

VALIDACIÓN DE UN INSTRUMENTO PARA MEDIR EL USO DE INTELIGENCIA ARTIFICIAL EN EL PROCESO DE ENSEÑANZA DE DOCENTES DE EDUCACIÓN BÁSICA

Validation of an Instrument to Measure the
use of Artificial Intelligence in the Teaching
Process of Basic Education Teachers

Alejandro García Mejía
Gerzon Yair Calle Álvarez
Universidad de Antioquia, Colombia

ALEJANDRO GARCÍA MEJÍA

MAESTRANDO EN EDUCACIÓN. ESTUDIANTE, UNIVERSIDAD DE ANTIOQUIA.
AGARCIA.MEJIA@UDEA.EDU.CO. [HTTPS://ORCID.ORG/0009-0000-2645-990X](https://orcid.org/0009-0000-2645-990X)

GERZON YAIR CALLE ÁLVAREZ

DOCTOR EN EDUCACIÓN. DOCENTE, UNIVERSIDAD DE ANTIOQUIA.
GERZON.CALLE@UDEA.EDU.CO. [HTTPS://ORCID.ORG/0000-0002-4083-6051](https://orcid.org/0000-0002-4083-6051)

RESUMEN

Objetivo: Validar por juicio de expertos un instrumento para medir el uso de inteligencia artificial en el proceso de enseñanza de docentes de educación básica.

Materiales y métodos: Estudio cuantitativo psicométrico de validez de contenido mediante juicio de expertos. Se consultó a 5 expertos con doctorado y experiencia en educación básica y universitaria. El grado de acuerdo de los jueces se dio a partir del cálculo V de Aiken como criterio de decisión para el instrumento 1 (Cuestionario de Avances y Limitaciones de la IA en los Procesos de Enseñanza), tomando como punto de referencia un valor $\leq 0,7$. Así mismo, para el instrumento 2 (Escala de Usabilidad de IA en los Procesos de Enseñanza - EUIAP-E), se empleó Kappa de Fleiss como indicador de fiabilidad para 5 niveles de uso en una escala Likert de 25 ítems.

Resultados: Los indicadores de claridad ($V = 0,94$), coherencia ($V = 0,89$) y relevancia ($0,94$), representados a través un coeficiente V de Aiken general de $0,92$, muestran que el instrumento presenta evidencia de validez sólida. De igual manera, el coeficiente de Kappa de Fleiss ($k = 0,417$) indica una fuerza de concordancia moderada entre los evaluadores, lo que sugiere un acuerdo superior al que se esperaría por azar.

Conclusiones: Los hallazgos encontrados muestran que tanto el Cuestionario de Avances y Limitaciones de IA en los Procesos de Enseñanza como la EUIAP-E constituyen herramientas válidas para la medición de los usos de la inteligencia artificial en los procesos de enseñanza. La labor docente, más que desaparecer o perdurar, debe preguntarse sobre los cómo, cuándo y por qué enseñar, lo cual requiere adentrarse en actualizaciones profundas con la llegada de nuevas tecnologías. Los instrumentos validados en este artículo permitirán analizar avances y posibilidades de mejora que permitirán integrar la IA en las dinámicas educativas desde una postura ética y sostenible.

PALABRAS CLAVE: docente, educación básica, enseñanza, inteligencia artificial, proceso de enseñanza, validación de contenido.

ABSTRACT

Research realities, characterized by a constant flow of information, are immersed in an ethical dilemma regarding reliable and rigorous processes in knowledge management. In this context, hyperstimulation mediated by Information and Communication Technologies, and especially AI, represents a phenomenon that cannot be compared to other technologies, as its impact has not yet been fully measured. Consequently, this reality has created a need in academia for reliable instruments that provide

theoretical and methodological support for research as a generator of knowledge for humanity.

Objective: *To validate, through expert judgment, an instrument for measuring the use of artificial intelligence in the teaching process of basic education teachers.*

Materials and methods: *The study employed a quantitative approach, using statistical tests of concordance for content validity. The psychometric study was based on content validity through expert judgment, involving five experts with doctorates in education and experience in basic and university education. Specific criteria were considered for their selection: being practicing teachers, having educational experience in basic education, using and integrating emerging technologies in teaching processes, and possessing extensive knowledge in the field of information and communication technologies. The degree of agreement among the judges was determined based on Aiken's V calculation as a decision criterion for instrument 1 (Questionnaire on Advances and Limitations of AI in Teaching Processes), which presented values ranging from 0.00 (no agreement) to 1.00 (total agreement), taking as a decision threshold a value ≤ 0.7 in the three indicators: clarity, coherence, and relevance. Regarding the scoring, the ratings were presented on a scale of 1 to 5, where 1 corresponded to "poor", 2 to "fair", 3 to "good", 4 to "very good", and 5 to "excellent". In parallel, instrument 2 (AI Usability Scale in Teaching Processes - EUIAP-E) used Fleiss's Kappa as a reliability indicator for five levels of AI use in teaching, all on a 25-item Likert scale—namely: Level 1: Resistant; Level 2: Exploratory; Level 3: Operational; Level 4: Integrative; Level 5: Critical. The Likert scale consisted of 25 items that were rated according to: Always (A), Almost always (AA), Sometimes (S), Almost never (AN), and Never (N). The strength of agreement for Fleiss's Kappa coefficient ranged from 0.00 to 1.0, being: poor (0.00), slight (0.1-0.20), acceptable (0.21-0.40), moderate (0.41-0.60), considerable (0.61-0.80), and almost perfect (0.81-1.0).*

Results: *The indicators of clarity ($V = 0.94$), consistency ($V = 0.89$), and relevance (0.94), represented by Aiken's overall V coefficient of 0.92, show that the AI Progress and Limitations in Teaching Processes Questionnaire is a solidly valid instrument. In detail, this instrument investigates the use of AI in the dimensions of teaching, learning, assessment, classroom management, educational risks, supports, obstacles, and teacher perception, all of which are inherent aspects of the teaching process. Meanwhile, the results for instrument 2 show that Fleiss's Kappa coefficient ($k = 0.417$) indicates a moderate strength of agreement among evaluators, suggesting an agreement greater than would be expected by mere chance. This test represents greater reliability than other statistical verifications such as mode, which is useful in this type of study.*

The results for the AI Usability Scale in Teaching Processes (EULAP-E) are considered statistically significant, as demonstrated by the Z-test value of 13.106 and a p-value of 0.000, which is below the standard significance threshold ($\alpha = 0.05$). The 95% asymptotic confidence interval for Kappa (0.354-0.480) indicates that the actual agreement in the population of judges is highly likely to fall within this range.

