Revista Interamericana de Psicología Ocupacional

Artículo de Investigación

Desarrollo y Validación de una Escala para Medir las Actitudes Hacia el Uso de la Inteligencia Artificial en la Selección de Personal

Rafael A. Blanco-Rovira[1*], Antonio Zapata Solá[1] & Carlos A. Andújar Rojas[2]

[1]Universidad Albizu, Recinto de San Juan (Puerto Rico)

[2]Universidad de Puerto Rico, Recinto de Arecibo (Puerto Rico)

*Corrrespondencia: rb0230039@sju.albizu.edu

Información Artículo

Recibido: 1-VIII-2026
Aceptado: 10-XI-2026

Palabras Clave:

Inteligencia Artificial, Selección de Personal, Modelamiento de Ecuaciones Estructurales Exploratorio (ESEM), Modelo Bifactor, Invarianza

RESUMEN

El presente estudio desarrolló y aportó evidencia inicial de validez y confiabilidad de la Escala de Actitudes hacia el Uso de la Inteligencia Artificial en los Procesos de Selección de Personal (ACT-IA), un instrumento fundamentado en el modelo tripartito de las actitudes (cognitivo, afectivo y conductual). Participaron 236 profesionales y estudiantes de Recursos Humanos y Psicología Industrial/Organizacional de Puerto Rico. La validez de contenido se evaluó mediante juicio de expertos con el método de Lawshe modificado por Tristán-López; la estructura interna se examinó con análisis factorial exploratorio y, posteriormente, mediante la comparación de modelos confirmatorios competitivos (unidimensional, tres factores oblicuos, bifactor-CFA y ESEM) estimados con un procedimiento robusto para datos ordinales. La versión final de 19 reactivos quedó mejor representada por un modelo bifactor con un factor general dominante (ECV = .731; ωH = .873) acompañado de facetas afectiva y conductual con varianza específica moderada; en la faceta afectiva, compuesta exclusivamente por reactivos de redacción inversa, esta varianza podría reflejar en parte un efecto de método. El modelo de tres factores correlacionados mostró invarianza configural, métrica y escalar por género; la escala presentó, además, consistencia interna excelente (α y ω entre .943 y .966) y una influencia mínima de la deseabilidad social. Se concluye que la puntuación total de la ACT-IA constituye el indicador más robusto del constructo y no se recomienda interpretar de forma independiente la subescala cognitiva. Dado que la depuración de reactivos y los análisis confirmatorios se realizaron con la misma muestra, estos hallazgos deben considerarse evidencia psicométrica inicial, pendiente de replicación en una muestra independiente.

Development and Validation of a Scale to Measure Attitudes Toward the Use of Artificial Intelligence in Personnel Selection

Keywords:

Artificial Intelligence, Personnel Selection, Exploratory Structural Equation Modeling (ESEM), Invariance

ABSTRACT

The present study developed and provided initial evidence of the validity and reliability of the Attitudes Toward the Use of Artificial Intelligence in Personnel Selection Scale (ACT-IA), an instrument grounded in the tripartite model of attitudes (cognitive, affective, and behavioral). Participants were 236 human resource professionals and students of Human Resource Management and Industrial/Organizational Psychology from Puerto Rico. Content validity was evaluated through expert judgment using Tristán-López's modification of Lawshe's content validity method. The internal structure was examined using exploratory factor analysis, followed by a comparison of competing confirmatory measurement models (unidimensional, three-factor oblique CFA, bifactor CFA, and ESEM) estimated with a robust procedure for ordinal data. The final 19-item version was best represented by a bifactor model characterized by a dominant general factor (ECV = .731; ωH = .873) accompanied by affective and behavioral specific factors with moderate unique variance; for the affective factor, composed exclusively of reverse-worded items, this variance may partly reflect a method effect. The correlated three-factor model demonstrated configural, metric, and scalar measurement invariance across gender; the scale also showed excellent internal consistency (α and ω ranging from .943 to .966), and minimal influence of social desirability. Overall, the findings support the use of the ACT-IA total score as the most robust indicator of the underlying construct, and independent interpretation of the cognitive subscale is not recommended. Because item purification and confirmatory analyses were conducted on the same sample, these findings should be regarded as initial psychometric evidence pending replication in an independent sample.

Citar como: Blanco-Rovira, R.A., Zapata Solá, A. & Andújar Rojas, C.A. (2026). Desarrollo y validación de una escala para medir las actitudes hacia el uso de la inteligencia artificial en la selección de personal. *Revista Interamericana de Psicología Ocupacional, 41*(1), 48-66. https://doi.org/10.71332/jj8kd256

Introducción

La inteligencia artificial (IA) se ha consolidado como un factor de transformación estratégica en las organizaciones y, de manera particular, en la función de recursos humanos (RR. HH.), donde su integración ha abierto un debate sobre el equilibrio entre la automatización y el juicio humano en decisiones tan delicadas como la selección de personal (Mariani et al., 2023; Mori et al., 2024). Se estima que entre el 35 % y el 45 % de las organizaciones ya han incorporado herramientas de IA en sus procesos de contratación, con beneficios documentados como reducciones de hasta un 30 % en el costo por contratación y una aceleración significativa de los tiempos de selección (SHRM, 2023). A través de algoritmos de aprendizaje automático, estas herramientas analizan currículos, evalúan entrevistas grabadas y clasifican candidatos según modelos predictivos (Islam et al., 2024; Qin et al., 2023).

No obstante, esta eficiencia no está exenta de riesgos. La falta de transparencia de los algoritmos puede reforzar sesgos preexistentes cuando los datos históricos de entrenamiento están contaminados por decisiones discriminatorias del pasado (Bennion, 2023; Sonderling & Kelley, 2023; Wilson & Caliskan, 2024). Por ello, la Sociedad de Psicología Industrial y Organizacional (SIOP) ha enfatizado la necesidad de validar estos sistemas conforme a los Principles for the Validation and Use of Personnel Selection Procedures (SIOP, 2023). Investigaciones recientes muestran, además, que la percepción de justicia de los candidatos resulta determinante para la aceptación de los sistemas automatizados: la opacidad algorítmica y la sensación de ser evaluado por una "caja negra" disminuyen la confianza y la intención de participar en procesos mediados por IA (Luo et al., 2025; Starke et al., 2022; Zhou et al., 2021).

Esta ambivalencia, en la que se valora la eficiencia, pero se desconfía de la equidad del sistema, sitúa a las actitudes hacia la IA como una variable psicológica central para comprender su adopción. Los modelos generales de aceptación tecnológica, como el Technology Acceptance Model (Davis, 1989) y la Unified Theory of Acceptance and Use of Technology (Venkatesh et al., 2003), explican la intención de uso a partir de la utilidad y la facilidad percibidas, pero no capturan la totalidad de los elementos emocionales, éticos y profesionales involucrados en contextos organizacionales reales (Islam et al., 2024). Las actitudes, entendidas desde el modelo tripartito como la integración de componentes cognitivo, afectivo y conductual (Breckler, 1984; Eagly & Chaiken, 1993; Rosenberg & Hovland, 1960), ofrecen un marco más completo: permiten distinguir entre lo que las personas piensan sobre la eficacia y la justicia de la IA, lo que sienten ante la automatización de las decisiones y su disposición a utilizarla. No obstante, al reflejar distintas facetas de una misma actitud hacia un mismo objeto (la IA en la selección de personal), cabría esperar que estos tres componentes estuvieran sustancialmente intercorrelacionados y compartieran una fuente común de varianza, lo cual anticipa conceptualmente la posibilidad de una estructura jerárquica, más que estrictamente tridimensional, como la explorada posteriormente mediante el modelo bifactor.

En los últimos años se han desarrollado instrumentos para medir actitudes generales hacia la IA, como la General Attitudes towards Artificial Intelligence Scale (Schepman & Rodway, 2020, 2023), validada en italiano por Cicero et al. (2025); la ATTARI-WHE (Gnambs et al., 2025); la escala de actitudes hacia la IA en el trabajo de Park et al. (2024); y, en población hispanohablante, la EVAIA-1 (Subía-Arellano et al., 2023). Estos instrumentos, sin embargo, evalúan actitudes hacia la IA en términos amplios y no fueron diseñados para captar las particularidades del contexto de selección de personal, donde confluyen la confianza en los algoritmos, la percepción de justicia organizacional y el temor a la deshumanización del proceso. A la fecha no se dispone de un instrumento psicométrico construido específicamente para medir las actitudes que generan las aplicaciones de IA en los procesos de reclutamiento y selección, y menos aún validado en población puertorriqueña. Estos elementos, confianza algorítmica, justicia percibida en el proceso de selección y temor a la deshumanización, no son abordados por escalas generales como la EVAIA-1 (Subía-Arellano et al., 2023) y, a diferencia de las actitudes generales que estas miden, son inherentes a una decisión que afecta directamente el acceso al empleo de la persona evaluada, por lo que ameritan un instrumento diseñado específicamente para ese contexto y no una adaptación de medidas de actitud hacia la IA en general.

El Presente Estudio

Para atender esta laguna, el presente estudio tuvo como propósito desarrollar y aportar evidencia inicial de validez y confiabilidad de la Escala de Actitudes hacia el Uso de la Inteligencia Artificial en los Procesos de Selección de Personal (ACT-IA), dirigida a profesionales y estudiantes de RR. HH. y Psicología Industrial/Organizacional (PIO). De manera específica, se buscó (a) generar evidencia de validez de contenido mediante juicio de expertos; (b)

examinar la estructura interna del instrumento a través de análisis factorial exploratorio y la comparación de modelos confirmatorios competitivos; (c) evaluar la confiabilidad y la precisión de las puntuaciones; (d) examinar la validez convergente y discriminante interna de las dimensiones, así como su relación con la deseabilidad social; y (e) examinar la invarianza de medición por género. Dado el carácter emergente del constructo, se prestó especial atención a determinar si las respuestas reflejan tres dimensiones diferenciadas o una actitud general dominante, para lo cual se contrastaron soluciones de tres factores oblicuos, bifactor y exploratorias estructurales (ESEM).

