la primera fase.

Esta articulación entre las dos fases del estudio, competencia docente medida psicométricamente y resultado estudiantil observado conductualmente constituye, a juicio de quienes desarrollan este trabajo, el aporte más original del diseño, aunque también su eslabón más frágil, la cadena causal completa, desde la competencia docente hasta el aprendizaje estudiantil, no fue modelada estadísticamente de forma directa, sino inferida a partir de la selección de docentes con puntuaciones altas.

Conviene precisar el alcance real de esta articulación para evitar lecturas causales que el diseño no respalda. Los datos de la fase instrumental, la estructura factorial, la fiabilidad y las puntuaciones en la ECD-IAES, describen exclusivamente el nivel de competencia declarado por el profesorado, no existe ninguna medición directa de cómo esa competencia se tradujo, sesión a sesión, en decisiones pedagógicas concretas dentro de la intervención. De modo semejante, los cambios observados en el rendimiento académico y en la percepción de aprendizaje, aun siendo estadísticamente significativos, constituyen evidencia empírica sobre lo ocurrido en un grupo sin condición de control, no una demostración de que la competencia docente en IA, medida por la escala, sea la causa de dichos cambios. Los dos conjuntos de resultados, el psicométrico y el cuasi-experimental, deben leerse entonces como evidencia preliminar y complementaria, articulada mediante un criterio de selección de participantes, y no como los dos extremos verificados de una cadena causal única.

El estudio presenta limitaciones que conviene reconocer explícitamente. En primer lugar, el muestreo no probabilístico por conveniencia, tanto en la fase docente como en la estudiantil, restringe la posibilidad de generalizar los resultados al conjunto del profesorado universitario ecuatoriano, incluso dentro de la propia provincia de Manabí. En segundo lugar, el diseño cuasi-experimental de un solo grupo sin condición de control impide aislar el efecto específico de la intervención pedagógica de otras explicaciones plausibles, como la maduración académica de los estudiantes durante el semestre o el efecto de novedad asociado a la experiencia. En tercer lugar, el carácter auto informado de la escala de competencias expone al instrumento a sesgos de deseabilidad social, especialmente en la dimensión ética, donde el profesorado podría sobrestimar su propio juicio crítico frente a la IA.

CONCLUSIONES

Este estudio se propuso diseñar y validar una escala multidimensional para evaluar las competencias en inteligencia artificial del profesorado de universidades públicas y privadas de Manabí, así como explorar la vinculación con una intervención pedagógica dirigida a estudiantes universitarios. El procedimiento seguido, juicio de expertos, análisis factorial exploratorio y confirmatorio, y estimación de fiabilidad mediante coeficientes complementarios, ofrece evidencias preliminares de que es posible construir un instrumento psicométricamente sólido para medir esta competencia emergente, estructurado en torno a cinco dimensiones que reflejan tanto los marcos internacionales disponibles como las particularidades del contexto universitario ecuatoriano.

La principal contribución conceptual del estudio reside en la incorporación explícita de la ingeniería de prompts educativos como dimensión diferenciada de la competencia docente en inteligencia artificial, un aspecto ausente en los instrumentos previos revisados en la literatura internacional. Esta decisión, respaldada por las evidencias factoriales propias del ejercicio metodológico desarrollado, sugiere que la capacidad de un docente para formular instrucciones efectivas a un sistema de inteligencia artificial generativa constituye un constructo distinguible de otras competencias pedagógicas, y no un simple corolario del manejo tecnológico general.

En el plano aplicado, la vinculación entre la escala docente y la intervención pedagógica dirigida a estudiantes universitarios abre una línea de indagación poco explorada hasta ahora, la posibilidad de establecer puentes empíricos entre lo que un docente sabe hacer con la inteligencia artificial y lo que efectivamente ocurre en el aprendizaje de sus estudiantes. Esta articulación, todavía incipiente en el diseño aquí propuesto, podría desarrollarse en investigaciones futuras mediante modelos estadísticos que incorporen la competencia docente como variable mediadora explícita, en lugar de como criterio de selección de los participantes.

Conviene precisar el alcance de esta conclusión, que el presente estudio no probó, en sentido estricto, que las competencias docentes en inteligencia artificial influyan sobre el aprendizaje estudiantil, dado que la muestra de docentes de la fase de intervención fue seleccionada precisamente por poseer niveles altos en dichas competencias, sin que existiera un grupo de comparación con docentes de competencia baja o media. Lo que el estudio aporta es evidencia preliminar, compatible con esa hipótesis, cuya confirmación exige diseños con mayor control, en particular estudios experimentales o cuasi-experimentales con grupo de control que contrasten intervenciones dirigidas por docentes con distintos niveles de competencia en la ECD-IAES, además de modelos de mediación estadística capaces de estimar el peso relativo de la competencia docente frente a otras variables que también pudieron influir en los resultados observados.