Conclusions: *The findings show that both instrument 1 (Questionnaire on Advances and Limitations of AI in Teaching Processes) and instrument 2 (EULAP-E) are valid tools for measuring the uses of artificial intelligence in teaching processes developed by basic education teachers. As such, these instruments represent a relevant alternative for strengthening and improving educational contexts.*

In this regard, the integration of artificial intelligence in education requires constant support to ensure safe and reliable processes with a cross-cutting approach. In an era characterized by major transformations, the role of teachers must adapt to social dynamics, which implies new forms of teaching, training, and professional development. Tool validation is therefore an act that allows us to view these instruments as valuable tools that guarantee processes of continuous improvement—in this case related to the use of AI in teaching, which is a fundamental part of driving innovation and modernization mechanisms in the education sector.

Likewise, information-gathering instruments enable informed discussion. Currently, the interaction between AI and education is giving rise to a series of ethical, moral, academic, social, and research debates. While the overwhelming rise of emerging technologies in many areas represents certain threats, it also offers valuable opportunities that are important to support through reliable instruments.

Based on the foregoing, it is clear that this is a constantly evolving issue. However, it is important to consider the postulates regarding its use in the field of education, especially given the repercussions of AI on the academic and institutional life of society. While it is true that teaching will be transformed in the coming years, far from disappearing or remaining stagnant, teachers must ask themselves how, when, and why to teach, which, although it has been a commendable task, must be updated with the arrival of various technologies. Ultimately, tools such as those presented in this article will allow for the analysis of advances and possibilities for improvement that will help integrate AI into educational dynamics from an ethical and sustainable standpoint.

KEYWORDS: *teacher, basic education, teaching, artificial intelligence, teaching process, content validation.*

INTRODUCCIÓN

Las mediciones en el campo educativo se apoyan de instrumentos no solo confiables, sino también válidos. Medir no es simplemente contar; en el ámbito educativo, medir implica interpretar, comprender y otorgar significado. La sociedad no necesita cifras vacías, sino herramientas que permitan leer con profundidad los procesos educativos. Por ello, la confiabilidad -aunque necesaria- no es suficiente. Validar es más que un atributo técnico, es una exigencia ética: implica preguntarse si realmente se mide lo que importa, lo que transforma. Reconocer la complejidad del ser humano obliga a evaluar constantemente los resultados, los métodos, las intenciones y los sentidos (Unesco, 2025). En este marco, la labor docente se convierte en un reto, pues enseñar no es transmitir, sino generar condiciones para el desarrollo integral, y ello genera, desde un punto de vista metodológico, un asunto de trascendencia significativa.

Es precisamente este desafío metodológico de validar el que se hace necesario debido a las influencias de la tecnología y en especial de la inteligencia artificial (IA, en adelante) en el proceso de enseñanza, pues es una labor crucial en la sociedad del conocimiento. Desde 1956, con la celebración de la Conferencia de Darmouth, ha existido un gran interés por el abordaje de la IA en el campo educativo (Santillán de la Torre et al., 2024; Douse, 2024; Su & Zhong, 2022; Gallent et al., 2024; Núñez et al., 2023), y especialmente desde el abordaje de la IA en el proceso de enseñanza. La personalización de estrategias metodológicas, la evaluación, la planeación, el currículo y la formación son algunas de las tendencias predominantes en la literatura científica (Douse, 2024; Chiu, 2022; Hu & Liao, 2020; Siamon et al., 2024; Bellas et al., 2023; Aparicio, 2023); esto ha permitido que la relación entre IA y docente sea un tema vigente en la educación investigativa.

De esta manera, se ha abierto un abanico de discusiones en el campo del conocimiento, pues la enseñanza apoyada por la IA da cuenta de un relato multifacético de experiencias que revelan un proceso de diseño y desarrollo. Gran parte de esta nueva realidad

se debe a un entorno complejo de influencias externas y desafíos locales (Dai et al., 2023). No obstante, la necesidad de un enfoque cauteloso se hace pertinente a fin de abordar los desafíos de implementación y transparencia que se dan en torno al uso de la IA en la educación (Santillán de la Torre et al., 2024). En otras palabras, mantener el sentido humano, formativo y social de la educación.

Respecto a esto, es fundamental el manejo ético. En sus estudios, Alam (2023) menciona que la AI se ha vuelto cada vez más avanzada y adoptada en la sociedad; sin embargo, es importante considerar la seguridad, ecuanimidad, la privacidad de datos, la interpretabilidad, interacción humano-IA e implicaciones éticas de estas tecnologías, pues esto puede repercutir en la planificación y diseño de los procesos de enseñanza. Pese a ello, no deja de ser una herramienta que diversifica el aprendizaje, proporciona asistencia, brinda automatización de tareas y analiza datos para mejorar la predicción del rendimiento estudiantil (Berrones et al., 2023). La incorporación didáctica y ética de la IA en los procesos de enseñanza puede ser una oportunidad para alcanzar mayores niveles de acceso de la educación para la población en general.

Este auge en la tecnología, y especialmente en la inteligencia artificial, trae consigo retos en la manera de enseñar y en la forma como los docentes se capacitan y adquieren las habilidades necesarias; por eso, Van Laar et al. (2017) mencionan que en el siglo XXI las competencias digitales son fundamentales para participar en la fuerza de trabajo basada en la sociedad del conocimiento. En ese sentido, el pensamiento crítico, la gestión de la información, la creatividad, la comunicación, el trabajo en equipo y la resolución de problemas son algunas competencias digitales que marcarán la diferencia en el proceso educativo del futuro (Harari, 2024). Estas competencias, como lo señala Selwin (2019), propiciarán que el rol del docente se diversifique y adopte una postura de orientación frente al uso de la IA, en lugar de asumir posturas negacionistas que prohíben su implementación en la academia y frenan el avance educativo al ritmo de la tecnología.