Método

Participantes

Participaron 236 personas mayores de 21 años matriculadas en programas de RR. HH. o PIO, a nivel de bachillerato o posgrado, o que se desempeñaban en funciones de RR. HH./PIO en organizaciones de Puerto Rico. En la Tabla 1 se puede apreciar que la muestra estuvo compuesta predominantemente por mujeres (61.9 %), seguidas por hombres (37.3 %) y un 0.8 % que prefirió no responder. El grupo etario más representado fue el de 31 a 50 años (40.7 %), seguido por el de 21 a 30 años (33.9 %) y el de 51 años o más (25.4 %). La mayoría reportó estudios graduados (40.7 % maestría y 26.3 % doctorado) y la totalidad se encontraba activamente empleada, mayoritariamente a tiempo completo (78.4 %) y en la empresa privada (61.0 %). El tamaño muestral fue consistente con las recomendaciones para el análisis factorial en el desarrollo de escalas (Boateng et al., 2018; Hair et al., 2019; Kline, 2016; MacCallum et al., 1999). Los criterios de inclusión exigían ser mayor de 21 años, residir y trabajar en Puerto Rico, saber leer y escribir en español, contar con al menos un año de experiencia laboral, disponer de un dispositivo con acceso a internet y estar vinculado académica o profesionalmente a RR. HH./PIO. La selección fue no probabilística por conveniencia, con apoyo de muestreo por redes (snowball).

Instrumentos

Hoja de Datos Sociodemográficos

Se administró una hoja de datos sociodemográficos que recopiló edad, género, nivel educativo, estatus laboral, tiempo en la organización, lugar de empleo y estado civil, y que operó además como filtro de elegibilidad para verificar el cumplimiento de los criterios de inclusión.

Tabla 1
Características sociodemográficas de los participantes

Variable	*f*	%
Género		
Femenino	146	61.9
Masculino	88	37.3
Prefirió no responder	2	0.8
Edad		
21–30 años	80	33.9
31–50 años	96	40.7
51 años o más	60	25.4
Preparación Académica		
Cursos universitarios	26	11.0
Bachillerato	47	19.9
Maestría	96	40.7
Doctorado	62	26.3
Otro	4	1.7
Estatus Laboral		
Tiempo completo	185	78.4
Tiempo parcial	51	21.6
Lugar de Empleo		
Empresa privada	144	61.0
Gobierno	46	19.5
Cuenta propia	34	14.4
ONG / Sin fines de lucro	7	3.0
Federal	3	1.3
Estado Civil		
Soltero(a)	112	47.5
Casado(a)	84	35.6
Divorciado(a)	32	13.6
Viudo(a)	1	0.4

Nota. n = 236. Algunas variables presentan casos con respuesta omitida.

Escala de Actitudes hacia el Uso de la IA en la Selección de Personal (ACT-IA)

El instrumento principal fue una escala de nueva creación diseñada para evaluar las actitudes hacia el uso de la IA en los procesos de selección de personal. La versión inicial estuvo compuesta por 45 reactivos en formato Likert de seis puntos, con anclajes desde 1 = “Totalmente en desacuerdo” hasta 6 = “Totalmente de acuerdo”. Su estructura se fundamentó en el modelo tripartito de las actitudes (Breckler, 1984; Eagly & Chaiken, 1993; Rosenberg & Hovland, 1960): la dimensión cognitiva recoge las creencias y evaluaciones racionales sobre la eficacia, utilidad y equidad de la IA; la dimensión afectiva aborda las respuestas emocionales (confianza, incomodidad, entusiasmo o escepticismo) ante la automatización; y la dimensión conductual examina la intención o disposición a utilizar, apoyar o participar en procesos

de selección mediados por IA. Dado que se trataba de un instrumento de nueva creación sin evidencia psicométrica previa en Puerto Rico, sus propiedades se estimaron en el presente estudio.

Escala de Deseabilidad Social – Versión Breve (EDS-VB)

Se utilizó la EDS-VB, una medida de cuatro reactivos en formato Likert de seis puntos diseñada para evaluar la tendencia a responder de manera socialmente deseable. Esta versión abreviada constituye una adaptación de la escala original de Rosario-Hernández y Rovira-Millán (2002) y fue validada recientemente con una muestra de 3,855 adultos en Puerto Rico, mostrando índices de ajuste sobresalientes que apoyaron su estructura unidimensional y la confiabilidad de la EDS-VB fue excelente mediante los coeficientes alfa de Cronbach y omega de McDonald fueron de .926 y .927, respectivamente (Rosario-Hernández et al., 2025). Su inclusión permitió examinar la validez divergente del nuevo instrumento.

Procedimiento

El reclutamiento se realizó mediante un anuncio electrónico difundido en plataformas digitales y redes sociales profesionales (Facebook, Instagram y LinkedIn) y a través de redes académicas de programas universitarios de RR. HH. y PIO, utilizando un enlace directo y un código QR. Todos los procedimientos fueron aprobados por la Junta de Revisión Institucional (IRB) de la Universidad Albizu. Al acceder al enlace, los participantes fueron dirigidos a la plataforma PsychData, donde se les presentó primero la hoja de consentimiento informado, que detallaba la naturaleza voluntaria de la participación, los riesgos mínimos, las garantías de confidencialidad y anonimato y la información de contacto. Solo quienes aceptaron participar continuaron con la hoja de datos sociodemográficos, que verificaba la elegibilidad; los participantes que no cumplían algún criterio fueron excluidos automáticamente. A continuación, los participantes cualificados completaron la ACT-IA y, finalmente, la EDS-VB. La participación tomó entre 10 y 15 minutos. No se recopiló información que permitiera identificar a los participantes y los datos se almacenaron de forma segura en PsychData.

Diseño y Análisis de Datos

El estudio empleó un diseño cuantitativo psicométrico orientado al desarrollo y la validación de un instrumento de medición (Muñiz, 2018). La validez de contenido se evaluó mediante un panel de tres jueces expertos en PIO y en construcción de pruebas, utilizando el método de Lawshe (1975) modificado por Tristán-López (2008); se retuvieron los reactivos que alcanzaron una razón de validez de contenido (RVC) igual o superior a .58. Se calcularon estadísticas descriptivas e índices de discriminación corregidos (rbis) por reactivo, adoptando .30 como criterio mínimo de discriminación (DeVellis & Thorpe, 2021; Spector, 1992).

La estructura interna se examinó en dos fases. Primero se condujeron análisis factoriales exploratorios (AFE) sobre la matriz de correlaciones de Pearson, aproximación razonable con reactivos de seis opciones de respuesta (Rhemtulla et al., 2012), usando la factorización de ejes principales como método de extracción y la técnica de rotación oblicua "Promax," la cual permite que los factores covarien como lo son las dimensiones de las actitudes (Fabrigar et al., 1999; Worthington & Whittaker, 2006); el número de factores se determinó mediante el criterio de Kaiser, el gráfico de sedimentación y la proporción de varianza explicada. Dado que el criterio de Kaiser y el gráfico de sedimentación pueden sobreestimar el número de factores a retener, esta decisión se complementó con un análisis paralelo (Horn, 1965; Hayton et al., 2004), procedimiento recomendado como criterio más robusto de retención factorial. El análisis paralelo se condujo sobre la misma matriz de correlaciones de Pearson de los 44 reactivos empleada en el AFE, generando 1,000 matrices de datos aleatorios de igual dimensión (n = 236) y comparando los autovalores empíricos con el promedio y el percentil 95 de los autovalores simulados, tanto en su versión de componentes principales como de factor común (ejes principales); como análisis de sensibilidad, el procedimiento se repitió sobre la matriz de correlaciones policóricas.

Posteriormente, sobre la versión depurada, se compararon cuatro modelos de medición competitivos: (a) unidimensional, (b) tres factores oblicuos mediante análisis factorial confirmatorio (CFA), (c) bifactor confirmatorio (bifactor-CFA), con un factor general y tres factores específicos no correlacionados (Reise, 2012; Rodriguez et al., 2016), y (d) un modelo exploratorio de ecuaciones estructurales (ESEM) de tres factores con rotación oblicua geomin, que permite cargas cruzadas estimadas en lugar de fijarlas en cero (Asparouhov & Muthén, 2009; Marsh et al., 2014; Morin et al., 2020). Dada la naturaleza ordinal de los reactivos, los modelos se estimaron con un procedimiento robusto de mínimos cuadrados ponderados (WLSMV/DWLS) basado en correlaciones policóricas, apropiado para indicadores categóricos (Brown, 2015), utilizando el paquete lavaan del programa R (Rosseel et al., 2025).

A diferencia de la práctica tradicional de interpretar la bondad de ajuste mediante puntos de corte fijos (e.g., Hu & Bentler, 1999), en el presente estudio los índices de ajuste se interpretaron utilizando el enfoque basado en percentiles propuesto por Howard et al. (2025). Este enfoque sitúa cada índice de ajuste (CFI, TLI, RMSEA y SRMR) dentro de la distribución empírica de valores observados en estudios publicados de análisis factorial, permitiendo clasificar el ajuste en cinco categorías ordinales delimitadas por los percentiles 10, 33, 66 y 90 de dicha distribución: Muy Débil (< P10), Débil (P10–P33), Moderado (P33–P66), Fuerte (P66–P90) y Muy Fuerte (> P90). En este marco, valores más altos de CFI y TLI, así como valores más bajos de RMSEA y SRMR, corresponden a percentiles superiores y reflejan un ajuste relativamente más favorable en comparación con la literatura disponible. A diferencia de los puntos de corte convencionales (p. ej., CFI ≥ .95; Hu & Bentler, 1999), que son fijos, Howard et al. (2025) reportan estos percentiles tanto para el conjunto de los análisis factoriales confirmatorios publicados como separados por terciles de grados de libertad (gl < 51, 51–163 y > 163), dado que los índices de ajuste, en particular los incrementales, son sensibles a los gl del modelo. Como los cuatro modelos comparados en este estudio tuvieron entre 117 y 152 gl, se emplearon los percentiles del tercil intermedio (Howard et al., 2025, Tabla 4). Los valores correspondientes a los percentiles 10, 33, 66 y 90 fueron .90, .94, .96 y .98 para el CFI; .89, .93, .96 y .98 para el TLI; .10, .08, .06 y .04 para el RMSEA; y .08, .06, .05 y .03 para el SRMR. Así, por ejemplo, un CFI = .948 —aceptable bajo los criterios de Hu y Bentler (1999)— se ubica en la banda Moderada, mientras que un RMSEA superior a .10 se clasifica como Muy Débil. Este procedimiento proporciona una evaluación más continua y contextualizada del ajuste que la interpretación dicotómica basada en puntos de corte convencionales. No obstante, este marco tiene una limitación relevante para el presente estudio: los percentiles de Howard et al. (2025) provienen de análisis factoriales confirmatorios publicados en revistas de ciencia organizacional y no se presentan diferenciados por estimador ni por tipo de indicador (continuo u ordinal), mientras que los índices obtenidos con DWLS/WLSMV no son directamente comparables con los de máxima verosimilitud (Xia & Yang, 2019). Por ello, las clasificaciones resultantes se emplearon como referencia orientativa para comparar los modelos entre sí, y no como juicios absolutos sobre su ajuste.

