

*Artículos científicos***Evaluar el pensamiento crítico en la era de la IA: construcción y validación de una escala de disposiciones en contextos educativos mediados por IAG***Evaluating Critical Thinking in the Age of AI: Construction and Validation of a Disposition Scale for GAI-Mediated Educational Contexts***Mónica del Rocío Rodríguez Sánchez**

Universidad Autónoma de Querétaro

mrrs_90@hotmail.com<https://orcid.org/0009-0008-5247-9523>**Resumen**

La incorporación de modelos de inteligencia artificial generativa (IAG) en la educación superior ha transformado las prácticas académicas, generando inquietudes sobre su impacto en el pensamiento crítico de los estudiantes. Ante este panorama, el estudio tuvo como objetivo construir y validar psicométricamente una escala para evaluar las disposiciones al pensamiento crítico en estudiantes universitarios que utilizan herramientas de IAG con fines académicos, tomando como base en el modelo de Ennis (1996). Se llevó a cabo un estudio cuantitativo con diseño instrumental. Se elaboró una versión inicial de 36 ítems, sometida a validación de contenido por juicio de expertos, obteniéndose un “Coeficiente de Validez de Contenido (CVC)” de .94. Posteriormente, la escala fue aplicada a una muestra de 176 estudiantes de licenciatura del área de Ciencias Sociales y Humanidades. Mediante un AFE por bloques, se reestructuraron las disposiciones al pensamiento crítico en cuatro factores emergentes con adecuada consistencia interna: rigor en el proceso de indagación ($\alpha = .848$), flexibilidad cognitiva ($\alpha = .778$), pensamiento reflexivo contextualizado ($\alpha = .863$) y evaluación crítica de la información ($\alpha = .887$). El modelo final mostró cargas factoriales entre .63 y .86, una varianza explicada del 55.3% y un alfa de Cronbach global de .951 (IC 95 %: .937–.965). Los resultados confirmaron la validez estructural y la consistencia interna del instrumento, que se presenta como una herramienta confiable para el diagnóstico de disposiciones cognitivas frente al uso académico de la IAG. Su aplicación puede orientar prácticas formativas centradas en la reflexión crítica, la autonomía intelectual y el uso ético de tecnologías emergentes en la educación superior.

Palabras clave: Validez de constructo, consistencia interna, diseño instrumental, juicio crítico, educación universitaria.

Abstract

The integration of generative artificial intelligence (GAI) tools into higher education has transformed academic practices, prompting questions about their impact on students' critical thinking. In response, this study aimed to develop and psychometrically validate a scale to assess dispositional associated with critical thinking in university students using GAI for academic purposes. The theoretical

foundation was based on Ennis' (1996) model of critical thinking dispositions. A quantitative, instrumental research design was adopted. An initial pool of 36 items was developed and evaluated by expert judgment, yielding a Content Validity Coefficient (CVC) of .94. The revised scale was administered to a sample of 176 undergraduate students from the social sciences and humanities. Exploratory factor analysis by conceptual blocks restructured the original twelve dispositions into four empirically supported factors: rigor in the inquiry process ($\alpha = .848$), cognitive flexibility ($\alpha = .778$), contextualized reflective thinking ($\alpha = .863$), and critical evaluation of information ($\alpha = .887$). The final model showed factor loadings ranging from .63 to .86, explained variance of 55.3%, and excellent overall reliability (Cronbach's $\alpha = .951$; 95% CI: .937–.965). Results confirmed the structural validity and internal consistency of the instrument, positioning it as a reliable diagnostic tool for assessing critical thinking dispositions in academic contexts mediated by GAI. Its use may inform the design of formative interventions that promote epistemic responsibility, reflective reasoning, and intellectual autonomy in technology-rich educational environments.

Key words: construct validity, internal consistency, instrumental design, critical judgment, university education

Fecha Recepción: Enero 2025

Fecha Aceptación: Mayo 2025

Introducción

La incorporación de modelos de Inteligencia artificial generativa (IAG), como *ChatGPT*, *Bing o Gemini*, está transformando las prácticas académicas en la educación superior. Estas herramientas no solo automatizan procesos de redacción y búsqueda de información, sino que también modifican profundamente la forma en que los estudiantes acceden al conocimiento, y elaboran juicios y toman decisiones académicas informadas (Shanto, Ahmed & Islam, 2024; Nguyen, Le & Nguyen, 2024). Actualmente, herramientas como ChatGPT han alcanzado más de 100 millones de usuarios mensuales, lo cual ha generado debates globales sobre su impacto en la autoría académica, la evaluación educativa y el desarrollo del pensamiento autónomo. Si bien su potencial pedagógico es amplio, la ausencia de una mediación crítica puede derivar en prácticas cognitivamente empobrecidas, como la dependencia excesiva o la superficialidad en el análisis.