De igual modo, el papel de la IA en la docencia dentro del proceso de educación básica ha permitido avances en esta materia (Casillas et al., 2024), especialmente en el surgimiento de plataformas de aprendizaje, estrategias didácticas, gestión efectiva del aula y el análisis de datos impulsados por la IA. Sin embargo, como plantea Martínez (2025), los docentes necesitan competencias digitales y pedagógicas específicas para usar la IA en los procesos de enseñanza, ya que no solo es dominar las herramientas, sino comprender y aplicar su potencial educativo.

De acuerdo con Gómez et al. (2025), es necesario validar instrumentos que permitan integrar la inteligencia artificial en la profesión docente. Algunos ejemplos son el AI-ED-SAT, utilizado para medir las competencias docentes en IA educativa (Sartor & Azevedo, 2026), y el TAICS, desarrollado por Chiu et al. (2025) como una escala de autoeficiencia de competencias en IA para docentes.

Pese a que la literatura menciona progresos en materia de enseñanza a través del uso de la inteligencia artificial, el abordaje de estas tensiones ha creado, basados en el estado del arte, la necesidad de validar un instrumento que permita medir los usos de la inteligencia artificial por parte de docentes de educación básica. Lo anterior como respuesta a los desafíos educativos del siglo XXI, que centran la atención en lograr el desarrollo de capacidades que permitan a los docentes integrarla en términos de transparencia, equidad y pertinencia. Lo mencionado posibilita visibilizarla como una herramienta y coagente en el aprendizaje y no como una amenaza al conocimiento y la misma sociedad. Así mismo, se han encontrado pocos estudios que aborden los usos de la IA por parte de los docentes de educación básica, siendo estos los encargados de formar las bases de jóvenes y adolescentes en los primeros años de vida. De igual manera, la literatura consultada no menciona los usos de la IA en la enseñanza; por lo cual, el abordaje del uso resistente, explorador, operativo, integrador y crítico funge como una oportunidad de actualización educativa. Por consiguiente, este estudio tiene como objetivo validar por juicio de expertos un instrumento que permita

medir el uso de la IA en el proceso de enseñanza de docentes de educación básica.

MATERIALES Y MÉTODOS

Este artículo empleó un enfoque cuantitativo, utilizando para ello métodos estadísticos de validez de contenido. Inicialmente, se utilizó la prueba V de Aiken para determinar la validez de contenido de las preguntas del instrumento 1 (*Cuestionario de Avances y Limitaciones de la IA en los Procesos de Enseñanza*), fundamentado en tres indicadores clave: *claridad, coherencia y relevancia*. Para el instrumento 2 (UEIAP-E) se utilizó la prueba Kappa de Fleiss, la cual analiza el grado de concordancia entre tres o más jueces expertos. Es decir, establecer si sus evaluaciones convergen o, por el contrario, presentan discrepancias.

Diseño de estudio psicométrico

La validez de contenido mediante juicio de expertos se realizó entre julio y agosto de 2025. Según Martínez (2020), el juicio de expertos se establece a partir de una opinión informada de personas con amplia experiencia en el campo de estudio, quienes con su conocimiento especializado están capacitados para emitir valoraciones, calificaciones y evidencias fundamentadas. Dado que Almanasreh (2019) recomienda entre 5 y 7 expertos para evaluar la validez de contenidos, el estudio seleccionó 5 expertos con doctorados en educación y áreas afines.

Muestra

El estudio estuvo integrado por 5 expertos con doctorado. Para su selección se consideraron criterios específicos: ser docentes en ejercicio, experiencia educativa en educación básica, utilización e integración de tecnologías emergentes en los procesos de enseñanza y poseer amplios conocimientos en el campo de las Tecnologías de la Información y la Comunicación (TIC). Estos criterios aseguraron que los expertos tuvieran las competencias teóricas y prácticas

necesarias para valorar rigurosamente los instrumentos. La tabla 1 resume los perfiles de los expertos.

Tabla 1. Perfil de expertos

Experto	Perfil	Experiencia
E1	Doctora en Literatura	Universitaria y Educación Básica
E2	Doctora en Educación	Universitaria y TIC
E3	Doctor en Educación	Educación Básica
E4	Doctor en Educación	Universitaria y TIC
E5	Doctora en Pedagogía	Educación Básica

Fuente: elaboración propia.

Instrumentos

Los instrumentos fueron elaborados como resultado de la investigación titulada *La inteligencia artificial en el proceso de enseñanza de docentes de educación básica en instituciones educativas del departamento de Antioquia*, compuesta de dos partes:

Instrumento 1. *Cuestionario de Avances y Limitaciones de la IA en los Procesos de Enseñanza*

Este instrumento se conformó mediante un cuestionario de 8 preguntas y múltiples opciones de respuesta que indagan sobre los avances y las limitaciones de la IA en los contextos educativos, especialmente desde los procesos de enseñanza; para ello, se tomó en cuenta diversas temáticas que se relacionan con la dinámica de la enseñanza por parte de los docentes (véase la siguiente figura). Cada pregunta tiene un enunciado y múltiples opciones de respuesta.



Fuente: elaboración propia.

Figura. Dimensiones instrumento 1

La anterior figura refleja los componentes del proceso de enseñanza de los docentes, cada uno de ellos responde de manera integral a diversas características propias del ejercicio (véase tabla 2).