Para el modelo bifactor se calcularon los índices auxiliares de la varianza común explicada (ECV), la proporción de correlaciones no contaminadas (PUC), el omega total (ωTG), el omega jerárquico general (ωH) y específico (ωHS), la replicabilidad del constructo (H) y la determinación factorial (FD), siguiendo las recomendaciones para evaluar la dimensionalidad y la fortaleza del factor general (Reise et al., 2013; Rodriguez et al., 2016; Zhang et al., 2021). Estos índices se calcularon a partir de las cargas estandarizadas del modelo bifactor-CFA mediante el paquete de R BifactorIndicesCalculator (Dueber, 2021).

La validez convergente y discriminante interna se examinó mediante la VME y el criterio de Fornell y Larcker (1981), y la relación con la deseabilidad social, mediante las correlaciones entre la puntuación total, las dimensiones y la EDS-VB. La invarianza de medición por género se evaluó de forma secuencial (configural, métrica y escalar) con el estimador WLSMV, adoptando los criterios de cambio de ΔCFI, ΔTLI, ΔRMSEA y ΔSRMR recomendados por Chen (2007) y Putnick y Bornstein (2016). Finalmente, la confiabilidad se estimó con el alfa de Cronbach y el omega de McDonald, y la precisión se evaluó con el error estándar de medición (EEM), calculado como $EEM = DE\sqrt{(1 - \omega)}$.

Resultados

Se presentan, en primer lugar, la evidencia de validez de contenido, el análisis de los reactivos, la estructura interna (AFE y comparación de modelos confirmatorios), los índices bifactoriales, la validez convergente y discriminante interna y la relación con la deseabilidad social, la invarianza de medición por género y la confiabilidad. Todos los análisis se realizaron con la muestra de 236 participantes.

Validez de Contenido

El conjunto de reactivos fue sometido a un panel de tres jueces expertos en PIO y en el desarrollo de pruebas. Los resultados del análisis de validez de contenido indicaron que los 45 reactivos alcanzaron niveles adecuados de pertinencia de acuerdo con el índice de razón de validez de contenido (RVC) calculado mediante la modificación del método de Lawshe propuesta por Tristán-López (2008). Cuarenta y cuatro reactivos obtuvieron un RVC de 1.00, reflejando consenso unánime entre los jueces acerca de su pertinencia para representar las dimensiones cognitiva, afectiva y conductual del constructo. Únicamente el ítem 6, correspondiente a la dimensión cognitiva, obtuvo un RVC de .67, debido a que uno de los jueces no lo consideró pertinente. No obstante, este valor superó el criterio mínimo de aceptación establecido para el número de jueces participantes

(RVC ≥ .58), por lo que el reactivo fue retenido para las fases posteriores de evaluación psicométrica.

De manera complementaria, el índice de validez de contenido preliminar fue de .978 para la dimensión cognitiva, 1.00 para las dimensiones afectiva y conductual, y .993 para la escala en su conjunto, lo que evidencia un elevado grado de acuerdo entre los jueces respecto a la pertinencia y representatividad del contenido del instrumento. Una vez establecida la validez de contenido de la versión inicial de la ACT-IA, se procedió a examinar el funcionamiento empírico de los reactivos mediante el análisis de sus estadísticas descriptivas y del índice de discriminación, con el propósito de evaluar la calidad psicométrica de cada reactivo antes de realizar los análisis de la estructura interna del instrumento.

Análisis Descriptivo y Discriminación de los Reactivos

Las Tablas 2, 3, y 4 presentan las estadísticas descriptivas y los índices de discriminación corregidos de los reactivos de las dimensiones cognitiva, afectiva y conductual. Las medias reflejaron una tendencia moderadamente favorable hacia el uso de la IA en la selección de personal, y las desviaciones estándar evidenciaron variabilidad adecuada. La mayoría de los reactivos superó ampliamente el criterio de discriminación de .30; los coeficientes más elevados se observaron en la dimensión conductual. En la dimensión cognitiva, el reactivo 11 mostró una correlación negativa (r_{bis} = −.499), lo que justificó su eliminación durante la depuración.

Tabla 2

Estadística descriptiva e índice de discriminación (rbis) de la dimensión cognitiva

Ítem	*Media*	*DE*	r_{bis}
1	4.55	1.47	.854
2	4.30	1.55	.800
3	4.43	1.49	.873
4	4.25	1.55	.852
5	4.40	1.50	.865
6	4.06	1.72	.832
7	4.31	1.51	.846
8	3.50	1.88	.548
9	4.22	1.55	.839
10	4.14	1.57	.851
11*	4.44	1.87	−.499
12	5.11	1.31	.554
13	4.58	1.39	.739
14	4.12	1.63	.506
15	3.51	1.94	.569

Nota. n = 236. * Reactivo eliminado. DE = desviación estándar.

Tabla 3

Estadística descriptiva e índice de discriminación (rbis) de la dimensión afectiva

Ítem	*Media*	*DE*	*rbis*
16	4.26	1.69	.834
17	4.08	1.63	.868
18	4.75	1.44	.707
19	4.17	1.70	.844
20	4.99	1.42	.575
21	4.12	1.68	.861
22	4.42	1.54	.817
23	4.61	1.53	.738
24	4.00	1.72	.810
25	3.74	1.84	.864
26	3.83	1.91	.830
27	3.79	1.93	.862
28	2.59	1.41	.450
29	3.67	1.92	.809
30	3.54	2.00	.798

Nota. n = 236. DE = desviación estándar.

Tabla 4

Estadística descriptiva e índice de discriminación (rbis) de la dimensión conductual

Ítem	*Media*	*DE*	*rbis*
31	4.67	1.43	.889
32	4.54	1.52	.880
33	4.53	1.51	.903
34	4.95	1.38	.802
35	4.59	1.56	.896
36	4.58	1.57	.892
37	4.81	1.46	.858
38	5.14	1.32	.702
39	5.19	1.29	.641
40	5.05	1.47	.709
41	4.19	1.74	.729
42	4.23	1.72	.778
43	3.84	1.81	.714
44	4.24	1.68	.695
45	3.87	1.88	.645

Nota. n = 236. DE = desviación estándar.

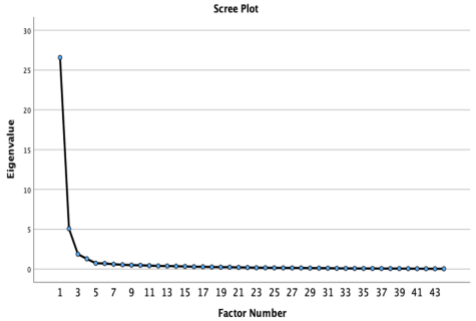
Estructura Interna: Análisis Factorial Exploratorio

En la primera fase se realizó un análisis factorial exploratorio (AFE) sobre los 44 reactivos utilizando el método de factorización de ejes principales con rotación oblicua Promax. El criterio de Kaiser identificó cuatro factores con autovalores superiores a 1.00, los cuales explicaron conjuntamente el 76.81 % de la varianza total. El primer factor explicó el 59.93 % de la varianza, mientras que los factores restantes

aportaron porcentajes considerablemente menores. Asimismo, el gráfico de sedimentación mostró un punto de inflexión consistente con una solución de cuatro factores (véase la Figura 1), respaldando la retención de dicha estructura factorial. No obstante, el análisis paralelo sugirió una solución más parsimoniosa de tres factores. En su versión de componentes principales (Horn, 1965), directamente comparable con el criterio de Kaiser, los autovalores empíricos superaron a los autovalores promedio simulados en los tres primeros factores (autovalores empíricos = 26.58, 5.07 y 1.87; autovalores simulados = 1.95, 1.84 y 1.76), pero no en el cuarto (1.28 frente a 1.69). En su versión de factor común, el cuarto autovalor empírico (0.757) superó apenas el promedio simulado (0.707), pero no el percentil 95 (0.761), y la repetición del análisis sobre la matriz policórica indicó igualmente tres factores (cuarto autovalor empírico = 0.596; promedio simulado = 0.708). Esta discrepancia es consistente con la tendencia del criterio de Kaiser a sobreestimar el número de factores (Hayton et al., 2004) y con el hecho de que el cuarto factor explicó apenas el 2.21 % de la varianza y estuvo definido principalmente por cargas cruzadas (reactivos 12, 13 y 39), con un único reactivo con carga primaria en él (reactivo 14; véase Tabla 5).

En consecuencia, la solución de cuatro factores se utilizó como punto de partida para identificar reactivos con funcionamiento problemático durante la depuración, mientras que el número de factores sugerido por el análisis paralelo coincide con el número de dimensiones del modelo teórico. Cabe precisar que esta coincidencia se limita al número de factores y no a su composición: en la solución inicial, los factores se organizaron en buena medida según la dirección de la redacción (los reactivos inversos de las tres dimensiones frente a los reactivos directos) más que según las dimensiones cognitiva, afectiva y conductual (véase la Tabla S1).

Figura 1. Gráfico de sedimentación del primer análisis de factores exploratorio.



Aunque la mayoría de los reactivos presentó sus cargas factoriales más elevadas en los factores teóricamente esperados, varios exhibieron cargas cruzadas, saturaciones en factores no previstos o patrones de carga difíciles de interpretar, lo que sugirió la necesidad de depurar la escala con el propósito de obtener una estructura factorial más simple, parsimoniosa e interpretable (véase Tabla 5). En particular, varios reactivos mostraron cargas cruzadas de magnitud considerable (e.g., reactivos 12, 13, 14, 18, 22, 23, 28, 31–36, 39 y 42), mientras que otros cargaron en factores distintos a los previstos conceptualmente, dificultando la obtención de una estructura factorial simple e interpretable.

