



Validación de contenido y confiabilidad del instrumento: ética e inteligencia artificial en estudiantes en salud

Content validity and reliability of an instrument on ethics and Artificial Intelligence in health sciences students

✉ **Francisco-Raúl Hernández-Burrola**
Universidad Autónoma de Chihuahua
México
<https://orcid.org/0009-0009-7482-7661>
rhernandez@uach.mx

Haydeé Parra-Acosta
Universidad Autónoma de Chihuahua
México
<https://orcid.org/0000-0003-1720-7182>
hparra@uach.mx

Recibido: 23-12-2025
Aceptado: 16-02-2026
Publicado: 09-03-2026

Cómo citar este artículo:

Hernández-Burrola, F. R., & Parra-Acosta, H. (2026). Validación de contenido y confiabilidad del instrumento: ética e inteligencia artificial en estudiantes en salud. *Revista Panamericana de Pedagogía*, 42, e3695. <https://doi.org/10.21555/rpp.3695>

Resumen

El uso creciente de la inteligencia artificial (IA) en la formación de estudiantes de ciencias de la salud plantea desafíos éticos relacionados con integridad académica, privacidad, sesgos algorítmicos y responsabilidad profesional. Se identificó la ausencia de instrumentos validados para evaluar estos factores en contextos universitarios. El objetivo fue diseñar y validar un instrumento para valorar los factores éticos vinculados al uso académico de la IA en estudiantes de ciencias de la salud. Se efectuó un estudio instrumental que incluyó el diseño inicial del instrumento, valoración de contenido mediante juicio de 13 expertos utilizando la Escala de Jueces Expertos (Centro Universitario CIFE, 2019), el cálculo del coeficiente V de Aiken, una prueba piloto aplicada a 32 estudiantes y la aplicación del instrumento a una muestra de 313 estudiantes. Los ítems mostraron una validez de contenido adecuada ($V \geq 0.75$) y se obtuvieron coeficientes de consistencia interna elevados ($\alpha = 0.886$; $\Omega = 0.89$). Los estudiantes reportaron claridad en las instrucciones, pertinencia

Este trabajo está bajo una licencia Creative Commons Attribution 4.0 International (CC BY 4.0).



de los contenidos y alta aceptabilidad del instrumento. Este instrumento constituye un aporte relevante para la evaluación pedagógica del uso ético de la inteligencia artificial en la educación superior en salud. En conclusión, presenta validez de contenido sólida y confiabilidad adecuada, lo que permite su uso en investigaciones posteriores orientadas a analizar los factores éticos en el uso académico de la inteligencia artificial en la educación en salud.

Palabras clave: Educación superior; Estudiantes de ciencias de la salud; Ética; Evaluación educativa; Inteligencia artificial.

Abstract

The increasing integration of artificial intelligence (AI) in health sciences education poses ethical challenges related to academic integrity, data privacy, algorithmic bias, and professional responsibility. Despite this growing relevance, there is a lack of validated instruments that comprehensively assess ethical factors associated with the academic use of AI among university students in health-related fields. The aim of this study was to design and validate an instrument to evaluate ethical factors linked to the academic use of artificial intelligence in health sciences students. An instrumental study was conducted that included the design of the instrument, content validation through expert judgment by 13 specialists using the Expert Judges Scale (Centro Universitario CIFE, 2019), calculation of Aiken's V coefficient, a pilot test applied to 32 students, and the final application of the instrument to a sample of 313 health sciences students. The items demonstrated adequate content validity ($V \geq 0.75$) and high internal consistency, with a Cronbach's alpha of 0.886 and a McDonald's Omega of 0.89. Students reported clear instructions, relevant content, and high acceptability of the instrument. In conclusion, the instrument shows solid content validity and adequate reliability, supporting its use in future research aimed at analyzing ethical factors related to the academic use of artificial intelligence in health sciences education.

Keywords: Artificial intelligence; Educational evaluation; Ethics; Health sciences students; Higher education.

INTRODUCCIÓN

La inteligencia artificial (IA) se define como sistemas computacionales capaces de procesar datos y tomar decisiones emulando facultades humanas como el razonamiento, el aprendizaje y la comunicación (Benítez et al., 2018; Unesco, 2022). Su desarrollo acelerado ha transformado múltiples sectores de la sociedad, incluyendo la educación y las ciencias de la salud. La ética aplicada a la IA busca alinear estos sistemas con principios y valores que prioricen el bienestar humano, la equidad y el respeto a los derechos fundamentales (IEEE, 2019), lo que plantea nuevos retos para la formación académica.

En el ámbito educativo, los estudiantes de ciencias de la salud recurren cada vez más a herramientas de IA para resolver dudas, elaborar trabajos, resumir artículos o generar ideas académicas (Sánchez-Mendiola, 2025; Rodríguez-Weber et al., 2024). Sin embargo, este uso no siempre se acompaña de lineamientos claros sobre honestidad, responsabilidad y transparencia, lo que abre debates en torno a la autoría, el plagio y la integridad académica (Avello-Sáez et al., 2024). Así mismo, emergen preocupaciones relacionadas con la privacidad de los datos (Melo, 2024), los sesgos algorítmicos y la equidad en el acceso a tecnologías de calidad (Vélez-Rivera et al., 2024).

De ahí la necesidad de contar con marcos institucionales que regulen el empleo de la IA y fortalezcan la autonomía, la responsabilidad y el juicio profesional (Rodríguez-Weber et al., 2024). En este sentido, resulta fundamental evaluar cómo los estudiantes perciben y utilizan la ética en la IA en su vida académica, y de qué manera esta interacción repercute en la calidad de sus aprendizajes y en su futura práctica profesional. La formación de profesionales de la salud implica no solo adquirir competencias técnicas, sino también consolidar un compromiso ético y crítico frente al uso de tecnologías emergentes.