Tabla 2. Dimensiones del proceso de enseñanza

Dimensión	Definición
Percepción docente	Creencias, actitudes, valoraciones que se dan en la práctica educativa, formas, modos y acciones que se emplean para enseñar.
Gestión del aula	Estrategias utilizadas para organizar el espacio (tiempo, recursos, normas).
Obstáculos educativos	Barreras o dificultades que enfrenta el proceso de enseñanza.
Riesgos educativos	Factores que comprometen la calidad educativa.
Apoyos educativos	Procesos que facilitan el proceso de enseñanza (tecnologías, materiales, capacitaciones, etc.).
Aprendizaje del estudiante	Proceso cognitivo, habilidades, competencias que se desarrollan en el estudiante.
Evaluación	Herramientas y métodos para medir, valorar o retroalimentar el aprendizaje.
Enseñanza docente	Estrategias metodológicas y didácticas, planeación, ejecución y reflexión de la labor docente.

Fuente: elaboración propia.

Instrumento 2. Escala de Usabilidad de IA en los Procesos de Enseñanza (EUIAP-E)

Este instrumento incluyó el EUIAP-E, de elaboración propia. Su construcción se fundamentó principalmente en una revisión de la literatura científica sobre herramientas que midieran el nivel de uso de la inteligencia artificial por parte de los docentes en sus procesos de enseñanza. Dado que no se encontró evidencia científica significativa que abordara estos aspectos, se identificó la necesidad de desarrollar un instrumento acorde con los contextos educativos y tecnológicos actuales.

Producto del análisis, se seleccionaron ocho categorías fundamentadas en el proceso de enseñanza y clasificadas en 5 niveles de uso. El diseño de las escalas Likert se basó en los postulados de Matas (2018), quien brinda recomendaciones para la construcción académica de estos instrumentos.

En consecuencia, el instrumento 2 se conformó por 25 ítems que miden los 5 niveles de uso de IA en los procesos de enseñanza (resistente, explorador, operativo, integrador y crítico). La redacción de los ítems tomó en cuenta criterios de claridad y precisión y uso sencillo del lenguaje, evitando sesgos y confusión para el lector (véase tabla 3).

Tabla 3. Niveles de uso en IA en los procesos de enseñanza

Nivel	Uso	Característica
Nivel 1	Uso resistente	Postura de rechazo o no utilización de la IA en los procesos de enseñanza. Percibe la IA como una amenaza a la autonomía docente.
Nivel 2	Uso explorador	Interés inicial por conocer sobre las implicaciones de la IA en los procesos de enseñanza, Valora la formación en IA para introducir paulatinamente nuevas tecnologías. Genera espacios de diálogo y discusión en torno a su aplicabilidad en el aula.
Nivel 3	Uso operativo	Aplicación básica de la IA para generar rápidamente contenidos, estrategias y recursos. Centra las tecnologías emergentes como una herramienta de apoyo que facilita la labor docente.

Continúa...

Nivel 4	Uso integrador	Emplea la IA en los procesos de enseñanza de acuerdo con las necesidades específicas. Explora y evalúa los contenidos. Transforma enfoques tradicionales y articula la tecnología con enfoque innovador.
Nivel 5	Uso crítico	Reflexión profunda y crítica de las implicaciones éticas, sociales y educativa de la IA. Evalúa sesgos, pertinencia y viabilidad de las herramientas. Se anticipa y considera aspectos claves de la IA en la enseñanza.

Fuente: elaboración propia.

El instrumento 2 se estructuró en 25 ítems, con 5 opciones de respuestas; el procedimiento consistió en marcar con una X, siendo Siempre (S), Casi siempre (CS), Algunas veces (AV), Casi nunca (CN) y Nunca (N).

Análisis estadístico

La validación fue aplicada utilizando Microsoft Excel para tabular la información y SPSS Statitics versión 27 para el análisis estadístico. El instrumento 1 incluyó el análisis de tres indicadores: claridad (sintaxis y semántica de cada ítem), coherencia (relación del ítem con lo que se mide) y relevancia (importante para medir el indicador). La valoración de los expertos osciló entre 1 y 5, siendo 1 Deficiente; 2 Regular; 3 Buena; 4 Muy buena y 5 Excelente. Se empleó V de Aiken para determinar el grado de acuerdo de los jueces; dicha prueba estadística es soportada en estudios de Penfield y Giacobbi (2007) y Merino y Livia (2009).

Este coeficiente presenta valores que oscilan entre 0,00 y 1,00, siendo 1,00 el valor máximo de acuerdo entre los expertos. Es decir, si $V = 0$, significa que hay un total desacuerdo en los ítems, mientras que si $V = 1,00$, significa que hay un total acuerdo en los ítems. El cálculo de la V de Aiken toma en cuenta las siguientes consideraciones, los cuales representan el grado de acuerdo entre distintos jueces.

$$V = \frac{(S-1)}{[n(c-1)]}$$

A partir de la siguiente fórmula:

- n = número de expertos que participan de la validación.
- c = número de niveles de la escala de valoración empleada.
- s = suma de valoraciones de todos los expertos por ítem o pregunta.

Como parámetro de decisión, para conservar el instrumento, se tomó 0,7 como valor mínimo del intervalo de confianza y 1,00 como límite máximo; un coeficiente menor da como resultado la anulación de la validación.

Asimismo, en el instrumento 2 se utilizó el Kappa de Fleiss (López y Pita, 1999), prueba estadística dedicada a determinar el grado de concordancia o acuerdo entre tres o más evaluadores. La fórmula es la siguiente:

$$K = \frac{Po - Pe}{[1 - Pe]}$$

De acuerdo con lo anterior, Po es el acuerdo observado de los calificadores, mientras que Pe es el acuerdo esperado de los calificadores. La interpretación de Kappa de Fleiss se expresa en un intervalo de 0 a 1,0, y su interpretación se sustenta en los planteamientos de Landis y Koch (1997) (véase tabla 4).

Tabla 4. Coeficiencia Kappa de Fleiss

Coeficiente Kappa de Fleiss	Fuerza de Concordancia
0,00	Pobre
0,1 - 0,20	Leve (ligeramente)
0,21 - 0,40	Aceptable (regular)
0,41 - 0,60	Moderada
0,61 - 0,80	Considerable (Sustancial)
0,81 - 1,0	Casi perfecta (casi perfecto)

Fuente: Tomado de Landis y Koch (1997).