Específicamente, se consideró la magnitud de las cargas factoriales, la presencia de cargas cruzadas, la coherencia teórica de cada reactivo con su dimensión propuesta y la interpretabilidad de la solución factorial. En la dimensión conductual, este criterio implicó privilegiar la simplicidad estructural sobre la sola magnitud de la discriminación: los reactivos 31, 32, 33, 35 y 36 mostraron índices de discriminación muy elevados (r_{bis} entre .880 y .903; véanse Tablas 4 y 5) pero presentaron cargas cruzadas relevantes en el primer AFE, por lo que fueron excluidos en favor de los reactivos 34, 37, 38, 39 y 40, cuya discriminación fue algo menor pero cuyo patrón de cargas resultó más simple e interpretable.

Tabla 5

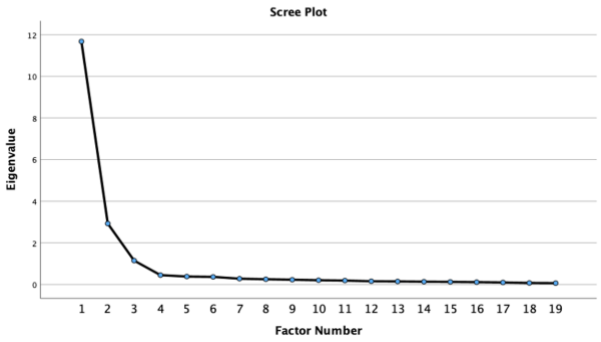
Primer análisis de factores exploratorio de los reactivos de la Escala de Actitudes hacia el Uso de la Inteligencia Artificial en los Procesos de Selección de Personal

Dimensión	Ítem	Factor				h^2
		1	2	3	4	
Cognitiva	1		.320	.656		.797
	2			**.779**		.669
	3			**.863**		.788
	4			**.806**		.771
	5			**.702**		.781
	6			**.957**		.787
	7			**.775**		.744
	8	.741				.599
	9			**.747**		.738
	10			**.915**		.836
	12		.667		.370	.585
	13		.343	.485	.427	.669
	14			.370	.546	.469
	15	.832				.736
Afectiva	16			.725		.783
	17			.867		.874
	18		.639	.349		.814
	19			.820		.830
	20		.840			.763
	21			.810		.850
	22		.301	.562		.792
	23		.428	.458		.726
	24	**.799**				.762
	25	**.826**				.863
	26	**.822**				.785
	27	**.839**				.851
	28	.594	−.432			.411
	29	**.863**				.792
	30	**.868**				.803
Conductual	31		.563	.402		.864
	32		.473	.478		.871
	33		.458	.452		.883
	34		**.858**			.810
	35		.455	.475		.919
	36		.501	.446		.902
	37		**.694**			.830
	38		**1.08**			.856
	39		**1.05**		.307	.828
	40		**.932**			.727
	41	.828				.739
	42	.819	.385			.779
	43	.853				.753
	44	.773	.324			.633
	45	.894				.734
Autovalor inicial		26.58	5.07	1.87	1.28	
Varianza Explicada		59.93	10.97	3.69	2.21	
Varianza Acumulada		59.93	70.90	74.59	76.81	

Nota. n = 236. Autovalor inicial = autovalor de la matriz de correlaciones (criterio de Kaiser); % de varianza = varianza explicada por cada factor tras la extracción por ejes principales.

A partir de los resultados obtenidos en el primer análisis factorial exploratorio, se procedió a evaluar el desempeño psicométrico de los reactivos considerando de manera conjunta la magnitud de las cargas factoriales, la presencia de cargas cruzadas, la coherencia conceptual con la dimensión teórica propuesta y la interpretabilidad de la solución factorial. Este proceso permitió identificar y eliminar aquellos reactivos que no cumplían satisfactoriamente con los criterios establecidos, dando lugar a una versión depurada de la escala compuesta por 19 reactivos (los reactivos eliminados, su patrón de cargas en el primer AFE y el criterio de eliminación de cada uno se presentan en la Tabla S1 del material suplementario). Posteriormente, se realizó un segundo análisis factorial exploratorio utilizando el mismo método de extracción y rotación con el propósito de examinar la claridad de la estructura interna de la versión refinada del instrumento. Como se observa en la Figura 2, el gráfico de sedimentación evidenció un punto de inflexión claramente definido después del tercer factor, respaldando la retención de una solución tridimensional en esta muestra.

Figura 2. Gráfico de sedimentación del segundo análisis de factores exploratorio.



Los resultados del segundo análisis factorial exploratorio evidenciaron una estructura simple de tres factores, consistente tanto con el modelo teórico propuesto como con la evidencia empírica obtenida durante el proceso de depuración (véase Tabla 6). En la Tabla 6, F1 corresponde a la dimensión afectiva, F2 a la cognitiva y F3 a la conductual. Los tres factores explicaron conjuntamente el 79.84 % de la varianza total. El primer factor explicó el 60.39 %de la varianza, el segundo aportó un 14.47 % adicional y el tercero explicó el 4.98 % restante. Asimismo, todos los reactivos presentaron cargas factoriales elevadas en la dimensión para la cual fueron diseñados, sin evidenciar cargas cruzadas de magnitud relevante, lo que refleja una estructura factorial simple y fácilmente interpretable. La carga del ítem 38 en el factor conductual superó ligeramente la unidad (λ = 1.035); en soluciones con rotación oblicua como Promax, las cargas del patrón son coeficientes de regresión parcial y no correlaciones, por lo que pueden exceder 1.00 sin implicar una solución impropia, y su comunalidad (.917) se mantuvo por debajo de 1.00. Las comunalidades oscilaron entre .697 y .917, indicando que una proporción sustancial de la varianza de cada reactivo fue explicada por los factores retenidos. Por otra parte, las correlaciones interfactoriales (r_{12} = .701; r_{13} = .357; r_{23} = .646) fueron positivas y de magnitud moderada a alta, justificando el uso de una rotación oblicua y sugiriendo que las dimensiones cognitiva, afectiva y conductual representan componentes diferenciados, aunque relacionados, de las actitudes hacia el uso de la inteligencia artificial en los procesos de selección de personal.

En conclusión, en esta muestra la versión depurada de 19 reactivos mostró una estructura interna clara e interpretable, con un número de factores consistente con el modelo teórico propuesto, que se examinó a continuación mediante la comparación de modelos confirmatorios y de ecuaciones estructurales exploratorias estimados con la misma muestra. Conviene señalar, no obstante, que esta estructura es, en parte, consecuencia del propio proceso de depuración: los reactivos retenidos en las dimensiones cognitiva y conductual son de redacción directa, mientras que los de la dimensión afectiva son exclusivamente de redacción inversa, por lo que la separación entre factores coincide también con la dirección de la redacción (véanse la Tabla S1 y la sección de Limitaciones).

Tabla 6

Segundo análisis factorial exploratorio de la versión depurada (19 reactivos) de la ACT-IA

Dim.	*Ítem*	*F₁*	*F₂*	*F₃*	*h²*
Cogn.	2		**.877**		.697
	3		**.959**		.819
	4		**.865**		.815
	5		**.710**		.799
	6		**.847**		.759
	7		**.748**		.733
	9		**.769**		.727
	10		**.829**		.830
Afec.	24	**.798**			.747
	25	**.867**			.871
	26	**.967**			.848
	27	**.910**			.881
	29	**.917**			.821
	30	**.919**			.806
Cond.	34			**.763**	.762
	37			**.680**	.789
	38			**1.035**	.917
	39			**.949**	.827
	40			**.833**	.722
Autovalor inicial		11.67	2.92	1.15	
% varianza		60.39	14.47	4.98	
% acumulado		60.39	74.86	79.84	

Nota. n = 236. Factorización de ejes principales con rotación Promax. F1 = afectiva; F2 = cognitiva; F3 = conductual. Correlaciones interfactoriales: r_{12} = .701; r_{13} = .357; r_{23} = .646. h^2 = comunalidad. Autovalor inicial = autovalor de la matriz de correlaciones (criterio de Kaiser); % de varianza = varianza explicada por cada factor tras la extracción por ejes principales.

Comparación de Modelos Competitivos

La Tabla 7 resume los cuatro modelos competitivos, interpretados conforme a los puntos de corte basados en percentiles de Howard et al. (2025). El modelo unidimensional mostró el ajuste menos favorable: aunque su CFI (.948) y TLI (.941) se ubicaron en la banda Moderada, su RMSEA (.192) y SRMR (.190) fueron muy superiores a los valores del percentil 10 (.10 y .08, respectivamente), es decir, menos favorables que los de nueve de cada diez modelos de referencia (Muy Débil), lo que indica que una sola dimensión no representa adecuadamente las respuestas.

El modelo de tres factores oblicuos (CFA) mejoró sustancialmente los índices incrementales, con un CFI (.980) en el límite de la banda Muy Fuerte y un TLI (.976) Fuerte, pero su SRMR (.063) se ubicó en la banda Débil y su RMSEA (.121) en la Muy Débil, evidenciando una aproximación todavía deficiente del error. El modelo bifactor-CFA exhibió el mejor desempeño global: el CFI (.993) y el TLI (.991) superaron el percentil 90 (Muy Fuerte), el SRMR (.045) se ubicó en la banda Fuerte y el RMSEA (.073) en la Moderada, configurando el ajuste más equilibrado del conjunto. El modelo ESEM también resultó competitivo, con CFI (.989), TLI (.984) y SRMR (.021) en la banda Muy Fuerte, aunque su RMSEA (.101) se ubicó en el límite de la banda Muy Débil.

Los modelos bifactor-ESEM y jerárquico de segundo orden no convergieron en soluciones admisibles, dificultad frecuente en estructuras altamente parametrizadas con indicadores ordinales y muestras moderadas (Eid et al., 2017; Morin et al., 2020). En conclusión, la evidencia favoreció la solución bifactorial confirmatoria, indicando un factor general dominante acompañado de facetas específicas residuales.