Desde una perspectiva optimista, Luckin, Holmes, Griffiths y Forcier (2016) sostienen que la IAG tiene el potencial de fortalecer el aprendizaje humano, siempre que su implementación se base en principios pedagógicos sólidos, centrados en el estudiante y orientados al desarrollo de habilidades complejas. Sin embargo, los autores advierten que, en ausencia de una orientación educativa clara, estas tecnologías podrían perpetuar enfoques tradicionales y mecanicistas que limitan el pensamiento crítico. En esta misma línea, Zawacki-Richter, Marín, Bond y Gouverneur (2019) subrayan que la investigación sobre IA en el ámbito de la educación superior ha tendido a centrarse en aspectos computacionales y técnicos, descuidando el análisis del rol docente y sus implicaciones pedagógicas. Esta omisión representa un obstáculo significativo para comprender de forma integral el fenómeno, especialmente en lo que respecta a la formación de competencias críticas en los estudiantes.

Como consecuencia de esta falta de orientación pedagógica, se han identificado riesgos significativos como la aceptación acrítica de información, la desinformación algorítmica y la disminución de la autoría intelectual, especialmente cuando se utilizan estas herramientas como fuentes únicas o definitivas (Ríos-Hernández, Mateus, Rivera & Ávila, 2024; Clemente Alcocer, Cabello Cabrera & Añorve García, 2024). Estas tensiones subrayan la necesidad urgente de formar a los estudiantes en competencias críticas que permitan evaluar, contrastar y contextualizar la información generada por sistemas de IAG.

Es por esto que, el pensamiento crítico cobra una relevancia central. Diversos autores coinciden en que esta competencia no solo implica habilidades cognitivas, sino también disposiciones personales que favorecen el pensamiento razonado, reflexivo y abierto (Ennis, 1996; Facione, 2011). Estas disposiciones, tales como la búsqueda de claridad, la evaluación de fuentes, la toma de postura informada y la voluntad de cambiar de opinión ante argumentos sólidos, son esenciales en contextos académicos mediados por tecnologías emergentes.

A pesar de los avances recientes, la mayoría de los estudios han privilegiado perspectivas instrumentales o éticas sobre la IAG (García-Peñalvo, 2023; Walczak & Cellary, 2023), sin explorar con suficiente profundidad cómo estas tecnologías afectan el desarrollo y activación de disposiciones críticas en los estudiantes. De manera particular, no se han identificado instrumentos psicométricos validados que permitan medir dichas disposiciones en el contexto del uso académico de la IAG. Esto limita la capacidad institucional de diseñar intervenciones pedagógicas basadas en evidencia.

Este vacío cobra especial relevancia si se consideran los hallazgos recientes de Ruiz-Rojas, Sánchez-González y Morales-Torres (2024), quienes señalan que los estudiantes tienden a utilizar herramientas de IAG como fuentes primarias sin contar con criterios claros de evaluación crítica, lo cual puede limitar su desarrollo intelectual. Ante esta laguna, el presente estudio tuvo como objetivo construir y validar una escala que evalúe las disposiciones cognitivas asociadas al pensamiento crítico en estudiantes universitarios que utilizan herramientas de IAG en sus actividades académicas. La escala se fundamenta teóricamente en el modelo propuesto por Ennis (1996) y fue sometida a un proceso riguroso de validación de contenido y AFE.

Este estudio responde a dicha necesidad mediante la construcción y validación de una escala que, además de atender criterios psicométricos rigurosos, permite interpretar el pensamiento crítico desde una perspectiva funcional, situada y adaptada a los desafíos de la educación superior en la era de la IAG.

El artículo se organiza en cinco apartados. En primer lugar, se describe la metodología empleada para la construcción y validación del instrumento; en segundo lugar, se presentan los resultados psicométricos obtenidos; posteriormente, se discuten los hallazgos a la luz del marco teórico; y finalmente, se plantean las conclusiones del estudio y las líneas futuras de investigación.

Metodología

Este estudio se enmarca en un enfoque cuantitativo, con un diseño instrumental orientado a la construcción y validación psicométrica de una escala para evaluar disposiciones al pensamiento crítico en contextos de uso académico de la IAG. La fundamentación teórica se basó en el modelo propuesto por Ennis (1996), el cual define doce componentes actitudinales que configuran una actitud crítica

consistente. Estas disposiciones, que sirvieron de base para la redacción de los ítems, se presentan en la Tabla 1.

Tabla 1 Disposiciones originales al pensamiento crítico según Ennis (1996)

Disposición	Descripción breve
Buscar claridad	Expresar ideas con precisión, sin ambigüedad.
Buscar razones	Fundamentar afirmaciones antes de aceptarlas.
Ser persistente en la búsqueda de la verdad	No conformarse con respuestas superficiales.
Mantener una mente abierta	Considerar puntos de vista distintos al propio.
Buscar explicaciones alternativas	Analizar otras posibles respuestas o causas.
Intentar ser bien informado	Buscar información suficiente y pertinente.
Considerar la situación total	Comprender el contexto de las afirmaciones.
Evaluar las implicaciones	Reflexionar sobre las consecuencias de las ideas.
Preocuparse por la exactitud	Verificar que los datos sean correctos y comprobables.
Ser sensible al nivel de detalle necesario	Ajustar la profundidad de la información según la necesidad del problema.
Usar fuentes creíbles	Basarse en información confiable y verificable.
Adaptarse a nuevas situaciones	Cambiar de opinión cuando haya razones válidas.

