

Competencias docentes para la inteligencia artificial generativa en educación superior: perfiles diferenciados y orientaciones para el DESARROLLO PROFESIONAL
FACULTY COMPETENCIES FOR GENERATIVE ARTIFICIAL INTELLIGENCE IN HIGHER EDUCATION: DIFFERENTIATED PROFILES AND ORIENTATIONS FOR PROFESSIONAL DEVELOPMENT

Pablo Rivarola Prados

Universidad Siglo 21, Argentina

[pablo.rivarola@ues21.edu.ar] [<https://orcid.org/0009-0001-7419-6080>]

Antonio Rafael Fernández Paradas

Universidad de Granada, España

[antonioparadas@ugr.es] [<https://orcid.org/0000-0003-3751-7479>]

Información del manuscrito:

Recibido/Received: 07/04/2026

Revisado/Reviewed: 05/06/2026

Aceptado/Accepted: 02/07/2026

RESUMEN

Palabras clave:

análisis de conglomerados, competencias docentes en IA, diseño instruccional, Educación Superior, inteligencia artificial generativa

La rápida adopción de la inteligencia artificial generativa (IAG) está transformando profundamente las prácticas de enseñanza en la educación superior. Este estudio presenta el desarrollo y validación del Cuestionario de Competencias Docentes para la IAG (CCDGAI) y analiza empíricamente los perfiles de competencia autopercibida del profesorado en tres dimensiones: tecnológica, pedagógica y ética. Se adoptó un diseño cuantitativo, no experimental y transversal con N=410 docentes universitarios de una institución de educación superior de gran escala. La validez de contenido se evaluó mediante el coeficiente V de Aiken con intervalos de confianza al 95% según el método de intervalo de Penfield & Giacobbini (2004) obtenido de un panel de cinco jueces expertos (V media=.91; rango .80–1.00). La consistencia interna resultó excelente (α total=.973; IC95: .968–.977). El análisis factorial exploratorio mostró evidencia preliminar de una estructura trifactorial (KMO=.968; Bartlett $p<.001$). Mediante análisis de conglomerados jerárquico-k-medias se identificaron tres perfiles diferenciados: alta competencia (n=140; 34.1%), competencia media (n=164; 40.0%) y baja competencia (n=106; 25.9%), con apoyo estadístico adecuado (silueta=.378; ARI-bootstrap=.821; $\eta^2>.67$). La competencia ética resultó la más desarrollada; las dimensiones tecnológica y pedagógica evidencian las brechas formativas más pronunciadas. Los hallazgos ofrecen un diagnóstico basado en evidencia para orientar el desarrollo profesional docente diferenciado y las estrategias institucionales de gobernanza de la IAG.

ABSTRACT

Keywords:

cluster analysis, faculty competencies in AI, generative

The rapid adoption of generative artificial intelligence (GAI) is profoundly transforming teaching practices in higher education, posing urgent challenges for instructional design, authentic assessment, and ethical governance. This study presents the

artificial intelligence, higher education, instructional design	development and validation of the Faculty Competencies for GAI Questionnaire (CCDGAI) and empirically analyzes faculty self-perceived competency profiles across three dimensions: technological, pedagogical, and ethical. A quantitative, non-experimental, cross-sectional design was employed with N=410 university instructors from a single large-scale higher education institution. Content validity was assessed via Aiken's V with 95% confidence intervals calculated following the interval method proposed by Penfield & Giacobbi (2004) for ordinal polytomous scales, from a panel of five expert judges (mean V=.91; range .80–1.00). Internal consistency was excellent across all subscales (total α=.973; 95%CI: .968–.977). Exploratory factor analysis yielded a three-factor solution preliminarily compatible with the proposed structure (KMO=.968; Bartlett p<.001); the bifactor model showed superior fit compared to alternative configurations. Hierarchical k-means clustering identified three distinct competency profiles: high competence (n=140; 34.1%), medium competence (n=164; 40.0%), and low competence (n=106; 25.9%), with adequate statistical support (silhouette=.378; bootstrap ARI=.821; η^2>.67). Ethical competence was the most developed dimension across all profiles, while technological and pedagogical competencies showed the most pronounced training gaps. Findings offer evidence-based diagnostics to support differentiated faculty professional development and responsible institutional GAI governance.
---	---

Introducción

La rápida expansión de la inteligencia artificial generativa (IAG), impulsada por la irrupción de los modelos de lenguaje de gran escala (LLM), está reconfigurando los fundamentos del trabajo académico en la educación superior. A través de la interacción conversacional en lenguaje natural, estos sistemas posibilitan la generación automatizada de textos, explicaciones, retroalimentación personalizada y materiales de aprendizaje adaptados, con implicaciones directas sobre el diseño instruccional, las prácticas evaluativas y la integridad académica (EDUCAUSE, 2025a, 2025b; Belkina et al., 2025). A diferencia de tecnologías educativas anteriores que operaban como herramientas puntuales, la IAG interviene de forma horizontal en todas las etapas del proceso de enseñanza-aprendizaje: desde la planificación curricular hasta la evaluación, pasando por la generación de contenido y la retroalimentación formativa (Hwang et al., 2020; Weng et al., 2024; Xia et al., 2024).

La investigación empírica acumulada indica que la integración educativa de la IAG depende menos de la disponibilidad tecnológica que de las competencias docentes para comprender críticamente estos sistemas, orquestarlos pedagógicamente y tomar decisiones instruccionales éticamente informadas (Mah & Groß, 2024; Burneo-Arteaga et al., 2025). Los estudios disponibles muestran niveles heterogéneos de preparación, con brechas pronunciadas en la integración pedagógica y en la toma de decisiones instruccionales (Baig & Yadegaridehkordi, 2025; Dringó-Horváth et al., 2025). Esta heterogeneidad ha llevado a organismos internacionales como la UNESCO y la OCDE a articular marcos de competencias explícitos que conceptualizan la competencia en IA para docentes como un constructo multidimensional, abarcando comprensión tecnológica, intencionalidad pedagógica y conciencia ética (Miao & Cukurova, 2024; UNESCO, 2024; OECD, 2025). Junto a estas oportunidades, una creciente literatura ha documentado riesgos de la adopción acrítica de la IAG: desafíos de integridad académica, generación de contenido impreciso, opacidad algorítmica y transformación no planificada de las prácticas evaluativas (Cotton et al., 2024; Xia et al., 2024).

