



Rodríguez-Moreno, J. y Santaella-Barcos, P. (2026). Capacitación docente en el uso educativo de la inteligencia artificial. *Revista Interuniversitaria de Formación del Profesorado*, 40(2), 233-248.

<https://doi.org/10.6018/rifop.707941>

Capacitación docente en el uso educativo de la inteligencia artificial

Teacher Training in the Educational Use of artificial intelligence

Javier Rodríguez-Moreno

Universidad de Jaén <https://orcid.org/0000-0002-5890-3654>

Pablo Santaella-Barcos

Universidad de Jaén <https://orcid.org/0009-0002-0334-0054>

Resumen

La inclusión de la inteligencia artificial (IA) en los centros docentes requiere un conocimiento preciso del nivel de formación del profesorado y de los factores que lo determinan. En el presente artículo, se ha analizado el perfil competencial del profesorado en relación con su nivel de realidad en el uso de la IA, además de calcular las relaciones significativas que se establecen con las variables sociodemográficas. Para ello, se ha planteado un diseño cuantitativo no experimental basado en el método de encuesta del cuestionario DOCENT-IA para un total de 398 docentes. Dicho análisis ha combinado una caracterización descriptiva del nivel competencial de los docentes con comparaciones entre grupos que incluyeron estimaciones del tamaño del efecto para evaluar la importancia de estas destrezas de las diferencias observadas. Los resultados dibujan un perfil competencial que se encuentra en un nivel neutro situado levemente bajo, con una clara heterogeneidad entre ítems: algunos reflejan una mayor conciencia y adopción de prácticas relacionadas con la IA, mientras que otros tienen una integración sistemática limitada. En el caso de las variables sociodemográficas, las diferencias fueron generalmente muy leves, y emergió sistemáticamente la formación específica en IA como el factor más vinculado a un nivel competencial más alto. En general, los hallazgos muestran la necesidad de fortalecer y reforzar la formación docente con itinerarios formativos más específicos y a largo plazo para promover una integración pedagógica efectiva de la IA.

Palabras clave: *Inteligencia artificial; Competencia docente; Formación del profesorado; Educación.*

Abstract

The integration of artificial intelligence (AI) in educational institutions requires a precise understanding of teachers' training levels and the factors that determine them. This article analyzes teachers' competency profiles in relation to their level of experience using artificial intelligence, as well as establishing significant relationships with sociodemographic variables. A non-experimental quantitative design was used, based on the DOCENT-IA questionnaire survey, administered to a total of 398 teachers. This analysis combined a descriptive characterization of teachers' competency levels with comparisons between groups, including effect size estimates to assess the significance of these skills and the observed differences. The results indicate a competency profile at a slightly low, neutral level, with clear heterogeneity among items: some reflect greater awareness and adoption of AI-related practices, while others show limited systematic integration. In the case of sociodemographic variables, the differences were generally very slight, and specific AI training consistently emerged as the factor most closely linked to a higher level of competence. Overall, the findings demonstrate the need to strengthen and reinforce teacher training with more specific and long-term training pathways to promote the effective pedagogical integration of AI.

Keywords: Artificial intelligence; Teacher competence; Professional development; Education.

Introducción

Una vez analizado el uso de la inteligencia artificial (IA) por parte de los docentes andaluces, se ha consolidado como una línea de investigación con gran prioridad dentro del campo de la tecnología educativa. En comparación con los enfoques del inicio, centrados en el desarrollo técnico de sistemas inteligentes, la literatura actual desprende un interés, el cual está creciendo, por comprender cómo los profesores utilizan realmente la IA, con qué finalidad y qué factores van a condicionar su adopción en la práctica educativa (Selwyn, 2019).

Existe un estudio que pone de manifiesto el predominio de los trabajos que están centrados en sistemas automatizados, y que también identifica un desplazamiento progresivo hacia el análisis del papel de los docentes como usuarios de la IA. Dicho estudio fue realizado por Zawacki-Richter et al. (2019), quienes analizaron de forma exhaustiva la investigación sobre aplicaciones de la IA en Educación Superior, siendo esta una de las revisiones sistemáticas más influyentes. Los resultados indican que los docentes emplean estas tecnologías principalmente como herramientas de apoyo a la evaluación, la analítica del aprendizaje y la personalización de contenidos, más que como sustitutos de la toma de decisiones pedagógicas.

Este patrón se confirma en revisiones más recientes centradas en los usos educativos de la IA. Según Crompton & Burke (2023), los docentes utilizan la IA de una forma instrumental, más concretamente para la generación de materiales didácticos, el apoyo a la retroalimentación formativa y la automatización de tareas repetitivas. No obstante, los autores advierten que estos usos suelen estar poco integrados en el diseño curricular y dependen, en gran medida, de la iniciativa individual del docente, lo que limita su impacto pedagógico a largo plazo.

Las investigaciones empíricas también señalan que la competencia profesional del profesorado constituye un factor clave en el uso de la IA. El informe de la OECD (2021) evidencia que los docentes con mayor nivel de competencia digital y experiencia en innovación educativa tienden a emplear la IA de forma más reflexiva, vinculándola a procesos de evaluación formativa, seguimiento del progreso del alumnado y toma de decisiones basadas en datos. Por otro lado, el profesorado que tiene menos formación suele limitar su uso a aplicaciones puntuales o muestran alguna reticencia ante dichas tecnologías.

En la investigación, la dimensión actitudinal ha sido ampliamente abordada. Según Selwyn et al. (2020), los docentes tienen una postura ambivalente hacia la IA, es decir, el profesorado sabe del potencial de la IA para mejorar la eficiencia y la personalización del aprendizaje, pero existe una

gran preocupación en cuanto a la pérdida de autonomía profesional, la opacidad de los algoritmos y de una posible deshumanización de la enseñanza. Este contraste repercute de una forma directa en la que los docentes deciden cómo integrar la IA en sus prácticas. Si se analiza desde una perspectiva cualitativa, los estudios de Celik et al., (2022) muestran que la utilización de la IA de manera pedagógica ocurre cuando los docentes la conciben como una herramienta de apoyo, según su juicio personal. En estos casos, el uso de la IA se lleva a cabo para atender la diversidad, en la evaluación y en la planificación, siendo así el docente el principal mediador del proceso de enseñanza-aprendizaje y garantizando el sentido educativo. En una línea similar, estudios recientes constatan actitudes cognitivas, emocionales, conductuales y éticas ambivalentes hacia la IA entre el alumnado universitario, lo que refuerza la necesidad de una formación docente que sepa acompañar y orientar estos procesos (Cabero et al., 2026).