Fuente: Adaptado de Ennis, R. H. (1996).

A partir de dichas disposiciones, se construyeron 36 ítems iniciales, redactados en primera persona y contextualizados al uso académico de herramientas como *ChatGPT*, *Bing* o *Gemini*. Cada disposición fue representada por tres ítems, formulados en lenguaje afirmativo y observacional, con el fin de facilitar la autorreflexión y la medición directa.

En la primera etapa del proceso se realizó una validación de contenido por juicio de expertos. Participaron tres académicas especialistas en evaluación educativa, pedagogía y formación docente, quienes evaluaron cada ítem en términos de claridad, congruencia teórica, pertinencia del lenguaje y relevancia. Para el análisis cuantitativo se empleó el Coeficiente de Validez de Contenido (CVC) propuesto por Hernández-Nieto (2002), obteniéndose un valor promedio de .94, considerado excelente. Con base en las sugerencias cualitativas se reformularon enunciados, eliminando ambigüedades y asegurando redacciones claras y unívocas.

La versión ajustada fue aplicada a una muestra no probabilística de 176 estudiantes de licenciatura del área de Ciencias Sociales y Humanidades, mediante un formulario digital administrado durante el primer trimestre de 2025. La participación fue voluntaria, anónima y precedida de consentimiento informado. Dos ítems fueron eliminados por presentar una correlación ítem-total inferior a .30, criterio psicométrico comúnmente aceptado para depurar instrumentos (Muñiz & Fonseca-Pedrero, 2019). Por tanto, el análisis psicométrico se realizó sobre un total de 34 ítems.

Se llevó a cabo un análisis factorial exploratorio (AFE) mediante el método de extracción de ejes principales con rotación oblimin, en atención a las correlaciones esperadas entre factores conceptualmente relacionados. Esta técnica resulta adecuada en fases iniciales de validación de

instrumentos, ya que permite identificar estructuras latentes sin imponer restricciones previas al modelo (Lloret-Segura et al., 2014; Ferrando & Lorenzo-Seva, 2017). Esta decisión se fundamenta en las recomendaciones de Lloret-Segura et al. (2014), quienes proponen este enfoque como el más adecuado para estructuras latentes correlacionadas. También se atendieron los lineamientos de Ferrando y Anguiano-Carrasco (2010), quienes sugieren evaluar empíricamente la unidimensionalidad de bloques conceptuales cuando se parte de modelos teóricos bien definidos. El AFE global arrojó un índice KMO de .933 y una prueba de esfericidad de Bartlett significativa ($p < .001$), lo cual confirmó la pertinencia del análisis al evidenciar relaciones robustas entre variables.

Sin embargo, la estructura bifactorial emergente no reflejaba adecuadamente la división teórica original en doce disposiciones. Ante esta discrepancia, se optó por realizar el AFE por bloques, siguiendo la recomendación metodológica de Méndez Martínez y Rondón Sepúlveda (2012), quienes indican esta técnica como apropiada cuando se parte de una estructura teórica previa y se busca evaluar empíricamente la unidimensionalidad específica de subdimensiones conceptuales, como es el caso del modelo de Ennis.

Los resultados permitieron reagrupar diez de las doce disposiciones originales en cuatro factores emergentes con afinidad conceptual y consistencia estadística. Las disposiciones “definir términos” y “preguntar y responder con precisión y relevancia” ueron excluidas debido a cargas factoriales bajas y comunalidades por debajo del umbral aceptable, indicando una aportación limitada al modelo. La Tabla 2 presenta la estructura final del instrumento, con cuatro factores emergentes derivados del análisis factorial. El alfa de Cronbach global de la escala fue de .951, indicando una excelente consistencia interna para los 21 ítems que la componen. Este coeficiente es uno de los indicadores más ampliamente aceptados para evaluar la fiabilidad interna de escalas en psicología y educación, especialmente en diseños instrumentales (Oviedo & Campo-Arias, 2005).

Tabla 2 Estructura final del instrumento validado: factores, disposiciones integradas, número de ítems y alfa de Cronbach

Factor emergente	Disposiciones de Ennis integradas	N.º de ítems	de Alfa de Cronbach
Rigor en el proceso de indagación	Buscar claridad, Buscar razones, Ser persistente en la búsqueda de la verdad	5	.848
Flexibilidad cognitiva	Mantener una mente abierta, Buscar explicaciones alternativas	4	.778
Pensamiento reflexivo contextualizado	Considerar la situación total, Evaluar implicaciones, Intentar ser bien informado	6	.863
Evaluación crítica de la información	Preocuparse por la exactitud, Ser sensible al nivel de detalle necesario	6	.887
Total escala	Diez disposiciones reagruparon de forma empírica	21	.951

Nota. El alfa de Cronbach global (.951) corresponde a la confiabilidad total de la escala compuesta por 21 ítems.

