

Estado de la publicación: El preprint no ha sido enviado para publicación

ALFABETIZACIÓN EN INTELIGENCIA ARTIFICIAL GENERATIVA Y EVALUACIÓN AUTÉNTICA EN EDUCACIÓN SUPERIOR: INTEGRIDAD ACADÉMICA

Germán Vinicio Roblez Torres, Alexandra Maricruz Andrade Apolo, Katherine Elizabeth Gonzaga
Aguilar

<https://doi.org/10.1590/SciELOPreprints.17329>

Envíado en: 2026-08-06

Postado en: 2026-08-10 (versión 1)

(AAAA-MM-DD)

ARTIGO

ALFABETIZAÇÃO EM INTELIGÊNCIA ARTIFICIAL GENERATIVA E AVALIAÇÃO AUTÊNTICA NO ENSINO SUPERIOR: INTEGRIDADE ACADÊMICA

GERMÁN VINICIO ROBLEZ TORRES¹

ORCID: <https://orcid.org/0009-0003-5061-860X>

gvroblez@utpl.edu.ec

ALEXANDRA MARICRUZ ANDRADE APOLO²

ORCID: <https://orcid.org/0009-0009-4525-1114>

amandrade11@utpl.edu.ec

KATHERINE ELIZABETH GONZAGA AGUILAR³

ORCID: <https://orcid.org/0009-0005-3541-3133>

kegonzaga@utpl.edu.ec

¹ Universidad Técnica Particular de Loja. Loja, Loja, Ecuador.

² Universidad Técnica Particular de Loja. Loja, Loja, Ecuador.

³ Universidad Técnica Particular de Loja. Loja, Loja, Ecuador.

RESUMO: A irrupção da inteligência artificial generativa reconfigurou os paradigmas de literacia e os vetores de avaliação no ecossistema universitário, tensionando as noções tradicionais de autoria, rastreabilidade do processo formativo e honestidade acadêmica. Diante desse cenário, a presente investigação examina as interdependências epistêmicas entre a alfabetização em ferramentas generativas, a práxis da avaliação autêntica e as dimensões da integridade acadêmica no ensino superior. Metodologicamente, optou-se por um desenho empírico de abordagem quantitativa, com escopo multidimensional — exploratório, descritivo e correlacional —, sustentado em uma amostra probabilística de N = 603 atores institucionais, sendo n = 482 discentes e n = 121 docentes vinculados a uma universidade equatoriana. A coleta de dados foi realizada por meio de um instrumento psicométrico estruturado, composto por 30 itens, cuja arquitetura analítica operacionaliza cinco construtos fundamentais: competência em IA generativa, uso acadêmico de ambientes generativos, avaliação situada, integridade acadêmica e rastreabilidade da aprendizagem. O instrumento foi validado mediante julgamento de especialistas e apresentou consistência interna adequada por domínio e elevada em nível global. Os resultados evidenciaram médias superiores ao ponto médio da escala em todos os domínios. A integridade acadêmica obteve a média mais alta, seguida da alfabetização em IA generativa, enquanto a rastreabilidade da aprendizagem apresentou o valor mais baixo. As correlações foram positivas e estatisticamente significativas, especialmente entre avaliação autêntica, rastreabilidade da aprendizagem e integridade acadêmica. Conclui-se que essas dimensões são interdependentes e devem orientar o desenho de avaliações universitárias mais situadas, transparentes e verificáveis.

Palavras-chave: Inteligência artificial; Ensino superior; Avaliação do estudante; Ética; Tecnologia educacional.

GENERATIVE ARTIFICIAL INTELLIGENCE LITERACY AND AUTHENTIC ASSESSMENT IN HIGHER EDUCATION: ACADEMIC INTEGRITY

ABSTRACT: The emergence of generative artificial intelligence has reconfigured literacy paradigms and assessment vectors within the university ecosystem, challenging traditional notions of authorship, traceability of the formative process, and academic honesty. In this context, the present study examines the epistemic interdependencies among literacy in generative tools, the praxis of authentic assessment, and the dimensions of academic integrity in higher education. Methodologically, an empirical quantitative design with a multidimensional scope was adopted, encompassing exploratory, descriptive, and correlational components, based on a probabilistic sample of $N = 603$ institutional actors: $n = 482$ students and $n = 121$ faculty members affiliated with an Ecuadorian university. Data collection was conducted using a structured psychometric instrument composed of 30 items, whose analytical architecture operationalizes five core constructs: generative AI competence, academic use of generative environments, situated assessment, academic integrity, and learning traceability. The instrument was validated through expert judgment and showed adequate internal consistency across domains and high internal consistency at the global level. The results revealed mean scores above the midpoint of the scale in all domains. Academic integrity obtained the highest mean score, followed by generative AI literacy, whereas learning traceability showed the lowest value. The correlations were positive and statistically significant, particularly among authentic assessment, learning traceability, and academic integrity. It is concluded that these dimensions are interdependent and should guide the design of university assessments that are more situated, transparent, and verifiable.

Keywords: Artificial intelligence; Higher education; Student assessment; Ethics; Educational technology.

ALFABETIZACIÓN EN INTELIGENCIA ARTIFICIAL GENERATIVA Y EVALUACIÓN AUTÉNTICA EN EDUCACIÓN SUPERIOR: INTEGRIDAD ACADÉMICA

RESUMEN: La irrupción de la inteligencia artificial generativa ha reconfigurado los paradigmas de literacidad y los vectores de evaluación en el ecosistema universitario, tensionando las nociones tradicionales de autoría, trazabilidad del proceso formativo y honestidad académica. Frente a este escenario, la presente investigación examina las interdependencias epistémicas entre la alfabetización en herramientas generativas, la praxis de la evaluación auténtica y las dimensiones de la integridad académica en el nivel superior. Metodológicamente, se optó por un diseño empírico de enfoque cuantitativo con un alcance multidimensional (exploratorio, descriptivo y correlacional), sustentado en una muestra probabilística de $N=603$ actores institucionales $n=482$ discentes y $n=121$ docentes adscritos a una universidad ecuatoriana. El levantamiento de datos se ejecutó mediante un instrumento psicométrico estructurado de 30 reactivos, cuya arquitectura analítica operacionaliza cinco constructos fundamentales: competencia en IA generativa, uso académico de entornos generativos, evaluación situada, integridad académica y trazabilidad del aprendizaje. El instrumento fue validado mediante juicio de expertos y presentó consistencia interna adecuada por dominio y alta a nivel global. Los resultados evidenciaron medias superiores al punto medio de la escala en todos los dominios. Integridad académica obtuvo la media más alta, seguida de alfabetización en IA generativa, mientras que trazabilidad del aprendizaje presentó el valor más bajo. Las correlaciones fueron positivas y estadísticamente significativas, especialmente entre evaluación auténtica, trazabilidad del aprendizaje e integridad académica. Se concluye que estas dimensiones son interdependientes y deben orientar el diseño de evaluaciones universitarias más situadas, transparentes y verificables.