RESULTADOS

El instrumento 1 estuvo compuesto por 8 preguntas, con múltiples opciones de respuesta, cada pregunta indagó sobre una dimensión del proceso de enseñanza. Teniendo claro esto, cada una de las ocho preguntas fue sometida a los indicadores de claridad, coherencia y relevancia. En la tabla 5 se describen las preguntas y las dimensiones que las conforman.

Tabla 5. Dimensiones y preguntas instrumento 1

Dimensiones	Preguntas del instrumento
Enseñanza docente	¿De qué manera ha utilizado la inteligencia artificial (IA) en los procesos de enseñanza en la educación básica?
Evaluación	¿De qué forma ha incorporado la inteligencia artificial (IA) en los procesos de evaluación del aprendizaje de sus estudiantes?
Gestión del aula	¿En qué aspectos ha considerado que la inteligencia artificial (IA) puede ser útil para mejorar la gestión del aula en su práctica docente?
Enseñanza-aprendizaje	¿Cómo ha incorporado la inteligencia artificial (IA) en el proceso de enseñanza-aprendizaje de los estudiantes de la educación básica?
Riesgos educativos	¿Cuáles son los riesgos de la IA que identificas en el proceso de enseñanza del docente?
Apoyos educativos	¿Qué tipo de apoyo ha considerado necesario para poder incorporar la IA en su labor docente?
Obstáculos educativos	¿Cuáles han sido los principales obstáculos para implementar IA en sus procesos enseñanza?
Percepción docente	¿Qué percepción personal le ha generado la implementación de la IA para la enseñanza en el escenario educativo?

Fuente: elaboración propia.

La evaluación global del instrumento 1 mostró que el coeficiente V de Aiken en los indicadores de *claridad*, *coherencia* y *relevancia* alcanzó un valor de 0,92 (véase tabla 6), es decir, las preguntas están redactadas de forma clara y comprensible, existe una

coherencia entre las preguntas y las dimensiones que miden, y los contenidos son relevantes y pertinentes para evaluar el constructo de interés.

Tabla 6. V de Aiken instrumento 1

V de Aiken		
Claridad	Coherencia	Relevancia
0,94	0,89	0,94
V de Aiken: 0,92		

Fuente: elaboración propia.

Este coeficiente expresa una ponderación de 0,92, lo que significa un alto acuerdo en el juicio de expertos. Las tablas 7, 8 y 9 muestran los resultados individuales de *claridad*, *coherencia* y *relevancia* en el instrumento 1, cuantificados mediante V de Aiken; como se observa, la gran mayoría de las evaluaciones contaron con valoraciones favorables.

Tabla 7. V de Aiken en el indicador claridad

Experto evaluador	Claridad							
	P1	P2	P3	P4	P5	P6	P7	P8
E1	4	4	4	4	3	4	3	4
E2	4	4	4	4	4	4	4	4
E3	4	4	4	4	4	4	4	4
E4	4	4	4	4	2	4	4	4
E5	3	4	3	2	3	4	4	3
Total	19	20	19	18	16	20	19	19
V de Aiken	0,95	1,00	0,95	0,90	0,80	1,00	0,95	0,95

Fuente: elaboración propia.

Los resultados para el indicador de *claridad* muestran aspectos favorables en la mayoría de las preguntas, destacando P2 y P6. El

coeficiente da cuenta de una redacción precisa y clara. Sin embargo, P5 obtuvo el puntaje más bajo (0,80), lo que muestra discrepancia en los evaluadores. Este resultado sugiere la necesidad de reformulación o ajuste gramatical, el cual fue ajustado en la tabla 5.

Tabla 8. V de Aiken en el indicador coherencia

Experto evaluador	Coherencia							
	P1	P2	P3	P4	P5	P6	P7	P8
E1	4	4	4	4	4	4	4	4
E2	3	4	3	3	3	4	4	4
E3	4	4	4	4	4	4	3	4
E4	4	4	2	2	2	4	4	4
E5	3	3	2	2	4	4	4	3
Total	18	19	15	15	17	20	19	19
V de Aiken	0,90	0,95	0,75	0,75	0,85	1,00	0,95	0,95

Fuente: elaboración propia.

El indicador de *coherencia* refleja que cinco preguntas alcanzaron valores superiores a 0,90. Sin embargo, P3 y P4 obtuvieron coeficientes de 0,75, por debajo del umbral aceptable, en ese sentido, fue necesario revisar su alineación teórica para fortalecer su coherencia conceptual. En la tabla 9 se observan los valores más altos para el indicador de *relevancia*, lo que indica que las ocho preguntas evidencian una valoración altamente favorable.

Tabla 9. V de Aiken en el indicador relevancia

Experto evaluador	Relevancia							
	P1	P2	P3	P4	P5	P6	P7	P8
E1	4	4	4	4	4	4	4	4
E2	4	4	2	4	4	4	4	4
E3	4	4	4	4	4	4	3	4
E4	4	4	4	4	4	4	4	4

Continúa...

E5	3	3	2	2	4	4	4	3
Total	19	19	16	18	20	20	19	19
V de Aiken	0,95	0,95	0,80	0,90	1,00	1,00	0,95	0,95

Fuente: elaboración propia.

Considerando el parámetro de 0,7 como valor mínimo del intervalo de confianza, cabe destacar que el instrumento 1 obtuvo un coeficiente V de Aiken superior al valor crítico, por lo cual se confirma la validación del instrumento para su posterior aplicación.

Por otro lado, con respecto a la validación del instrumento 2, se tomó en cuenta el Kappa de Fleiss (López y Pita, 1999). Inicialmente, se procedió a la tabulación de las valoraciones asignadas por cada experto a los 25 ítems de la escala Likert. Es importante destacar que cada evaluador emitió su juicio de valor en función de su experiencia y dominio del área temática. Concretamente, los 5 expertos consultados debieron analizar cada ítem de la escala y asignarle un nivel de uso, tal como se muestra a continuación.