Esta investigación resalta la importancia que tiene el contexto institucional. En las orientaciones que ofrece la UNESCO (2023), el profesorado utiliza la IA cuando existen unos marcos éticos claros, con una política educativa definida y con programas de formación continua. Al no existir dichas condiciones, el uso de la IA se produce de una manera fragmentada y dependiente de la iniciativa individual, con lo que así se dificulta su asentamiento como práctica educativa sostenible. En este sentido, el análisis documental de buenas prácticas en el uso de la IA generativa en las universidades confirma que su implementación efectiva depende de marcos institucionales definidos y de programas de formación continua del profesorado (Ares et al., 2026).

Por último, diversos estudios advierten que el uso docente de la IA está inevitablemente vinculado a consideraciones éticas y profesionales. Floridi et al. (2018) destacan la necesidad de que los docentes comprendan los riesgos asociados a los sesgos algorítmicos, la gestión de datos y la transparencia de los sistemas automatizados. En este sentido, la investigación coincide en que el profesorado no actúa como un mero usuario técnico de la IA, sino como un agente pedagógico cuyas decisiones están mediadas por valores, creencias profesionales y contextos institucionales. Por todo ello, el objetivo que se planteó para esta investigación fue analizar el perfil competencial de los docentes en el uso de la inteligencia artificial (IA). En esta dirección, el desarrollo del pensamiento crítico mediante modelos didácticos que incorporan la IA en la educación superior se ha propuesto como una vía complementaria para que el profesorado afronte estos retos éticos desde una perspectiva pedagógica (Area & González, 2026). Además, la rápida expansión de la IA generativa en los procesos de enseñanza y aprendizaje, con impactos ya constatados en la carga cognitiva y los comportamientos de lectura del alumnado (Sadiq et al., 2026), pone de manifiesto necesidades emergentes de capacitación docente que la presente investigación pretende contribuir a caracterizar.

Método

El estudio utilizó una metodología cuantitativa, con un diseño no experimental de muestreo no probabilístico (autoselección de participantes), el tamaño de la muestra se estimó como guía de una forma orientativa a través de la fórmula para poblaciones finitas, para poder tener una referencia mínima de N.

En cualquier caso, este cálculo no ha de interpretarse como garantía de representatividad, por lo que los resultados se refieren únicamente a la muestra analizada. Para la recolección de datos, se utilizó el método de encuesta, estructurado y diseñado específicamente para dar respuesta al objetivo planteado en esta investigación. El cuestionario incluyó varios ítems que abordaron la realidad docente en relación con la utilización de la IA en el ámbito educativo, evaluando las percepciones sobre el uso, las aplicaciones de la IA, y los desafíos o barreras que enfrentan los docentes.

Muestra

La muestra del estudio estuvo compuesta por 398 participantes, los cuales son profesores del sistema educativo andaluz en el curso 2024-2025 (N= 106.613). De forma complementaria, el tamaño muestral se estimó con carácter orientativo mediante la fórmula para poblaciones finitas, a modo de referencia del mínimo de respuestas, sin que ello implique representatividad

poblacional. También se tuvo en cuenta ciertas condiciones para la selección de muestra, su acceso y su aceptación, con la cumplimentación de forma libre de un cuestionario mediante un correo electrónico a todos los centros pertenecientes a la comunidad andaluza para un proceso de recogida de datos de una forma óptima y eficaz.

Tal y como se puede observar en la Tabla 1, el análisis sociodemográfico de los participantes en el estudio revela varios aspectos relevantes que pueden influir en sus percepciones y competencias relacionadas con el uso de la inteligencia artificial. La edad media estimada de los docentes a partir de su año de nacimiento es de 45.7 años, lo que refleja una población de docentes con experiencia profesional moderada. La antigüedad media como docente es de 11.85 años, lo que indica que la muestra está compuesta por profesionales que han acumulado una cantidad significativa de experiencia en la enseñanza.

En cuanto al género, la muestra está compuesta en su mayoría por mujeres (65.83%), lo que es consistente con la tendencia observada en la Educación Primaria y Secundaria, y con la distribución del profesorado en estas etapas; este desequilibrio se consideró en los análisis comparativos posteriores. La distribución en etapas educativas muestra que la mayoría de los docentes se encuentra en Educación Primaria (39.20%), seguida por Educación Secundaria (13.07%):

Tabla 1
Rasgos sociodemográficos y otras variables independientes

Variable		M (DT), min.- max.	
Año nacimiento		1981 (8.99), 1960-2000	
Antigüedad como docente		11.85 (7.7), 2.5-21	
Variable		N (%)	
Sexo	Mujer	262	65.83%
	Hombre	136	34.17%
Etapas/ educativas en la que ejerce actualmente	Educación Primaria	156	39.20%
	Educación Secundaria	52	13.07%
	Formación Profesional	44	11.06%
	Educación Infantil	42	10.55%
	Educación Infantil y Educación Primaria	46	11.56%
	Educación Secundaria y Bachillerato	24	6.03%
	Educación Primaria y Educación Secundaria	10	2.51%
	Educación Secundaria y Formación Profesional	8	2.01%
	Educación de adultos	8	2.01%
	Educación Secundaria, Bachillerato y Formación Profesional	8	2.01%
Tipo de centro en el que ejerce	CEIP (Educación Infantil y Primaria)	210	52.90%
	IES (Instituto de Educación Secundaria)	129	32.50%
	CPR (Colegio Público Rural)	20	5.03%
	SEMI-D (Educación Primaria y Primer ciclo de ESO)	20	5.03%
	Otros	18	4.50%
Seleccione la Provincia donde ejerce como docente	Granada	80	20.10%
	Jaén	73	18.34%
	Almería	59	14.82%
	Málaga	52	13.07%
	Cádiz	39	9.80%
	Sevilla	38	9.55%
	Córdoba	35	8.79%
Horas de formación recibidas sobre Inteligencia artificial	Huelva	22	5.53%
	No he recibido formación	215	54.0%
	Menos de 30 horas	91	22.90%
	Entre 31 y 60 horas	57	14.30%
	Entre 61 y 90 horas	16	4.00%
	Entre 91 y 120 horas	14	3.50%
	Más de 120 horas	5	1.30%

Instrumento de Recogida de Datos

Para la recogida de datos, se ha utilizado el método por encuesta a través del cuestionario sobre la autopercepción del profesorado en relación con la realidad y el uso de la IA "DOCENT-IA". Se trata de un instrumento más amplio que está dividido en 5 dimensiones, pero en este artículo nos centraremos en la Dimensión "Realidad docente a nivel de uso de la inteligencia artificial", con 10 ítems de respuesta tipo Likert, cuyas opciones de respuesta están comprendidas entre 1 y 6.