Revisiones sistemáticas y metaanálisis recientes confirman que los efectos de la IAG en los resultados de aprendizaje no son uniformes, sino que están mediados por las decisiones pedagógicas del docente (Crompton & Burke, 2023; Deng et al., 2024; Liang et al., 2025; Zawacki-Richter et al., 2019). Sin embargo, la evidencia empírica sobre cómo los perfiles diferenciados de competencia docente condicionan las prácticas instruccionales concretas sigue siendo escasa (Belkina et al., 2025; Garzón et al., 2025). La literatura disponible se centra predominantemente en actitudes, percepciones o patrones generales de uso, sin ofrecer diagnósticos estructurados que puedan orientar decisiones de desarrollo profesional o políticas institucionales. Este artículo responde a esa laguna mediante el desarrollo y validación del CCDGAI y la identificación empírica de perfiles diferenciados de competencia en una muestra de profesores universitarios.

El estudio persigue tres objetivos: (O1) desarrollar el CCDGAI y aportar evidencias iniciales de validez y fiabilidad para el diagnóstico de competencias docentes en IAG; (O2) identificar perfiles diferenciados de competencia mediante análisis de conglomerados con criterios complementarios de validación interna; y (O3) discutir las implicaciones de los perfiles para el diseño instruccional, el desarrollo profesional docente y la gobernanza institucional de la IAG.

Contribución teórica y posicionamiento del estudio

La presente investigación articula cuatro contribuciones diferenciadas. Primero, presenta el desarrollo y validación inicial del Cuestionario de Competencias Docentes

para la IAG (CCDGAI), un instrumento de autoinforme con 28 ítems organizados en tres dimensiones teóricamente fundamentadas y con evidencia empírica preliminar. Segundo, aplica un protocolo de validación de contenido robusto que combina el coeficiente V de Aiken con intervalos de confianza al 95% calculados mediante el método de intervalo propuesto por Penfield y Giacobbi (2004), junto con medidas complementarias de acuerdo entre jueces. Tercero, propone un análisis factorial exploratorio con rotación oblimin y un análisis factorial confirmatorio comparativo de tres modelos alternativos (unidimensional, trifactorial y bifactor), permitiendo identificar la estructura interna que mejor se ajusta a los datos. Cuarto, deriva una tipología empírica de perfiles docentes mediante análisis de conglomerados jerárquico-k-medias con validación multi-índice y estabilidad por bootstrap del Adjusted Rand Index, articulando los perfiles con orientaciones formativas e institucionales contextualizadas.

Método

Diseño de investigación

Se adoptó un diseño cuantitativo, no experimental y transversal orientado a la descripción y clasificación de perfiles de competencia docente para el uso educativo de la IAG. El componente de análisis de conglomerados —de carácter exploratorio— responde al objetivo de identificar tipologías latentes que van más allá de la descripción univariada. El reporte sigue las recomendaciones de la guía STROBE (von Elm et al., 2007) adaptadas a estudios transversales con instrumentos de autoinforme.

Participantes

La muestra estuvo compuesta por N=410 docentes con cuestionarios completamente válidos, provenientes de n=421 participantes de una institución de educación superior de gran escala con modalidad distribuida. Los 11 casos excluidos presentaron respuestas omitidas en al menos un ítem (tasa de exclusión=2.6%). Dado que la modalidad de distribución no permite verificar con precisión el número total de docentes contactados, no es posible calcular una tasa de respuesta global; se reporta en su lugar la tasa de completitud de los cuestionarios recibidos (97.4%).

La distribución por género fue: femenino 56.8% (n=233), masculino 42.2% (n=173), prefiere no responder 1.0% (n=4). La distribución etaria se concentró en las franjas de 45–54 años (38.3%) y 35–44 años (34.4%). El nivel máximo de formación se distribuyó entre grado/licenciatura (52.2%), posgrado (46.3%) y otras titulaciones (1.5%). Los participantes representaron un amplio espectro disciplinar (ciencias de la salud, ciencias jurídicas, ingeniería, ciencias de la educación, ciencias económicas). El tamaño muestral satisface los criterios recomendados para análisis de conglomerados (N>200; Hair et al., 2019) y análisis factorial (ratio $\geq 10:1$ sujetos/ítem; Comrey & Lee, 1992).

Instrumento: CCDGAI

El Cuestionario de Competencias Docentes para la IAG (CCDGAI) es un instrumento de autoinforme desarrollado ad hoc, fundamentado en marcos de competencia digital docente (Redecker, 2017) y en literatura empírica reciente (Miao & Cukurova, 2024; Mah & Groß, 2024; Ng et al., 2021). Conceptualiza la competencia docente en IAG como constructo tridimensional: (a) Competencia tecnológica (10 ítems): comprensión del funcionamiento de los sistemas de IAG, identificación de herramientas, evaluación de fiabilidad y conocimiento sobre sesgos algorítmicos. (b) Competencia pedagógica (10 ítems): integración intencional de la IAG en el diseño instruccional, recursos evaluativos,

retroalimentación formativa y apoyo a la autorregulación. (c) Competencia ética (8 ítems): conciencia crítica en relación con privacidad, consentimiento, transparencia, equidad de acceso y límites del apoyo algorítmico. Los 28 ítems del instrumento final se formularon como afirmaciones declarativas en escala Likert de cinco puntos (1=totalmente en desacuerdo; 5=totalmente de acuerdo). Un ítem adicional fue eliminado durante la validación de contenido ($V=.68$; $IC95: .57-.77$).

Validación de contenido

El CCDGAI fue sometido a validación de contenido mediante juicio de expertos (Sireci & Faulkner-Bond, 2014), cuantificada con el coeficiente V de Aiken (Aiken, 1985). El panel estuvo integrado por cinco especialistas con perfiles complementarios en psicología y metodología cuantitativa, didáctica, ingeniería en IA y tecnología educativa. Cada juez evaluó los 29 ítems en cuatro criterios: claridad, relevancia, pertinencia y coherencia (escala 1–5). Los intervalos de confianza al 95% se calcularon mediante el método de intervalo de Penfield & Giacobbi (2004) para el coeficiente V de Aiken con escalas ordinales politómicas. El criterio de aceptación fue $V \geq .75$ o que el límite inferior del $IC95\%$ superara $.70$. Como medidas complementarias se calcularon el porcentaje de aceptación unánime por unidad ítem-criterio, el coeficiente de correlación intraclass $ICC(2,k)$ y el kappa de Fleiss sobre la decisión dicotómica aceptar (≥ 4) versus rechazar (< 4).