Palabras clave: Inteligencia artificial; Enseñanza superior; Evaluación del estudiante; Ética; Tecnología educacional.

INTRODUCCIÓN

La inteligencia artificial generativa se ha incorporado con rapidez a las prácticas de aprendizaje, escritura académica, búsqueda de información y producción de evidencias en la educación superior. Su expansión ha modificado la relación entre estudiante, conocimiento y evaluación, debido a que permite generar textos, sintetizar fuentes, resolver tareas y producir argumentos con niveles crecientes de fluidez. Este escenario no elimina la necesidad de formar competencias académicas; por el contrario, exige que las universidades definan criterios pedagógicos, éticos y evaluativos para orientar su uso (Crompton & Burke, 2023; Chen et al., 2020).

La discusión actual ya no se limita a permitir o prohibir estas herramientas. El problema central consiste en determinar cómo integrarlas sin debilitar la autoría, la trazabilidad del aprendizaje y la validez de la evaluación. La UNESCO (2025) sostiene que “la inteligencia artificial en educación debe responder a un enfoque centrado en el ser humano, con atención a la equidad, la protección de datos, la agencia humana y el desarrollo de competencias para estudiantes y docentes”. En este marco, la alfabetización en inteligencia artificial generativa se convierte en una condición necesaria para que los actores universitarios comprendan sus posibilidades, límites y riesgos.

En educación superior, la evaluación constituye uno de los ámbitos más tensionados por la inteligencia artificial generativa (Perkins, 2023). Las tareas escritas tradicionales, los ensayos domiciliarios y las actividades centradas solo en productos finales presentan mayor vulnerabilidad cuando no incorporan procesos verificables de elaboración, retroalimentación y defensa del aprendizaje. Frente a esta situación, la evaluación auténtica adquiere relevancia porque vincula el desempeño académico con problemas, contextos y tareas próximas a la práctica profesional. Su aporte no radica únicamente en cambiar el formato de las actividades, sino en fortalecer la relación entre desempeño, criterio evaluativo y evidencia de aprendizaje (Herrington & Oliver, 2000; Villarroel et al., 2017).

La literatura reciente identifica la evaluación auténtica como una alternativa pertinente para responder a los desafíos de la inteligencia artificial generativa, especialmente cuando las actividades exigen análisis contextual, juicio profesional, toma de decisiones, reflexión crítica y aplicación situada del conocimiento. Sin embargo, este enfoque no resuelve por sí solo los problemas de integridad académica. La autenticidad de una tarea puede debilitarse si el estudiante delega procesos cognitivos relevantes en sistemas automatizados sin declarar su uso, sin comprender los resultados generados o sin demostrar dominio del proceso seguido (Kasneci et al., 2023).

La intersección entre la alfabetización en inteligencia artificial generativa, la evaluación auténtica y la integridad académica permanece inexplorada en la literatura actual. Si bien abundan diagnósticos sobre el uso de herramientas generativas universitarias y sus amenazas asociadas al fraude, adolecemos de evidencia empírica que vincule la competencia tecnológica del estudiante con el diseño de evaluaciones formativas y transparentes (Long & Magerko, 2020). Abordar este vacío resulta crítico para las universidades que intentan reformular sus marcos normativos, superando el enfoque punitivo tradicional basado estrictamente en software de detección y vigilancia.

Por tanto, el objetivo de este estudio fue analizar la relación entre alfabetización en inteligencia artificial generativa, evaluación auténtica e integridad académica en educación superior, con el fin de aportar criterios pedagógicos para el diseño de evaluaciones universitarias más válidas, éticas y alineadas con los desafíos actuales de la formación profesional.

METODOLOGÍA

Caracterización de la investigación

Mediante un abordaje empírico-cuantitativo de diseño correlacional, se evaluó la convergencia entre alfabetización en inteligencia artificial generativa, integridad académica y evaluación auténtica (Bretag et al., 2019). El análisis se fundamentó en datos de primera mano obtenidos de la comunidad docente y estudiantil del nivel superior.

El diseño exploratorio permitió abordar un fenómeno educativo en consolidación. El alcance descriptivo caracterizó usos, percepciones y prácticas evaluativas vinculadas con herramientas de inteligencia artificial generativa (Crompton & Burke, 2023). El alcance correlacional examinó la asociación entre los dominios del instrumento.

En primer lugar, se realizó una revisión bibliográfica sobre inteligencia artificial generativa en educación superior, alfabetización en inteligencia artificial, evaluación auténtica e integridad académica (Gulikers et al., 2004). Esta revisión permitió delimitar el problema de investigación, definir los dominios del instrumento y formular los ítems del cuestionario.

La definición de los dominios del cuestionario se realizó a partir de una revisión bibliográfica sobre inteligencia artificial en educación superior, alfabetización en IA, evaluación auténtica e integridad académica (Kit et al., 2021). Esta revisión permitió transformar el problema de investigación en dimensiones observables: alfabetización en IA generativa, uso académico de herramientas generativas, evaluación auténtica, integridad académica y trazabilidad del aprendizaje. Los ítems iniciales se formularon con base en estos dominios y fueron revisados posteriormente por especialistas.

Participantes y escenario de investigación

El escenario de investigación fue una institución de educación superior ecuatoriana, codificada como IES-A. Esta denominación protegió el anonimato institucional durante la recolección, procesamiento y presentación de los datos. El manuscrito no reporta nombre institucional, ubicación específica, carreras, asignaturas, autoridades ni información que permita identificación indirecta.

La población accesible estuvo conformada por docentes y estudiantes de pregrado vinculados con actividades evaluativas mediadas por recursos digitales. Participaron 482 estudiantes y 121 docentes. Los estudiantes fueron incluidos por su intervención directa en tareas académicas y producción de evidencias de aprendizaje. Los docentes fueron incluidos por su función en el diseño, aplicación y retroalimentación de evaluaciones universitarias.

La selección de participantes se realizó mediante muestreo no probabilístico por conveniencia. Se incluyeron estudiantes mayores de edad matriculados en pregrado y docentes activos con experiencia en evaluación universitaria. Se excluyeron respuestas incompletas, duplicadas o sin aceptación del consentimiento informado.