Tabla 7

Índices de ajuste de los modelos de medición competitivos de la ACT-IA

Modelo	*χ^2 (gl)*	*SRMR*	*RMSEA [IC 90%]*	*CFI*	*TLI*
Unidimensional	1,465 (152)	.190	.192 [.183, .201]	.948	.941
3 factores oblicuos (CFA)	664 (149)	.063	.121 [.112, .131]	.980	.976
Bifactor-CFA	300 (133)	.045	.073 [.062, .084]	.993	.991
3 factores oblicuos (ESEM)	397 (117)	.021	.101 [.090, .112]	.989	.984

Nota. n = 236. El modelo bifactor-CFA (en negrita) presentó el mejor ajuste. Índices interpretados según los percentiles de Howard et al. (2025, Tabla 4; tercil de 51–163 gl): Muy Débil (< P10), Débil (P10–P33), Mod (P33–P66), Fuerte (P66–P90) y Muy Fuerte (> P90). Clasificación (CFI/TLI/RMSEA/SRMR): unidimension Moderado/Moderado/Muy Débil/Muy Débil; CFA de tres factores = Muy Fuerte/Fuerte/Muy Débil/Débil; bi CFA = Muy Fuerte/Muy Fuerte/Moderado/Fuerte; ESEM = Muy Fuerte/Muy Fuerte/Muy Débil/Muy Fuerte. corresponden a los valores escalados del estimador WLSMV.

Con el propósito de comparar las soluciones factoriales que mostraron el mejor desempeño psicométrico, se presentan los resultados de los tres modelos que obtuvieron los índices de ajuste más favorables durante la evaluación de la estructura interna del instrumento: el modelo confirmatorio de tres factores oblicuos (CFA), el modelo exploratorio de ecuaciones estructurales de tres factores (ESEM) y el modelo bifactor confirmatorio. Mientras que los modelos CFA y ESEM permitieron evaluar la estabilidad y consistencia de la estructura tridimensional propuesta, el modelo bifactor permitió examinar la magnitud de la varianza atribuible a un factor general y determinar la conveniencia de interpretar la ACT-IA mediante una puntuación total además de sus dimensiones específicas. La Tabla 8 presenta las cargas factoriales estandarizadas de los tres modelos, así como la varianza común explicada por ítem (I-ECV), el sesgo relativo absoluto del parámetro (ARPB) y los principales índices bifactoriales utilizados para evaluar el grado de unidimensionalidad esencial del instrumento.

Como puede observarse, las cargas factoriales obtenidas mediante el modelo CFA fueron consistentemente elevadas en las tres dimensiones, con valores comprendidos entre .864 y .986**,** proporcionando evidencia de una adecuada representación de los reactivos por sus respectivos factores latentes. De manera similar, el modelo ESEM reprodujo el mismo patrón factorial, manteniendo cargas principales elevadas y cargas cruzadas de magnitud reducida, lo que resulta consistente con la estructura tridimensional identificada en el AFE con la misma muestra. Cabe señalar que cinco cargas del modelo ESEM superaron ligeramente el valor de 1.00: los ítems 2 y 3 en la dimensión cognitiva (λ = 1.040 y 1.026), los ítems 26 y 27 en la afectiva (λ = 1.071 y 1.008) y el ítem 38 en la conductual (λ = 1.004). En soluciones con rotación oblicua, como la rotación geomin empleada, las cargas estandarizadas son coeficientes de regresión parcial y no correlaciones, por lo que pueden exceder la unidad cuando los factores están correlacionados y el reactivo presenta cargas cruzadas negativas, sin implicar necesariamente una solución impropia (Marsh et al., 2014). Para descartar un caso Heywood se examinó la matriz de varianzas residuales del modelo: todas las varianzas residuales estimadas fueron positivas (rango = .044–.221), incluidas las de estos cinco reactivos (ítem 2 = .221; ítem 3 = .138; ítem 26 = .091; ítem 27 = .059; ítem 38 = .044); además, el modelo convergió sin advertencias y la matriz de correlaciones entre factores fue definida positiva. Por tanto, no hubo evidencia de soluciones impropias, y las cargas superiores a 1.00 se interpretan como un efecto de la rotación oblicua y no como un caso Heywood.

Tabla 8

Cargas factoriales estandarizadas e índices bifactoriales de los modelos CFA de tres factores, ESEM de tres factores y bifactor CFA de la ACT-IA

Dimensión	Ítem	Modelo Tres Factores Oblicuos (CFA)			Modelo Tres Factores Oblicuos (ESEM)			Modelo Bifactor (CFA)					
		Cog.	Afec.	Cond.	Cog.	Afec.	Cond.	FG	Cog.	Afec.	Cond.	I-ECV	ARPB
Cognitiva	ia2	.864			**1.040**	-.192	-.013	.844	.339			.861	.017
	ia3	.914			**1.026**	-.176	.057	.904	.211			.948	.019
	ia4	.929			**.930**	-.065	.078	.915	.251			.930	.011
	ia5	.926			**.760**	.058	.166	.925	.078			.993	.039
	ia6	.912			**.915**	.056	-.075	.917	-.067			.995	.043
	ia7	.890			**.827**	.001	.091	.892	.043			.998	.040
	ia9	.901			**.854**	.050	.008	.907	-.104			.987	.040
	ia10	.946			**.900**	.083	-.037	.957	-.138			.980	.040
Afectiva	ia24		.914		-.019	**.892**	.069	.744		.515		.676	.159
	ia25		.958		.064	**.919**	-.041	.768		.566		.649	.209
	ia26		.946		-.164	**1.071**	.031	.715		.628		.565	.292
	ia27		.971		-.052	**1.008**	.010	.762		.600		.617	.253
	ia29		.935		.013	**.956**	-.075	.717		.604		.585	.263
	ia30		.929		.040	**.963**	-.151	.696		.628		.552	.292
Conductual	ia34			.927	-.015	.271	**.802**	.749			.552	.648	.165
	ia37			.986	.017	.372	**.726**	.821			.480	.746	.080
	ia38			.955	-.045	.007	**1.004**	.600			.775	.375	.553
	ia39			.935	.044	-.086	**.960**	.579			.757	.369	.548
	ia40			.879	.082	.010	**.836**	.628			.637	.493	.288
Índices Bifactoriales													
ECV								.731					
PUC								.690					
ω								.986	.971	.970	.963		
ωH/ωHS								.873	.006	.360	.435		
H								.982	.223	.766	.810		
FD								.991	.809	.965	.975		

Nota. n = 236. CFA = análisis factorial confirmatorio; ESEM = modelado exploratorio de ecuaciones estructurales; FG = factor general; Cog. = cognitiva; Afec. = afectiva; Cond. = conductual; I-ECV = varianza común explicada por el ítem; ARPB = sesgo relativo absoluto del parámetro, calculado como $|\lambda U - \lambda FG| / \lambda FG$, donde λU es la carga en el modelo unidimensional y λFG la carga en el factor general del modelo bifactor. En ESEM se presentan las cargas estandarizadas de los tres factores; las cargas principales se resaltan en negritas. En el modelo bifactor se presentan las cargas del factor general y las cargas específicas por dimensión. ω = coeficiente omega total; ωH = omega jerárquico del factor general; ωHS = omega jerárquico específico; H = replicabilidad del constructo; FD = determinación factorial.

Los resultados del modelo bifactor mostraron que las cargas del factor general fueron elevadas en la mayoría de los reactivos (λ = .579–.957), mientras que las cargas correspondientes a los factores específicos fueron considerablemente menores, particularmente en la dimensión cognitiva, donde varios reactivos presentaron cargas cercanas a cero o incluso negativas una vez controlada la influencia del factor general. En concordancia con este patrón, los valores de I-ECV fueron elevados para la mayoría de los reactivos, especialmente en la dimensión cognitiva, indicando que una proporción importante de la varianza común de dichos reactivos fue explicada por el factor general. El sesgo relativo absoluto del parámetro (ARPB) se calculó como la diferencia absoluta entre la carga de cada reactivo en el modelo unidimensional y su carga en el factor general del modelo bifactor, dividida por esta última (Rodriguez et al., 2016). Los valores fueron mínimos en la dimensión cognitiva (ARPB = .011–.043), lo que indica que tratar estos reactivos como unidimensionales apenas sesga sus cargas, en concordancia con su escasa varianza específica. En cambio, la mayoría de los reactivos de las dimensiones afectiva (ARPB = .159–.292) y conductual (ARPB = .080–.553) superó el umbral de 10–15 % considerado aceptable (Rodriguez et al., 2016), particularmente los ítems 38 y 39 (ARPB = .553 y .548, respectivamente); el ARPB promedio de la escala fue de .176. Este patrón indica que, aunque el factor general es dominante, un modelo estrictamente unidimensional sesgaría de manera apreciable las cargas de los reactivos afectivos y conductuales (el ARPB, al definirse como valor absoluto, informa la magnitud del sesgo, pero no su dirección), lo cual es coherente con la varianza específica que conservan estas dimensiones (véanse los valores de ωHS y H), aspecto que se retoma en la discusión.

Los índices bifactoriales complementaron esta evidencia. El índice de varianza común explicada (ECV = .731) indicó que aproximadamente el 73 % de la varianza común fue atribuible al factor general, mientras que el porcentaje de correlaciones no contaminadas (PUC = .690) reflejó una proporción moderada de covarianzas influenciadas exclusivamente por dicho factor. La confiabilidad del puntaje total fue excelente (ω = .986), y el omega jerárquico del factor general (ωH = .873) indicó que la

mayor parte de la varianza confiable de las puntuaciones puede atribuirse al factor general. En contraste, la dimensión cognitiva presentó una contribución específica prácticamente nula (ωHS = .006**;** H = .223), lo que sugiere que aporta escasa varianza confiable independiente una vez controlado el factor general. Por su parte, las dimensiones afectiva (ωHS = .360**;** H = .766) y conductual (ωHS = .435**;** H = .810) conservaron una cantidad moderada de varianza específica, indicando que mantienen cierto grado de diferenciación conceptual, aunque gran parte de su confiabilidad continúa siendo explicada por el factor general. Finalmente, los elevados índices de determinación factorial del factor general (FD = .991**)** y de las dimensiones afectiva (FD = .965) y conductual (FD = .975) indican que las puntuaciones factoriales estimadas representan adecuadamente los constructos latentes correspondientes.