Procedimiento de recogida de datos

La recogida de datos se realizó durante el segundo semestre académico de 2024 mediante cuestionario en línea. La participación fue estrictamente voluntaria y anónima; el consentimiento informado fue obtenido electrónicamente mediante una declaración explícita al inicio del cuestionario, antes de acceder a los ítems. Los datos fueron tratados de forma confidencial y agregada. El estudio fue conducido de acuerdo con los principios éticos de la American Psychological Association (APA, 2017) para la investigación con participantes humanos en ciencias sociales, incluyendo voluntariedad, anonimato y uso exclusivo de los datos para fines científicos. El protocolo de investigación recibió aprobación del comité de ética institucional correspondiente; los datos específicos del comité están disponibles para los revisores bajo solicitud.

Análisis estadístico

El análisis de datos se desarrolló en tres fases en Python 3.11 (pandas, numpy, scipy, scikit-learn, factor_analyzer). *Fase 1 — Fiabilidad:* se calculó el alpha de Cronbach con $IC95\%$ bootstrap no paramétrico ($B=2000$; Efron & Tibshirani, 1993), el coeficiente omega de McDonald y correlaciones ítem-total corregidas. *Fase 2 — Análisis factorial exploratorio:* se verificaron los supuestos KMO (Kaiser, 1974) y Bartlett, y se aplicó AFE de tres factores (máxima verosimilitud, rotación oblimin); las comunales se calcularon a partir del producto de las matrices de patrón y estructura. Para examinar la estructura interna se compararon tres modelos confirmatorios (CFA): unidimensional, trifactorial correlacionado y bifactor, con estimación ML mediante la biblioteca semopy. *Fase 3 — Análisis de conglomerados:* las puntuaciones dimensionales fueron estandarizadas y se aplicó clustering jerárquico Ward + refinamiento k-medias ($n_{init}=500$; semilla=42). La selección de k se fundamentó en los índices de silueta (Rousseeuw, 1987), Calinski-Harabasz y Davies-Bouldin; la estabilidad se estimó con ARI bootstrap ($B=300$; Steinley, 2004). Las diferencias entre perfiles se evaluaron con ANOVA y eta cuadrado (η^2 ; Cohen, 1988).

Resultados

Análisis descriptivo

La Tabla 1 presenta los estadísticos descriptivos de los 28 ítems. A nivel dimensional, la competencia ética obtuvo las puntuaciones más elevadas ($M=3.18$; $DE=1.08$), seguida por la tecnológica ($M=2.87$; $DE=0.99$) y la pedagógica ($M=2.86$; $DE=1.08$), situando a la muestra en un nivel de competencia moderado-bajo ($M_{\text{total}}=2.95$; $DE=0.97$). En la dimensión tecnológica, los ítems de mayor puntuación correspondieron a identificar modelos disponibles (T05) y usar IAG para docencia (T08), mientras que los de comprensión técnica profunda —LLMs (T02; $CV=53.4\%$) y redes neuronales (T06; $CV=50.1\%$)— presentaron los valores más bajos, evidenciando que la familiaridad con la interfaz de uso supera la comprensión de los mecanismos subyacentes. En la dimensión pedagógica, los ítems de mayor puntuación correspondieron a la generación de materiales personalizados mediante prompts (P01) y la generación de preguntas de evaluación alineadas a objetivos (P04), mientras que los de aplicación pedagógica compleja —actividades gamificadas (P06; $CV=47.3\%$) y autorregulación académica (P08; $CV=45.7\%$)— mostraron las puntuaciones más bajas. En la dimensión ética, los ítems sobre límites de la IAG frente al apoyo humano (E05) y análisis crítico de respuestas (E08) fueron los más altos, con mayor variabilidad relativa en privacidad estudiantil (E03; $CV=49.3\%$).

Tabla 1
Estadísticos descriptivos del CCDGAI por ítem (N=410)

Dimensión	Cód.	Ítem	M	DE	Med	IQR	CV%
Tecnológica	T01	Diferencia entre IAG y aprendizaje automático	3.05	1.34	3	[2–4]	44.0
	T02	Conoce qué es un LLM y cómo funciona	2.44	1.30	2	[1–3]	53.4
	T03	Describe cómo el modelo transforma entradas en salidas	2.81	1.29	3	[2–4]	46.0
	T04	Entiende el entrenamiento con grandes conjuntos de datos	2.91	1.30	3	[2–4]	44.6
	T05	Identifica modelos disponibles (GPT, Claude, Gemini)	3.47	1.27	4	[3–4]	36.7
	T06	Sabe cómo funcionan las redes neuronales profundas	2.40	1.20	2	[1–3]	50.1
	T07	Comprende las implicaciones del sesgo algorítmico	2.67	1.32	3	[1–4]	49.6
	T08	Usa herramientas IAG para docencia y	3.15	1.13	3	[2–4]	35.9

Competencias docentes para la inteligencia artificial generativa en educación superior: perfiles diferenciados y orientaciones para el desarrollo profesional

Dimensión	Cód.	Ítem	M	DE	Med	IQR	CV%
Pedagógica		desarrollo profesional					
	T09	Diferencia modelos de código abierto y propietarios	2.67	1.31	3	[2-4]	49.0
	T10	Evalúa la fiabilidad de resultados generados por IAG	3.11	1.26	3	[2-4]	40.5
	P01	Usa prompts para generar materiales personalizados	3.20	1.31	3	[2-4]	40.8
	P02	Diseña prompts para casos o simulaciones	2.92	1.30	3	[2-4]	44.4
	P03	Usa IAG para planificación y estructuración de clases	2.89	1.31	3	[2-4]	45.2
	P04	Genera preguntas de evaluación alineadas a objetivos	2.98	1.34	3	[2-4]	45.1
	P05	Usa IAG para corrección y retroalimentación	2.71	1.27	3	[2-4]	46.8
	P06	Diseña actividades gamificadas con IAG	2.65	1.25	3	[2-4]	47.3
	P07	Genera ejercicios adaptados a distintos niveles	2.84	1.28	3	[2-4]	44.9
Ética	P08	Apoya la autorregulación académica con IAG	2.71	1.24	3	[2-4]	45.7
	P09	Diseña actividades de pensamiento crítico	2.91	1.27	3	[2-4]	43.8
	P10	Crea ejercicios interactivos mediante prompts	2.77	1.27	3	[2-4]	45.9
	E01	Reconoce el sesgo en IAG al evaluar estudiantes	3.23	1.30	3	[2-4]	40.2
	E02	Reconoce implicaciones éticas de la	3.34	1.28	4	[2-4]	38.3

Dimensión	Cód.	Ítem	M	DE	Med	IQR	CV%
		evaluación automatizada					
	E03	Garantiza la privacidad estudiantil en el uso de IAG	2.52	1.24	2	[1-3]	49.3
	E04	Informa sobre los datos recolectados y el consentimiento	2.73	1.32	3	[2-4]	48.4
	E05	Comprende los límites de la IAG frente al apoyo humano	3.50	1.33	4	[3-5]	38.1
	E06	Promueve el uso inclusivo para reducir la brecha digital	3.43	1.24	4	[3-4]	36.0
	E07	Selecciona recursos IAG que sean inclusivos	3.19	1.34	3	[2-4]	42.0
	E08	Fomenta el análisis crítico de respuestas de IAG	3.50	1.30	4	[3-5]	37.2

Nota. M = media; DE = desviación estándar; Med = mediana; IQR = rango intercuartil (Q1-Q3); CV% = coeficiente de variación. IAG = inteligencia artificial generativa.