Esta herramienta está diseñada específicamente en base al marco competencial de IA para docentes (UNESCO, 2024). En una primera fase inicial de la investigación, se ha llevado a cabo una validación de contenidos mediante juicios de expertos (Escobar & Cuervo, 2008). Para ello, se revisaron los ítems del cuestionario para analizar su coherencia, claridad, relevancia y suficiencia que presenta el cuestionario. Posteriormente, se realizó un estudio piloto para identificar posibles errores o problemas para solucionarlos antes de la implementación definitiva. Para ello, fueron seleccionados un total de 49 docentes, los cuales otorgaron su consentimiento informado para la participación en el estudio. Dichos participantes fueron excluidos de la aplicación final del instrumento. En esta línea, estudios recientes han contribuido a actualizar la fundamentación de instrumentos de evaluación de competencias y conocimientos vinculados a la IA en contextos educativos, mediante el uso del coeficiente de competencia experta para el diseño y validación de instrumentos sobre IA generativa (Colomo et al., 2026) y la adaptación transcultural de escalas sobre el conocimiento tecnológico pedagógico del contenido en IA (Ceballos et al., 2026).

Por último, para examinar la estructura interna y la robustez del cuestionario, se llevó a cabo un análisis exploratorio mediante Análisis de Componentes Principales (PCA), con el objetivo de analizar la dimensionalidad del instrumento e identificar las dimensiones subyacentes que explican la variabilidad de las respuestas. La versión final del instrumento quedó compuesta por 47 ítems distribuidos en cinco dimensiones: (1) Realidad docente a nivel del uso de la inteligencia artificial; (2) Ética y uso de la inteligencia artificial; (3) Aplicaciones de la inteligencia artificial en las aulas; (4) Enseñanza asistida por inteligencia artificial para la transformación pedagógica; y (5) Desarrollo profesional de la inteligencia artificial como habilitadora del aprendizaje profesional continuo. Los indicadores de validez convergente (AVE) y fiabilidad compuesta (CR) se derivan de una evaluación psicométrica previa del instrumento (modelo de medida) y se informan en este estudio con un objetivo descriptivo para poder así, contextualizar la calidad de la dimensión analizada, sin derivarlas del análisis por componentes principales. En cuanto a la dimensión objeto de estudio en este artículo, "Realidad docente a nivel del uso de la inteligencia artificial", la consistencia interna fue elevada, alcanzándose así el coeficiente alfa de Cronbach con un valor de 0.918, muy superior al umbral convencional. Asimismo, la varianza media extraída fue de 0.551, lo que indica una evidencia de validez convergente, y el indicador de fiabilidad compuesta alcanzó un valor de 0.916, lo que indica una alta confiabilidad del conjunto de ítems.

Análisis de Datos

El tratamiento estadístico se realizó con SPSS (IBM SPSS Statistics v28). Se efectuó un análisis descriptivo y exploratorio (tendencia central, dispersión y forma) de la dimensión y de sus ítems, incluyendo intervalos y valores extremos.

Para el análisis inferencial se emplearon contrastes no paramétricos: U de Mann-Whitney en comparaciones entre dos grupos y H de Kruskal-Wallis cuando existían tres o más categorías, con comparaciones post-hoc por pares y ajuste de Holm. Se reportaron tamaños del efecto (r_{rb} y ϵ^2) para valorar la magnitud de las diferencias.

Resultados

Los resultados del ítem RDIA05 muestran una tendencia hacia puntuaciones altas, con una media de 4.61. La mediana es de 5.0, el cuartil 1 (P25) es 2.0, mientras que el cuartil 3 (P75) es 6.0, sugiriendo un predominio de puntuaciones elevadas, aunque con elevada heterogeneidad (IQR=4) y con una desviación estándar de 1.77.

Por otra parte, el ítem RDIA03 presenta una media más baja de 1.32. Su mediana es 1.0, pero es importante destacar el bajo rango intercuartílico (IQR) de 1.0, indicando baja variabilidad en las respuestas y se observan respuestas extremas puntuales, aunque la mayoría se concentra en valores bajos. En cuanto a RDIA06 y a RDIA07, presentan medias de 2.09 y 1.95 respectivamente; lo que muestra opiniones dispares con rangos IQR de 2.0 y desviaciones estándar cercanas a 1.13 y 1.45, lo que alude a una mayor dispersión en las respuestas.

El ítem RDIA01 presenta una media de 2.81 y una mediana de 3.0, con una tendencia neutra-baja por debajo del punto medio, aunque con una variabilidad moderada, como lo muestra la desviación estándar de 1.28 y un IQR de 2.0. El ítem RDIA02, presenta una media de 3.02, mostrando una tendencia similar, con una mediana de 4.0, aunque con un mayor rango intercuartílico (IQR) de 2.0, lo que nos indica una dispersión elevada en las respuestas.

Por otro lado, el ítem RDIA04, tiene una media de 3.61 y mediana de 4.0, indica una percepción mayoritariamente positiva con un rango intercuartílico de 3.0, lo que va a sugerir una mayor dispersión en las respuestas, mientras que el ítem RDIA08, con una media de 2.08 y mediana de 2.0, muestra una tendencia moderada o de carácter neutro en las respuestas, con una desviación estándar de 1.67 y un IQR de 3.0, lo que llega a indicar una considerable variabilidad en las opiniones de los encuestados.

Finalmente, RDIA09 y RDIA10, con medias de 1.72 y 2.44 respectivamente, indican puntuaciones más bajas, lo que sugiere una heterogeneidad, con un rango intercuartílico de 3.0 y 2.0 respectivamente, lo que refleja una alta dispersión en las respuestas.

Tabla 2
Análisis exploratorio y descriptivo

N= 398	Centralidad				Variabilidad			
	Media (M)	Mediana (P50)	P25	P75	Desviación Estándar	Valor Mínimo	Valor Máximo	Rango Intercuartil
RDIA01	2.81	3	2	4	1.28	1	6	2.0
RDIA02	3.02	4	2	4	1.36	1	6	2.0
RDIA03	1.32	1	1	2	1.04	1	6	1.0
RDIA04	3.61	4	2	5	1.66	1	6	3.0
RDIA05	4.61	5	2	6	1.77	1	6	4.0
RDIA06	2.09	2	1	3	1.13	1	6	2.0
RDIA07	1.95	2	1	3	1.45	1	6	2.0
RDIA08	2.08	2	1	3	1.67	1	6	3.0
RDIA09	1.72	1	1	4	1.62	1	6	3.0
RDIA10	2.44	3	2	4	1.94	1	6	2.0

En los siguientes apartados, se procederá a detallar el perfil competencial de los docentes dentro de la dimensión estudiada, así como las posibles relaciones con las variables sociodemográficas que forman parte de este análisis.