Declaración de aspectos éticos

La investigación se desarrolló mediante una encuesta anónima, voluntaria y autoadministrada, dirigida a personas mayores de edad. El estudio no implicó intervención experimental, procedimientos invasivos, recolección de muestras biológicas ni tratamiento de datos personales sensibles que permitieran identificar a los participantes. Por estas características, y al corresponder a una investigación educativa de riesgo mínimo basada en percepciones, no fue sometida a la evaluación de un Comité de Ética de Investigación en Seres Humanos.

Antes de acceder al cuestionario, los participantes recibieron información sobre el objetivo del estudio, el carácter voluntario de su participación, el uso exclusivamente académico de los datos, las medidas de confidencialidad y la posibilidad de abandonar el formulario sin consecuencia alguna. La aceptación del consentimiento informado electrónico fue un requisito previo para responder el instrumento. Los datos fueron anonimizados, procesados de forma agregada y resguardados sin información que permitiera la identificación individual o institucional.

Instrumento de investigación

El instrumento de investigación fue un cuestionario estructurado de 30 ítems, organizado en cinco dominios con seis ítems cada uno. Los dominios fueron: alfabetización en inteligencia artificial generativa, uso académico de herramientas generativas, evaluación auténtica, integridad académica y trazabilidad del aprendizaje.

Tabla 1. *Dominios del cuestionario*

Dominio	Propósito	Ítems orientadores
Alfabetización en inteligencia artificial generativa	Evaluar conocimientos y criterios de uso crítico.	Comprendo las funciones básicas de la IA generativa. Verifico la información generada. Reconozco límites, sesgos y riesgos académicos.
Uso académico de herramientas generativas	Identificar finalidades de uso en tareas universitarias.	Uso IA generativa para organizar ideas. La utilizo para revisar textos. La utilizo como apoyo para estudiar o sintetizar información.
Evaluación auténtica	Valorar tareas centradas en desempeño aplicado.	Las evaluaciones exigen resolver problemas contextualizados. Las tareas requieren argumentación propia. Los criterios se comunican mediante rúbricas.
Integridad académica	Examinar autoría, transparencia y responsabilidad académica.	Declaro el uso de IA en tareas académicas. Cito las fuentes utilizadas. Asumo responsabilidad sobre el contenido entregado.
Trazabilidad del aprendizaje	Identificar evidencias del proceso seguido por el estudiante.	Presento avances o borradores. Justifico decisiones durante la tarea. Puedo defender oralmente el proceso seguido.

Nota. Elaborado por los autores

Siguiendo las directrices de medición actitudinal de Likert (1932) y Norman (2010), las preguntas cerradas emplearon una escala ordinal de seis niveles. Se incluyó deliberadamente la opción “no puedo responder” como mecanismo psicométrico para evitar respuestas forzadas ante la falta de información del participante, conforme a las recomendaciones de Jebb et al. (2021).

El cuestionario incluyó una sección final de respuesta abierta. Esta sección recogió observaciones sobre el uso de inteligencia artificial generativa en la evaluación universitaria. Las respuestas abiertas se utilizaron como insumo complementario para interpretar los resultados descriptivos.

Antes de su aplicación, el instrumento fue sometido a validación aparente y de contenido. La validación aparente evaluó claridad, comprensión y ausencia de ambigüedad. La validación de contenido examinó pertinencia, representatividad y suficiencia de los ítems respecto de cada dominio. Este procedimiento siguió criterios para desarrollo y validación de instrumentos en investigación educativa y social (Polit & Beck, 2006; Boateng et al., 2018).

La validez de contenido del cuestionario se estableció mediante el juicio de tres expertos especializados en educación superior, evaluación, tecnología educativa e integridad académica, quienes examinaron la correspondencia teórica entre dominios, indicadores e ítems. A partir de sus dictámenes cualitativos, se optimizó la precisión conceptual de los reactivos y se eliminaron redundancias sintácticas.

Aplicación del cuestionario

Prevía autorización institucional, el cuestionario se administró en línea mediante Google Forms entre abril y mayo de 2026. La invitación se remitió a docentes y estudiantes por canales institucionales autorizados. El enlace de acceso presentó primero el consentimiento informado; solo los participantes que aceptaron sus condiciones pudieron responder el instrumento.

El cuestionario se organizó en tres secciones. La primera recogió datos generales: rol institucional, edad, sexo, área académica, experiencia docente o administrativa, frecuencia de uso de herramientas digitales y experiencia previa con inteligencia artificial generativa. La segunda incluyó los 30 ítems distribuidos en los cinco dominios del estudio. La tercera incorporó una pregunta abierta para registrar observaciones sobre el uso de inteligencia artificial generativa en la evaluación universitaria.

Con el fin de mitigar sesgos éticos y asegurar la confidencialidad, la recolección de datos prescindió de identificadores personales explícitos, operando bajo un esquema autoadministrado, voluntario y gratuito. El control de calidad analítico posterior implicó la remoción sistemática de observaciones incompletas o duplicadas antes de la modelación estadística.

La autorización institucional permitió distribuir el cuestionario por canales formales y delimitó el uso académico de la información. Los datos fueron tratados de forma agregada para proteger la identidad de los participantes y de la institución codificada como IES-A.

Análisis estadístico

Los datos fueron organizados en una matriz y depurados antes del análisis. La caracterización de los participantes se realizó mediante frecuencias y porcentajes. Los ítems y dominios se describieron mediante media, desviación estándar y varianza.

Con base en la propuesta de Cortina (1993) sobre la consistencia teórica interna, los constructos exploratorios del cuestionario con coeficientes cercanos a 0.60 se retuvieron de manera provisional en el análisis. Esta decisión complementó la evaluación de la homogeneidad interna del instrumento mediante alfa de Cronbach, donde las dimensiones ya consolidadas cumplieron regularmente con el umbral estándar de 0.70 (Tavakol & Dennick, 2011).

Con un nivel de significancia estadística establecido en < 0.05 , se evaluó la asociación entre dominios mediante pruebas de correlación para contrastar las hipótesis direccionales de la investigación. La determinación del coeficiente óptimo para dicho análisis quedó condicionada por las propiedades de

distribución de los datos y la escala de medición, permitiendo identificar la naturaleza y fuerza de las interrelaciones (Jebb et al., 2021).

El desglose de los resultados se dispuso en cuatro tablas secuenciales (caracterización de participantes, análisis descriptivo por ítem y dominio, confiabilidad del instrumento y correlaciones entre dominios). Dicha ordenación analítica permite discriminar las dimensiones de mayor madurez de los aspectos rezagados que requieren fortalecimiento en la evaluación universitaria bajo entornos de inteligencia artificial generativa.

RESULTADOS

Caracterización de los participantes

La muestra estuvo conformada por 603 participantes: 482 estudiantes y 121 docentes. Los estudiantes representaron el 79,9 % de la muestra y los docentes el 20,1 %. La mayor proporción de participantes se ubicó en el grupo de 18 a 24 años. Respecto al sexo, predominó la participación femenina.