Validación de contenido: V de Aiken con IC 95%

La Tabla 2 presenta los resultados. El coeficiente V de Aiken global fue de .91 (rango: .80–1.00), indicando acuerdo inter-jueces excelente. De los 29 ítems de la versión preliminar, 28 superaron el criterio de aceptación. Un ítem de la dimensión ética no alcanzó el umbral en su formulación original ($V=.68$; IC95: .57–.77) por ambigüedad y solapamiento conceptual, y fue eliminado. A partir de las observaciones cualitativas de los jueces, se redactó una nueva formulación que fue incorporada como E04 en la versión final ($V=.80$; IC95: .70–.87). Aunque E04 superó el criterio de aceptación, sus valores en pertinencia ($V=.70$) y coherencia ($V=.70$) se ubican en el límite inferior del rango aceptable, característica que deberá considerarse en la interpretación de los resultados en fases posteriores. La versión final del instrumento quedó compuesta por 28 ítems.

El análisis por dimensión revela patrones diferenciados. La dimensión Tecnológica obtuvo una V media de .91 (rango .84–1.00); la Pedagógica presentó la mayor homogeneidad ($V=.93$; rango .85–1.00); la Ética mostró la mayor variabilidad ($V=.89$; rango .80–1.00), resultado esperable dada la complejidad conceptual de los constructos éticos asociados al uso de IAG en educación superior. Como medidas complementarias, el porcentaje de aceptación unánime por unidad ítem-criterio fue del 60.2%. Los coeficientes dependientes de variabilidad entre valoraciones —ICC(2,k)=.118 y kappa de Fleiss— resultaron bajos pese a una concordancia observada del 89.8%, un resultado atribuible a la paradoja de alta concordancia: cuando la mayoría de los jueces asigna puntuaciones en el rango superior, la restricción de rango deflacta artificialmente los índices basados en varianza. Estos indicadores se informan como análisis complementarios; la V de Aiken con IC95% constituye el criterio central de decisión.

Tabla 2

Validez de contenido: coeficiente V de Aiken con IC 95% (N=5 jueces; k=5 categorías; c=4)

D	Cód.	V Clar.	V Relev.	V Pert.	V Coher.	V media	IC 95%	Dec.
T	T01	0.80	0.90	0.80	0.95	0.86	[0.77-0.92]	✓
T	T02	0.90	0.95	1.00	1.00	0.96	[0.90-0.99]	✓
T	T03	1.00	1.00	1.00	1.00	1.00	[0.95-1.00]	✓
T	T04	0.90	0.80	0.85	0.80	0.84	[0.74-0.90]	✓
T	T05	0.95	0.95	0.95	0.95	0.95	[0.88-0.98]	✓
T	T06	0.95	0.80	0.85	0.80	0.86	[0.76-0.91]	✓
T	T07	0.80	0.95	0.95	0.95	0.91	[0.83-0.96]	✓
T	T08	1.00	1.00	0.95	1.00	0.99	[0.93-1.00]	✓
T	T09	0.90	0.85	0.85	0.90	0.88	[0.79-0.93]	✓
T	T10	0.90	0.80	0.90	0.80	0.85	[0.76-0.91]	✓
P	P01	0.95	0.95	0.95	0.95	0.95	[0.88-0.98]	✓
P	P02	1.00	1.00	1.00	1.00	1.00	[0.95-1.00]	✓
P	P03	0.90	0.95	0.95	0.95	0.94	[0.86-0.97]	✓
P	P04	1.00	0.90	0.80	0.95	0.91	[0.83-0.96]	✓
P	P05	0.95	0.95	0.95	0.95	0.95	[0.88-0.98]	✓
P	P06	0.75	0.95	0.95	0.95	0.90	[0.82-0.95]	✓
P	P07	0.80	0.90	0.85	0.90	0.86	[0.77-0.92]	✓
P	P08	0.88	0.88	0.88	0.88	0.88	[0.77-0.94]	✓
P	P09	0.85	0.85	0.85	0.85	0.85	[0.75-0.93]	✓
P	P10	0.90	0.85	0.85	0.90	0.88	[0.79-0.93]	✓
E	E01	1.00	1.00	1.00	1.00	1.00	[0.95-1.00]	✓
E	E02	1.00	1.00	1.00	1.00	1.00	[0.95-1.00]	✓
E	E03	0.95	0.95	0.95	0.95	0.95	[0.88-0.98]	✓
E	E04*	1.00	0.80	0.70	0.70	0.80	[0.70-0.87]	✓
E	E05	1.00	1.00	1.00	1.00	1.00	[0.95-1.00]	✓
E	E06	0.95	0.80	0.80	0.90	0.86	[0.77-0.92]	✓
E	E07	1.00	1.00	1.00	1.00	1.00	[0.95-1.00]	✓
E	E08	0.90	0.80	0.80	0.80	0.83	[0.73-0.89]	✓
E	Eliminado	0.65	0.70	0.70	0.65	0.68	[0.57-0.77]	✗

Nota. D = dimensión (T=Tecnológica, P=Pedagógica, E=Ética). IC95% calculado según el método de intervalo propuesto por Penfield & Giacobbi (2004) para el coeficiente V de Aiken con escalas ordinales politómicas (k=5 categorías; c=k-1=4). Criterio de aceptación: $V \geq .75$ o LI del IC95% $\geq .70$ (Aiken, 1985; Penfield & Giacobbi, 2004). ✓ = ítem aceptado; ✗ = ítem eliminado. E04* = ítem reformulado tras revisión cualitativa de observaciones de los jueces.

Fiabilidad

La Tabla 3 presenta los índices de consistencia interna. Todos los coeficientes alpha superaron el umbral de .70 (Nunnally & Bernstein, 1994): la escala total alcanzó $\alpha=.973$ [IC95: .968-.977] y las subescalas oscilaron entre $\alpha=.929$ (tecnológica) y $\alpha=.967$

(pedagógica). Las estimaciones omega resultaron congruentes. Las correlaciones ítem-total corregidas oscilaron entre .608 (T05) y .855 (P04), todas superiores al umbral de .30 (DeVellis, 2017). El análisis de alpha-si-se-elimina-el-ítem mostró que la eliminación de cualquier ítem no produciría incrementos sustanciales ($\Delta\alpha$ máximo=.002).