A este respecto existen diversos instrumentos para medir actitudes hacia la tecnología y la ética en la educación como el de Chauca-Saavedra et al. (2025), que se enfoca en analizar la percepción del impacto de la IA en estudiantes de ciencias de la salud, pero no en la relación directa entre ética e inteligencia artificial. Al-Qererm et al. (2023) analizan las percepciones éticas de los estudiantes y su conocimiento sobre implicaciones éticas de la IA en medicina, pero no evalúa directamente el desempeño académico en relación con la ética en la IA. La propuesta de Holmes et al. (2023) tiene un enfoque ético y reflexivo, no cuantitativo ni de rendimiento; sin embargo, sí explora cómo los aspectos éticos de la IA pueden influir en la experiencia y formación del estudiante de diversas disciplinas universitarias vinculadas a la educación a distancia, incluyendo docencia, administración y gestión institucional. Ahmad et al. (2023) exploran la relación entre la ética, la IA y el comportamiento educativo, pero no mide resultados académicos (rendimiento o desempeño), es decir no muestra cómo la dependencia tecnológica puede deteriorar el pensamiento crítico y la autonomía moral del estudiante. Por su parte el aporte de Sallam et al. (2023) sí relaciona la ética con la adopción tecnológica, muestra que la percepción del riesgo ético afecta la disposición de los estudiantes a usar ChatGPT y, por ende, puede influir en su proceso de aprendizaje y desempeño futuro, pero no mide rendimiento académico. De lo anterior se puede observar que, pocos son los aportes instrumentales que se enfocan específicamente en estudiantes de ciencias de la salud y en la relación directa entre ética en inteligencia artificial y el desempeño académico, por lo que ante este vacío de información, el objetivo del presente estudio fue diseñar un instrumento que permita analizar de manera integral los dilemas éticos asociados al uso académico de la IA y generar evidencia empírica que constituya una base de lineamientos institucionales orientados a una integración crítica, responsable y formativa de estas tecnologías en la educación superior en salud.

METODOLOGÍA

Tipo de estudio

Se realizó un estudio de tipo instrumental, de acuerdo con Montero y León (2005), que se orienta al desarrollo, adaptación y validación de instrumentos, así como al análisis de sus propiedades psicométricas. Este tipo de estudio se caracteriza por centrarse en el diseño metodológico de instrumentos que permiten medir constructos de manera válida y confiable, mediante procesos sistemáticos de elaboración, juicio de expertos, análisis de validez de contenido y pruebas de confiabilidad (Juárez-Hernández y Tobón, 2018). Específicamente en el presente se realizó un estudio instrumental orientado al diseño y validación de un instrumento para evaluar los factores éticos relacionados con el uso de la inteligencia artificial (IA) en la formación académica de estudiantes de ciencias de la salud.

Diseño del instrumento

El constructo central fue la ética en el uso de la inteligencia artificial en la educación en salud, entendida como el conjunto de principios, valores y conductas que guían el uso responsable, honesto y crítico de sistemas inteligentes en los procesos formativos, priorizando el bienestar humano, la transparencia, la privacidad y la integridad académica (UNESCO, 2022; IEEE, 2019; Sánchez-Mendiola, 2025).

El instrumento que se diseñó se denominó “*Ética en inteligencia artificial en la formación de estudiantes de salud*”, y fue diseñado de acuerdo con los referentes conceptuales mencionados. Integra diez dimensiones: frecuencia y acceso a la IA; tipo de herramientas de IA utilizadas; aplicabilidad académica de la IA; percepción sobre tu rendimiento académico; percepción sobre tu cumplimiento y responsabilidad; percepción sobre la calidad del trabajo; integridad académica; privacidad y datos; sesgos y equidad; y por último la responsabilidad profesional. Estas dimensiones fueron derivadas de los estudios de Robles-Zeas et al. (2024), Rodríguez-Weber et al. (2024) y Avello-Sáez et al. (2024), así como de las directrices éticas de la UNESCO (2022).

El instrumento se conformó por 51 ítems (tabla 1) distribuidos en diez dimensiones, además de una sección de datos sociodemográficos (sexo, edad, semestre, promedio, etc.). Se empleó una escala de respuesta tipo Likert de cinco puntos para los primeros 50 ítems y 1 ítem de selección múltiple con casillas de verificación. La escala tipo Likert esta graduada de la siguiente manera: 1 = Nunca / Deficiente / Nada; 2 = Casi nunca / Poco satisfecho / Muy poco; 3 = Algunas veces / Bueno / Poco; 4 = Frecuentemente / Muy bueno / Mucho y 5 = Siempre / Excelente / Totalmente.

El instrumento adopta una aproximación integral para el análisis de la ética en el uso de la inteligencia artificial, incorporando dimensiones relacionadas con las prácticas de uso, las percepciones académicas y los principios éticos que orientan dichas prácticas. Desde esta perspectiva, la ética se concibe como un eje transversal que permite interpretar el uso responsable de la inteligencia artificial en la formación de estudiantes de ciencias de la salud.