En conclusión, estos hallazgos indican que, aunque la ACT-IA conserva cierto grado de multidimensionalidad, limitado a las dimensiones afectiva y conductual (la dimensión cognitiva carece prácticamente de varianza específica; ωHS = .006), las actitudes hacia el uso de la inteligencia artificial en los procesos de selección de personal están fuertemente dominadas por un factor general. En consecuencia, los resultados respaldan la utilización de una puntuación total como indicador principal del constructo, sin excluir la posibilidad de interpretar la dimensión conductual y, con las reservas que se discuten más adelante, la afectiva, cuando el interés de la investigación requiera un nivel de análisis más específico.

Validez Convergente y Discriminante Interna y Relación con la Deseabilidad Social

Con el propósito de obtener evidencia de la validez convergente y discriminante de la ACT-IA, se calcularon la varianza media extraída (VME), la máxima varianza compartida (VMC), la varianza promedio compartida (VPC) y las correlaciones entre los factores latentes estimados mediante el modelo CFA de tres factores oblicuos. La Tabla 9 presenta estos resultados de acuerdo con el criterio propuesto por Fornell y Larcker (1981). Dado que el estudio no incluyó una medida externa teóricamente afín a la ACT-IA, la evidencia de validez convergente se limita a la VME de cada dimensión. Por otra parte, el criterio de Fornell y Larcker apoya la validez discriminante de la dimensión cognitiva (la raíz cuadrada de su VME, .911, supera sus correlaciones con las dimensiones afectiva, .790, y conductual, .765), lo que contrasta con el modelo bifactor, según el cual esta dimensión carece de varianza específica fiable una vez controlado el factor general (ωHS = .006). Esta discrepancia sugiere que el criterio de Fornell y Larcker, basado en un modelo de factores correlacionados, es poco sensible a la presencia de un factor general dominante; índices complementarios como la razón heterorrasgo-monorrasgo (HTMT; Henseler et al., 2015) deberían incorporarse en estudios futuros.

Tabla 9

Validez convergente y discriminante de la ACT-IA

Dimensión	VME	VMC	VPC	1	2	3
1. Cognitiva	.829	.624	.605	(.911)		
2. Afectiva	.888	.624	.453	.790	(.942)	
3. Conductual	.878	.585	.433	.765	.530	(.937)

Nota. Los valores en la diagonal (entre paréntesis) corresponden a la raíz cuadrada del VME. VME = varianza media extraída; VMC = varianza máxima compartida; VPC = varianza promedio compartida. Existe evidencia de validez convergente cuando VME ≥ .50 y de validez discriminante cuando la VME es mayor que la VMC y la VPC, y la raíz cuadrada del VME excede las correlaciones interfactoriales (Fornell & Larcker, 1981).

Adicionalmente, se correlacionaron las puntuaciones de la ACT-IA y sus dimensiones entre sí y con las puntuaciones de la EDS-VB. En la Tabla 10 se puede apreciar que la puntuación total correlacionó alta y positivamente con las dimensiones cognitiva (r = .941), afectiva (r = .861) y conductual (r = .733), todas significativas ($p < .01$). Estas correlaciones reflejan relaciones parte-todo, por lo que no constituyen evidencia de validez convergente en sentido estricto. Las correlaciones interdimensionales fueron moderadas a altas, destacando cognitiva–afectiva (r = .706). Respecto a la deseabilidad social, se observaron correlaciones negativas y bajas con la puntuación total ($r = −.199$, $p < .01$) y con las dimensiones cognitiva ($r = −.226$, $p < .01$) y afectiva ($r = −.151$, $p < .05$), mientras que la asociación con la dimensión conductual no fue significativa (r = −.111). Este patrón de correlaciones bajas y negativas constituye evidencia útil de que las respuestas a la ACT-IA no están fuertemente contaminadas por la tendencia a responder de forma socialmente deseable; sin embargo, por sí sola no constituye una demostración completa de validez divergente, la cual requeriría además la comparación con constructos teóricamente distintos y no solo con un sesgo de respuesta.

Tabla 10

Matriz de correlaciones entre la ACT-IA, sus dimensiones y la deseabilidad social

Escala/Dimensión	1	2	3	4	5
1. ACT-IA (total)	—				
2. Cognitiva	.941**	—			
3. Afectiva	.861**	.706**	—		
4. Conductual	.733**	.662**	.382**	—	
5. Deseabilidad social	−.199**	−.226**	−.151*	−.111	—

Nota. n = 236. * $p < .05$. ** $p < .01$.

Invarianza de Medición por Género

Aunque el modelo bifactor CFA presentó el mejor ajuste global entre los modelos evaluados, su propósito principal fue examinar la dimensionalidad del instrumento y determinar la contribución relativa del factor general y de los factores específicos, más que constituir el modelo de medición destinado a la interpretación rutinaria de las puntuaciones (Eid et al., 2017; Reise, 2012). Por otra parte, la especificación bifactor ESEM, que permite una representación más flexible de la estructura factorial al incorporar cargas cruzadas, no convergió hacia una solución admisible, por lo que no fue posible utilizarla en los análisis multigrupo. En consecuencia, la invarianza de medición se evaluó utilizando el modelo confirmatorio de tres factores oblicuos, el cual representa la estructura conceptual propuesta para la ACT-IA y constituye el modelo que sustenta la interpretación de las puntuaciones de sus dimensiones (Brown, 2015; Millsap, 2011). Debe precisarse, por tanto, que la evidencia de invarianza que se presenta corresponde al modelo de factores correlacionados y no directamente al modelo que sustenta la puntuación total recomendada en este estudio. La comparabilidad de la puntuación total entre hombres y mujeres se infiere de la equivalencia de las cargas y los umbrales de los 19 reactivos que la componen, pero no se evaluó mediante un modelo unidimensional o bifactor multigrupo, lo que deberá examinarse en estudios futuros. Con el propósito de determinar si la estructura factorial de la ACT-IA es equivalente entre hombres y mujeres, se llevó a cabo un análisis factorial confirmatorio multigrupo (MG-CFA).

De manera secuencial, se estimaron los modelos configural, métrico y escalar, comparando el ajuste de cada modelo con el inmediatamente anterior mediante los cambios en los índices de ajuste (ΔCFI, ΔTLI, ΔRMSEA y ΔSRMR), de acuerdo con las recomendaciones de Cheung y Rensvold (2002) y Chen (2007). Los resultados se presentan en la Tabla 11.

El modelo configural mostró índices incrementales favorables (CFI = .984; TLI = .982; SRMR = .065), aunque su RMSEA (.114) se ubica en la banda Muy Débil del marco de Howard et al. (2025): con 298 gl corresponde el tercil de más de 163 gl, cuyo percentil 10 es .08. Por tanto, la invarianza se evaluó sobre un modelo de referencia con ajuste absoluto limitado y con un grupo masculino de 88 personas, lo que obliga a interpretar estos resultados con cautela. Posteriormente, la imposición de restricciones de igualdad sobre las cargas factoriales (modelo métrico) produjo cambios mínimos en los índices de ajuste (ΔCFI = +.001, ΔTLI = +.002, ΔRMSEA = −.006 y ΔSRMR = +.002), lo que respalda la equivalencia de las cargas factoriales entre hombres y mujeres. Dado que la diferencia simple entre los valores χ^2 robustos WLSMV no se distribuye como χ^2 con estimadores robustos y variables categóricas, la Tabla 11 no reporta pruebas de $\Delta\chi^2$, y la decisión sobre invarianza se fundamentó en los índices incrementales (ΔCFI, ΔTLI, ΔRMSEA, ΔSRMR). Finalmente, el modelo escalar también mantuvo cambios triviales respecto al modelo métrico (ΔCFI = .000, ΔTLI = +.003, ΔRMSEA = −.010 y ΔSRMR = −.002), proporcionando evidencia de invarianza escalar. En conjunto, estos resultados indican que la ACT-IA presenta equivalencia de medición según género, lo que, con las salvedades señaladas, apoya la comparación de las medias latentes de las dimensiones entre hombres y mujeres (Meredith, 1993; Cheung & Rensvold, 2002; Chen, 2007; Millsap, 2011).

Confiabilidad y Estadística Descriptiva de la ACT-IA

La Tabla 12 presenta las estadísticas descriptivas, los índices de confiabilidad y el error estándar de medición. La puntuación total mostró una media de 81.81 (DE = 24.03) en un rango de 19 a 114. La consistencia interna fue excelente tanto para la puntuación total (α = .963; ω = .962) como para las subescalas (α y ω entre .943 y .966), con intervalos de confianza estrechos. Los valores del EEM fueron bajos (ACT-IA = 4.68), respaldando la precisión de las puntuaciones. No obstante, conforme a la evidencia bifactorial, conviene recordar que coeficientes elevados de α y ω total pueden sobreestimar la interpretabilidad independiente de las subescalas cuando existe un factor general dominante (Reise et al., 2013; Rodriguez et al., 2016).

Tabla 11

Resultados de la invarianza de medición de la Escala de Actitudes hacia el Uso de la Inteligencia Artificial en los Procesos de Selección de Personal (ACT-IA) según género

	χ^2 (*gl*)	SRMR	RMSEA	CFI	TLI	Modelo de Referencia	ΔSRMR	ΔRMSEA	ΔCFI	ΔTLI
1. Configural	749 (298)	.065	.114	.984	.982	-----	-----	-----	-----	-----
2. Métrica	735 (314)	.067	.108	.985	.984	1	+.002	-.006	+.001	+.002
3. Escalar	821 (387)	.065	.098	.985	.987	2	-.002	-.010	.000	+.003

Nota. n = 236. Δ se calcula respecto al modelo inmediatamente anterior. No se reportan pruebas de $\Delta\chi^2$, dado que la diferencia simple de los valores robustos WLSMV no se distribuye como χ^2 con variables categóricas; la decisión sobre invarianza se basó en los índices incrementales (ΔCFI, ΔTLI, ΔRMSEA, ΔSRMR).

Tabla 12

Estadística descriptiva, confiabilidad y error estándar de medición de la ACT-IA

Escala/Dimensión	*Ítems*	*M*	*DE*	*α [IC 95%]*	*ω [IC 95%]*	*EEM*
ACT-IA (total)	19	81.81	24.03	.963 [.955, .971]	.962 [.951, .969]	4.68
Cognitiva	8	34.10	11.08	.962 [.951, .970]	.962 [.952, .970]	2.16
Afectiva	6	22.57	10.46	.965 [.955, .974]	.966 [.956, .974]	1.93
Conductual	5	25.14	6.27	.944 [.921, .962]	.943 [.918, .962]	1.50

Nota. n = 236. M = media; DE = desviación estándar; EEM = error estándar de medición, calculado en todos los casos como $DE\sqrt{(1-\omega)}$; IC = intervalo de confianza.