Fuente: Elaboración propia con base en el AFE (JASP, 2025).

Con base en esta estructura refinada, se construyó la versión final del instrumento, conformada por 21 ítems distribuidos en cuatro dimensiones: rigor en el proceso de indagación, flexibilidad cognitiva, pensamiento reflexivo contextualizado y evaluación crítica de la información. El análisis factorial final mostró un KMO de .941, cargas factoriales entre .63 y .86, y una varianza explicada del 55.3 %. La confiabilidad interna global del instrumento, medida mediante el coeficiente alfa de Cronbach, fue de .951 (IC 95 %: .937–.965), con valores por subescala entre .778 y .887, lo que indica una consistencia interna excelente tanto a nivel global como factorial.

La Tabla 3 detalla los ítems que componen la versión final del instrumento, indicando para cada uno el código asignado, su redacción, la disposición crítica que evalúa y la categoría emergente a la que pertenece.

Tabla 3 Reactivos finales del instrumento agrupados por factor emergente y disposición original de Ennis

Factor emergente	Ítem Enunciado del reactivo	Disposición de Ennis
Rigor en el proceso de indagación	R4 Si las respuestas de la IAG no son claras o carecen de fundamentos, recorro a fuentes confiables para verificar o ampliar la información.	Buscar razones
	R5 Verifico si la IAG incluye fundamentos, ejemplos o referencias que respalden su respuesta.	
	R26 No me conformo con la primera respuesta de la IAG; indago con mayor profundidad para comprender mejor el tema.	
	R27 Cuando la IAG me da respuestas contradictorias, realizo una investigación adicional para comprender mejor el tema.	Ser persistente en la búsqueda de la verdad
	R28 Antes de dar por válida una respuesta de la IAG, consulto diferentes fuentes para asegurarme de su veracidad.	
Flexibilidad cognitiva	R16 Si la respuesta de la IAG difiere de mis ideas, la contrasto con otras fuentes antes de aceptarla o rechazarla.	Mantener una mente abierta
	R21 No acepto de inmediato lo que dice la IAG; primero considero otras explicaciones para analizar diferentes puntos de vista.	
	R22 Analizo distintas perspectivas antes de integrar como válida la conclusión que me da la IAG	Buscar explicaciones alternativas
	R23 Reviso si hay otras explicaciones más claras o completas que las proporcionadas por la IAG.	
Pensamiento reflexivo contextualizado	R3 Me aseguro de que la redacción y el sentido de lo que dice la IAG sean comprensibles y coherentes.	Buscar claridad

Factor emergente	Ítem Enunciado del reactivo	Disposición de Ennis
	R7 Cuando la información de la IAG no tiene suficiente profundidad, recorro a textos especializados para ampliar mi comprensión.	Intentar ser bien informado
	R8 Reviso toda la información que proporciona la IAG para asegurarme de entenderla bien.	
	R12 Analizo si la información que proporciona la IAG es pertinente y útil para mi contexto específico.	
	R13 Analizo si la información que ofrece la IAG es coherente con el contexto y los enfoques de mi disciplina.	Considerar la situación total
	R14 Modifico o complemento la información que me da la IAG si no se ajusta a las particularidades de mi disciplina.	
Evaluación crítica de la información	R18 Reviso que la información generada por la IAG sea precisa y esté libre de errores.	
	R19 Contrasto los datos o cifras proporcionados por la IAG con otras fuentes verificadas.	Preocuparse por la exactitud
	R20 Verifico que la información proporcionada por la IAG sea específica y precisa.	
	R29 Evalúo si la respuesta de la IAG incluye el nivel de detalle requerido para abordar adecuadamente el contenido.	
	R30 Reflexiono si la información proporcionada por la IAG es suficiente o si requiero ampliar mi búsqueda.	Ser sensible al nivel de detalle necesario
	R31 Evalúo si necesito más información que la que da la IAG.	

Fuente: Elaboración propia con base en los resultados del análisis factorial y la reorganización teórica del instrumento (2025).

Este proceso metodológico buscó preservar el fundamento teórico del modelo original, a Este proceso metodológico buscó preservar el fundamento teórico del modelo original, a la par de garantizar una estructura empírica robusta y coherente. La investigación forma parte de un proyecto doctoral en curso en el marco del Doctorado en Innovación en Tecnología Educativa de la Universidad Autónoma de Querétaro, y fue aprobada conforme a los lineamientos éticos institucionales.

Resultados

El AFE final evidenció una estructura de cuatro factores empíricamente coherente y estadísticamente robusta. El índice KMO obtenido fue de .941 y la prueba de esfericidad de Bartlett resultó significativa ($\chi^2 = 2465.37$, $gl = 210$, $p < .001$), donde gl representa los grados de libertad, es decir, la cantidad de parámetros que pueden variar sin infringir las restricciones del modelo estadístico. Estos resultados

confirmaron la idoneidad de la matriz de correlaciones para el análisis y respaldan el uso del AFE como técnica válida para identificar estructuras latentes, especialmente cuando se parte de un marco teórico definido, como recomiendan Méndez Martínez y Rondón Sepúlveda (2012).