- 1 = Nivel 1. Uso Resistente
- 2 = Nivel 2. Uso Explorador
- 3 = Nivel 3. Uso Operativo
- 4 = Nivel 4. Uso Integrador
- 5 = Nivel 5. Uso Crítico

Las valoraciones asignadas por cada experto (véase tabla 10) representa un nivel de uso específico para cada ítem. Si bien una medida descriptiva como la moda podría identificar el nivel de mayor frecuencia de acuerdo, esta aproximación es limitada para establecer la concordancia entre jueces. Por ello, la aplicación de la prueba Kappa de Fleiss resulta útil para este tipo de instrumentos, ya que es una herramienta robusta que garantiza que las escalas Likert sean confiables y permitan una evaluación consistente y precisa.

**Tabla 10. Tabulación respuestas instrumento
2 por juicio de expertos**

Ítem	Indicador	E1	E2	E3	E4	E5
1	Reflexiono sobre los sesgos algorítmicos presentes en las herramientas de IA aplicadas a la educación.	5	5	5	5	2
2	Consulto la IA en busca de respuestas rápidas dentro del ejercicio profesional.	3	3	4	3	3
3	Opto por mantener prácticas de enseñanza convencionales que experimentar con tecnologías emergentes en IA.	1	1	1	1	2
4	Verifico la precisión de los contenidos educativos sugeridos por la IA antes de aplicarlos en el aula.	4	4	4	4	4
5	Recurro a la IA para generar soluciones didácticas que acontecen en el aula de clase.	3	3	3	4	3
6	Examino a profundidad los riesgos e implicaciones éticas de la IA en la educación.	5	5	5	5	2
7	Considero que la IA representa una amenaza para la autonomía docente y el mejoramiento de la enseñanza en educación.	1	1	1	1	3
8	Participo de espacios de formación o charlas encaminadas al uso introductorio de la IA en la educación.	2	2	2	2	1
9	Introduzco progresivamente la inteligencia artificial como recurso para innovar y fortalecer mis procesos formativos como docente.	4	2	4	3	3

Continúa...

10	Me intereso por indagar ventajas y aplicaciones de la IA en la enseñanza de la educación básica.	4	4	5	4	2
11	Incentivo espacios de formación docente sobre el uso consciente y contextualizado de la inteligencia artificial.	2	5	5	5	1
12	Indago sobre diferentes herramientas de IA con fines educativos, sin incorporarla a la práctica docente.	3	2	2	2	2
13	Empleo la IA para eliminar actividades repetitivas de mi labor educativa.	3	3	3	3	2
14	Evito el uso de herramientas basadas en IA debido a aspectos éticos y de privacidad.	1	1	1	1	2
15	Utilizo la IA en aspectos puntuales de enseñanza, sin analizar sus riesgos y efectos en el aprendizaje	4	2	4	3	4
16	Analizo datos generados por sistemas de IA para tomar decisiones pedagógicas informadas en el contexto educativo.	4	5	4	4	3
17	Analizo críticamente el impacto de la IA en la labor pedagógica del docente	5	5	5	5	2
18	Indago aplicaciones de la IA con el fin de mejorar procesos evaluativos y de retroalimentación.	3	4	4	4	3
19	Utilizo la IA como una herramienta de apoyo a la labor educativa.	2	2	3	3	2
20	Comparto opiniones con docentes sobre experiencias iniciales en el uso de la IA en el ámbito educativo.	2	3	2	2	2
21	Desconfío de los resultados presentados por la IA especialmente en aspectos educativos.	1	1	1	1	2

Continúa...

22	Evalúo la pertinencia y viabilidad de soluciones propuestas por IA antes de aplicarlas al contexto educativo.	5	4	5	5	4
23	Incorporo herramientas de inteligencia artificial para diversificar estrategias metodológicas, adaptándolas a las necesidades y características de los estudiantes.	4	4	4	4	3
24	Empleo de manera básica la inteligencia artificial en mi actividad pedagógica.	3	3	3	3	2
25	Rechazo la incorporación de la IA en la enseñanza por no aportar valor significativo a los procesos de enseñanza.	1	1	1	1	2

Fuente: elaboración propia.

El coeficiente de Kappa de Fleiss ($k = 0,417$) (véase tabla 11) indica una fuerza de concordancia moderada entre los evaluadores, lo que sugiere un acuerdo superior al que se esperaría por mero azar en la calificación de los ítems de la EUIAP-E. Este resultado se considera estadísticamente significativo, como lo demuestra el valor de la prueba Z ($Z = 13,106$) y un P -valor de 0,000, el cual es inferior al umbral de significancia estándar ($\alpha = 0,05$). Finalmente, el intervalo de confianza asintótico del 95 % para Kappa (0,354 - 0,480) indica que el acuerdo real en la población de jueces se encuentra con alta probabilidad dentro de este rango.

Tabla 11. Kappa de Fleiss instrumento 2

Kappa general					
Kappa de Fleiss	Fuerza de concordancia	Z	P Valor	Límite inferior al 95%	Límite superior al 95%
,417	Moderada	13,106	,000	,354	,480

Fuente: elaboración propia.

Así mismo, de acuerdo con la tabla 12, la prueba Kappa de Fleiss por experto muestra valores moderados en tres de los cinco expertos. Considerando el experto 2 y 3 con una valoración aceptable en la prueba estadística.

Tabla 12. Kappa de Fleiss instrumento 2 por experto

Kappa general					
Experto	Probabilidad condicional	Kappa	P Valor	Límite inferior al 95%	Límite superior al 95%
1	,682	,614	,000	,490	,738
2	,389	,221	,000	,071	,344
3	,482	,333	,000	,209	,457
4	,576	,448	,000	,324	,572
5	,600	,524	,000	,400	,648

Fuente: elaboración propia.

DISCUSIÓN

Los resultados obtenidos a partir del análisis estadístico muestran que tanto el instrumento 1 como el 2 cuentan con una fuerte validez de contenido. Respecto al instrumento 1, los criterios de *claridad* ($V = 0,94$) y *relevancia* ($V = 0,94$) indican un consenso alto entre los jueces, lo que expresa preguntas correctamente elaboradas, comprensibles y pertinentes para el instrumento.