Tabla 3*Índices de fiabilidad por subescala (N=410)*

Subescala	k	α	IC 95% Bootstrap	ω McDonald	r_it rango
Tecnológica	10	0.929	[0.917-0.939]	.936	0.601-0.801
Pedagógica	10	0.967	[0.961-0.972]	.963	0.807-0.883
Ética	8	0.930	[0.917-0.941]	.929	0.691-0.806
Total (28 ítems)	28	0.973	[0.968-0.977]	.977	0.608-0.855

Nota. k = n.º ítems. α = alpha de Cronbach. IC95% = bootstrap no paramétrico (B=2000). ω = coeficiente omega de McDonald basado en CFA unidimensional por subescala. r_it = correlaciones ítem-total corregidas. $\alpha > .90$ = consistencia interna muy alta (Cicchetti, 1994).

Análisis factorial exploratorio

Los supuestos resultaron óptimos: KMO=.968 y Bartlett $\chi^2(378)=10984.8$, $p<.001$. Se aplicó un AFE de tres factores (ML, oblimin). La varianza común explicada fue del 55.3% (F1=25.6%; F2=15.5%; F3=14.2%). El Factor 1 agrupa los ítems de la dimensión pedagógica; el Factor 2, los de la dimensión tecnológica; y el Factor 3, los de la dimensión ética, configuración mayormente coherente con la estructura teórica propuesta, aunque algunos ítems presentaron cargas moderadas o de frontera conceptual (T05, T08, T10 en la dimensión tecnológica; E03 en la dimensión ética). Las cargas factoriales por ítem se presentan en la Tabla 4. Las comunales oscilaron entre $h^2=.382$ y $h^2=.809$, sin soluciones de Heywood. Las correlaciones interfactoriales fueron elevadas ($\phi_{F1F2}=.676$; $\phi_{F1F3}=.715$; $\phi_{F2F3}=.563$). El análisis factorial confirmatorio comparativo mostró que el modelo bifactor (G + tres factores específicos) presenta el mejor ajuste relativo ($\chi^2=1309.3$, gl=322, CFI=.910, RMSEA=.087), superando al trifactorial correlacionado (CFI=.847) y al unidimensional (CFI=.772). No obstante, los índices absolutos no alcanzan los umbrales convencionales (CFI $\geq$.95; RMSEA $\leq$.06; Hu & Bentler, 1999), por lo que la solución debe interpretarse como evidencia preliminar de la estructura subyacente, requiriendo verificación confirmatoria con estimador WLSMV en muestras independientes (Redecker, 2017; Ng et al., 2021).

Tabla 4

Cargas factoriales del AFE oblicuo (3 factores, ML + oblimin; N=410)

D	Cód.	F1	F2	F3	h ²
T	T01	0.237	0.511	0.028	0.508
T	T02	0.114	0.761	-0.054	0.658
T	T03	0.163	0.664	0.021	0.634
T	T04	0.035	0.768	0.058	0.683
T	T05	0.163	0.343	0.199	0.382
T	T06	0.057	0.870	-0.149	0.691
T	T07	-0.169	0.812	0.211	0.688
T	T08	0.321	0.324	0.210	0.567
T	T09	0.063	0.642	0.075	0.538
T	T10	0.164	0.359	0.337	0.564
P	P01	0.693	0.054	0.113	0.665
P	P02	0.764	0.064	0.082	0.757
P	P03	0.754	-0.044	0.171	0.729
P	P04	0.739	0.063	0.143	0.794
P	P05	0.901	-0.013	-0.072	0.710
P	P06	0.841	0.065	-0.052	0.722
P	P07	0.910	0.035	-0.068	0.785
P	P08	0.921	0.006	-0.072	0.766
P	P09	0.785	-0.033	0.172	0.798
P	P10	0.838	0.041	0.045	0.809
E	E01	0.001	0.246	0.640	0.648
E	E02	0.033	0.136	0.684	0.631
E	E03	0.287	0.346	0.242	0.589
E	E04	0.140	0.246	0.434	0.522
E	E05	-0.051	0.009	0.886	0.732
E	E06	0.112	-0.061	0.776	0.680
E	E07	0.251	-0.035	0.679	0.730
E	E08	0.070	0.000	0.775	0.684

Nota. F1-F3 = factores rotados (oblimin). h² = comunalidad calculada como producto de matrices patrón y estructura. La varianza total explicada fue del 55.3% (F1=25.6%; F2=15.5%; F3=14.2%). Cargas > .40 son prácticamente significativas (Tabachnick & Fidell, 2019). Correlaciones interfactoriales (ϕ): F1-F2=.676; F1-F3=.715; F2-F3=.563.

Perfiles de competencia docente

La Tabla 5 muestra los índices de evaluación para k=2 a k=6. El índice de silueta alcanzó su valor máximo en k=2 (.474); sin embargo, esa solución no permite capturar la heterogeneidad multidimensional esperada desde el modelo teórico ni distinguir el perfil intermedio con implicaciones formativas diferenciadas. K=3 fue seleccionado por criterio combinado estadístico-teórico-pragmático: mantiene una silueta moderada (.378; Rousseeuw, 1987) con un Calinski-Harabasz de 535.6 y Davies-Bouldin de .895, mientras que las soluciones k≥4 muestran caída monótona en silueta y aumento progresivo de Davies-Bouldin; teóricamente, k=2 mezclaría perfiles intermedios y extremos, ocultando estratificación formativa relevante; pragmáticamente, k=3 explica el 72.5% de la varianza

inter-grupos y produce tres perfiles interpretables (Milligan & Cooper, 1987). La estabilidad fue adecuada: ARI bootstrap=.821 [IC95: .567-.992; B=300]. El perfil de alta competencia (n=140; 34.1%) exhibe puntuaciones elevadas y consistentes en las tres dimensiones (M_T=3.90; M_P=4.02; M_E=4.12). El perfil de competencia media (n=164; 40.0%) presenta el gradiente ética (M=3.17) > tecnológica (M=2.80) ≈ pedagógica (M=2.84), revelando docentes con conciencia ética desarrollada pero fluidez tecnológica y pedagógica limitada. El perfil de baja competencia (n=106; 25.9%) muestra puntuaciones por debajo de 2 en la dimensión pedagógica (M=1.53). El ANOVA confirmó diferencias estadísticamente significativas entre los tres perfiles en las tres dimensiones ($F>423$; $p<.001$; $\eta^2>.67$), con diferencias significativas entre todos los pares (Games-Howell, $p<.001$). Dado que los perfiles se derivaron de las propias puntuaciones dimensionales, estos valores de η^2 deben interpretarse como evidencia descriptiva de separación entre conglomerados, no como validación externa independiente. No se observó asociación entre perfil y género ($\chi^2=1.88$; $gl=2$; $p=.390$; se excluyeron los cuatro casos de la categoría "prefiere no responder" por frecuencias esperadas insuficientes).