Perfil competencial

El análisis del promedio de la dimensión sugiere un nivel global neutro-bajo, con evidencia de desviación de la normalidad. La asimetría negativa indica mayor concentración relativa de puntuaciones altas, aunque persiste variabilidad entre participantes.

La prueba de normalidad realizada (prueba de Kolmogorov-Smirnov) arroja un p-valor de $p < .001$, lo que confirma que los datos no siguen una distribución normal, lo que justifica el uso de pruebas no paramétricas para el análisis inferencial.

En términos de centralidad, la media de las puntuaciones promedio es 2.98, lo que refleja que, en general, los encuestados se encuentran en una posición neutra-baja, por debajo del punto medio.

La mediana, con un valor de 2.6, refuerza esta interpretación, sugiriendo que el 50% de los encuestados están en una posición neutra-baja.

La dispersión observada es moderada, coherente con la coexistencia de perfiles docentes heterogéneos en el uso de la IA.

Tabla 3
Análisis exploratorio y descriptivo

Variables N= 398	Exploración: Forma			Centralidad		Variabilidad	
	Asimetría	Curtosis	Kolmogorov -Smirnov (p)	Media	Mediana	Desv. estándar	Rango IQR
RdIA (Promedio)	-0.42	0.89	< .001	2.98	2.6	1.33	2.0

En la Figura 2, se muestra un gráfico de radar que muestra las puntuaciones medias de los ítems de la dimensión “Realidad docente a nivel de uso de la inteligencia artificial” (RDIA), con los ejes representando cada ítem (RDIA01 a RDIA10).

Figura 2
Puntuaciones medias



Relaciones significativas con variables sociodemográficas

Se evaluó la dimensión analizando las puntuaciones acumuladas de cada ítem en función de diferentes segmentos sociodemográficos. A través de una serie de pruebas no paramétricas, se buscó determinar si existían diferencias significativas entre los segmentos de interés.

Según año de nacimiento

A continuación, se examinan las diferencias en el promedio de la dimensión según año de nacimiento. Dado que se comparan más de dos grupos independientes, se aplicó la prueba H de Kruskal-Wallis, complementada con comparaciones por pares post-hoc cuando procedía.

Respecto al año de nacimiento, descriptivamente se aprecia una ligera tendencia a puntuaciones promedio más altas en las cohortes más recientes (especialmente 1981-1990 y 1991-2000) frente al grupo 1960-1970. Sin embargo, la H de Kruskal-Wallis no mostró diferencias estadísticamente significativas entre grupos ($H = 5.90$; $gl = 3$; $p = .116$) y el tamaño del efecto fue insignificante ($\epsilon^2 = .007$). En consecuencia, no puede sostenerse una relación robusta entre el año de nacimiento y el promedio de la dimensión; además, la diferencia puntual observada en alguna comparación por pares no se mantiene tras la corrección por múltiples comparaciones.

Tabla 4
Contraste de hipótesis según año de nacimiento

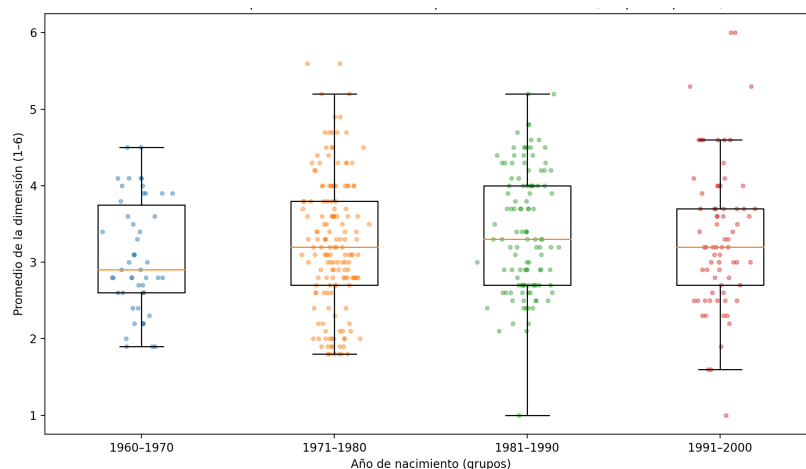
	Descriptivos			Test de contraste				
	Media	Desv. estándar	Mediana	H	p-valor	gl	ϵ^2	Interpretación
1960-1970	3.07	0.72	2.90	5.90	.116	3	.007	Insignificante
1971-1980	3.23	0.87	3.20					
1981-1990	3.40	0.80	3.30					
1991-2000	3.33	0.94	3.20					

Nota: Dado que el contraste global no resulta significativo ($p \geq .05$), los análisis post-hoc se reportan únicamente con fines exploratorios y no inferenciales.

Tabla 5
Comparaciones por pares (post-hoc exploratorio: U de Mann-Whitney con corrección Holm)

Comparación	U	p	p ajustada (Holm)	Dif. medias	r _{rb} (efecto)
1960-1970 vs 1971-1980	3528.0	.2307	.6922	-0.165	-0.112
1960-1970 vs 1981-1990	2203.0	.0171	.1029	-0.330	-0.234
1960-1970 vs 1991-2000	1524.5	.1220	.6101	-0.263	-0.165
1971-1980 vs 1981-1990	8167.0	.1317	.6101	-0.165	-0.107
1971-1980 vs 1991-2000	5525.0	.5578	.9278	-0.098	-0.048
1981-1990 vs 1991-2000	4464.0	.4639	.9278	0.067	0.063

Figura 3
Distribución del promedio de la dimensión por año de nacimiento



En la figura anteriormente planteada, se muestran pequeñas diferencias en la práctica debido al solapamiento presente entre hombres y mujeres en la distribución del promedio.

Según la antigüedad como docente

En este apartado, se presenta el análisis entre la antigüedad como docente y el promedio de la dimensión. Al tratarse de más de dos grupos independientes, se empleó la prueba H de Kruskal-Wallis, complementada con comparaciones post-hoc cuando fue necesario, con el objetivo de identificar posibles diferencias entre los niveles de antigüedad.