Sin embargo, la coherencia para las 8 preguntas mostró V de Aiken = 0,89, lo cual refleja que, si bien existe una conexión fuerte, este valor podría significar una revisión minuciosa a las preguntas con las puntuaciones más bajas, especialmente P3 y P4. En ese sentido, la prueba presenta una oportunidad para ajustar y perfeccionar el instrumento desde la redacción y la claridad con el fin de garantizar que la información recopilada aporte al desarrollo educativo desde el uso de la inteligencia artificial en los procesos de enseñanza,

lo cual fue previamente subsanado. Pruebas estadísticas como la V de Aiken se relacionan en estudios de Torres et al. (2022) y García et al. (2024), en los que se enfocan criterios de claridad, coherencia y relevancia; este tipo de estudios puede representar utilidad en las investigaciones educativas.

En relación con el instrumento 2, los expertos 1, 4 y 6 muestran niveles de acuerdo considerablemente más altos, con valores Kappa de 0,614, 0,448 y 0,524, respectivamente. Estos valores denotan una comprensión y aplicación más consistente en todos los indicadores de la escala Likert. De igual modo, los expertos 2 y 3 reflejaron valores más bajos, lo que sugiere la necesidad de mejora en estudios posteriores.

La aplicación del Kappa de Fleiss para el instrumento 2 demuestra que es una técnica reconocida por su viabilidad en contextos educativos y médicos, lo cual se expresa en el grado de acuerdo entre múltiples evaluadores (Loyo et al., 2024; Moons & Vanderviere, 2023). Los resultados obtenidos en este estudio ($k = 0,417$) exponen una concordancia moderada. Se establece un punto de partida para nuevas investigaciones que quieran profundizar sobre la relación entre educación y tecnologías emergentes.

Finalmente, la validación de los instrumentos psicométricos presentados permitirá no solo medir la efectividad y nivel de uso de la IA, sino que, desde la perspectiva aplicada, este artículo postula bases para la articulación de políticas educativas que desarrollen programas de formación para maestros, enfocados en la adquisición de competencias digitales para las nuevas realidades de los procesos de enseñanza.

CONCLUSIONES

La validación por juicio de expertos representa una oportunidad robusta de validación en la época actual, máxime cuando los contextos dinámicos de la sociedad del conocimiento requieren instrumentos que reflejen realidades académicas e investigativas lejos de la improvisación, la falta de análisis y las ambigüedades a las que se somete el conocimiento hoy en día.

Los resultados obtenidos en el instrumento 1 muestran validez sólida y consistente, ya que los indicadores de *claridad*, *coherencia* y *relevancia* superan el umbral de aprobación, reflejado a través de un coeficiente V de Aiken de 0,92.

La contribución de este estudio consiste en la validación del instrumento 2, una escala Likert que mide el nivel de uso de la IA en los procesos de enseñanza de la educación básica. Este tipo de instrumentos contribuyen al avance investigativo de los contextos latinoamericanos, pues se articula a las nuevas realidades educativas del siglo XXI. Es de mencionar que, aunque los resultados son confiables y aplicables en posteriores estudios, existen márgenes de mejora con el fin de alcanzar niveles de confiabilidad más altos. Futuras investigaciones podrán incrementar el uso de expertos especializados en el uso de la tecnología y la IA. La validación empleada reflejó observaciones en la redacción de ítems y preguntas, aspectos que ya fueron ajustados e incorporados en la versión final del instrumento.

A partir de los resultados obtenidos en las pruebas Kappa de Fleiss y V de Aiken, se valida el contenido de los instrumentos 1 y 2, respectivamente. La integración de la IA en la educación básica enfrenta retos y desafíos importantes que merecen un abordaje transversal. La incorporación de una validación por juicio de expertos aporta a la comprensión de cómo los docentes integran, perciben y utilizan la IA en las prácticas pedagógicas. Hecho que orienta políticas de formación, reflexión y uso consciente de tecnologías emergentes en el aula.

REFERENCIAS

- Alam, A. (2023). Developing a Curriculum for Ethical and Responsible AI: A University Course on Safety, Fairness, Privacy, and Ethics to Prepare Next Generation of AI Professionals. In *Intelligent Communication Technologies and Virtual Mobile Networks* (pp. 879-894). Springer Nature Singapore.

- Almanasreh, E., Moles, R. y Chen, T. (2019). Evaluación de métodos para estimar la validez de contenido. *Investigación en farmacia social y administrativa*, 15 (2), 214-221.
- Aparicio, G. W. O. (2023). La inteligencia artificial y su incidencia en la educación: Transformando el aprendizaje para el siglo XXI. *Revista Internacional de Pedagogía e Innovación Educativa*, 3(2), 217-230.
- Bellas, F., Guerreiro-Santalla, S., Naya, M., & Duro, R. J. (2023). AI curriculum for European high schools: An embedded intelligence approach. *International Journal of Artificial Intelligence in Education*, 33(2), 399-426.
- Berrones Yaulema, L. P. y Salgado Oviedo, S. A. (2023). La aplicación de la inteligencia artificial para mejorar la enseñanza y el aprendizaje en el ámbito educativo. *Esprint Investigación*, 2(1), 52-60. <https://doi.org/10.61347/ei.v2i1.52>.
- Casillas González, A. C., López Hernández, C. A. y Del Pilar Ortega, R. D. (2024). Hacia un Aprendizaje Avanzado: La Integración de la Inteligencia Artificial en la Educación Básica. *Ciencia Latina Revista Científica Multidisciplinar*, 8(4), 9702-9714. https://doi.org/10.37811/cl_rcm.v8i4.13126.
- Chiu, T. K. (2022). Designing Inclusive and Diverse Artificial Intelligence (AI) Curriculum for K-12 Education. In *Digital Communication and Learning: Changes and Challenges* (pp. 33-45). Springer Nature Singapore.
- Chiu, T. K. F., Ahmad, Z., & Çoban, M. (2025). Development and validation of teacher artificial intelligence (AI) competence self-efficacy (TAICS) scale. *Educ Inf Technol*, 30, 6667-6685. <https://doi.org/10.1007/s10639-024-13094-z>
- Dai, Y., Liu, A., Qin, J., Guo, Y., Jong, M. S. Y., Chai, C. S., & Lin, Z. (2023). Collaborative construction of artificial intelligence curriculum in primary schools. *Journal of Engineering Education*, 112(1), 23-42. doi: <https://doi.org/10.1002/jee.20503>.
- Douse, M. (2024). Educational Planning and Artificial Intelligence. *Educational Planning*, 31(2), 9-21.
- Gallent-Torres, C., Romero, B. A., Adillón, M. V. y Foltýnek, T. (2024). Inteligencia Artificial en educación: entre riesgos y potencialidades. *Praxis Educativa*, 19, 1-29.