Tabla 1
Estructura del instrumento

Dimensión	Descripción	N. Ítems	Ítems	Tipo de respuesta
Frecuencia y acceso	Evalúa la frecuencia con que los estudiantes utilizan herramientas de IA, los dispositivos desde los que acceden y el grado de integración en sus métodos de estudio.	5	1. Utilizas herramientas de IA para tus actividades académicas semanalmente? 2. Accedes a plataformas de IA desde tu computadora personal para asuntos académicos? 3. Usas IA en tu celular para resolver dudas académicas rápidas? 4. Recurras a la IA como primera fuente de consulta antes que a la biblioteca digital, revistas electrónicas y buscadores como Google? 5. Integras constantemente la IA en tus métodos de estudio?	Escala Likert 1-5
Tipo de herramientas de IA utilizadas	Identifica las herramientas más empleadas (chatbots, generadores de texto, IA para imágenes o redacción) y su finalidad académica.	5	6. Empleas chatbots (como ChatGPT, Gemini, DeepSeek) para obtener explicaciones sobre temas complejos? 7. Utilizas generadores de texto (como ChatGPT, Gemini, DeepSeek) para ayudar a estructurar tus ensayos y reportes? 8. Has interactuado con tutores virtuales (como ChatGPT, Gemini, DeepSeek) basados en IA para prepararte para un examen? 9. Usas IA para generar imágenes o diagramas (como ChatGPT, Gemini, Canva) para tus presentaciones académicas? 10. Utilizas software de IA (como ChatGPT, Gemini, DeepSeek) para parafrasear o mejorar la redacción de tus trabajos?	Escala Likert 1-5
Aplicabilidad académica de la IA	Analiza las tareas específicas en las que los estudiantes aplican la IA: búsqueda de información, traducción, redacción o generación de ideas.	6	11. Usas la IA para generar ideas iniciales para un proyecto de investigación? 12. Empleas la IA para resumir artículos científicos? 13. Utilizas herramientas de IA para resolver problemas o ejercicios prácticos de tus asignaturas? 14. Te apoyas en la IA para traducir textos académicos de otros idiomas? 15. Usas la IA para crear borradores de trabajos de investigación formales dirigidos a tus profesores? 16. Usas la IA para buscar referencias válidas para tus trabajos de investigación?	Escala Likert 1-5
Percepción sobre tu rendimiento cuantitativo	Explora cómo los estudiantes perciben el impacto de la IA en su rendimiento académico medible (calificaciones y aprobaciones).	4	17. Tus calificaciones han mejorado desde que utilizas herramientas de IA? 18. La calificación que obtienes en tus ensayos y trabajos escritos es generalmente alta (9 o 10)? 19. La IA te ayuda a obtener mejores resultados en los exámenes de opción múltiple? 20. Has logrado aprobar asignaturas que considerabas difíciles gracias al apoyo de la IA?	Escala Likert 1-5
Percepción sobre tu cumplimiento y responsabilidad	Evalúa hábitos de estudio, organización, puntualidad y compromiso con el aprendizaje académico.	5	21. Entregas tus tareas y proyectos académicos consistentemente antes de la fecha límite? 22. Tu asistencia y participación en clases son regulares y activas? 23. Organizas tu tiempo de estudio de manera eficaz para cumplir con todas sus obligaciones? 24. Revisas cuidadosamente las rúbricas y los requisitos de evaluación de cada materia? 25. Asumes la responsabilidad de tu propio proceso de aprendizaje?	Escala Likert 1-5

Dimensión	Descripción	N. Ítems	Ítems	Tipo de respuesta
Percepción sobre la calidad de tu trabajo	Examina la percepción sobre la originalidad, claridad y profundidad de los productos académicos elaborados con apoyo de IA.	5	<p>26. Te esfuerzas por asegurar que tus trabajos presenten un alto grado de originalidad?</p> <p>27. Tus argumentos en los trabajos escritos están bien fundamentados?</p> <p>28. Compruebas que tus escritos sean comprensibles, coherentes y fáciles de seguir antes de presentarlos?</p> <p>29. La realimentación de tus profesores suele destacar la calidad y el pensamiento crítico de tus trabajos?</p> <p>30. Tus trabajos reflejan una comprensión integral de los temas abordados, más allá de un resumen de fuentes de información?</p>	Escala Likert 1-5
Integridad académica	Mide la conciencia sobre honestidad, autoría y responsabilidad ética en el uso de IA dentro del ámbito académico.	5	<p>31. Presentar un texto generado por IA como si fuera propio es una forma de deshonestidad académica?</p> <p>32. Es éticamente incorrecto usar IA para completar un examen?</p> <p>33. Los estudiantes deben citar o declarar siempre que una IA ha sido utilizada como herramienta de apoyo en un trabajo?</p> <p>34. El uso de IA para generar ideas no vulnera la integridad académica, siempre que el desarrollo sea propio?</p> <p>35. Confiar ciegamente en la información de la IA sin verificarla es académicamente irresponsable?</p>	Escala Likert 1-5
Privacidad y datos	Indaga sobre la preocupación y cuidado respecto al manejo de información personal por parte de plataformas de IA.	5	<p>36. Te preocupas por la cantidad de información personal que las plataformas de IA recopilan?</p> <p>37. Lees los términos y condiciones sobre el uso de tus datos antes de utilizar una nueva herramienta de IA?</p> <p>38. Evitas introducir información sensible o personal en los chats de IA?</p> <p>39. Crees que las universidades deberían regular qué herramientas de IA se pueden usar para proteger los datos de los estudiantes?</p> <p>40. Consideras que el uso de herramientas de IA justifica ceder parte de tu privacidad?</p>	Escala Likert 1-5
Sesgos y equidad	Examina la percepción de los estudiantes sobre los sesgos algorítmicos y la equidad en el acceso a herramientas de IA.	5	<p>41. Estás consciente de que las respuestas de la IA pueden reflejar sesgos culturales, de género, etc.?</p> <p>42. Has notado que la información proporcionada por una IA a veces es inexacta o está desactualizada?</p> <p>43. Piensas que confiar demasiado en la inteligencia artificial puede disminuir la capacidad del estudiante para explorar diferentes formas de pensar o resolver problemas?</p> <p>44. Consideras que es importante cuestionar y verificar críticamente la información que una IA proporciona?</p> <p>45. Crees que el acceso desigual a herramientas de IA de calidad puede generar inequidad entre los estudiantes?</p>	Escala Likert 1-5
Responsabilidad profesional	Analiza cómo los estudiantes asocian el uso ético de la IA con su formación y futuro desempeño profesional en salud.	5	<p>46. El uso excesivo de IA en tu formación podría disminuir tu capacidad de razonamiento crítico?</p> <p>47. Es una responsabilidad ética como futuro profesional de la salud dominar los fundamentos teóricos sin depender de la IA?</p> <p>48. La IA debe ser una herramienta para aumentar la capacidad humana y no para reemplazar el juicio profesional?</p> <p>49. La dependencia de la IA afectará negativamente la relación profesional de la salud-paciente en el futuro?</p> <p>50. Como estudiante de ciencias de la salud, tienes la responsabilidad de usar la IA de manera que beneficie el aprendizaje?</p>	Escala Likert 1-5

Dimensión	Descripción	N. Ítems	Ítems	Tipo de respuesta
Tipo de herramientas de IA	Identifica las herramientas de IA más utilizadas por los estudiantes.	1	Selecciona las herramientas de IA que has usado.	Opción múltiple con casillas de verificación

Nota. El instrumento se compone de 51 ítems distribuidos en 10 dimensiones. Los ítems se responden mediante una escala tipo Likert de cinco puntos (1 = nunca / deficiente / nada; 5 = siempre / excelente / totalmente), excepto un ítem de selección múltiple con casillas de verificación.