Tabla 5

Perfiles de competencia docente: tamaño, medias e índices de validación (N=410)

Perfil / k	n (%) / Silueta	M Tec. / C-H	M Ped. / D-B	M Ética	Decisión
k = 2	Sil = 0.474	C-H = 566.8	D-B = 0.754	—	Máx. silueta
k = 3	Sil = 0.378	C-H = 535.6	D-B = 0.895	—	Seleccionado*
k = 4	Sil = 0.336	C-H = 478.0	D-B = 0.965	—	—
k = 5	Sil = 0.322	C-H = 452.8	D-B = 1.037	—	—
k = 6	Sil = 0.316	C-H = 422.7	D-B = 1.055	—	—
Alta competencia	140 (34.1%)	3.90 (0.57)	4.02 (0.49)	4.12 (0.59)	—
Competencia media	164 (40.0%)	2.80 (0.49)	2.84 (0.55)	3.17 (0.57)	—
Baja competencia	106 (25.9%)	1.73 (0.53)	1.53 (0.46)	1.80 (0.69)	—
ANOVA F(2,407)	p < .001	504.42 / $\eta^2=.713$	721.65 / $\eta^2=.780$	434.86 / $\eta^2=.681$	—

Nota. C-H = Calinski-Harabász; D-B = Davies-Bouldin. *k=3 seleccionado por criterio combinado estadístico-teórico. M = media (DE). ARI-bootstrap=.821 [IC95: .567-.992; B=300]. η^2 = eta cuadrado (Cohen, 1988): grande = $\eta^2 > .14$.

Discusión y conclusiones

Los resultados aportan evidencia empírica sobre la heterogeneidad de las competencias docentes para el uso educativo de la IAG en la educación superior. Cuatro hallazgos centrales articulan la discusión.

El primero concierne a la asimetría sistemática entre la competencia ética —la más desarrollada en todos los perfiles— y las competencias tecnológica y pedagógica. Este patrón, coherente con hallazgos recientes en contextos europeos y asiáticos (Dringó-

Horváth et al., 2025; Baig & Yadegaridehkordi, 2025), puede interpretarse como producto de la circulación del discurso ético sobre la IA en los espacios institucionales y mediáticos: el profesorado ha sido expuesto a debates sobre sesgos algorítmicos, privacidad y riesgos antes de haber desarrollado la competencia técnica para comprenderlos o la pedagógica para actuar sobre ellos. Esta asimetría constituye tanto un activo —una base normativa sobre la que construir competencia instrumental— como un riesgo: el compromiso ético declarativo, sin sustrato tecnológico y pedagógico, puede permanecer en la retórica sin traducirse en prácticas instruccionales efectivas (Cotton et al., 2024; Inter-American Development Bank, 2025). Que incluso el perfil de baja competencia global presente valores éticos superiores a los tecnológicos y pedagógicos sugiere que esta asimetría no es un artefacto de la muestra, sino un patrón recurrente en la adopción institucional de la IAG.

El segundo hallazgo se refiere a las brechas en la comprensión técnica profunda. Los ítems sobre LLMs (T02; CV=53.4%) y redes neuronales profundas (T06; CV=50.1%) son los de puntuación más baja en todo el instrumento, con coeficientes de variación superiores al 50%, indicativos de marcada heterogeneidad intramuestral. La brecha entre conocimiento de uso —saber operar la herramienta— y conocimiento epistémico —comprender cómo funciona— es determinante para la evaluación crítica de outputs de IAG y para el diseño de actividades que promuevan la agencia intelectual del estudiante en lugar de la delegación pasiva (Kalantzis & Cope, 2025; Kasneci et al., 2023). Sin competencia técnica subyacente, la evaluación crítica de las respuestas de la IAG y el diseño de tareas que aprovechen sus capacidades de forma pedagógicamente intencional pueden verse significativamente limitados.

El tercer hallazgo son los tres perfiles diferenciados e interpretables ($ARI=.821$; $\eta^2>.67$). El perfil de competencia media —el más numeroso (40.0%)— merece atención especial: su configuración (ética>tecnológica≈pedagógica) lo posiciona como el principal destinatario de intervenciones que vinculen el compromiso ético con la práctica instruccional concreta. El perfil de baja competencia (25.9%) requiere alfabetización básica en IAG antes de abordar la integración pedagógica sofisticada.

El cuarto hallazgo es la ausencia de asociación estadísticamente significativa entre el perfil de competencia y el género del docente ($\chi^2=1.88$; $gl=2$; $p=.390$; $n=406$, excluidos cuatro casos de "prefiere no responder"). Como análisis exploratorio complementario, el modelo trifactorial confirmatorio se estimó por separado en mujeres ($n=233$) y hombres ($n=173$), observándose índices de ajuste comparables aunque insuficientes para establecer invarianza de medición (mujeres: $CFI=.827$, $RMSEA=.112$; hombres: $CFI=.838$, $RMSEA=.120$); la evaluación formal de invarianza configural, métrica y escalar queda pendiente para estudios confirmatorios. Este resultado contrasta con algunos estudios sobre competencia digital general (Dringó-Horváth et al., 2025), pero es consistente con investigaciones recientes sobre IAG en educación superior que no encuentran efecto significativo de género cuando se controla por área disciplinar y acceso a formación (Ayyoub et al., 2025). Las implicaciones son directas para el diseño de programas de desarrollo profesional: los datos no justifican una diferenciación de las intervenciones basada en género, sino en el diagnóstico previo del perfil de competencia de cada docente.

La estructura bifactor, que mostró el mejor ajuste relativo entre los modelos comparados (factor general G + tres factores específicos), sugiere que, en la fase actual de penetración de la IAG en la educación superior, la competencia docente autopercibida se experimenta como un constructo en gran parte unitario: los docentes que se perciben competentes en lo tecnológico tienden a reportar también mayor competencia pedagógica y ética. Esto puede deberse a que la familiarización con la IAG se da típicamente como una experiencia integradora donde lo técnico, didáctico y reflexivo

ocurren simultáneamente. Sin embargo, la solución exploratoria y la comparación de modelos sugieren que las dimensiones teóricas conservan diferenciación empírica preliminar. Esta configuración tiene implicaciones formativas: las intervenciones que solo abordan una dimensión podrían tener menor efectividad que aquellas que articulan las tres de manera integrada.