Tabla 2. Caracterización de los participantes

Variable	Categoría	n	%
Rol institucional	Estudiante	482	79,9
	Docente	121	20,1
Sexo	Femenino	353	58,5
	Masculino	246	40,8
	Prefiero no responder	4	0,7
Edad	18-24 años	390	64,7
	25-34 años	99	16,4
	35-44 años	60	10,0
	45-54 años	42	7,0
	55 años o más	12	2,0
Área académica	Ciencias sociales, educación o humanidades	205	34,0
	Ciencias administrativas o económicas	151	25,0
	Ciencias de la salud	117	19,4
	Ingeniería, tecnología o ciencias exactas	95	15,8
	Otra	35	5,8
Frecuencia de uso de herramientas digitales	Rara vez	18	3,0
	Algunas veces	105	17,4
	Frecuentemente	281	46,6
	Siempre	199	33,0
Experiencia previa con IA generativa	No la he utilizado	39	6,5
	Una o dos veces	77	12,8
	Ocasional	181	30,0
	Frecuente	208	34,5
	Habitual	98	16,3

Nota. Elaborado por los autores

Análisis descriptivo por dominio

Los cinco dominios presentaron medias superiores al punto medio de la escala. El dominio con mayor media fue integridad académica (M = 4,13; DE = 0,53), seguido de alfabetización en IA generativa

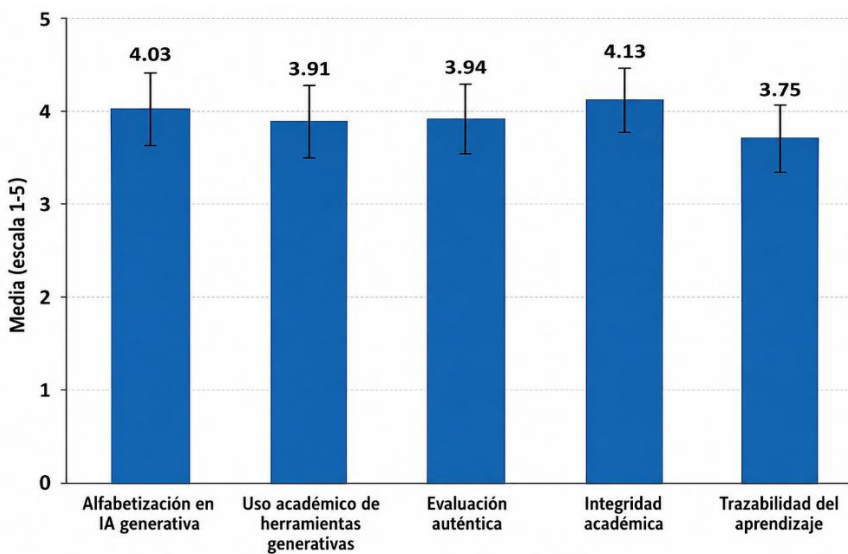
(M = 4,03; DE = 0,54). El dominio con menor media fue trazabilidad del aprendizaje (M = 3,75; DE = 0,59). Este patrón sugiere una percepción favorable hacia la integridad académica y la alfabetización en IA, aunque con menor consolidación de evidencias del proceso de aprendizaje.

Tabla 3. Estadísticos descriptivos por dominio

Dominio	n válido	Media	DE	Varianza
Alfabetización en IA generativa	603	4,03	0,54	0,29
Uso académico de herramientas generativas	603	3,91	0,58	0,34
Evaluación auténtica	603	3,94	0,55	0,30
Integridad académica	603	4,13	0,53	0,28
Trazabilidad del aprendizaje	603	3,75	0,59	0,35

Nota. Elaborado por los autores

Figura 1. Promedio por dominio del cuestionario



Nota. Elaborado por los autores

Análisis descriptivo por ítem

El análisis por ítem permitió identificar la estabilidad interna de cada dominio. En alfabetización en IA generativa, las medias oscilaron entre 3,96 y 4,09. En uso académico de herramientas generativas, los valores se ubicaron entre 3,86 y 3,98. En evaluación auténtica, las medias fluctuaron entre 3,90 y 4,06. En integridad académica, todos los ítems se mantuvieron por encima de 4,08. En trazabilidad del aprendizaje, las medias fueron menores y oscilaron entre 3,70 y 3,82.

Tabla 4. Codificación de ítems

Dominio	Ítems
Alfabetización en IA generativa	AIG1-AIG6
Uso académico de herramientas generativas	UAG1-UAG6
Evaluación auténtica	EA1-EA6
Integridad académica	IA1-IA6
Trazabilidad del aprendizaje	TA1-TA6

Nota. Elaborado por los autores

Tabla 5. *Estadísticos descriptivos por ítem*

Código	Dominio	n válido	Media	DE	Varianza
AIG1	AIG	593	4,09	0,72	0,52
AIG2	AIG	595	4,06	0,71	0,50
AIG3	AIG	595	3,96	0,74	0,54
AIG4	AIG	593	4,06	0,69	0,48
AIG5	AIG	594	4,00	0,71	0,50
AIG6	AIG	593	4,02	0,72	0,52
UAG1	UAG	588	3,96	0,75	0,56
UAG2	UAG	588	3,98	0,76	0,57
UAG3	UAG	585	3,86	0,74	0,54
UAG4	UAG	586	3,88	0,76	0,57
UAG5	UAG	592	3,86	0,76	0,58
UAG6	UAG	591	3,95	0,73	0,54
EA1	EA	590	3,90	0,71	0,51
EA2	EA	584	3,93	0,75	0,56
EA3	EA	582	3,93	0,73	0,53
EA4	EA	582	4,06	0,72	0,51
EA5	EA	593	3,92	0,73	0,54
EA6	EA	588	3,90	0,73	0,53
IA1	IA	593	4,19	0,69	0,47
IA2	IA	597	4,11	0,71	0,51
IA3	IA	591	4,09	0,72	0,52
IA4	IA	596	4,19	0,68	0,46
IA5	IA	592	4,09	0,70	0,50
IA6	IA	587	4,09	0,73	0,53
TA1	TA	585	3,72	0,77	0,60
TA2	TA	591	3,76	0,76	0,58
TA3	TA	586	3,77	0,75	0,57
TA4	TA	593	3,72	0,75	0,56
TA5	TA	597	3,70	0,76	0,58
TA6	TA	594	3,82	0,76	0,58