Las implicaciones prácticas del estudio apuntarían hacia la necesidad de que las universidades públicas y privadas de la provincia prioricen la ingeniería de prompts y la dimensión ética dentro de sus programas de formación docente continua, en lugar de tratar la competencia en inteligencia artificial como un bloque homogéneo que se adquiere de manera uniforme. La heterogeneidad esperable entre dimensiones, algunas más desarrolladas que otras según el patrón documentado en la literatura internacional revisada, sugiere que las políticas de capacitación institucional ganarían en eficacia si se diseñaran de manera diferenciada por competencia, en lugar de ofrecer talleres genéricos sobre inteligencia artificial aplicada a la docencia.

REFERENCIAS

Aiken, L. R. (1980). Content validity and reliability of single items or questionnaires. Educational and Psychological Measurement, 40(4), 955–959. https://doi.org/10.1177/001316448004000419

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

Chiu, T. K. F., Ahmad, Z., & Çoban, M. (2025). Development and validation of teacher artificial intelligence (AI) competence self-efficacy (TAICS) scale. Education and Information Technologies, 30(5), 6667–6685. https://doi.org/10.1007/s10639-024-13094-z

Dai, W., Tsai, Y.-S., Lin, J., Aldino, A., Jin, H., Li, T., Gašević, D., & Chen, G. (2024). Assessing the proficiency of large language models in automatic feedback generation: An evaluation study. Computers and Education: Artificial Intelligence, 7, 100299. https://doi.org/10.1016/j.caeai.2024.100299

Laupichler, M. C., Aster, A., Haverkamp, N., & Raupach, T. (2023). Development of the “Scale for the Assessment of Non-Experts’ AI Literacy” (SNAIL): An exploratory factor analysis. Computers in Human Behavior Reports, 12, 100338. https://doi.org/10.1016/j.chbr.2023.100338

Ledesma, R. D., Ferrando, P. J., & Tosi, J. D. (2019). Uso del análisis factorial exploratorio en RIDEP: Recomendaciones para autores y revisores. Revista Iberoamericana de Diagnóstico y Evaluación, e Avaliação Psicológica, 52(3), 173–180. https://doi.org/10.21865/RIDEP52.3.13

Lee, D., & Palmer, E. (2025). Prompt engineering in higher education: A systematic review to help inform curricula. International Journal of Educational Technology in Higher Education, 22, 7. https://doi.org/10.1186/s41239-025-00503-7

Miao, F., & Cukurova, M. (2024). AI competency framework for teachers. UNESCO. https://doi.org/10.54675/ZJTE2084

Mikeladze, T., Meijer, P. C., & Verhoeff, R. P. (2024). A comprehensive exploration of artificial intelligence competence frameworks for educators: A critical review. European Journal of Education, 59(3), Article e12663. https://doi.org/10.1111/ejed.12663

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. Computers and Education: Artificial Intelligence, 2, 100041. https://doi.org/10.1016/j.caeai.2021.100041

Redecker, C. (2017). European framework for the digital competence of educators: DigCompEdu. Publications Office of the European Union. https://doi.org/10.2760/159770

Sumba Arévalo, V. M., & Segarra Figueroa, O. P. (2026). Competencias digitales e inteligencia artificial en docentes que cursan un posgrado en educación, Ecuador. European Public & Social Innovation Review, 11, 1–19. https://doi.org/10.31637/epsir-2026-2439

Ventura-León, J. L., & Caycho-Rodríguez, T. (2017). El coeficiente Omega: Un método alternativo para la estimación de la confiabilidad [Carta al editor]. Revista Latinoamericana de Ciencias Sociales, Niñez y Juventud, 15(1), 625–627.

Wang, B., Rau, P.-L. P., & Yuan, T. (2023). Measuring user competence in using artificial intelligence: Validity and reliability of artificial intelligence literacy scale. Behaviour & Information Technology, 42(9), 1324–1337. https://doi.org/10.1080/0144929X.2022.2072768

Wang, S., Wang, F., Zhu, Z., Wang, J., Tran, T., & Du, Z. (2025). Towards responsible artificial intelligence in education: A systematic review on identifying and mitigating ethical risks. Humanities and Social Sciences Communications, 12, 1111. https://doi.org/10.1038/s41599-025-05252-6