Implicaciones para el diseño instruccional y el desarrollo profesional

Los perfiles tienen implicaciones diferenciadas. Para el perfil de alta competencia: intervenciones de diseño instruccional avanzado (orquestración de tareas complejas con IAG, simulaciones, retroalimentación adaptativa, evaluación auténtica) y liderazgo pedagógico como facilitadores del desarrollo de sus pares. Para el perfil de competencia media: intervenciones que traduzcan la conciencia ética en práctica instruccional concreta, con énfasis en el diseño de actividades con intencionalidad curricular explícita y rediseño de tareas evaluativas. Para el perfil de baja competencia: itinerarios de alfabetización básica en IAG antes de abordar la integración pedagógica. A nivel institucional, estos hallazgos refuerzan la necesidad de diagnósticos de punto de partida y rutas de desarrollo diferenciadas, en lugar de estrategias uniformes (Liu & Bates, 2025).

Limitaciones

Este estudio presenta varias limitaciones. Primera, el diseño de autoinforme es susceptible al sesgo de deseabilidad social y no captura prácticas instruccionales reales; el constructo medido es la competencia autopercebida, no la competencia demostrada. Segunda, el diseño transversal impide inferencias causales o temporales. Tercera, la muestra se circunscribe a una única institución de educación superior de gran escala, lo que limita la generalización a otros contextos institucionales y culturales. Cuarta, aunque el modelo bifactor mostró buen ajuste relativo (CFI=.910; RMSEA=.087), las elevadas correlaciones interfactoriales (ϕ entre .563 y .715) sugieren que las puntuaciones específicas comparten una proporción sustancial de varianza con el factor general; la interpretación dimensional debe complementarse con evidencia adicional de validez convergente y discriminante en estudios independientes. Además, la interpretación del modelo bifactor debe considerarse preliminar, dado que no se reportan índices específicos de bifactorialidad (omega jerárquico, ECV, PUC), cuyo cálculo se reserva para la fase confirmatoria con muestras independientes. Asimismo, los modelos CFA se estimaron sobre la misma muestra utilizada para el AFE, lo que limita su interpretación como confirmación independiente de la estructura; la verificación confirmatoria propiamente dicha requiere muestras independientes. Quinta, dado que la convocatoria se realizó mediante plataforma digital abierta, no es posible calcular una tasa de respuesta global. Sexta, el panel de validación estuvo conformado por cinco expertos en metodología, didáctica, ingeniería en IA y tecnología educativa, sin incluir docentes universitarios en ejercicio como usuarios directos del instrumento; futuros estudios podrían ampliar el panel para incluir esa perspectiva profesional complementaria. Séptima, los elevados coeficientes de consistencia interna, especialmente en la escala total ($\alpha=.973$) y en la dimensión pedagógica ($\alpha=.967$), sugieren alta homogeneidad; futuras versiones podrían examinar posible redundancia de contenido y valorar una forma abreviada del instrumento.

Líneas futuras de investigación

La agenda de investigación derivada de este estudio incluye varias líneas prioritarias. Primera, la verificación confirmatoria del CCDGAI mediante AFC con estimador WLSMV apropiado para datos ordinales en muestras independientes y multi-

institucionales, incluyendo comparación formal de los modelos competidores (unidimensional, trifactorial correlacionado, bifactor y jerárquico de Schmid-Leiman). Segunda, el establecimiento de evidencia de validez convergente y discriminante mediante correlaciones con marcos consolidados como DigCompEdu (Redecker, 2017), TPACK e instrumentos de IA literacy específicos (Ng et al., 2021; Ayyoub et al., 2025). Tercera, el análisis exhaustivo de invarianza de medición métrica, escalar y estricta entre subgrupos sociodemográficos (género, edad, área disciplinar, antigüedad docente; Vandenberg & Lance, 2000). Cuarta, estudios longitudinales que evalúen la sensibilidad del instrumento al cambio tras intervenciones formativas y la estabilidad temporal de los perfiles identificados. Quinta, la triangulación con medidas objetivas de desempeño — productos didácticos generados con IAG, observación de prácticas en aula, evaluaciones realizadas por estudiantes— que permitan establecer la correspondencia entre competencia autopercebida y competencia demostrada. Sexta, la replicación del estudio en muestras multinacionales del espacio iberoamericano para fortalecer la generalización del instrumento y de la tipología de perfiles.

En síntesis, el CCDGAI aporta evidencias preliminares de validez de contenido ($V=.91$) y consistencia interna excelente ($\alpha=.973$) para el diagnóstico de competencias docentes en IAG. Los tres perfiles diferenciados, con tamaños del efecto grandes ($\eta^2 \geq .681$), demuestran que la preparación del profesorado es un espacio heterogéneo con necesidades formativas diferenciadas, y la asimetría entre dimensión ética y dimensiones tecnológica y pedagógica identifica un patrón recurrente con implicaciones directas para el desarrollo profesional, la planificación curricular y la gobernanza responsable de la IAG en la educación superior.

Agradecimientos

Los autores agradecen la predisposición y participación del profesorado de la institución participante. Esta investigación no recibió financiación externa.

Referencias

- Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurement*, 45(1), 131–142. <https://doi.org/10.1177/0013164485451012>
- American Psychological Association. (2017). *Ethical principles of psychologists and code of conduct* (2002, amended effective June 1, 2010, and January 1, 2017). <https://www.apa.org/ethics/code>
- Ayyoub, A. M., Khlaif, Z. N., Shamali, M., Abu Eideh, B., Assali, A., Hattab, M. K., Barham, K. A., & Bsharat, T. R. K. (2025). Advancing higher education with GenAI: Factors influencing educator AI literacy. *Frontiers in Education*, 10, Article 1530721. <https://doi.org/10.3389/feduc.2025.1530721>
- Baig, M. I., & Yadegaridehkordi, E. (2025). Factors influencing academic staff satisfaction and continuous usage of generative artificial intelligence (GenAI) in higher education. *International Journal of Educational Technology in Higher Education*, 22, Article 5. <https://doi.org/10.1186/s41239-025-00506-4>
- Belkina, M., Daniel, S., Nikolic, S., Haque, R., Lyden, S., Neal, P., Grundy, S., & Hassan, G. M. (2025). Implementing generative AI (GenAI) in higher education: A systematic