Análisis de validez de contenido

El proceso de validación de contenido contó con la participación de trece jueces expertos (tabla 2) seleccionados con base en los criterios propuestos por Juárez-Hernández y Tobón (2018) y el modelo de validación del Centro Universitario CIFE (2019).

Tabla 2
Jueces expertos

Descripción general de los jueces	Resultados
Sexo	7 hombres, 6 mujeres
Grado académico predominante	Mayormente Doctorado o Maestría
Experiencia profesional (años)	De 5 a más de 30 años
Actividad docente	100% realiza docencia
Actividad investigadora	100% realiza investigación
Experiencia como investigador (años)	1 a 29 años
Artículos publicados (rango estimado)	2 a más de 40 artículos
Media	24.46
Desviación estándar	21.71
Libros publicados	0 a 6 libros
Media	2.31
Desviación estándar	3.02
Capítulos de libros (rango)	0 a 16 capítulos
Media	4.77
Desviación estándar	4.94
Ponencias presentadas (rango estimado)	2 a más de 30 ponencias
Media	11.46
Desviación estándar	10.23
Experiencia en validación de instrumentos	100% tiene experiencia
Años de experiencia en validación (rango)	1 a más de 10 años
Media	4.40
Desviación estándar	3.34

Nota. Los datos corresponden a trece jueces expertos con experiencia en docencia, investigación y validación de instrumentos. Los valores se presentan como frecuencias, rangos, medias y desviaciones estándar según la variable.

El proceso se desarrolló con un enfoque cuali-cuantitativo, utilizando la Escala de Validación de Jueces Expertos (Centro Universitario CIFE, 2019), la cual permite evaluar: pertinencia del ítem (grado en que refleja el constructo) y la redacción (claridad y coherencia semántica).

Análisis estadístico

Se aplicó el coeficiente V de Aiken (Penfield y Giacobbi, 2004) para estimar la validez de contenido con intervalos de confianza del 95%. Se consideraron aceptables los ítems con valores $V \geq 0.75$, tal como establecen Bulger y Housner (2007), y $V_i \geq 0.50$ para el límite inferior. Para lo anterior se utilizó la plantilla de Juárez-Hernández (2018).

Prueba piloto

Después del proceso de validación de contenido con jueces expertos y una vez incorporadas las modificaciones sugeridas, se aplicó una prueba piloto con 32 estudiantes de ciencias de la salud (tabla 3). Esta etapa tuvo como propósito evaluar la comprensión semántica de los ítems, la claridad de las instrucciones, así como la aceptabilidad y satisfacción del instrumento utilizando la escala propuesta por el Centro Universitario CIFE (2018).

Tabla 3

Prueba Piloto

Variable	Descripción
Total de alumnos	32
Género	Femenino: 20 (62.5%) Masculino: 12 (37.5%)
Edad	Media: 19.91 años DE: 1.96 Rango: 18–27 años
Semestre	Primero: 18 Segundo: 10 Tercero: 3 Cuarto: 1
Carrera	Ingeniería Biomédica: 17 Médico Cirujano y Partero: 15
Promedio general	Media: 8.65 DE: 0.68 Rango: 7.0–9.7
Materias no aprobadas	0: 21 estudiantes 1: 5 estudiantes 2: 5 estudiantes 3: 1 estudiante

Nota. DE = desviación estándar. Los datos se presentan como frecuencias absolutas y porcentajes o como medidas descriptivas según corresponda.

Análisis de confiabilidad

Asimismo, una vez concluida la prueba piloto, se examinó el desempeño del cuestionario en términos de confiabilidad mediante el coeficiente Omega de McDonald y el coeficiente alfa de Cronbach. El tamaño de muestra empleado es de 313 estudiantes de ciencias de la salud, que se consideró adecuado conforme a lo indicado por Carpenter (2018), quien sugiere que entre 5 y 100 participantes son suficientes para esta fase dentro de los estudios instrumentales. Estos objetivos permitieron valorar la calidad práctica y psicométrica del instrumento.

Tabla 4

Muestra

Variable	Cantidad	%
Género femenino	206	65.8
Género masculino	101	32.3
Prefiero no responder	5	1.6
No binario	1	0.3
Médico Cirujano y Partero	148	47.3
Ingeniería Biomédica	74	23.6
Terapia Física	69	22.0
Salud Pública	22	7.0

Nota. Edad media = 20.2 años (DE = 2.32). Cantidad media de materias no aprobadas = 1.27 (DE = 3.91).

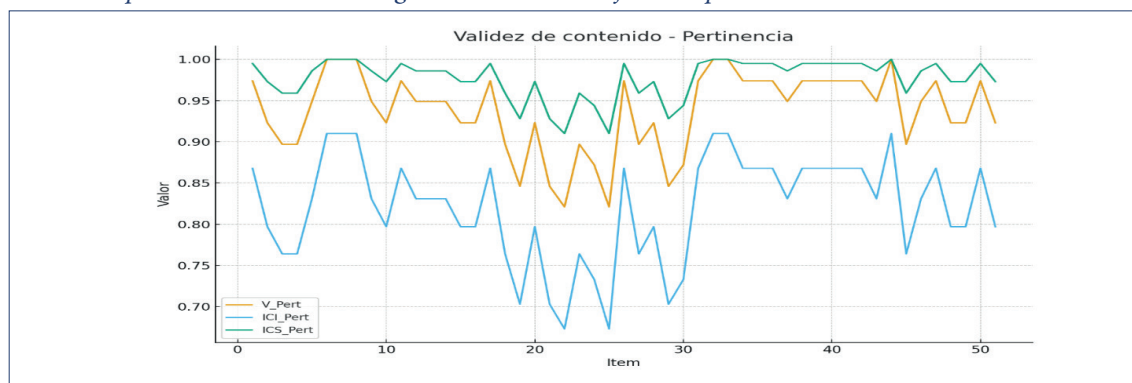
RESULTADOS

Cada juez calificó ambos criterios (pertinencia y redacción) mediante una escala de cuatro niveles (nulo, bajo, aceptable y alto). Las observaciones cualitativas fueron consideradas para reformular o eliminar ítems ambiguos o redundantes.