Las cargas factoriales oscilaron entre .63 y .86, con unicidades por debajo de .50, lo cual sugiere que más del 50 % de la varianza de cada ítem es explicada por el factor correspondiente, indicando así una apropiada contribución de cada ítem a su dimensión. La varianza total explicada por los cuatro factores fue de 55.3 %, lo que se considera aceptable para estudios en ciencias sociales con constructos complejos como las disposiciones al pensamiento crítico.

En términos de confiabilidad, el coeficiente alfa de Cronbach global fue de .951 (IC 95 %: .937–.965), lo que evidencia una consistencia interna excelente. A nivel de subescalas, los resultados por subescala también indicaron una confiabilidad adecuada: .848 para rigor en el proceso de indagación, .778 para flexibilidad cognitiva, .863 para pensamiento reflexivo contextualizado, y .887 para evaluación crítica de la información.

Los hallazgos permiten concluir que el instrumento validado presenta una estructura interna estable y consistente. La reagrupación de diez de las disposiciones originales de Ennis en cuatro factores emergentes no solo mantuvo su integridad conceptual, sino que también optimizó el comportamiento psicométrico de la escala, consolidándola como una herramienta confiable para el diagnóstico de disposiciones al pensamiento crítico en contextos de uso académico de IAG.

Como complemento a los indicadores de validez estructural y confiabilidad interna, se calcularon estadísticos descriptivos por cada dimensión emergente del instrumento, con el objetivo de ofrecer normas preliminares de interpretación. Estos valores permiten identificar tendencias generales en las respuestas del estudiantado y ofrecen un punto de partida para su uso con fines diagnósticos o formativos.

Tabla 4. Estadísticos descriptivos y percentiles por dimensión emergente del instrumento

Dimensión	Cant. ítems	Media	DE	Mín.	Máx.	Percentil 25	Percentil 50	Percentil 75
Evaluación crítica de la información	5	4.208	0.876	3.873	4.352	4.225	4.268	4.324
Flexibilidad cognitiva	4	3.947	0.936	3.704	4.07	3.926	4.007	4.028
Pensamiento reflexivo contextualizado	6	4.324	0.83	3.93	4.493	4.3	4.408	4.434
Rigor en el proceso de indagación	5	4.062	1.003	3.761	4.31	3.972	4.113	4.155

Nota. Los valores corresponden a los promedios agregados por dimensión, calculados con base en las respuestas de 176 estudiantes de licenciatura. Estos datos se presentan con fines interpretativos preliminares y deberán complementarse con estudios normativos posteriores en muestras ampliadas y diversificadas.

Discusión

Los resultados obtenidos respaldan la validez estructural y la consistencia interna del instrumento diseñado para evaluar disposiciones al pensamiento crítico en el uso académico de IAG. La reorganización de diez disposiciones del modelo de Ennis (1996) en cuatro factores emergentes evidencia una adaptación teórica coherente con los patrones de respuesta observados, sin perder la esencia conceptual del marco original (Facione, 2011; Muñiz & Fonseca-Pedrero, 2019).

El agrupamiento de disposiciones como la búsqueda de claridad, la búsqueda de razones y la persistencia sugiere que los estudiantes tienden a activar estilos cognitivos integrados en entornos mediados por IAG, lo cual coincide con hallazgos recientes que subrayan la importancia de comprender el pensamiento crítico como una práctica situada y contextual (Ríos-Hernández & Mateu-Martín, 2024; Gallent-Torres et al., 2023). En particular, la dimensión de evaluación crítica de la información adquiere relevancia en contextos donde los estudiantes podrían aceptar sin cuestionamiento las respuestas automatizadas, lo que plantea la necesidad de reforzar sus criterios de valoración epistémica (Floridi, 2013; Clemente Alcocer et al., 2024).

Asimismo, los factores asociados con la flexibilidad cognitiva y el pensamiento reflexivo contextualizado muestran que el juicio crítico frente a la IAG no solo exige precisión lógica, sino también apertura al cambio, lectura contextualizada de la información y disposición a reconsiderar posturas frente a nuevos argumentos (Paul & Elder, 2008; García-Peñalvo, 2023). Esta perspectiva multidimensional se alinea con concepciones contemporáneas del pensamiento crítico que lo entienden como una actitud activa y situada, no reducida al dominio de habilidades formales.

Desde esta perspectiva, la escala propuesta constituye un aporte significativo al enfocarse no en competencias técnicas de uso de la IA, sino en las disposiciones profundas que sostienen el pensamiento crítico frente a tecnologías emergentes (Ríos-Hernández & Mateu-Martín, 2024; Gouia-Zarrad & Gunn, 2024). En esta línea, estudios como el de Nguyen y Kim (2024) han demostrado que las herramientas de escritura asistida por IAG pueden fomentar procesos críticos significativos, siempre que se integren en marcos pedagógicos que promuevan la reflexión activa.