- García, A., Corona, K., Hernández, M., Hernández, R. y Villalpa, H. (2024). Diseño y validación de contenido por juicio de expertos de un instrumento para evaluar la efectividad del proceso educativo relacionado al emprendimiento en estudiantes de educación superior. *Revista Espacios*, 45(4), 179-189. <https://doi.org/10.48082/espacios-a24v45n04p14>.
- Gómez-García, M., Ruiz-Palmero, J., Boumadan-Hamed, M. y Soto-Varela, R. (2025). Percepciones de futuros docentes y pedagogos sobre uso responsable de la IA. Un instrumento de medida. *RIED-Revista Iberoamericana de Educación a Distancia*, 28(2), 105-130.
- Harari, Y. (2024). *Nexus. Una breve historia de las redes de información desde la Edad de Piedra hasta la IA*. Editorial Debate.
- Hu, Y., & Liao, Z. (2020). Design of knowledge base and curriculum planning based on artificial intelligence teaching system. In *Proceedings of the 2020 International Conference on Computers, Information Processing and Advanced Education* (pp. 327-330).
- Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1), 159-174. <https://doi.org/10.2307/2529310>.
- López, G. y Pita, F. S. (1999). Medidas de concordancia: el índice de Kappa. *Cad Aten Primaria*, 6, 169-171.
- Loyo, O. M., Fuentes, M. Q., Pérez, G. L. y Zavaleta, J. G. (2024). Validez de contenido por medio de juicio de expertos para evaluar la inteligencia turística municipal. *593 Digital Publisher CEIT*, 9(4), 575-585.
- Martínez-Bonilla, I. (2025). Inteligencia Artificial en los procesos de enseñanza-aprendizaje: retos y oportunidades. *Psicología Educativa*, 13, 1-8. <https://doi.org/10.22201/fpsic.24486957e.2025.13.146>.
- Martínez-Corona, J. I., Palacios-Almón, G. E. y Juárez-Hernández, L. G. (2020). Diseño y Validación del instrumento: Enfoque Directivo en la Gestión para Resultados en la Sociedad del Conocimiento. *Espacios*, 10(19), 153-165. <https://doi.org/10.17163/ret.n19.2020.09>.
- Matas, A. (2018). Diseño del formato de escalas tipo Likert: un estado de la cuestión. *Revista Electrónica de Investigación Educativa*, 20(1), 38-47. <https://doi.org/10.24320/redie.2018.20.1.1347>.

- Merino, C. y Livia, J. (2009). Intervalos de confianza asimétricos para el índice de validez de contenido: un programa visual basic para la V de Aiken. *Anales de Psicología/Annals of Psychology*, 25(1), 169-171. <https://revistas.um.es/analesps/article/view/71631>.
- Moons, F. y Vandervieren, E. (2023). Medición de la concordancia entre varios evaluadores que clasifican a los sujetos en una o más categorías nominales (jerárquicas). Una generalización del índice kappa de Fleiss. *Preimpresión de arXiv*:2303.12502.
- Núñez-Michuy, C. M., Agualongo-Chela, L. M., Vistin, J. M. y López Quincha, M. (2023). La Inteligencia Artificial en la pedagogía como modelo de enseñanza. *Magazine de las Ciencias: Revista de Investigación e Innovación*, 8(2), 120-135. <https://doi.org/10.33262/rmc.v8i1.2932>.
- Penfield, R. y Giacobbi, Jr., P. R. (2004). Aplicación de un intervalo de confianza de puntuación al índice de relevancia del contenido de los ítems de Aiken. *Meas. Phys. Educ. Ejerc. Sci*, 8, 213-225.
- Santillán de la Torre, C. A., Silva López, M. de los Ángeles, Limongi Basantes, D. S. y Criollo Chango, E. A. (2024). El papel de la Inteligencia Artificial en la adaptación curricular: Perspectivas para la mejora continua de la educación digital. *Revista Imaginario Social*, 7(2). <https://doi.org/10.59155/is.v7i2.1869>.
- Sartor-Harada, A. y Azevedo-Gomes, J. (2026). AI-ED-SAT: diseño y validación de un cuestionario para autoevaluar competencias docentes en IA educativa. *RIED-Revista Iberoamericana de Educación a Distancia*, 29(1), 79-110. <https://doi.org/10.5944/ried.4541>.
- Selwyn, N. (2019). *Should Robots Replace Teachers?* Ediciones Morata.
- Su, J., & Zhong, Y. (2022). Artificial Intelligence (AI) in early childhood education: Curriculum design and future directions. *Computers and Education: Artificial Intelligence*, 3, 100072.
- Torres-Malca, J. R., Vera-Ponce, V. J., Zuzunaga-Montoya, F. E., Talavera, J. E. y Cruz-Vargas, J. A. (2022). Validez de contenido por juicio de expertos de un instrumento para medir conocimientos, actitudes y prácticas sobre el consumo de sal en la población peruana. *Revista de la Facultad de Medicina Humana*, 22(2), 273-279.
- Van Laar, E., Van Deursen, A. J., Van Dijk, J. A., & De Haan, J. (2017). The relation between 21st-century skills and digital skills: A systematic literature review. *Computers in Human Behavior*, 72, 577-588.