Los resultados mostraron valores globales de V de Aiken > 0.75 para todos los ítems lo que indica una adecuada correspondencia entre los ítems y las dimensiones teóricas del constructo. En la figura 1 se presentan los valores correspondientes a la pertinencia de los ítems, donde el trazo asociado al coeficiente V de Aiken, representa la estimación puntual del grado de acuerdo entre jueces. El intervalo de confianza inferior (ICI), indica el valor mínimo esperado de dicha validez bajo un criterio conservador, mientras que el intervalo de confianza superior (ICS) muestra el límite máximo de estimación. La cercanía entre el V observado y sus intervalos de confianza refleja estabilidad y precisión en la valoración de los jueces respecto a la adecuación conceptual de los ítems.

Figura 1

Criterio de pertinencia de los ítems según la evaluación de jueces expertos

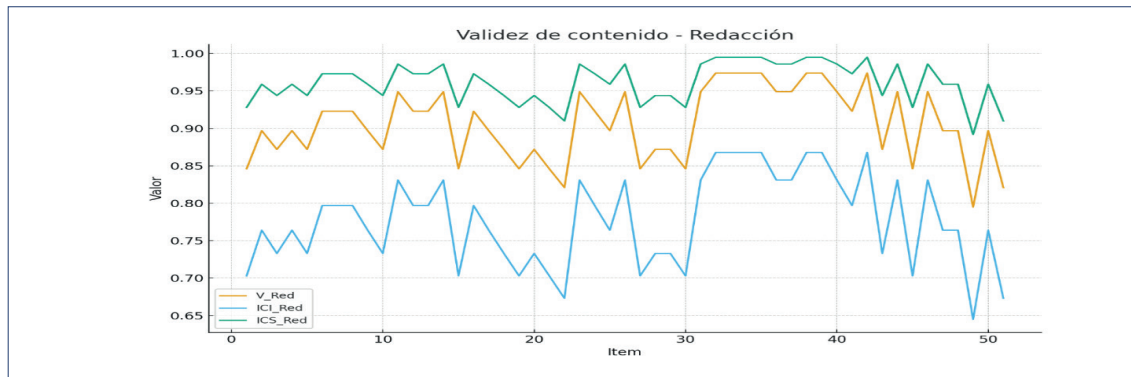


Nota. Los valores corresponden al coeficiente V de Aiken (color naranja), calculado con intervalos de confianza del 95 %, para evaluar la pertinencia de los ítems respecto al constructo teórico. El ICI está identificado por el color azul y el ICS por el color verde.

Por su parte, la figura 2 muestra estos mismos coeficientes aplicados al criterio de redacción. En esta gráfica se observa un patrón similar, con valores elevados de V y rangos estrechos entre el ICI y el ICS, lo que sugiere que los ítems presentan claridad, coherencia lingüística y adecuada formulación. La consistencia visual entre ambas figuras permite apreciar que tanto el contenido como la redacción del instrumento cumplen con criterios psicométricos aceptables para su uso en estudios posteriores.

Figura 2

Criterio de redacción de los ítems según la evaluación de jueces expertos



Nota. Los valores corresponden al coeficiente V de Aiken (color naranja), calculado con intervalos de confianza del 95 %, para evaluar la claridad y coherencia semántica de los ítems. El ICI está identificado por el color azul y el ICS por el color verde.

En cuanto al pilotaje, la evaluación de satisfacción fue aplicada a los 32 estudiantes participantes en la prueba piloto. En relación con la comprensión de las instrucciones, el 46.9 % reportó un grado excelente, el 37.5 % un grado bueno y el 15.6 % un grado aceptable. En cuanto a la comprensión de los ítems, el 53.1 % indicó un grado excelente, el 34.4 % un grado bueno y el 12.5 % un grado aceptable. La satisfacción general con el instrumento presentó una distribución favorable, con un 50 % en la categoría excelente y un 40.6 % en la categoría buena. Asimismo, la relevancia percibida de las preguntas se calificó como excelente en el 53.1 % y buena en el 37.5 % de los casos. Finalmente, la mayoría de los participantes (78.1 %) no reportó sugerencias de mejora, y los comentarios cualitativos señalaron que el instrumento es claro, pertinente y fácil de contestar. Los estudiantes reportaron una alta aceptabilidad del instrumento; no obstante, también realizaron algunas sugerencias de mejora. Entre ellas destacaron la necesidad de clarificar ciertos términos relacionados con la inteligencia artificial y ajustar la redacción para evitar ambigüedades. Estas observaciones fueron consideradas para fortalecer la claridad semántica del instrumento en versiones posteriores.

Una vez concluidas las etapas anteriores y aplicados los ajustes derivados de las mismas, el instrumento fue administrado en su versión final a una muestra de 313 estudiantes de ciencias de la salud, con el propósito de estimar su confiabilidad y consistencia interna en condiciones reales de aplicación. Los coeficientes obtenidos reflejaron valores adecuados. El alfa de Cronbach global fue $\alpha = 0.886$, dentro del rango considerado óptimo para estudios instrumentales (≥ 0.80). Por su parte, el Omega de McDonald fue $\Omega = 0.89$ con-

firmando la solidez de la consistencia interna del instrumento y respaldando la estabilidad de la medición del constructo evaluado.