Además, investigaciones como las de Luckin et al. (2016) advierten que la IAG solo potencia el aprendizaje humano cuando está mediada por una intencionalidad pedagógica clara, centrada en el desarrollo de habilidades complejas. Sin dicha mediación, el uso de estas tecnologías corre el riesgo de reproducir prácticas pasivas o descontextualizadas. Del mismo modo, Zawacki-Richter et al. (2019) han señalado que la mayoría de los estudios sobre IA en educación se han centrado en dimensiones técnicas, dejando de lado las implicaciones pedagógicas y la participación activa del profesorado.

En este sentido, el instrumento propuesto no solo tiene valor diagnóstico, sino que también proporciona una base empírica para el diseño de intervenciones educativas con enfoque crítico, orientadas al fortalecimiento de la autonomía intelectual y el juicio reflexivo en entornos mediados por IAG.

Más allá de su validación inicial, esta herramienta permite articular la evaluación de disposiciones críticas con procesos formativos situados, brindando información clave para diseñar experiencias de aprendizaje que favorezcan la reflexión epistémica, la autorregulación cognitiva y la toma de decisiones informadas en el uso académico de IA.

Esta proyección resulta especialmente relevante en un contexto donde el despliegue de tecnologías generativas exige que el pensamiento crítico se consolide como una competencia transversal, que no solo resista la automatización, sino que promueva el discernimiento ético y la apropiación responsable del conocimiento.

Así, esta investigación contribuye a posicionar el pensamiento crítico como eje articulador de una educación superior orientada al juicio reflexivo, la alfabetización tecnológica crítica y el desarrollo de ciudadanías cognitivamente activas en la era de la inteligencia artificial.

Limitaciones y recomendaciones metodológicas

Si bien los hallazgos obtenidos permiten confirmar la validez estructural y la consistencia interna del instrumento, es necesario reconocer una serie de limitaciones metodológicas que deben ser consideradas al interpretar los resultados y que orientan las futuras fases del proyecto.

Una primera limitación se relaciona con la naturaleza no probabilística de la muestra, compuesta exclusivamente por estudiantes del área de Ciencias Sociales y Humanidades. Aunque esta decisión respondió a criterios logísticos y de disponibilidad institucional, restringe la validez externa del instrumento, al no permitir la generalización a otros campos disciplinares donde las disposiciones críticas podrían manifestarse de manera distinta. Se recomienda, por tanto, ampliar la aplicación del instrumento a disciplinas STEM, salud e ingenierías, con el fin de explorar diferencias entre campos y avanzar hacia un modelo de invarianza factorial multigrupo.

Aunque el AFE arrojó resultados sólidos, con índices estadísticos adecuados y estructura coherente con el modelo teórico, no se incluyó en esta fase un análisis factorial confirmatorio (AFC). Dado que el presente estudio se centró en la validación exploratoria inicial, el AFC se contempla como un paso siguiente en una nueva fase, con una muestra independiente, para evaluar la estabilidad de la estructura mediante índices de ajuste más rigurosos.

Otra limitación importante es la falta de evidencias de validez convergente y discriminante, ya que no se aplicaron instrumentos adicionales con los cuales establecer correlaciones. Esta omisión se debió a la escasez de escalas validadas en español que evalúen disposiciones críticas específicamente en contextos de uso académico de IAG. En futuras fases, se planea emplear escalas estandarizadas como el California Critical Thinking Disposition Inventory (CCTDI) o el Watson-Glaser Critical Thinking Appraisal, para robustecer la validez externa y teórica del instrumento.

Tampoco se evaluó la estabilidad temporal mediante un test-retest. Considerando que las disposiciones al pensamiento crítico pueden verse influenciadas por experiencias académicas o intervenciones formativas, se proyecta implementar esta estrategia de seguimiento con un intervalo de tres a cuatro semanas, a fin de estimar la confiabilidad temporal y observar posibles variaciones vinculadas al contexto formativo.

Por último, se reconoce la ausencia de normas preliminares de interpretación que faciliten su aplicación diagnóstica. Como primer paso hacia su interpretación pedagógica, se han calculado medidas de tendencia central y dispersión por dimensión. No obstante, será necesario contar con una muestra más amplia y diversa para construir baremos comparativos útiles en contextos educativos diferenciados.

Finalmente, se realizó un análisis de poder estadístico post hoc, que confirmó la suficiencia del tamaño muestral utilizado ($n = 176$) en función del número de ítems y factores emergentes, cumpliendo con los criterios establecidos por la literatura metodológica para análisis factoriales (entre 5 y 10 sujetos por ítem).

Estas recomendaciones metodológicas no solo permiten delimitar el alcance de los hallazgos presentados, sino que trazan una ruta clara para el fortalecimiento del instrumento como una herramienta psicométrica confiable, válida y aplicable en escenarios educativos mediados por IAG.