Tabla 6
Contraste de hipótesis según antigüedad como docente

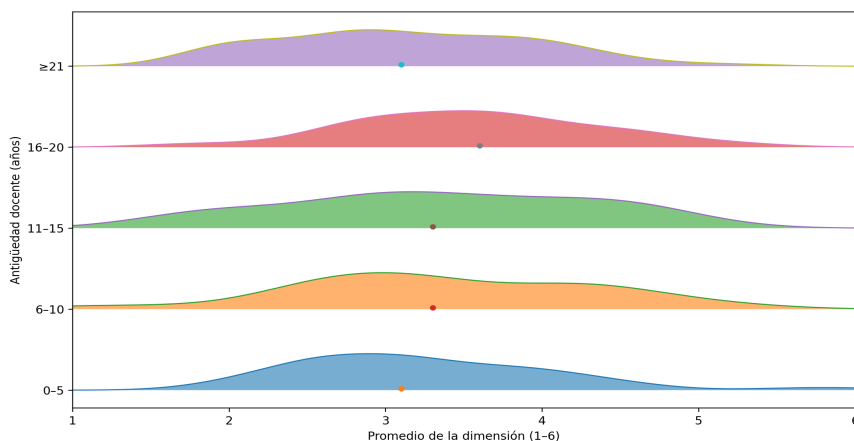
	Descriptivos			Test de contraste				Interpretación
	Media	Desv. estándar	Mediana	H	p-valor	gl	ϵ^2	
Entre 0 y 5 años	3.25	.82	3.10	11.61	.020	4	.019	Diferencias significativas (efecto pequeño)
Entre 6 y 10 años	3.38	.94	3.30					
Entre 11 y 15 años	3.33	.93	3.30					
Entre 16 y 20 años	3.55	.76	3.60					
21 años o más	3.12	.81	3.10					

Tabla 7
Comparaciones por pares (post-hoc confirmatorio: U de Mann-Whitney con corrección Holm)

Comparación	U	p	p ajustada (Holm)	Dif. medias	r _{rb}
Entre 0 y 5 años vs. Entre 11 y 15 años	1822.0	.404	> .999	-0.083	-0.095
Entre 0 y 5 años vs. Entre 16 y 20 años	2272.0	.008	.0780	-0.305	-0.255
Entre 0 y 5 años vs. Entre 6 y 10 años	3964.0	.153	> .999	-0.130	-0.122
Entre 11 y 15 años vs. Entre 16 y 20 años	726.0	.358	> .999	-0.223	-0.120
Entre 11 y 15 años vs. Entre 6 y 10 años	1209.0	.938	> .999	-0.048	-0.010
Entre 16 y 20 años vs. Entre 6 y 10 años	2072.0	.258	> .999	0.175	0.120
21 años o más vs. Entre 0 y 5 años	6750.0	.404	> .999	-0.122	-0.062
21 años o más vs. Entre 11 y 15 años	1688.0	.243	> .999	-0.205	-0.133
21 años o más vs. Entre 16 y 20 años	2084.0	.002	.0265	-0.427	-0.294
21 años o más vs. Entre 6 y 10 años	3604.0	.0420	.3358	-0.252	-0.175

Los resultados descriptivos del promedio de la dimensión por antigüedad docente muestran diferencias entre grupos que, a nivel inferencial, se confirman mediante la prueba H de Kruskal-Wallis, evidenciando diferencias estadísticamente significativas entre los tramos considerados ($H = 11.61$; $gl = 4$; $p = .020$). Sin embargo, el tamaño del efecto es bajo ($\epsilon^2 = 0.019$), lo que muestra que la antigüedad explica una pequeña proporción de la variabilidad en el constructo. En las comparaciones por pares, la diferencia más consistente se da entre el grupo de “21 años o más” y “16-20 años” (p ajustada Holm = .0265), lo que indica puntuaciones relativamente más altas que en el tramo de 16-20 años, aunque dentro de un patrón general de superposición entre distribuciones.

Figura 4
Ridgeline plot del promedio de la dimensión según la antigüedad docente



La figura permite comparar la distribución del promedio por tramos de antigüedad, apreciándose variaciones moderadas entre algunos grupos y un solapamiento general considerable.

Según sexo

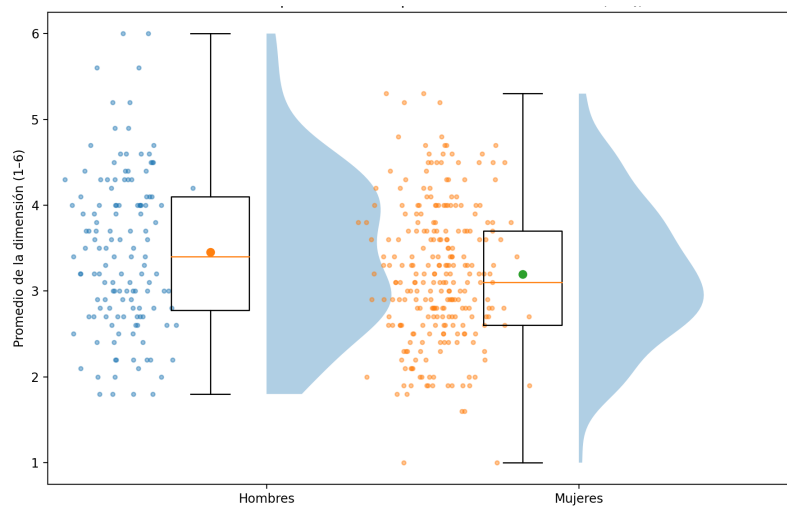
A continuación, se examinan las posibles diferencias en el promedio de la dimensión en función del sexo del profesorado. Dado que se comparan dos grupos independientes y que la variable no cumple plenamente los supuestos de normalidad, se aplicó la prueba U de Mann-Whitney para evaluar la existencia de diferencias estadísticamente significativas entre hombres y mujeres.

Tabla 8
Contraste de hipótesis según sexo

	Descriptivos			Test de contraste				
	Media	Desv. estándar	Mediana	U	p-valor	Dif. Medias	Tam. Efecto	Interpretación
Hombres	3.45	0.91	3.40	20465.0	.012	0.261	.153	Pequeño
Mujeres	3.19	0.80	3.10					

Como se presentaba en la Tabla 8, la distribución por sexo no es homogénea, con una mayor representación de mujeres (65.83%; $n = 262$) frente a hombres (34.17%; $n = 136$). En el promedio de la dimensión, los hombres presentan puntuaciones ligeramente superiores ($M = 3.45$; $DE = 0.91$; $Me = 3.40$) respecto a las mujeres ($M = 3.19$; $DE = 0.80$; $Me = 3.10$). El contraste no paramétrico mediante U de Mann-Whitney indica que estas diferencias son estadísticamente significativas ($U = 20465.0$; $p = .012$), si bien el tamaño del efecto es pequeño ($r_{rb} = .153$).