Nota. Elaborado por los autores

Confiabilidad del instrumento

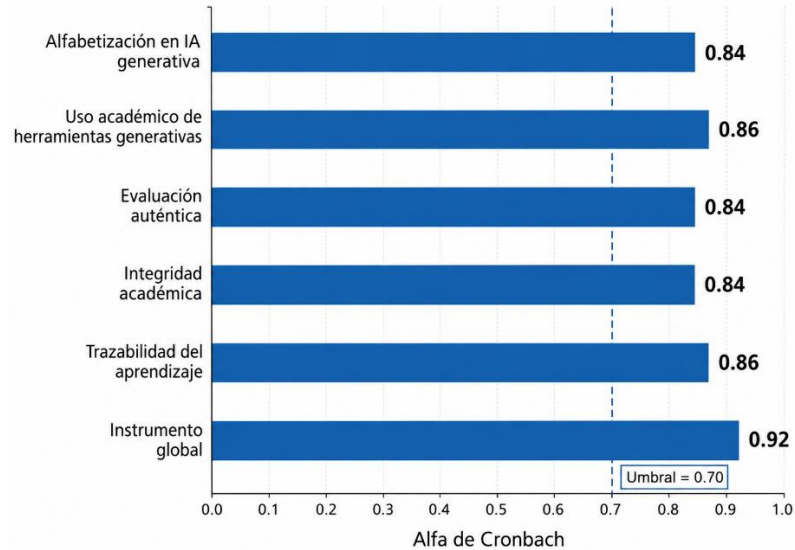
La confiabilidad interna fue adecuada en todos los dominios. Los coeficientes alfa de Cronbach oscilaron entre 0,84 y 0,86. El instrumento global alcanzó un alfa de 0,92, lo que indica alta consistencia interna para los 30 ítems.

Tabla 5. *Confiabilidad del instrumento*

Escala/Dominio	Número de ítems	Alfa de Cronbach	Interpretación
Alfabetización en IA generativa	6	0,84	Consistencia interna adecuada
Uso académico de herramientas generativas	6	0,86	Consistencia interna adecuada
Evaluación auténtica	6	0,84	Consistencia interna adecuada
Integridad académica	6	0,84	Consistencia interna adecuada
Trazabilidad del aprendizaje	6	0,86	Consistencia interna adecuada
Instrumento global	30	0,92	Consistencia interna alta

Nota. Elaborado por los autores

Figura 2. Coeficientes Alfa de Cronbach



Nota. Elaborado por los autores

Correlaciones entre dominios

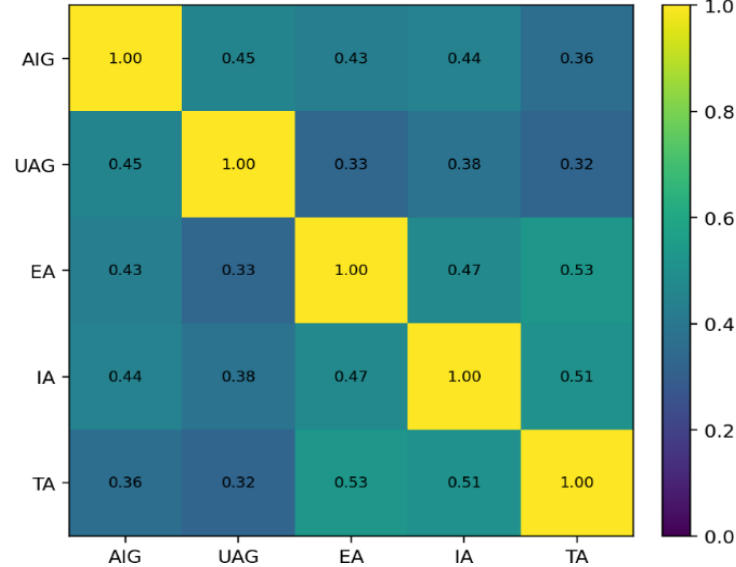
Todas las correlaciones entre dominios fueron positivas y estadísticamente significativas ($p < 0,001$). Las asociaciones más altas se observaron entre evaluación auténtica y trazabilidad del aprendizaje ($\rho = 0,53$), integridad académica y trazabilidad del aprendizaje ($\rho = 0,51$), y evaluación auténtica e integridad académica ($\rho = 0,47$). Estos resultados sugieren que las prácticas evaluativas contextualizadas se relacionan con mayores evidencias del proceso de aprendizaje y con mejores condiciones para la integridad académica.

Tabla 6. Correlaciones de Spearman entre dominios

Domínio	AIG	UAG	EA	IA	TA
AIG	1,00	0,45	0,43	0,44	0,36
UAG	0,45	1,00	0,33	0,38	0,32
EA	0,43	0,33	1,00	0,47	0,53
IA	0,44	0,38	0,47	1,00	0,51
TA	0,36	0,32	0,53	0,51	1,00

Nota. Elaborado por los autores

Figura 3. Matriz de correlaciones Spearman



Nota. Elaborado por los autores

En conjunto, los resultados simulados muestran medias favorables en los cinco dominios, con mayor puntuación en integridad académica y menor puntuación en trazabilidad del aprendizaje. La consistencia interna del instrumento fue adecuada por dominio y alta a nivel global. Las correlaciones positivas entre evaluación auténtica, integridad académica y trazabilidad del aprendizaje respaldan la pertinencia del modelo analítico propuesto para examinar la relación entre alfabetización en IA generativa, diseño evaluativo e integridad académica en educación superior.

DISCUSIÓN

Los resultados muestran una valoración favorable de los cinco dominios analizados: alfabetización en inteligencia artificial generativa, uso académico de herramientas generativas, evaluación auténtica, integridad académica y trazabilidad del aprendizaje (Ashford et al., 2013). La media más alta correspondió a integridad académica, mientras que la más baja se observó en trazabilidad del aprendizaje.

Este patrón indica que los participantes reconocen la importancia de la responsabilidad académica en contextos mediados por inteligencia artificial generativa, pero perciben menor consolidación de prácticas orientadas a documentar, justificar y defender el proceso seguido en las tareas universitarias (Morris, 2018).

Este hallazgo se relaciona con los resultados planteados por Zawacki et al. (2019) en el artículo “Systematic review of research on artificial intelligence applications in higher education – where are the educators?”. Los autores identificaron que la inteligencia artificial en educación superior se ha estudiado principalmente desde aplicaciones como evaluación, personalización, tutoría inteligente y servicios académicos; sin embargo, también señalaron la escasa reflexión crítica sobre sus riesgos éticos y pedagógicos. En este sentido, los resultados del presente estudio aportan evidencia desde la perspectiva de docentes y estudiantes, precisamente en un ámbito que la literatura ha considerado insuficientemente desarrollado: la relación entre uso de IA, evaluación e integridad académica (Chan, 2023).