Tabla 5

Tabla resumida del alfa de Cronbach

Característica	Valor
Número total de ítems	50
Tamaño de la muestra (n)	313
Coefficiente alfa de Cronbach (α)	0.886
Rango de correlaciones ítem-total corregidas	≥ 0.30
Incremento del alfa al eliminar ítems	No se observó incremento sustantivo

Nota. La tabla 5 presenta un resumen del análisis de confiabilidad interna del instrumento, calculado mediante el coeficiente alfa de Cronbach a partir de los 50 ítems que lo integran. El valor obtenido ($\alpha = 0.886$) indica un nivel alto de consistencia interna, de acuerdo con los criterios comúnmente aceptados en estudios instrumentales, lo que sugiere que los reactivos presentan una adecuada homogeneidad en la medición del constructo.

Asimismo, las correlaciones ítem-total corregidas mostraron valores aceptables en la mayoría de los reactivos, lo que evidencia una contribución positiva de los ítems a la puntuación total del instrumento. El análisis del alfa de Cronbach si el ítem se elimina no evidenció incrementos sustantivos en el coeficiente global, por lo que se decidió conservar la totalidad de los ítems, atendiendo tanto a criterios estadísticos como a su relevancia teórica en la evaluación ética del uso de la inteligencia artificial en la formación de estudiantes de salud.

En conjunto, estos resultados respaldan la confiabilidad del instrumento en esta fase inicial de validación, sin que ello implique aún una validación psicométrica completa, la cual requerirá análisis adicionales en estudios posteriores.

DISCUSIÓN

El presente estudio tuvo como propósito desarrollar y validar un instrumento para evaluar los factores éticos asociados al uso académico de la inteligencia artificial (IA) en estudiantes de ciencias de la salud, el cual es un tema que adquiere especial relevancia ante el uso creciente de tecnologías generativas en los procesos formativos. La ética en la IA se ha convertido en un eje prioritario debido a los riesgos emergentes relacionados con privacidad, sesgos algorítmicos, integridad académica y responsabilidad profesional, particularmente en áreas donde las decisiones requieren juicio clínico y fundamentos bioéticos sólidos. Aunque existen estudios que abordan percepciones sobre IA en contextos educativos o clínicos, la revisión de literatura muestra una ausencia de instrumentos específicos que integren simultáneamente aspectos éticos, académicos y formativos, lo cual limita las posibilidades de evaluación sistemática en programas de salud. Por ello, la creación de un instrumento propio se justifica plenamente como una respuesta a este vacío conceptual y metodológico.

Los instrumentos previos identificados en la literatura tienden a centrarse en competencias digitales, alfabetización en IA o creencias sobre su uso, pero no profundizan

en sus implicaciones éticas, ni contemplan dimensiones como integridad académica, responsabilidad profesional o sesgos algorítmicos. Esta falta de abordaje integral constituye una limitación importante, previamente señalada por Avello-Sáez et al. (2024) y Holmes et al. (2023), quienes destacan que la IA no puede evaluarse únicamente desde la eficacia o funcionalidad, sino desde la capacidad del estudiante para utilizarla de manera crítica, segura y ética. En este contexto, el instrumento propuesto aporta un enfoque novedoso al integrar dimensiones cognitivas, conductuales y éticas, ofreciendo una herramienta sólida para el análisis educativo.

Respecto a la validez de contenido, los resultados muestran valores de V de Aiken superiores a 0.75 en todos los ítems, con intervalos de confianza aceptables y límites inferiores por encima de 0.50. Esto indica que los jueces consideraron los reactivos pertinentes, coherentes y alineados con los constructos teóricos. Además, la selección de jueces expertos, con experiencia en docencia, investigación, diseño y validación de instrumentos, fortalece la calidad del proceso de validación, destacándose su carácter mixto (evaluación cualitativa y cuantitativa), como recomiendan Juárez-Hernández y Tobón (2018).

En relación con la prueba piloto, los estudiantes manifestaron altos niveles de comprensión y satisfacción con las instrucciones, los ítems y la relevancia de las preguntas. La ausencia de valoraciones negativas y la predominancia de respuestas en las categorías “excelente” y “bueno” indican una adecuada aceptabilidad del instrumento. Estos hallazgos concuerdan con lo planteado por Carvajal et al. (2011), quienes enfatizan que una prueba piloto exitosa constituye evidencia importante de validez aparente y usabilidad. Los comentarios cualitativos refuerzan esta percepción positiva al señalar claridad, pertinencia y facilidad en el proceso de respuesta.

Asimismo, los coeficientes de confiabilidad obtenidos mediante el alfa de Cronbach ($\alpha = 0.886$) y el Omega de McDonald ($\Omega = 0.89$) evidencian una consistencia interna elevada del instrumento. En términos psicométricos, valores iguales o superiores a 0.80 suelen considerarse adecuados para escalas en desarrollo, mientras que coeficientes por encima de 0.90 reflejan una consistencia interna excelente (Nunnally y Bernstein, 1995; George y Mallery, 2019). En el caso del Omega, su valor cercano a 0.90 confirma la coherencia entre los ítems bajo supuestos menos restrictivos que los del alfa, por lo que se reconoce como un estimador más robusto de la confiabilidad compuesta (McDonald, 2013; Zinbarg et al., 2005). Estos resultados respaldan el uso del instrumento en investigaciones posteriores y lo ubican dentro de los rangos recomendados para estudios instrumentales.

LIMITACIONES DEL ESTUDIO

Es importante señalar que algunos ítems del instrumento, particularmente aquellos relacionados con el compromiso ético y la integridad académica, podrían estar influidos por la deseabilidad social o la tendencia a la aquiescencia. No obstante, su inclusión resulta pertinente para explorar la conciencia ética declarada de los estudiantes frente al uso de la inteligencia artificial. Futuras investigaciones podrán fortalecer el instrumento mediante la incorporación de ítems invertidos o escalas específicas de control de deseabilidad social, así como mediante aplicaciones en contextos diversos.

Si bien los resultados obtenidos permiten afirmar que el instrumento presenta una adecuada validez de contenido y niveles satisfactorios de confiabilidad, para fortalecer la exploración psicométrica del instrumento, será necesario incorporar en estudios posteriores análisis de validez de constructo, tales como análisis factorial exploratorio y confirmatorio, así como procedimientos correlacionales que permitan evaluar con mayor profundidad la estructura interna y el comportamiento del constructo evaluado.