Figura 5
Raincloud plot – Sexo vs. promedio de la dimensión



Para apoyar los resultados obtenidos a partir de los test de contraste, la Figura 5 presenta un raincloud plot del promedio de la dimensión según sexo. Visualmente, se aprecia una ligera tendencia a puntuaciones más altas en hombres frente a mujeres, aunque con un amplio solapamiento entre distribuciones. Este patrón respalda la significación estadística del contraste, pero es coherente con un tamaño del efecto pequeño.

Según el tipo de centro

A continuación, se analizan las posibles diferencias medias de la dimensión en función del tipo de centro educativo. Para este fin, se utilizó un contraste no paramétrico adecuado para poder

comparar varios grupos independientes y determinar si la naturaleza o titularidad del centro se asociaba con diferencias significativas en el nivel competencial del profesorado.

Tabla 9
Contraste de hipótesis según tipo de centro

	Descriptivos			Test de contraste				
	Media	Desv. estándar	Mediana	H	p-valor	gl	ϵ^2	Interpretación
CEIP (Educación Infantil y Primaria)	3.312	.887	3.300	10.40	.003	4	.016	Diferencias significativas (efecto pequeño)
IES (Instituto de Educación Secundaria)	3.279	.850	3.200					
CPR (Colegio Público Rural)	3.590	.497	3.600					
SEMI-D (Educación Primaria y Primer ciclo de ESO)	2.890	.618	2.750					
Otros	3.089	.875	3.000					

Tabla 10
Comparaciones por pares (post-hoc confirmatorio: U de Mann-Whitney con corrección Holm)

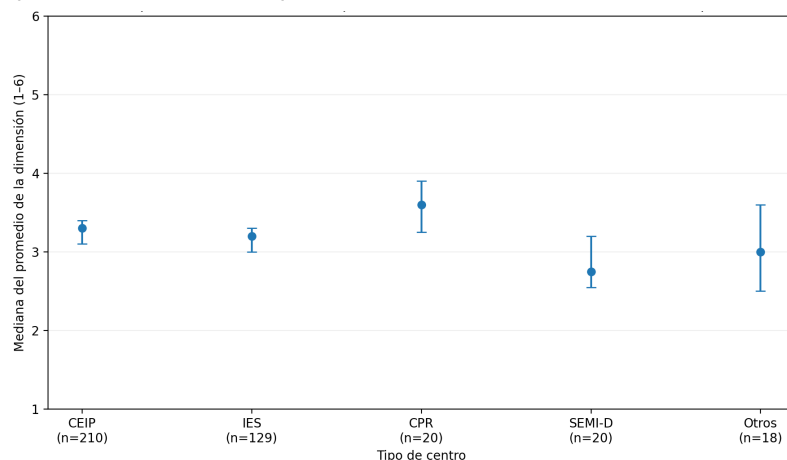
Comparación	U	p	p ajustada (Holm)	Dif. medias	r _{rb}
CEIP (Educación Infantil y Primaria) vs. IES (Instituto de Educación Secundaria)	13834.0	0.7417	> .999	0.033	0.021
CEIP (Educación Infantil y Primaria) vs. CPR (Colegio Público Rural)	1584.0	0.0695	0.4070	-0.278	-0.246
CEIP (Educación Infantil y Primaria) vs. SEMI-D (Educación Primaria y Primer ciclo de ESO)	2722.0	0.0287	0.2292	0.422	0.296
CEIP (Educación Infantil y Primaria) vs. Otros	2178.0	0.2839	> .999	0.223	0.152
IES (Instituto de Educación Secundaria) vs. CPR (Colegio Público Rural)	962.0	0.0678	0.4070	-0.311	-0.254
IES (Instituto de Educación Secundaria) vs. SEMI-D (Educación Primaria y Primer ciclo de ESO)	1657.0	0.0410	0.2868	0.389	0.284
IES (Instituto de Educación Secundaria) vs. Otros	1334.0	0.3075	> .999	0.190	0.149
CPR (Colegio Público Rural) vs. SEMI-D (Educación Primaria y Primer ciclo de ESO)	326.0	0.0006	0.0063	0.700	0.630
CPR (Colegio Público Rural) vs. Otros	262.0	0.0169	0.1523	0.501	0.456
SEMI-D (Educación Primaria y Primer ciclo de ESO) vs. Otros	168.0	0.7361	> .999	-0.199	-0.067

Las tablas anteriores muestran que el promedio de la dimensión presenta variaciones entre los distintos tipos de centro, con valores centrales y dispersión comparables en la mayoría de las categorías. El contraste global mediante Kruskal-Wallis evidencia diferencias estadísticamente significativas entre grupos ($H = 10.40$; $gl = 4$; $p = .003$); sin embargo, el tamaño del efecto es pequeño ($\epsilon^2 = .016$), lo que indica que el tipo de centro explica una proporción reducida de la variabilidad del constructo. Así, para la variable “tipo de centro” existen diferencias significativas, aunque estas deben interpretarse apoyándose en comparaciones por pares (post-hoc), para establecer, realmente, en qué categorías existían diferencias.

En la presente figura, se aprecian pequeñas diferencias y amplios solapamientos entre categorías, coherentes con efectos de baja magnitud.

Figura 6

Tipo de centro: mediana del promedio de la dimensión con IC 95%



Según las horas de formación relacionadas con la IA

Con el fin de analizar la posible influencia de las horas de formación en inteligencia artificial sobre el nivel competencial del profesorado, se examinaron las diferencias en el promedio de la dimensión en función de los distintos tramos formativos. Dado el carácter ordinal de la variable y la distribución de las puntuaciones, se aplicaron contrastes no paramétricos para comparar los grupos independientes, complementados con estadísticos descriptivos que permiten contextualizar la magnitud y el sentido de las diferencias observadas.