Discusión

El presente estudio desarrolló y aportó evidencia inicial de validez y confiabilidad de la ACT-IA, un instrumento dirigido a medir las actitudes hacia el uso de la IA en los procesos de selección de personal en población puertorriqueña. La evidencia de validez de contenido fue favorable, ya que la totalidad de los reactivos superó el criterio de Lawshe modificado por Tristán-López (2008), y los análisis de reactivos evidenciaron índices de discriminación elevados, particularmente en la dimensión conductual. El proceso de depuración condujo a una versión parsimoniosa de 19 reactivos organizados en tres factores que explicaron casi el 80 % de la varianza. Este porcentaje, no obstante, probablemente está inflado, dado que se obtuvo sobre un conjunto de reactivos seleccionado con la misma muestra; además, la separación de las tres dimensiones coincide en parte con la dirección de la redacción de los reactivos retenidos, por lo que su coherencia conceptual debe matizarse (véase Limitaciones).

El hallazgo central provino de la comparación de modelos competitivos. Aunque las soluciones de tres factores oblicuos (CFA) y ESEM mostraron índices aceptables a fuertes, el modelo bifactor-CFA presentó el mejor ajuste global. La adopción del marco percentilar de Howard et al. (2025) resultó decisiva para esta conclusión: al situar cada índice respecto a la distribución de valores publicados, fue posible diferenciar entre modelos que, bajo los cortes fijos tradicionales, habrían sido catalogados indistintamente como “buenos”. Por ejemplo, tanto el CFA de tres factores como el bifactor superaban el umbral clásico de CFI > .95; sin embargo, el enfoque de percentil mostró que el modelo bifactor presentó el patrón de ajuste más favorable, al combinar índices incrementales por encima del percentil 90 con un SRMR en la banda Fuerte y un RMSEA en una banda intermedia, mientras que los modelos CFA y ESEM conservaron un RMSEA en una banda de ajuste menos favorable. Esta lectura más matizada evita las inferencias erróneas que pueden derivarse de la aplicación mecánica de cortes dicotómicos, y es consistente con la advertencia de que un buen ajuste de un modelo multidimensional no implica, por sí mismo, que las subescalas deban interpretarse como

dimensiones independientes (Reise et al., 2013; Rodriguez et al., 2016). Es preciso reconocer, no obstante, que la superioridad del modelo bifactor debe matizarse a la luz de dos observaciones. Primero, cinco cargas ESEM superaron ligeramente 1.00 (ítems 2, 3, 26, 27 y 38); aunque no se observaron varianzas residuales negativas ni otros indicios de soluciones impropias, este patrón refleja la elevada correlación entre factores bajo rotación oblicua. Segundo, los valores de ARPB superiores al umbral de 10–15 % en la mayoría de los reactivos afectivos y conductuales (particularmente los ítems 38 y 39) indican que una solución estrictamente unidimensional sesgaría sus cargas, lo que refuerza que estas dimensiones conservan varianza específica más allá del factor general. Estas observaciones no invalidan la comparación de modelos, pero ameritan verificación adicional en futuras réplicas. Además, la superioridad en ajuste del modelo bifactor debe leerse a la luz de su mayor propensión a ajustarse a los datos en comparación con modelos de factores correlacionados (Bonifay et al., 2017). Asimismo, el patrón de cargas específicas cercanas a cero o negativas en la dimensión cognitiva corresponde a una de las anomalías descritas por Eid et al. (2017) para los modelos con factor general; un modelo bifactor-(S−1) que tome la dimensión cognitiva como dominio de referencia podría ofrecer una representación más interpretable y constituye una línea de trabajo futura. Por último, dado que los percentiles de Howard et al. (2025) no se diferencian por estimador ni por tipo de indicador (Xia & Yang, 2019), las clasificaciones de ajuste deben entenderse como una referencia orientativa para comparar los modelos entre sí.

La superioridad del modelo bifactorial no invalida la propuesta tridimensional, sino que indica que la ACT-IA posee una estructura jerárquica dominada por un factor general. Los índices auxiliares reforzaron esta interpretación: el ECV = .731 y el ωH = .873 mostraron que la mayor parte de la varianza común y de la confiabilidad del puntaje total se atribuye al factor general, lo que respalda una interpretación esencialmente unidimensional del instrumento (Reise, 2012; Rodriguez et al., 2016). La dimensión cognitiva quedó prácticamente absorbida por el factor general (ωHS = .006; H = .223), mientras que las facetas afectiva y conductual conservaron varianza específica moderada y replicabilidad adecuada. En el caso de la faceta conductual, esta varianza específica podría aportar validez incremental en contextos aplicados (Zhang et al., 2021). En el caso de la faceta afectiva, en cambio, sus reactivos son exclusivamente de redacción inversa y expresan preocupación, ansiedad o desconfianza, por lo que su varianza específica (ωHS = .360) podría reflejar, en parte, un efecto de método asociado a la redacción inversa (véase Limitaciones). En consecuencia, esta subescala se describe con mayor precisión como una medida de afecto negativo o desconfianza hacia el uso de la IA en la selección de personal, y no debería atribuírsele validez incremental hasta que se separe empíricamente el contenido afectivo de la dirección de la redacción. En conjunto, se recomienda interpretar la ACT-IA principalmente mediante su puntuación total, reservar la subescala conductual para usos complementarios y limitar la subescala afectiva a usos exploratorios, condicionados a dicha verificación. En cambio, dada su contribución específica prácticamente nula una vez controlado el factor general (ωHS = .006; H = .223), no se recomienda interpretar o reportar la subescala cognitiva de manera independiente.

La escala mostró, además, evidencia de invarianza configural, métrica y escalar por género, lo que sugiere que mide el constructo de manera equivalente en personas masculinas y femeninas y habilita, con cautela, comparaciones entre grupos en las dimensiones; la comparabilidad de la puntuación total se infiere de este resultado, pero no se evaluó directamente (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education [AERA] et al., 2014; Putnick & Bornstein, 2016). La consistencia interna fue excelente y el error estándar de medición, bajo, lo que respalda la precisión de las puntuaciones. Finalmente, las correlaciones negativas y de baja magnitud con la deseabilidad social sugieren que las respuestas a la ACT-IA no están fuertemente influidas por una tendencia a responder de forma socialmente favorable (Paulhus, 2002; Podsakoff et al., 2003).

Limitaciones e Investigación Futura

Los resultados deben interpretarse a la luz de varias limitaciones. En primer lugar, todos los análisis se realizaron con una sola muestra, lo que impidió la validación cruzada mediante submuestras de calibración y validación e incrementa el riesgo de capitalización del azar (Brown, 2015; Fokkema & Greiff, 2017). Cabe destacar que el análisis factorial exploratorio, la depuración de reactivos y los modelos confirmatorios posteriores se realizaron con esta misma muestra, por lo que los índices de ajuste reportados para los modelos CFA, ESEM y bifactor deben interpretarse como evidencia inicial de estructura interna, y no como una confirmación definitiva, hasta que se repliquen en una muestra independiente. Asimismo, el muestreo no probabilístico por conveniencia, apoyado en la técnica de redes (snowball) y en la difusión a través de redes

sociales, limita la representatividad de la muestra y la generalización de los hallazgos a la población general de profesionales de RR. HH. y PIO de Puerto Rico. Relacionado con lo anterior, el tamaño muestral, adecuado para el AFE y el CFA inicial, se volvió considerablemente más ajustado al dividirse por género para las pruebas de invarianza (aproximadamente 146 mujeres y 88 hombres), lo que pudo reducir la estabilidad de las estimaciones del modelo escalar y debe considerarse una limitación adicional.

En segundo lugar, no se examinó la estabilidad temporal del instrumento mediante procedimientos de prueba-reprueba, consideración relevante dado que las actitudes hacia tecnologías emergentes pueden evolucionar con la familiaridad y la experiencia de uso (Glikson & Woolley, 2020; Venkatesh et al., 2003). En tercer lugar, los intentos de estimar modelos más complejos (bifactor-ESEM y jerárquico de segundo orden) no produjeron soluciones admisibles. Asimismo, la validez de contenido se evaluó con un panel de tres jueces, el mínimo admisible para el método de Lawshe modificado por Tristán-López (2008), y la interpretación de los índices de ajuste se basó en percentiles derivados de estudios que no se diferencian por estimador ni por tipo de indicador, lo que limita su aplicación directa a modelos estimados con WLSMV (Xia & Yang, 2019).

En cuarto lugar, el primer AFE agrupó en un mismo factor los reactivos de redacción inversa de las tres dimensiones (véase la Tabla S1), y la dimensión afectiva retenida quedó compuesta exclusivamente por reactivos de redacción inversa que expresan preocupación, ansiedad o desconfianza, mientras que los reactivos afectivos de redacción directa se agruparon con los cognitivos y fueron eliminados. En consecuencia, en la versión final el contenido afectivo y la dirección de la redacción están confundidos, por lo que con estos datos no es posible determinar en qué medida la varianza específica de la dimensión afectiva ($\omega HS = .360$) refleja una distinción sustantiva, el afecto negativo hacia la IA como componente diferenciable de las creencias, o un efecto de método asociado a los reactivos inversos (Marsh, 1996; Weijters et al., 2013). Futuros estudios deberían incluir reactivos afectivos redactados en ambas direcciones y modelar explícitamente un factor de método para los reactivos inversos, por ejemplo, mediante un modelo CT-C(M−1) o un modelo de intercepto aleatorio (Eid et al., 2017; Maydeu-Olivares & Coffman, 2006), lo que permitiría separar ambas fuentes de varianza.

Investigaciones futuras deberían replicar la estructura en muestras independientes y de mayor tamaño, evaluar la confiabilidad temporal y acumular evidencia de validez predictiva y de criterio frente a variables como la confianza en algoritmos, la aversión al algoritmo y la disposición a adoptar sistemas automatizados (Dietvorst et al., 2015). Esta evidencia adicional debería preceder a cualquier recomendación de uso individual o de alto impacto en decisiones de selección de personal.