CONCLUSIONES

El instrumento desarrollado demuestra validez de contenido sólida, alta aceptabilidad por parte de los estudiantes y confiabilidad consistente, lo que lo posiciona como una herramienta adecuada para evaluar los factores éticos del uso de IA en la formación en salud. Sus aportes se alinean con las recomendaciones internacionales de organismos como la Unesco (2022) y responden a la necesidad de contar con evaluaciones integrales que permitan orientar políticas educativas, estrategias formativas y prácticas académicas responsables en un entorno donde la IA continuará creciendo.

Desde una perspectiva pedagógica, este instrumento permite a las instituciones de educación superior y a los docentes contar con una herramienta sistemática para identificar prácticas, percepciones y dilemas éticos asociados al uso académico de la inteligencia artificial. Su aplicación facilita la toma de decisiones formativas orientadas al desarrollo del pensamiento crítico, la integridad académica y la responsabilidad profesional en la formación de estudiantes de ciencias de la salud, contribuyendo así a una integración educativa más reflexiva, ética y contextualizada de estas tecnologías emergentes.

CONSIDERACIONES ÉTICAS

El estudio cumplió con los principios éticos establecidos por el Código de Núremberg (Comisión Nacional de Bioética, 1947), la Declaración de Helsinki (Conbioética, 1964), el Informe Belmont (1979) y la Ley Federal de Protección de Datos Personales en Posesión de los Particulares (Cámara de Diputados, 2025). Todos los participantes, jueces expertos y estudiantes, recibieron un consentimiento informado elaborado específicamente para este proyecto, en el cual se explicó el propósito del estudio, la naturaleza voluntaria de su participación, los posibles riesgos y beneficios, y las medidas de confidencialidad.

El presente estudio se condujo conforme a los principios de integridad de la investigación científica establecidos en la Declaración de Singapur (2010) sobre la integridad en la investigación, asegurando la honestidad intelectual, la transparencia en el diseño metodológico, la rigurosidad en el análisis de los datos y la responsabilidad ética en la comunicación de los resultados. Los autores declaran que la información presentada es original, veraz y obtenida mediante procedimientos metodológicos éticamente aceptables, sin incurrir en fabricación, falsificación o manipulación indebida de los datos.

FINANCIACIÓN Y ORIGEN DEL ESTUDIO

El presente estudio no contó con financiamiento externo y fue desarrollado como parte de un proyecto académico de investigación en el ámbito de la educación superior en ciencias de la salud.

CONFLICTO DE INTERESES

Los autores declaran que no existieron intereses financieros, institucionales o comerciales que pudieran influir en el diseño, desarrollo, análisis o interpretación de los resultados.

DISPONIBILIDAD DE DATOS

Los conjuntos de datos generados y analizados durante el presente estudio no se encuentran disponibles públicamente debido a consideraciones éticas relacionadas con la confidencialidad y protección de la información de los participantes. No obstante, los datos podrán ser solicitados de manera razonable al autor de correspondencia, previa evaluación ética y con fines exclusivamente académicos.

CONTRIBUCIÓN DE LOS AUTORES

Francisco-Raúl Hernández-Burrola: Administración del proyecto; Análisis formal; Conceptualización; Curación de datos; Escritura - borrador original; Escritura - revisión y edición; Investigación; Metodología; Recursos; Software; Validación; Visualización; Adquisición de fondos.

Haydeé Parra-Acosta: Análisis formal; Conceptualización; Escritura - revisión y edición; Metodología; Recursos; Supervisión; Validación; Visualización; Adquisición de fondos.

REFERENCIAS

- Ahmad, S. F., Han, H., Alam, M. M., Rehmat, M. K., Irshad, M., Arraño-Muñoz, M., & Ariza-Montes, A. (2023). Impact of artificial intelligence on human loss in decision making, laziness and safety in education. *Humanities and Social Sciences Communications*, 10(1), Artículo 311. <https://doi.org/10.1057/s41599-023-01787-8>
- Al-Qerem, W., Eberhardt, J., Jarab, A., Al Bawab, A. Q., Hammad, A., Alasmari, F., Alazab, B., Husein, D. A., Alazab, J., & Al-Beool, S. (2023). Exploring knowledge, attitudes, and practices towards artificial intelligence among health professions' students in Jordan. *BMC Medical Informatics and Decision Making*, 23(1), Artículo 288. <https://doi.org/10.1186/s12911-023-02403-0>