Tabla 11

Contraste de hipótesis según las horas de formación

	Descriptivos			Test de contraste				
	Media	Desv. estándar	Mediana	H	p-valor	gl	ϵ^2	Interpretación
No he recibido formación	3.070	.86	3.000	44.58	< .001	5	.101	Diferencias significativas (efecto mediano)
Menos de 30 horas	3.395	.73	3.400					
Entre 31 y 60 horas	3.532	.75	3.300					
Entre 61 y 90 horas	3.800	.69	4.000					
Entre 91 y 120 horas	4.271	.63	4.300					
Más de 120 horas	3.200	.98	3.200					

Tabla 12

Comparaciones por pares (post-hoc confirmatorio: U de Mann-Whitney con corrección Holm)

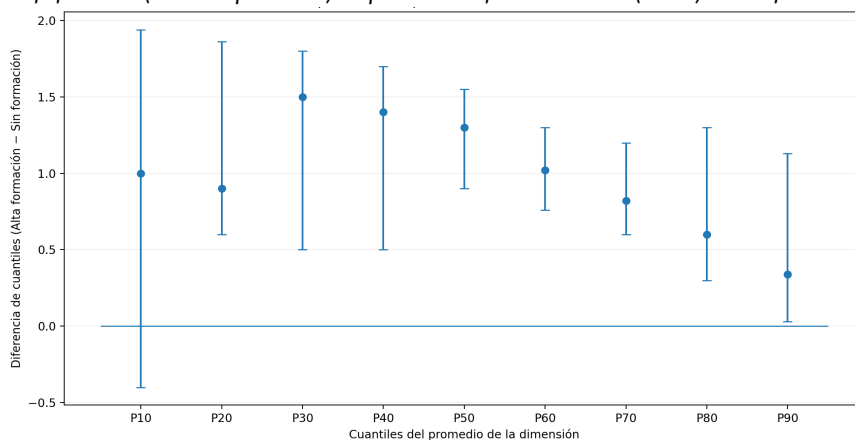
Comparación	U	p	p ajustada (Holm)	Dif. medias	r _{rb}
No he recibido formación vs. Menos de 30 horas	7430.5	.0011	.0122	-0.324	-0.237
No he recibido formación vs. Entre 31 y 60 horas	4176.5	.0003	.0033	-0.461	-0.315
No he recibido formación vs. Entre 61 y 90 horas	868.5	.0010	.0122	-0.730	-0.493
No he recibido formación vs. Entre 91 y 120 horas	409.0	< .001	< .001	-1.201	-0.727
No he recibido formación vs. Más de 120 horas	462.5	.6069	> .999	-0.130	-0.136

Menos de 30 horas vs. Entre 31 y 60 horas	2311.0	.2658	> .999	-0.137	-0.109
Menos de 30 horas vs. Entre 61 y 90 horas	457.0	.0179	.1435	-0.405	-0.372
Menos de 30 horas vs. Entre 91 y 120 horas	209.0	< .001	< .001	-0.877	-0.672
Menos de 30 horas vs. Más de 120 horas	218.0	.8818	> .999	0.195	-0.042
Entre 31 y 60 horas vs. Entre 61 y 90 horas	351.5	.1646	.9874	-0.268	-0.229
Entre 31 y 60 horas vs. Entre 91 y 120 horas	176.0	.0013	.0126	-0.740	-0.559
Entre 31 y 60 horas vs. Más de 120 horas	154.0	.7756	> .999	0.332	0.081
Entre 61 y 90 horas vs. Entre 91 y 120 horas	66.0	.0574	.4015	-0.471	-0.411
Entre 61 y 90 horas vs. Más de 120 horas	53.0	.2961	> .999	0.600	0.325
Entre 91 y 120 horas vs. Más de 120 horas	62.0	.0131	.1180	1.071	0.771

En horas de formación en IA, se observan diferencias globales y un patrón ascendente de puntuaciones hasta el tramo de 91-120 horas, mientras que el grupo >120 horas debe interpretarse con cautela por su reducido n.

Figura 7

Shift function (bootstrap IC 95%): impacto de la formación alta ($\geq 91h$) vs. sin formación



La Figura 7 compara cuantiles del promedio entre niveles de formación: las mayores diferencias se concentran en los tramos centrales y altos de la distribución, apoyando que la formación se asocia a mejoras consistentes en el perfil competencial.

Discusión y conclusión

Desde una perspectiva competencial, las puntuaciones globales neutras o moderadamente favorables indican una adopción aún incipiente y desigual de la IA en la práctica docente, con áreas específicas donde el uso resulta más limitado (Zawacki-Richter et al., 2019; Crompton & Burke, 2023).

Las variables sociodemográficas mostraron, en general, muy poca variación, lo que invita a que el perfil no depende significativamente del género, la edad o el contexto institucional, aunque puede haber ciertos matices según la trayectoria profesional y el tipo de centro (Redecker & Punie, 2017).

Las puntuaciones más altas se asocian a la formación específica en IA, lo que muestra las necesidades de itinerarios formativos estructurados, que se apliquen y sostengan en el tiempo (Celik et al., 2022; Mishra & Koehler, 2006). Estos hallazgos convergen con investigaciones recientes que sitúan la formación específica en IA como el principal factor asociado a mayores niveles competenciales y a actitudes más favorables hacia su integración pedagógica (Cabero et al., 2026; Area & González, 2026).

Desde la perspectiva práctica, estos resultados refuerzan la necesidad de diseñar e implementar programas de formación para docentes en IA que no se limiten a un enfoque técnico, sino que incorporen dimensiones pedagógicas, éticas y críticas, tal como recomiendan organismos internacionales como UNESCO (2024). Estas necesidades formativas se ven reforzadas por el análisis de buenas prácticas en la implementación de la IA generativa en las universidades, que subraya el papel de la formación continua del profesorado como condición para una adopción institucional sostenible (Ares et al., 2026). Así, la formación puede ejercer como una llave para reducir desigualdades competenciales y favorecer una integración educativa de la IA que sea más sostenible y equitativa.

En cuanto al perfil competencial de los docentes en el uso de la IA, este nos muestra un nivel medio, con diferencias significativas entre participantes. Las diferencias relacionadas con las variables sociodemográficas, como “tipo de centro” y “antigüedad docente”, aunque son estadísticamente significativas, sugieren un impacto práctico menor. Por otro lado, en cuanto a la “formación específica en IA” se confirma como el factor más determinante, asociado a diferencias consistentes. Estos resultados respaldan la necesidad de priorizar políticas y programas de desarrollo profesional docente centrados en IA, alineados con marcos internacionales de competencia digital y orientados a una integración pedagógica, ética y reflexiva de estas tecnologías en el ámbito educativo (Selwyn, 2019; Floridi et al., 2018).

En cuanto a las limitaciones de este estudio podemos mencionar el uso de un muestreo no probabilístico, lo que restringe la generalización de los resultados y la falta de análisis longitudinal. En futuras líneas de investigación, sería recomendable implementar estudios longitudinales y enfoques mixtos, que integren tanto análisis cuantitativos como cualitativos, para explorar más a fondo las barreras y oportunidades de la IA en la educación y las necesidades formativas de los docentes.

Consideraciones éticas de la investigación y uso de inteligencia artificial

Los autores declaran que, siguiendo las directrices de la Declaración de Helsinki y el resto de normativas éticas y legales aplicables a la investigación en seres humanos, y de acuerdo con la normativa española de protección de datos de carácter personal, todas las personas participantes en este estudio fueron informadas debidamente sobre los objetivos de la investigación y el tratamiento de sus datos mediante el consentimiento informado.