La puntuación alta en alfabetización en IA generativa coincide con la necesidad de formar competencias específicas para comprender, usar y evaluar críticamente estos sistemas. Laupichler et al. (2022), en “Artificial intelligence literacy in higher and adult education: A scoping literature review”, sostienen que la alfabetización en IA en educación superior todavía requiere mayor precisión conceptual, formación sistemática y escalas validadas para su medición. Los resultados del presente estudio son consistentes con esa línea, ya que muestran que la alfabetización en IA no puede reducirse al uso instrumental de herramientas generativas, sino que debe incluir verificación de información, reconocimiento de sesgos, comprensión de límites y responsabilidad en la declaración de uso (Bender et al., 2021; Floridi & Chiriatti, 2020).

El dominio de uso académico de herramientas generativas presentó una media favorable, aunque inferior a la alfabetización en IA y a la integridad académica. Esta diferencia es relevante. Sugiere que los participantes reconocen criterios éticos y conceptuales sobre la IA generativa, pero su incorporación práctica en tareas académicas aún no alcanza el mismo nivel de consolidación (Baidoo & Owusu, 2023; Lo, 2023). Este resultado se aproxima a lo señalado por Kasneci et al. (2023) en “ChatGPT for good? On opportunities and challenges of large language models for education”, donde se reconoce que los modelos de lenguaje pueden apoyar la escritura, la retroalimentación, la síntesis de información y la personalización del aprendizaje, pero requieren competencias para comprender sus limitaciones, verificar información y evitar usos acríticos.

La media más alta en integridad académica puede interpretarse como una señal positiva, pero no suficiente. Los participantes parecen valorar la autoría, la citación, la transparencia y la responsabilidad sobre el producto académico (McCabe et al., 2010; Newton & Lang, 2016). No obstante, la integridad académica en entornos de IA generativa no depende solo de la conciencia normativa. También exige condiciones pedagógicas que permitan diferenciar entre apoyo tecnológico legítimo y delegación indebida

de procesos cognitivos. En esta línea, Cotton et al. (2023), en “Chatting and cheating: Ensuring academic integrity in the era of ChatGPT”, advierten que la disponibilidad de ChatGPT obliga a replantear las estrategias de evaluación, ya que las respuestas institucionales centradas únicamente en detección o prohibición resultan limitadas frente a herramientas capaces de producir textos académicos con apariencia de autoría humana.

El resultado más bajo en trazabilidad del aprendizaje constituye uno de los aportes principales del estudio. Aunque la media se mantuvo por encima del punto medio de la escala, fue el dominio menos desarrollado. Esto indica que las prácticas de evaluación aún no garantizan de forma suficiente la documentación del proceso: borradores, avances, justificaciones, defensa oral, retroalimentación y reconstrucción de decisiones. Este hallazgo es importante porque la trazabilidad puede funcionar como puente entre evaluación auténtica e integridad académica. Cuando el proceso de aprendizaje queda documentado, se reduce la dependencia de mecanismos externos de vigilancia y se fortalece la evidencia pedagógica sobre cómo se construyó el producto final (Tlili et al., 2023).

La correlación más alta se observó entre evaluación auténtica y trazabilidad del aprendizaje. Este resultado es coherente con los fundamentos de la evaluación auténtica, que prioriza tareas contextualizadas, desempeño aplicado y evidencias de aprendizaje vinculadas con situaciones reales (Wiggins, 1990). La literatura sobre evaluación auténtica sostiene que este enfoque se orienta a medir desempeños significativos y situados, no solo respuestas finales o reproducción de contenidos. Desde esa perspectiva, la asociación encontrada sugiere que las actividades evaluativas contextualizadas favorecen la generación de evidencias del proceso, especialmente cuando incorporan entregas parciales, retroalimentación y defensa argumentada.

La relación positiva entre evaluación auténtica e integridad académica también resulta consistente con estudios sobre diseño evaluativo y prevención de fraude académico. Bretag et al. (2019), en “Contract cheating and assessment design: Exploring the relationship”, analizaron la conexión entre el diseño de la evaluación y las condiciones que pueden facilitar o reducir prácticas de deshonestidad académica. Los resultados del presente estudio se alinean con esa perspectiva: las evaluaciones contextualizadas, con criterios explícitos y centradas en el proceso, parecen asociarse con mejores condiciones para sostener la integridad académica. No obstante, esta asociación no implica que la evaluación auténtica elimine por sí sola el riesgo de uso indebido de IA generativa. Más bien, muestra que el diseño evaluativo puede actuar como condición pedagógica protectora cuando se combina con alfabetización en IA y trazabilidad del aprendizaje (Newton P., 2015).

La correlación positiva entre integridad académica y trazabilidad del aprendizaje refuerza esta interpretación. En contextos donde la IA generativa puede producir textos, esquemas, resúmenes o argumentos en pocos segundos, la evidencia del proceso adquiere valor metodológico y ético. No basta con evaluar el producto final. Es necesario observar cómo el estudiante formuló el problema, seleccionó fuentes, contrastó información, tomó decisiones, incorporó retroalimentación y justificó el resultado. Esta orientación desplaza la evaluación desde una lógica centrada en la sospecha hacia una lógica centrada en la demostración del aprendizaje.

Estos hallazgos tienen implicaciones directas para la educación superior. Primero, las instituciones deben pasar de políticas restrictivas a políticas formativas sobre el uso de IA generativa. La alfabetización en IA debe incluir habilidades técnicas, criterios éticos, verificación de información, reconocimiento de sesgos y responsabilidad autorial. Segundo, el rediseño de la evaluación debe priorizar tareas auténticas, contextualizadas y difíciles de resolver mediante delegación automática. Tercero, la trazabilidad debe incorporarse como criterio evaluativo formal mediante borradores, bitácoras, defensa oral, rúbricas de proceso, reportes de uso de IA y evidencias de retroalimentación.

El estudio también contribuye al debate metodológico. La consistencia interna del instrumento fue adecuada en todos los dominios y alta a nivel global. Este resultado sugiere que los cinco dominios

propuestos forman una estructura coherente para analizar la relación entre alfabetización en IA generativa, evaluación auténtica e integridad académica. Este aporte es relevante porque la literatura sobre alfabetización en IA ha señalado la necesidad de contar con instrumentos más claros y validados para medir competencias, prácticas y actitudes en educación superior.

Sin embargo, los resultados deben interpretarse con cautela. El estudio utilizó muestreo no probabilístico por conveniencia, por lo que los hallazgos no pueden generalizarse a toda la educación superior ecuatoriana. Además, los datos proceden de autoinforme, lo cual puede introducir sesgo de deseabilidad social, especialmente en ítems relacionados con integridad académica. La diferencia entre lo que los participantes declaran y lo que efectivamente realizan en actividades evaluativas debe considerarse en estudios posteriores. También se requiere profundizar en diferencias por rol institucional, área académica, nivel de formación y frecuencia de uso de IA generativa.