- review of case studies. *Computers and Education: Artificial Intelligence*, 8, 100407. <https://doi.org/10.1016/j.caeai.2025.100407>
- Burneo-Arteaga, P., Lira, Y., Murzi, H., Balula, A., & Costa, A. P. (2025). Capability-based training framework for generative AI in higher education. *Frontiers in Education*, 10, Article 1594199. <https://doi.org/10.3389/feduc.2025.1594199>
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6(4), 284–290. <https://doi.org/10.1037/1040-3590.6.4.284>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum.
- Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis* (2nd ed.). Lawrence Erlbaum.
- Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 61(2), 228–239. <https://doi.org/10.1080/14703297.2023.2190148>
- Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education*, 20, Article 22. <https://doi.org/10.1186/s41239-023-00392-8>
- Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2024). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. *Computers & Education*, 227, 105224. <https://doi.org/10.1016/j.compedu.2024.105224>
- DeVellis, R. F. (2017). *Scale development: Theory and applications* (4th ed.). SAGE.
- Dringó-Horváth, I., Rajki, Z., & T. Nagy, J. (2025). University teachers' digital competence and AI literacy: Moderating role of gender, age, experience, and discipline. *Education Sciences*, 15(7), Article 868. <https://doi.org/10.3390/educsci15070868>
- EDUCAUSE. (2025a). *2025 EDUCAUSE AI landscape study: Into the digital AI divide*. <https://library.educause.edu/resources/2025/2/2025-educause-ai-landscape-study>
- EDUCAUSE. (2025b). *2025 EDUCAUSE Horizon Report: Teaching and learning edition*. <https://library.educause.edu/resources/2025/5/2025-educause-horizon-report-teaching-and-learning-edition>
- Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. Chapman & Hall.
- Garzón, J., Patiño, E., & Marulanda, C. (2025). Systematic review of artificial intelligence in education: Trends, benefits, and challenges. *Multimodal Technologies and Interaction*, 9(8), Article 84. <https://doi.org/10.3390/mti9080084>
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Hwang, G.-J., Xie, H., Wah, B. W., & Gašević, D. (2020). Vision, challenges, roles and research issues of artificial intelligence in education. *Computers and Education: Artificial Intelligence*, 1, 100001. <https://doi.org/10.1016/j.caeai.2020.100001>
- Inter-American Development Bank. (2025). *AI and education: Building the future through digital transformation* (IDB-TN-3122). <https://publications.iadb.org/publications/english/document/AI-and-Education-Building-the-Future-Through-Digital-Transformation.pdf>
- Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, 39(1), 31–36. <https://doi.org/10.1007/BF02291575>

- Kalantzis, M., & Cope, B. (2025). Literacy in the time of artificial intelligence. *Reading Research Quarterly*, 60(1), e591. <https://doi.org/10.1002/rrq.591>
- Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., . . . Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Liang, J., Stephens, J. M., & Brown, G. T. L. (2025). A systematic review of the early impact of artificial intelligence on higher education curriculum, instruction, and assessment. *Frontiers in Education*, 10, Article 1522841. <https://doi.org/10.3389/feduc.2025.1522841>
- Liu, D. Y. T., & Bates, S. (2025). *Generative AI in higher education: Current practices and ways forward*. APRU. <https://www.apru.org/wp-content/uploads/2025/01/APRU-Generative-AI-in-Higher-Education-Whitepaper-Jan-2025.pdf>
- Mah, D.-K., & Groß, N. (2024). Artificial intelligence in higher education: Exploring faculty use, self-efficacy, distinct profiles, and professional development needs. *International Journal of Educational Technology in Higher Education*, 21(1), Article 58. <https://doi.org/10.1186/s41239-024-00490-1>
- Miao, F., & Cukurova, M. (2024). *AI competency framework for teachers*. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000391104>
- Milligan, G. W., & Cooper, M. C. (1987). Methodology review: Clustering methods. *Applied Psychological Measurement*, 11(4), 329–354. <https://doi.org/10.1177/014662168701100401>
- Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). AI literacy: Definition, teaching, evaluation and ethical issues. *Proceedings of the Association for Information Science and Technology*, 58(1), 504–509. <https://doi.org/10.1002/pra2.487>
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.
- OECD. (2025). *What should teachers teach and students learn in a future of powerful AI?* https://www.oecd.org/en/publications/what-should-teachers-teach-and-students-learn-in-a-future-of-powerful-ai_ca56c7d6-en
- Penfield, R. D., & Giacobbi, P. R. (2004). Applying a score confidence interval to Aiken's item content-relevance index. *Measurement in Physical Education and Exercise Science*, 8(4), 213–225. https://doi.org/10.1207/s15327841mpee0804_3
- Redecker, C. (2017). *European framework for the digital competence of educators: DigCompEdu* (Y. Punie, Ed.; EUR 28775 EN). Joint Research Centre, European Commission. <https://publications.jrc.ec.europa.eu/repository/handle/JRC107466>
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
- Sireci, S. G., & Faulkner-Bond, M. (2014). Validity evidence based on test content. *Psicothema*, 26(1), 100–107. <https://doi.org/10.7334/psicothema2013.256>
- Steinley, D. (2004). Properties of the Hubert-Arabie adjusted Rand index. *Psychological Methods*, 9(3), 386–396. <https://doi.org/10.1037/1082-989X.9.3.386>
- Tabachnick, B. G., & Fidell, L. S. (2019). *Using multivariate statistics* (7th ed.). Pearson.
- UNESCO. (2024). *What you need to know about UNESCO's new AI competency frameworks for students and teachers*. <https://www.unesco.org/en/articles/what-you-need-know-about-unescos-new-ai-competency-frameworks-students-and-teachers>
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for

- organizational research. *Organizational Research Methods*, 3(1), 4–70. <https://doi.org/10.1177/109442810031002>
- von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C., Vandenbroucke, J. P., & STROBE Initiative. (2007). The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: Guidelines for reporting observational studies. *PLoS Medicine*, 4(10), e296. <https://doi.org/10.1371/journal.pmed.0040296>
- Weng, X., Xia, Q., Gu, M., Rajaram, K., & Chiu, T. K. F. (2024). Assessment and learning outcomes for generative AI in higher education: A scoping review on current research status and trends. *Australasian Journal of Educational Technology*, 40(6), 37–55. <https://doi.org/10.14742/ajet.9540>
- Xia, Q., Weng, X., Ouyang, F., Lin, T. J., & Chiu, T. K. F. (2024). A scoping review on how generative artificial intelligence transforms assessment in higher education. *International Journal of Educational Technology in Higher Education*, 21. <https://doi.org/10.1186/s41239-024-00468-z>
- 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. <https://doi.org/10.1186/s41239-019-0171-0>