Implicaciones prácticas de uso

La escala validada representa una herramienta útil para el diagnóstico formativo de disposiciones al pensamiento crítico en el contexto del uso académico de IAG. Su aplicación práctica puede extenderse a diversos escenarios educativos:

- Como instrumento diagnóstico inicial para estudiantes de nuevo ingreso en programas de licenciatura, con el fin de identificar necesidades formativas en pensamiento crítico
- Como insumo para cursos de alfabetización digital crítica, en donde se analice el uso ético y reflexivo de herramientas de IAG
- Como parte de procesos de evaluación formativa en asignaturas centradas en habilidades de pensamiento superior.

Además, esta escala permite orientar el diseño de estrategias pedagógicas diferenciadas, personalizando las intervenciones según los niveles de disposición identificados en cada estudiante. Por ejemplo, puntajes bajos en el factor “evaluación crítica de la información” podrían detonar actividades específicas de análisis de fuentes o talleres de verificación de datos, mientras que niveles medios en “reflexión contextualizada” podrían abordarse con estudios de caso o debates. Asimismo, puede facilitar procesos de retroalimentación individual o grupal y servir como insumo empírico en proyectos institucionales enfocados en el desarrollo de una ciudadanía digital crítica, ética y autónoma frente a los desafíos emergentes de la IA en entornos educativos.

Conclusiones

La validación psicométrica del instrumento desarrollado representa un aporte sustantivo a la medición de disposiciones cognitivas asociadas al pensamiento crítico en contextos educativos mediados por IAG. Mediante un AFE riguroso, se reorganizaron diez disposiciones del modelo de Ennis (1996) en cuatro factores emergentes que demostraron coherencia teórica, solidez estadística y aplicabilidad educativa. Esta reformulación permite una lectura más operativa de las disposiciones críticas, ajustada a los desafíos cognitivos que plantea el uso de herramientas de IAG en la educación superior.

A diferencia de escalas centradas exclusivamente en actitudes hacia la tecnología o competencias instrumentales, esta herramienta se enfoca en disposiciones intelectuales profundas que articulan juicio, reflexión y evaluación crítica. Su estructura flexible permite tanto el diagnóstico general como el análisis por dimensiones, lo que habilita su uso en procesos de formación universitaria, programas de inducción, cursos de alfabetización digital crítica y estrategias de retroalimentación personalizada. En este sentido, el instrumento no solo aporta una herramienta de evaluación confiable y válida, sino que también habilita intervenciones pedagógicas orientadas al fortalecimiento de la autonomía

intelectual y el juicio ético en entornos mediados por IA. En un escenario educativo en transformación, contar con instrumentos como este resulta clave para diseñar prácticas formativas informadas, contextualizadas y socialmente responsables.

Futuras líneas de investigación

Una primera línea de continuidad consiste en desarrollar un instrumento complementario que permita evaluar también las 16 habilidades del pensamiento crítico propuestas por Ennis (1996), con el propósito de integrar la dimensión disposicional con la dimensión cognitiva. Estas habilidades abarcan procesos como inferencia, análisis, evaluación, explicación y autorregulación, fundamentales para operacionalizar el pensamiento crítico desde una perspectiva funcional. Esta integración permitiría una comprensión más holística del pensamiento crítico en escenarios educativos mediados por IAG, fortaleciendo tanto su medición como su intervención pedagógica.

En segundo lugar, resulta prioritario explorar la validez y confiabilidad del instrumento en contextos educativos diversos, incluyendo disciplinas técnicas (STEM), niveles educativos distintos: media superior o posgrado, así como regiones culturales variadas. Esto permitiría evaluar la invarianza factorial y la aplicabilidad del instrumento más allá del contexto original, contribuyendo a su generalización y robustez teórica.

Otra línea prometedora es la realización de estudios longitudinales que evalúen la estabilidad de las disposiciones críticas a lo largo del tiempo y analicen el impacto de intervenciones educativas mediadas por IA en el fortalecimiento de estas disposiciones. Esto aportaría evidencia sobre la sensibilidad del instrumento para detectar cambios en procesos formativos sostenidos.

Asimismo, siguiendo las sugerencias de Vázquez-Parra et al. (2024), futuras investigaciones podrían indagar en la relación entre las percepciones estudiantiles sobre la adopción y formación en herramientas de IA, y el desarrollo de sus disposiciones críticas. Sería relevante examinar cómo estas percepciones varían en función del nivel educativo, el tipo de herramienta utilizada o el acompañamiento pedagógico recibido.

Finalmente, la implementación de esta escala puede servir como base empírica para diseñar prácticas educativas orientadas al pensamiento crítico, promoviendo una relación ética, reflexiva y transformadora con las tecnologías emergentes. Esto permitiría sustentar propuestas curriculares más pertinentes ante los desafíos actuales de la educación superior en la era de la inteligencia artificial.