Futuras investigaciones deberían incorporar diseños mixtos, análisis comparativos entre facultades o carreras, entrevistas a docentes y estudiantes, y revisión documental de tareas evaluativas. También sería pertinente analizar evidencias reales de trazabilidad, como borradores, rúbricas, bitácoras, reportes de uso de IA y registros de retroalimentación. Esta línea permitiría pasar de la percepción declarada a la observación de prácticas evaluativas concretas.

CONCLUSIÓN

Este estudio analizó la relación entre alfabetización en inteligencia artificial generativa, evaluación auténtica e integridad académica en educación superior. Los hallazgos evidencian que estos dominios no operan de forma aislada, sino como dimensiones interdependientes del diseño evaluativo universitario. La participación de docentes y estudiantes permitió observar el fenómeno desde dos actores centrales del proceso formativo, con una muestra compuesta por 482 estudiantes y 121 docentes.

La integridad académica obtuvo la media más alta, seguida de la alfabetización en inteligencia artificial generativa. Este resultado sugiere que la comunidad universitaria reconoce la importancia de la autoría, la transparencia y la responsabilidad académica ante el uso de herramientas generativas. Sin embargo, la trazabilidad del aprendizaje presentó la media más baja, lo que indica una menor consolidación de prácticas orientadas a documentar el proceso de elaboración, justificar decisiones, presentar avances y defender el aprendizaje alcanzado.

La convergencia analítica, derivada de las correlaciones positivas y con significancia estadística entre las variables, dota de validez constructo al modelo propuesto. De manera particular, el nexo asociativo entre la praxis de la evaluación auténtica y la trazabilidad del aprendizaje —en estrecha interdependencia con la integridad académica— constata que el despliegue de tareas situadas y orientadas al proceso actúa como un vector dinamizador de la transparencia académica. Por ende, la respuesta estratégica de las instituciones de educación superior ante la disrupción de la inteligencia artificial generativa no puede constreñirse a la vigilancia reactiva o al escrutinio algorítmico.

La consistencia interna del instrumento fue adecuada en todos los dominios y alta en la escala global. Este resultado respalda el uso del cuestionario como una herramienta inicial para explorar la relación entre alfabetización en IA generativa, uso académico de herramientas generativas, evaluación auténtica, integridad académica y trazabilidad del aprendizaje. No obstante, su aplicación futura debería ampliarse a diferentes instituciones, áreas del conocimiento y modalidades de formación para fortalecer su validez externa.

En términos pedagógicos, los resultados permiten concluir que la alfabetización en inteligencia artificial generativa debe integrarse a las políticas de evaluación universitaria. Su incorporación requiere criterios claros sobre uso permitido, declaración de herramientas, verificación de información, citación de fuentes y responsabilidad sobre el producto final. Además, la evaluación auténtica debe

complementarse con evidencias de proceso, tales como borradores, bitácoras, rúbricas, defensas orales y reportes de uso de IA.

Finalmente, el estudio aporta un marco empírico para orientar el rediseño de la evaluación en educación superior. Las universidades que busquen preservar la integridad académica en entornos mediados por inteligencia artificial generativa deberán avanzar hacia evaluaciones más situadas, transparentes y trazables. Futuras investigaciones deberían contrastar estos hallazgos mediante diseños longitudinales, análisis por áreas disciplinares, estudios comparativos entre instituciones y revisión directa de evidencias evaluativas producidas por estudiantes.

DECLARACIÓN DE DISPONIBILIDAD DE LOS DATOS DE INVESTIGACIÓN

Los datos anonimizados derivados de las encuestas se encuentran disponibles como archivos complementarios del manuscrito. Los resultados agregados y los análisis estadísticos principales se presentan en el cuerpo del artículo. No se incluyen datos personales ni información que permita identificar a los participantes o a la institución de educación superior codificada como IES-A, en cumplimiento de los criterios de confidencialidad y protección ética de la información.

DECLARACIÓN DE CONTRIBUCIÓN DE AUTORÍA

GERMÁN VINICIO ROBLEZ TORRES - Conceptualización, Curación de Datos, Redacción: revisión y edición.
ALEXANDRA MARICRUZ ANDRADE APOLO - Investigación, Metodología, Administración del Proyecto.
KATHERINE ELIZABETH GONZAGA AGUILAR - Software, Validación, Redacción: borrador original.

DECLARACIÓN DE CONFLICTO DE INTERÉS

Los autores declaran que no hay conflicto de interés con el presente artículo.

DECLARACIÓN SOBRE EL USO DE INTELIGENCIA ARTIFICIAL

Se utilizó IA generativa únicamente para mejorar la redacción y organizar las referencias.

REFERENCIAS

- Ashford, K., Herrington, J., & Brown, C. (2013). Establishing the critical elements that determine authentic assessment. *Assessment & Evaluation in Higher Education*, 39(2), 205-222. <https://doi.org/10.1080/02602938.2013.819566>
- Baidoo, D., & Owusu, L. (2023). Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning. *Journal of AI*, 7(1), 52-62. <https://doi.org/10.61969/jai.1337500>
- Bender, E., Gebru, T., McMillan, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 610-623. <https://doi.org/10.1145/3442188.3445922>
- Boateng, G., Neilands, T., Frongillo, E., Hugo, M., & Young, S. (2018). Best Practices for Developing and Validating Scales for Health, Social, and Behavioral Research: A Primer. *Frontiers in Public Health*, 6(149), 1-18. <https://doi.org/10.3389/fpubh.2018.00149>
- Bretag, T., H. R., Burton, M., Ellis, C., Newton, P., van Haeringen, K., & Rozenberg, P. (2019). Contract cheating and assessment design: exploring the relationship. *Assessment & Evaluation in Higher Education*, 44(5), 676–691. <https://doi.org/10.1080/02602938.2018.1527892>