Las herramientas de inteligencia artificial se utilizaron únicamente para apoyo en la redacción y mejora del estilo del manuscrito. La responsabilidad del contenido, análisis y conclusiones recae íntegramente en los autores.

Agradecimientos y financiación

Este estudio no recibió financiación externa.

Conflicto de intereses

Los autores declaran no tener ningún conflicto de intereses.

Contribuciones de los autores

Conceptualización, J.R.M. y P.S.B.; metodología, J.R.M. y P.S.B.; software, J.R.M. y P.S.B.; validación, J.R.M. y P.S.B.; análisis formal, J.R.M. y P.S.B.; investigación, J.R.M. y P.S.B.; recursos, J.R.M. y P.S.B.; análisis de datos, J.R.M. y P.S.B.; redacción del borrador original, P.S.B.; redacción, revisión y edición, J.R.M.; supervisión, J.R.M.; administración de proyectos, J.R.M.

Referencias

- Area, M., & González, M. (2026). Análisis de un caso práctico de un modelo didáctico para el pensamiento crítico con la inteligencia artificial en la educación superior. *Pixel-Bit. Revista de Medios y Educación*, 76, Art. 1. <https://doi.org/10.12795/pixelbit.116639>
- Ares, M., De la Cruz, J. C., Morodo, A., & Robert, S. (2026). Buenas prácticas en el uso de la IA generativa en las universidades. Un análisis documental. *Innoeduca. International Journal of Technology and Educational Innovation*, 12(1), 5-20. <https://doi.org/10.24310/ijtei.121.2026.22152>
- Cabero, J., Barroso, J., Loaiza, M. I., & Palacios, A. (2026). Artificial intelligence in higher education: Cognitive, emotional, behavioral and ethical attitudes of Latin American students. *Journal of New Approaches in Educational Research*, 15(1). <https://doi.org/10.1007/s44322-026-00063-2>
- Ceballos, E. A., Cuellar, E. P., & Correa, J. E. (2026). Escala sobre el conocimiento tecnológico pedagógico del contenido en inteligencia artificial: Adaptación transcultural y propiedades psicométricas de la versión en español. *Pixel-Bit. Revista de Medios y Educación*, 76, Art. 5. <https://doi.org/10.12795/pixelbit.117639>
- Celik, I., Dindar, M., Muukkonen, H., y Järvelä, S. (2022). The promises and challenges of artificial intelligence for teachers: A systematic review of research. *TechTrends*, 66(4), 616–630. <https://doi.org/10.1007/s11528-022-00715-y>
- Colomo, E., Cabero, J., Ruiz, J., & Palacios, A. (2026). Ethical and responsible use of generative artificial intelligence by university students: design and validation of an instrument using the expert competence coefficient. *IJERI: International Journal of Educational Research and Innovation*, (25), 1-20. <https://doi.org/10.46661/ijeri.13253>
- Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education*, 20, 22. <https://doi.org/10.1186/s41239-023-00392-8>
- Crompton, H., Jones, M. V., & Burke, D. (2024). Affordances and challenges of artificial intelligence in K-12 education. A systematic review. *Journal of Research on Technology in Education*, 56(3), 248–268. <https://doi.org/10.1080/15391523.2022.2121344>
- Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, 25(1), 7-29. <https://doi.org/10.1177/0956797613504966>
- Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. Chapman & Hall/CRC. <https://doi.org/10.1007/978-1-4899-4541-9>
- Escobar, J., & Cuervo, A. (2008). Validez de contenido y juicio de expertos: una aproximación a su utilización. *Avances en Medición*, 6(1), 27-36.
- Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society. *Minds and Machines*, 28(4), 689-707. <https://doi.org/10.1007/s11023-018-9482-5>
- Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., & Altman, D. G. (2016). Statistical tests, P values, confidence intervals, and power: A guide to misinterpretations. *European Journal of Epidemiology*, 31(4), 337-350. <https://doi.org/10.1007/s10654-016-0149-3>
- Ho, J., Tumkaya, T., Aryal, S., Choi, H., & Claridge-Chang, A. (2019). Moving beyond P values: Data analysis with estimation graphics. *Nature Methods*, 16(7), 565-566. <https://doi.org/10.1038/s41592-019-0470-3>
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2), 65-70.
- Kerby, D. S. (2014). The simple difference formula: An approach to teaching nonparametric

- correlation. *Comprehensive Psychology*, 3, Article 11. <https://doi.org/10.2466/11.IT.3.1>
- Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, 47(260), 583-621. <https://doi.org/10.1080/01621459.1952.10483441>
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4, 863. <https://doi.org/10.3389/fpsyg.2013.00863>
- Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, 18(1), 50-60. <https://doi.org/10.1214/aoms/1177730491>
- Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record*, 108(6), 1017-1054. <https://doi.org/10.1111/j.1467-9620.2006.00684.x>
- OECD. (2021). OECD digital education outlook 2021: Pushing the frontiers with artificial intelligence, blockchain and robots. OECD Publishing. <https://doi.org/10.1787/589b283f-en>
- Redecker, C., & Punie, Y. (2017). *European framework for the digital competence of educators: DigCompEdu* (EUR 28775 EN). Publications Office of the European Union. <https://doi.org/10.2760/159770>
- Rousselet, G. A., Pernet, C. R., & Wilcox, R. R. (2017). Beyond differences in means: Robust graphical methods to compare two groups in neuroscience. *European Journal of Neuroscience*, 46(2), 1738-1748. <https://doi.org/10.1111/ejn.13610>
- Sadiq, G., Li, Y., Jing, D., & Abbas, Z. (2026). Inteligencia Artificial generativa en la lectura académica: Impactos en la carga cognitiva y los comportamientos de lectura de los estudiantes. Innoeduca. *International Journal of Technology and Educational Innovation*, 12(1), 21-39. <https://doi.org/10.24310/ijtei.121.2026.22461>
- Selwyn, N. (2019). *Should robots replace teachers? AI and the future of education*. Polity.
- Selwyn, N., Hillman, T., Eynon, R., Ferreira, G., Knox, J., Macgilchrist, F., & Sancho-Gil, J. M. (2020). What's next for Ed-Tech? Critical hopes and concerns for the 2020s. *Learning, Media and Technology*, 45(1), 1-6. <https://doi.org/10.1080/17439884.2020.1694945>
- UNESCO (2023). *Guidance for generative AI in education and research*. UNESCO Publishing.
- UNESCO (2024). *AI competency framework for teachers*. UNESCO.
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education. *International Journal of Educational Technology in Higher Education*, 16(1), 39. <https://doi.org/10.1186/s41239-019-0171-0>