Agradecimientos

La presente investigación forma parte del proyecto doctoral desarrollado en el marco del Doctorado en Innovación en Tecnología Educativa de la Universidad Autónoma de Querétaro. Se agradece el valioso apoyo de las académicas expertas que participaron en el proceso de validación de contenido del instrumento, así como la disposición y colaboración de las y los estudiantes de la Facultad de Ciencias Sociales y Humanidades de la Universidad La Salle Bajío, quienes conformaron la muestra del estudio. Su participación fue fundamental para la construcción de este instrumento.

Referencias

- Clemente Alcocer, A. A., Cabello Cabrera, A., & Añorve García, E. (2024). La inteligencia artificial en la educación: desafíos éticos y perspectivas hacia una nueva enseñanza. *LATAM Revista Latinoamericana de Ciencias Sociales y Humanidades*, 5(6), 464–472. <https://doi.org/10.56712/latam.v5i6.3019>
- Ennis, R. H. (1996). *Critical thinking dispositions: Their nature and assessability*. *Informal Logic*, 18(2), 165–182. <https://doi.org/10.22329/il.v18i2.2378>
- Facione, P. A. (2011). *Critical thinking: What it is and why it counts* (PDF Ed.). Insight Assessment. <https://www.researchgate.net/publication/251303244>
- Ferrando, P. J., & Anguiano-Carrasco, C. (2010). El análisis factorial como técnica de investigación en psicología. *Papeles del Psicólogo*, 31(1), 18–33. <https://www.redalyc.org/articulo.oa?id=77812441003>
- Ferrando, P. J., & Lorenzo-Seva, U. (2017).** Program FACTOR at 10: Origins, development and future directions. *Psicothema*, 29(2), 236–240. <https://doi.org/10.7334/psicothema2016.304>
- Floridi, L. (2013). *The ethics of information*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199641321.001.0001>
- Gallent, C.; A. Zapata; J. Ortego (2023). El impacto de la inteligencia artificial generativa en educación superior: una mirada desde la ética y la integridad académica. *RELIEVE - Revista Electrónica de Investigación y Evaluación Educativa*, 29(2). <https://doi.org/10.30827/relieve.v29i2.29134>
- García-Peñalvo, F. J., Llorens-Largo, F., & Vidal, J. (2024). The new reality of education in the face of advances in generative artificial intelligence. *RIED. Revista Iberoamericana de Educación a Distancia*, 27(1), 11–30. <https://doi.org/10.5944/ried.27.1.37716>
- Gouia-Zarrad, R.; C. Gunn (2024). Enhancing students' learning experience in mathematics class through ChatGPT. *International Electronic Journal of Mathematics Education*, 19(3), em0781. <https://doi.org/10.29333/iejme/14614>
- Lloret-Segura, S., Ferreres-Traver, A., Hernández-Baeza, A., & Tomás-Marco, I. (2014).** El análisis factorial exploratorio de los ítems: una guía práctica, revisada y actualizada. *Anales de Psicología*, 30(3), 1151–1169. <https://doi.org/10.6018/analesps.30.3.199361>
- Muñiz, J., & Fonseca-Pedrero, E. (2019). Diez pasos para la construcción de un test. *Psicothema*, 31(1), 7–16. <https://doi.org/10.7334/psicothema2018.153>
- Nguyen, T.; N. Lai; Q. Nguyen (2024). Artificial Intelligence (AI) in Education: A Case Study on ChatGPT's Influence on Student Learning Behaviors. *Educational Process: International Journal*, 13(2), 105-121. <https://doi.org/10.22521/edupij.2024.132.7>
- Oviedo, H. C., & Campo-Arias, A. (2005).** Aproximación al uso del coeficiente alfa de Cronbach. *Revista Colombiana de Psiquiatría*, 34(4), 572–580. <https://www.redalyc.org/pdf/806/80634409.pdf>
- Paul, R., & Elder, L. (2008). *The miniature guide to critical thinking: Concepts and tools*. Foundation for Critical Thinking Press.

- Ríos Hernández, I. N., Mateus, J. C., Rivera, D., & Ávila, L. (2024). *Percepciones de estudiantes latinoamericanos sobre el uso de la inteligencia artificial en la educación superior*. *Austral Comunicación*, 13(1), e01302. <https://doi.org/10.26422/aucom.2024.1301.rio>
- Ruiz-Rojas, L.; L. Salvador-Ullauri; P. Acosta-Vargas (2024). Collaborative Working and Critical Thinking: Adoption of Generative Artificial Intelligence Tools in Higher Education. *Sustainability*, 16(13), 5367. <https://doi.org/10.3390/su16135367>
- Shanto, S. S., Ahmed, Z., & Islam, A. (2024). Enriching the learning process with generative AI: A proposed framework to cultivate critical thinking in higher education using ChatGPT. *Journal of Propulsion Technology*. <https://www.researchgate.net/publication/377931524>
- Walczak, K., & Cellary, W. (2023). *Challenges for higher education in the era of widespread access to Generative AI*. *Economics and Business Review*, 9(2), 71–100. <https://doi.org/10.18559/ebr.2023.2.743>