- Avello-Sáez, D., Lucero-González, N., & Villagrán, I. (2024). Desarrollo de una declaración de uso de inteligencia artificial con una perspectiva de integridad académica en educación médica y ciencias de la salud. *Revista Médica Clínica Las Condes*, 35(5–6), 412–420. <https://doi.org/10.1016/j.rmclc.2024.06.003>
- Benítez, R., Cencerrado, A., Gerard, B., & Kanaan, E. S. (2018). *Inteligencia artificial avanzada*. Editorial UOC. <https://openaccess.uoc.edu/server/api/core/bitstreams/509e0233-eefd-44e0-acf7-58582fa65f7a/content>
- Bulger, S. M., & Housner, L. D. (2007). Modified Delphi investigation of exercise science in physical education teacher education. *Journal of Teaching in Physical Education*, 26(1), 57–80. <https://doi.org/10.1123/jtpe.26.1.57>
- Cámara de Diputados del H. Congreso de la Unión. (2025, 20 de marzo). *Ley Federal de protección de datos personales en posesión de los particulares*. Secretaría de Servicios Parlamentarios. <https://www.diputados.gob.mx/LeyesBiblio/pdf/LFPDPPP.pdf>
- Carpenter, S. (2018). Ten steps in scale development and reporting: A guide for researchers. *communication methods and measures*, 12(1), 25–44. <https://doi.org/10.1080/19312458.2017.1396583>
- Carvajal, A., Centeno, C., Watson, R., Martínez, M., & Sanz Rubiales, Á. (2011). ¿Cómo validar un instrumento de medida de la salud? *Anales del Sistema Sanitario de Navarra*, 34(1), 63–72. <https://recyt.fecyt.es/index.php/ASSN/article/view/10317>
- Centro Universitario CIFE. (2018). *Cuestionario de satisfacción con el instrumento (plantilla)*. Google Forms. <https://docs.google.com/forms/d/e/1FAIpQLSc8-jOiWYwG64QbnhRyGAgl1ELTggq2aP1XiSg45pyN9XLbXNQ/viewform>
- Centro Universitario CIFE. (2019). *Escala Jueces Expertos (plantilla)*. <https://bit.ly/4c2yZ8t>
- Chauca Saavedra, C. L., Arones Mayuri, M. E., Quispe Nombresas, V. C., & Olivera Machado, S. H. (2025). Percepción del impacto del aprendizaje de la Inteligencia Artificial en la formación de estudiantes universitarios de la salud. *Revista Información Científica*. <https://doi.org/10.5281/zenodo.15392016>
- Comisión Nacional de Bioética. (1947). *Código de Nuremberg*. https://www.conbioetica-mexico.salud.gob.mx/descargas/pdf/normatividad/normatinternacional/2.INTL._Cod_Nuremberg.pdf
- Conbioética. (1964). *Declaración de Helsinki de la AMM*. https://conbioetica-mexico.salud.gob.mx/descargas/pdf/Declaracion_Helsinki_Brasil.pdf
- Declaración de Singapur sobre la integridad en la investigación*. (2010). <https://www.conicyt.cl/fondap/files/2014/12/DECLARACION%20C3%93N-SINGAPUR.pdf>
- George, D., & Mallery, P. (2019). *IBM SPSS Statistics 26 step by step: A simple guide and reference*. Routledge. <https://doi.org/10.4324/9780429056765>
- Holmes, W., Iniesto, F., Anastopoulou, S., & Boticario, J. G. (2023). Stakeholder perspectives on the ethics of AI in distance-based higher education. *The International Review of Research in Open and Distributed Learning*, 24(2), 96–117. <https://doi.org/10.19173/irrodl.v24i2.6089>
- IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. (2019). *Ethically aligned design: A vision for prioritizing human well-being with artificial intelligence and autonomous systems*. IEEE.

- Informe Belmont: Principios y guías éticos para la protección de los sujetos humanos de investigación.* (1979). Comisión Nacional para la Protección de los Sujetos Humanos de Investigación Biomédica y del Comportamiento. https://www.conbioetica-mexico.salud.gob.mx/descargas/pdf/normatividad/normatinternacional/10._INTL_Informe_Belmont.pdf
- Juárez-Hernández, L. G. (2018). *Plantilla para el cálculo de la V de Aiken e intervalos de confianza.*
- Juárez-Hernández, L. G., & Tobón, S. (2018). Análisis de los elementos implícitos en la validación de contenido de un instrumento de investigación. *Revista Espacios*, 39(53), 23–30. <https://www.revistaespacios.com/cited2017/cited2017-23.pdf>
- McDonald, R. P. (2013). *Test theory: A unified treatment.* Psychology Press. <https://doi.org/10.4324/9781410601087>
- Melo, Z. (2024). Inteligencia artificial en salud: desafíos éticos para lograr la aplicación de las tecnologías a la salud del paciente. *Trayectorias Humanas Trascontinentales*, (18). <https://doi.org/10.25965/trahs.6349>
- Montero, I., y León, O. G. (2005). Sistema de clasificación del método en los informes de investigación en psicología. *International Journal of Clinical and Health Psychology*, 5(1), 115–127. <https://www.redalyc.org/pdf/337/33701007.pdf>
- Nunnally, J., y Bernstein, I. (1995). *Psychometric theory* (3.ª ed.). McGraw-Hill.
- Penfield, R. D., & Giacobbi, P. R., Jr. (2004). Applying a score confidence interval to Aiken's item content-relevance index. *Measurement in Physical Education and Exercise Science*, 8(4), 213–225. https://doi.org/10.1207/s15327841mpee0804_3
- Robles Zeas, R. M., Serrano Paredes, K. L., & Cruz Gavilanes, T. M. (2024). Uso de inteligencia artificial como soporte para el aprendizaje en las ciencias de la salud. *Imaginario Social*, 7(1). <http://revista-imaginariosocial.com/index.php/es/index>
- Rodríguez Weber, F. L., Portela Ortiz, J. M., & Enríquez Barajas, A. (2024). La inteligencia artificial (IA) en la medicina y su aprendizaje. *Acta Médica Grupo Ángeles*, 22(3), 261–263. <https://doi.org/10.35366/117512>
- Sallam, M., Salim, N. A., Barakat, M., Al-Mahzoum, K., Al-Tammemi, A. B., Malaeb, D., Hallit, R., & Hallit, S. (2023). Assessing health students' attitudes and usage of ChatGPT in Jordan: Validation study. *JMIR Medical Education*, 9(1), Artículo e48254. <https://doi.org/10.2196/48254>
- Sánchez Mendiola, M. (2025). ¿Autonomía en riesgo? Ética y la dependencia de la inteligencia artificial generativa en la formación médica. *Investigación en Educación Médica*, 14(53), 5–7. <https://doi.org/10.22201/fm.20075057e.2025.53.24670>
- UNESCO. (2022). *Recomendación sobre la ética de la inteligencia artificial.* https://unesdoc.unesco.org/ark:/48223/pf0000381137_spa
- Vélez-Rivera, R., Muñoz-Álvarez, D., Leal-Orellana, P., & Ruiz-Garrido, A. (2024). Uso de Inteligencia Artificial en educación superior y sus implicancias éticas: Mapeo sistemático de literatura. *Hachetetepe: Revista Científica de Educación y Comunicación*, (28), Artículo 1105. <https://doi.org/10.25267/Hachetetepe.2024.i28.1105>
- Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α , Revelle's β , and McDonald's ω_H : Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika*, 70(1), 123–133. <https://doi.org/10.1007/s11336-003-0974-7>