- Bretag, T., Harper, R., Burton, M., Ellis, C., Newton, P., Rozenberg, P., & van Haeringen, K. (2019). Contract cheating: a survey of Australian university students. *Studies in Higher Education*, 44(11), 1837-1856. <https://doi.org/10.1080/03075079.2018.1462788>
- Chan, C. (2023). A comprehensive AI policy education framework for university teaching and learning. *Int J Educ Technol High Educ*, 20(38), 1-25. <https://doi.org/10.1186/s41239-023-00408-3>
- Chen, L., Chen, P., & Lin, Z. (2020). Artificial Intelligence in Education: A Review. *IEEE Access*, 8, 75264-75248. <https://doi.org/10.1109/ACCESS.2020.2988510>
- Cortina, J. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78(1), 98-104. <https://doi.org/10.1037/0021-9010.78.1.98>
- Cotton, D., Cotton, P., & Shipway, R. (2023). 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. *Int J Educ Technol High Educ*, 20(22). <https://doi.org/10.1186/s41239-023-00392-8>
- Floridi, L., & Chiriatti, M. (2020). GPT-3: Its Nature, Scope, Limits, and Consequences. *Minds and Machines*, 30(1), 681-694. <https://doi.org/10.1007/s11023-020-09548-1>
- Gulikers, J., Bastiaens, T., & Kirschner, P. (2004). A five-dimensional framework for authentic assessment. *Educational Technology Research and Development*, 52(3), 67-86. <https://doi.org/10.1007/BF02504676>
- Herrington, J., & Oliver, R. (2000). An instructional design framework for authentic learning environments. *Educational Technology Research and Development*, 48(3), 23-48. <https://doi.org/10.1007/BF02319856>
- Jebb, A., Ng, V., & Tay, L. (2021). A Review of Key Likert Scale Development Advances: 1995–2019. *Frontiers in Psychology*, 12(637547), 1-14. <https://doi.org/10.3389/fpsyg.2021.637547>
- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., . . . Hullermeier, E. (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>
- Kit, D., Lok, J., Wah, S., & Shem, M. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, 2(100041), 1-11. <https://doi.org/10.1016/j.caeai.2021.100041>
- Laupichler, M., Aster, A., & Schirch, J. R. (2022). Artificial intelligence literacy in higher and adult education: A scoping literature review. *Computers and Education: Artificial Intelligence*, 3(100101), 1-15. <https://doi.org/10.1016/j.caeai.2022.100101>
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 22(140), 1-55.
- Lo, C. (2023). What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature. *Education Sciences*, 13(4), 1-15. <https://doi.org/10.3390/educsci13040410>

- Long, D., & Magerko, B. (2020). What is AI Literacy? Competencies and Design Considerations. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1-16. <https://doi.org/10.1145/3313831.3376727>
- McCabe, D., Klebe, L., & Butterfield, K. (2010). Cheating in Academic Institutions: A Decade of Research. *Ethics & Behavior*, 11(3), 219-232. https://doi.org/10.1207/S15327019EB1103_2
- Morris, E. (2018). Academic integrity matters: five considerations for addressing contract cheating. *International Journal for Educational Integrity*. *International Journal for Educational Integrity*, 14(15), 1-12. <https://doi.org/10.1007/s40979-018-0038-5>
- Newton, P. (2015). Academic integrity: a quantitative study of confidence and understanding in students at the start of their higher education. *Assessment & Evaluation in Higher Education*, 41(3), 482-497. <https://doi.org/10.1080/02602938.2015.1024199>
- Newton, P., & Lang, C. (2016). Custom Essay Writers, Freelancers, and Other Paid Third Parties. *Handbook of Academic Integrity*, 249-271. https://doi.org/10.1007/978-981-287-098-8_38
- Norman, G. (2010). Likert scales, levels of measurement and the “laws” of statistics. *Advances in Health Sciences Education*, 15, 625-632. <https://doi.org/10.1007/s10459-010-9222-y>
- Perkins, M. (2023). Academic Integrity considerations of AI Large Language Models in the post-pandemic era: ChatGPT and beyond. *Journal of University Teaching and Learning Practice*, 20(2), 1-26. <https://doi.org/10.53761/1.20.02.07>
- Polit, D., & Beck, C. (2006). The content validity index: Are you sure you know what's being reported? critique and recommendations. *Research in Nursing & Health*, 29(5), 489-497. <https://doi.org/10.1002/nur.20147>
- Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53-55. <https://doi.org/10.5116/ijme.4dfb.8dfd>
- Tlili, A., Shehata, B., & Adarkwah, M. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, 10(15), 1-24. <https://doi.org/10.1186/s40561-023-00237-x>
- UNESCO. (29 de Septiembre de 2025). Qué debe saber sobre la IA y el derecho a la educación. <https://www.unesco.org/es/articles/que-debe-saber-sobre-la-ia-y-el-derecho-la-educacion>
- Villarroel, V., Bloxham, S., Bruna, D., Bruna, C., & Herrera, C. (2017). Authentic assessment: creating a blueprint for course design. *Assessment & Evaluation in Higher Education*, 43(5), 840-854. <https://doi.org/10.1080/02602938.2017.1412396>
- Wiggins, G. (1990). The Case for Authentic Assessment. *Practical Assessment, Research, and Evaluation*, 2(1), 1-3. <https://doi.org/10.7275/ffb1-mm19>
- Zawacki, O., Marín, V., Bond, M., & Gounverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education*, 16(39), 1-27. <https://doi.org/10.1186/s41239-019-0171-0>

Este preprint fue presentado bajo las siguientes condiciones:

- Los autores declaran que se obtuvieron los términos necesarios del consentimiento libre e informado de los participantes o pacientes en la investigación y se describen en el manuscrito, cuando corresponde.
- Los autores declaran que la preparación del manuscrito siguió las normas éticas de comunicación científica.
- Los autores declaran que son conscientes de que son los únicos responsables del contenido del preprint y que el depósito en SciELO Preprints no significa ningún compromiso por parte de SciELO, excepto su preservación y difusión.
- Los autores declaran que los datos, las aplicaciones y otros contenidos subyacentes al manuscrito están referenciados.
- El manuscrito depositado está en formato PDF.
- Los autores declaran que la investigación que dio origen al manuscrito siguió buenas prácticas éticas y que las aprobaciones necesarias de los comités de ética de investigación, cuando corresponda, se describen en el manuscrito.
- Los autores declaran que una vez que un manuscrito es postado en el servidor SciELO Preprints, sólo puede ser retirado mediante solicitud a la Secretaría Editorial deSciELO Preprints, que publicará un aviso de retracción en su lugar.
- Los autores aceptan que el manuscrito aprobado esté disponible bajo licencia [Creative Commons CC-BY](#).
- El autor que presenta el manuscrito declara que las contribuciones de todos los autores y la declaración de conflicto de intereses se incluyen explícitamente y en secciones específicas del manuscrito.
- Los autores declaran que el manuscrito no fue depositado y/o previamente puesto a disposición en otro servidor de preprints o publicado en una revista.
- Si el manuscrito está siendo evaluado o siendo preparando para su publicación pero aún no ha sido publicado por una revista, los autores declaran que han recibido autorización de la revista para hacer este depósito.
- El autor que envía el manuscrito declara que todos los autores del mismo están de acuerdo con el envío a SciELO Preprints.