Implicaciones

Desde una perspectiva aplicada, la ACT-IA constituye una herramienta prometedora, con evidencia psicométrica inicial favorable, para evaluar las actitudes hacia la integración de la IA en los procesos de reclutamiento y selección. Podría utilizarse, a nivel grupal, para explorar niveles de apertura, confianza o resistencia entre profesionales de RRHH durante procesos de transformación digital, así como para focalizar intervenciones de capacitación orientadas a fortalecer la confianza y promover una adopción ética y responsable de los sistemas automatizados (Marler & Boudreau, 2017; Strohmeier, 2020).

A la luz del predominio del factor general, la puntuación total se recomienda como el indicador más robusto y defendible del constructo. En términos concretos, un profesional de RR. HH. podría emplear la puntuación total de la ACT-IA (rango 19–114) para clasificar de manera preliminar a los equipos o unidades con actitudes menos favorables hacia la IA (p. ej., puntuaciones en el tercio inferior de la distribución observada, $M = 81.81$, $DE = 24.03$) como prioritarios para intervenciones de sensibilización, comunicación transparente sobre el funcionamiento de los sistemas automatizados y capacitación previa a la implementación de herramientas de IA en el proceso de selección; puntuaciones más altas, en cambio, podrían orientar hacia una adopción más rápida acompañada de mecanismos de supervisión humana. Dado que aún no se dispone de bandas de interpretación normativas ni de puntos de corte validados, estas aplicaciones deben entenderse como orientaciones generales de gestión del cambio organizacional, y no como criterios diagnósticos individuales, hasta que se acumule evidencia adicional de validez de criterio y estabilidad temporal.

Conclusión

La ACT-IA aporta a la literatura un instrumento en español, dirigido específicamente al contexto de selección de personal, con evidencia inicial favorable de validez de contenido, estructura interna, confiabilidad e invarianza por género. La interpretación de los índices de ajuste mediante el enfoque de percentiles de Howard et al. (2025)

permitió una valoración más matizada de los modelos competitivos y respaldó la conclusión de que las actitudes hacia la IA en la selección de personal se organizan en torno a un factor general dominante con facetas específicas residuales. Por ello, se recomienda priorizar la puntuación total de la ACT-IA sobre la interpretación aislada de sus subescalas, en particular la dimensión cognitiva y, mientras no se separe el efecto de la redacción inversa, la afectiva. Estos hallazgos sientan las bases para futuras investigaciones que repliquen y amplíen la evidencia de validez del instrumento.

Referencias

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling, 16*(3), 397–438. https://doi.org/10.1080/10705510903008204

Bennion, B. (2023). *Seductive allure of neuroscience and artificial intelligence in hiring* [Tesis de maestría, Bowling Green State University]. OhioLINK Electronic Theses and Dissertations Center. http://rave.ohiolink.edu/etdc/view?acc_num=bgsu1678806399885144

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health, 6*, 149. https://doi.org/10.3389/fpubh.2018.00149

Bonifay, W., Lane, S. P., & Reise, S. P. (2017). Three concerns with applying a bifactor model as a structure of psychopathology. *Clinical Psychological Science, 5*(1), 184–186. https://doi.org/10.1177/2167702616657069

Breckler, S. J. (1984). Empirical validation of affect, behavior, and cognition as distinct components of attitude. *Journal of Personality and Social Psychology, 47*(6), 1191–1205. https://doi.org/10.1037/0022-3514.47.6.1191

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2.ª ed.). Guilford Press.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14*(3), 464-504. https://doi.org/10.1080/10705510701301834

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 9*(2), 233-255. https://doi.org/10.1207/S15328007SEM0902_5

Cicero, L., Russo, A., Di Stefano, G., & Zammitti, A. (2025). The General Attitudes towards Artificial Intelligence Scale (GAAIS): Validation and psychometric properties analysis in the Italian context. *BMC Psychology, 13*(1), 641. https://doi.org/10.1186/s40359-025-02935-2

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340. https://doi.org/10.2307/249008

DeVellis, R. F., & Thorpe, C. T. (2021). *Scale development: Theory and applications* (5.ª ed.). Sage.

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General, 144*(1), 114–126. https://doi.org/10.1037/xge0000033

Dueber, D. M. (2021). *BifactorIndicesCalculator: Bifactor indices calculator* (Version 0.2.2) [R package]. https://doi.org/10.32614/CRAN.package.BifactorIndicesCalculator.

Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes*. Harcourt Brace.

Eid, M., Geiser, C., Koch, T., & Heene, M. (2017). Anomalous results in G-factor models. *Psychological Methods, 22*(3), 541–562. https://doi.org/10.1037/met0000083

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*(3), 272–299. https://doi.org/10.1037/1082-989X.4.3.272

Fokkema, M., & Greiff, S. (2017). How performing PCA and CFA on the same data equals trouble. *European Journal of Psychological Assessment, 33*(6), 399-402.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with

unobservable variables and measurement error. *Journal of Marketing Research, 18*(1), 39-50. https://doi.org/10.1177/002224378101800104

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals, 14*(2), 627–660.

Gnambs, T., Stein, J.-P., Appel, M., Griese, F., & Zinn, S. (2025). An economical measure of attitudes towards artificial intelligence in work, healthcare, and education (ATTARI-WHE). *Computers in Human Behavior: Artificial Humans, 3*(1), 100106. https://doi.org/10.1016/j.chbah.2024.100106

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8.ª ed.). Cengage Learning.

Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods, 7*(2), 191–205. https://doi.org/10.1177/1094428104263675

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science, 43*(1), 115–135. https://doi.org/10.1007/s11747-014-0403-8

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*(2), 179–185. https://doi.org/10.1007/BF02289447

Howard, M. C., Boudreaux, M., Cogswell, J., Manix, K. G., & Oglesby, M. T. (2025). A literature review of model fit and model comparisons with confirmatory factor analysis: Formalizing the informal in organizational science. Applied Psychology, 74(1), e12592. https://doi.org/10.1111/apps.12592

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55. https://doi.org/10.1080/10705519909540118

Islam, M., Rahman, M. M., Taher, M. A., Quaosar, G. M. A. A., & Uddin, M. A. (2024). Using artificial intelligence for hiring talents in a moderated mechanism. *Future Business Journal, 10*(1), 1–16. https://doi.org/10.1186/s43093-024-00303-x

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4.ª ed.). Guilford Press.

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology, 28*(4), 563–575. https://doi.org/10.1111/j.1744-6570.1975.tb01393.x

Luo, W., Zhang, Y., & Mu, M. (2025). Why might AI-enabled interviews reduce candidates' job application intention? The role of procedural justice and organizational attractiveness. *Humanities and Social Sciences Communications, 12*(1), 1-12. https://doi.org/10.1057/s41599-025-05607-z

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods, 4*(1), 84–99. https://doi.org/10.1037/1082-989X.4.1.84

Mariani, M. M., Machado, I., Magrelli, V., & Dwivedi, Y. K. (2023). Artificial intelligence in innovation research: A systematic review, conceptual framework, and future research directions. *Technovation, 122*, 102623. https://doi.org/10.1016/j.technovation.2022.102623

Marler, J. H., & Boudreau, J. W. (2017). An evidence-based review of HR analytics. *The International Journal of Human Resource Management, 28*(1), 3–26.

Marsh, H. W. (1996). Positive and negative global self-esteem: A substantively meaningful distinction or artifactors? *Journal of Personality and Social Psychology, 70*(4), 810–819. https://doi.org/10.1037/0022-3514.70.4.810

Marsh, H. W., Morin, A. J. S., Parker, P. D., & Kaur, G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology, 10*, 85–110. https://doi.org/10.1146/annurev-clinpsy-032813-153700

Maydeu-Olivares, A., & Coffman, D. L. (2006). Random intercept item factor analysis. *Psychological Methods, 11*(4), 344–362. https://doi.org/10.1037/1082-989X.11.4.344

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*(4), 525–543.

Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. Routledge.

Mori, M., Sassetti, S., Cavaliere, V., & Bonti, M. (2024). A systematic literature review on

artificial intelligence in recruiting and selection: A matter of ethics. *Personnel Review*. https://doi.org/10.1108/PR-03-2023-0257

Morin, A. J. S., Myers, N. D., & Lee, S. (2020). Modern factor analytic techniques: Bifactor models, exploratory structural equation modeling, and bifactor-ESEM. En G. Tenenbaum & R. C. Eklund (Eds.), *Handbook of sport psychology* (4.ª ed.). Wiley.

Muñiz, J. (2018). *Introducción a la psicometría: Teoría clásica y TRI*. Pirámide.

Park, J., Woo, S. E., & Kim, J. (2024). Attitudes towards artificial intelligence at work: Scale development and validation. *Journal of Occupational and Organizational Psychology, 97*(3), 920–951. https://doi.org/10.1111/joop.12502

Paulhus, D. L. (2002). Socially desirable responding: The evolution of a construct. En H. I. Braun, D. N. Jackson, & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 49–69). Lawrence Erlbaum Associates.

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879–903.

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting. *Developmental Review, 41*, 71–90.

Qin, C., Zhang, L., Zha, R., Shen, D., Zhang, Q., Sun, Y., Zhu, C., Zhu, H., & Xiong, H. (2023). *A comprehensive survey of artificial intelligence techniques for talent analytics*. *arXiv*. https://doi.org/10.48550/arXiv.2307.03195

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research, 47*(5), 667-696. https://doi.org/10.1080/00273171.2012.715555

Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment, 95*(2), 129-140.

Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods, 17*(3), 354–373. https://doi.org/10.1037/a0029315

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods, 21*(2), 137–150. https://doi.org/10.1037/met0000045

Rosario-Hernández, E., & Rovira-Millán, L. V. (2002). Desarrollo y validación de una escala para medir las actitudes hacia el retiro. *Revista Puertorriqueña de Psicología, 13*, 47-63.

Rosario-Hernández, E., Rovira-Millán, L. V., & Blanco-Rovira, R. A. (2025). Review of the psychometric properties of the Social Desirability Scale and the development of a short-form: Social Desirability Scale–Brief Version (SDS-BV). *Revista Caribeña de Psicología, 9*, 1-18. https://doi.org/10.37226/rcp.v9i1.13313

Rosenberg, M. J., & Hovland, C. I. (1960). Cognitive, affective, and behavioral components of attitudes. En M. J. Rosenberg, C. I. Hovland, W. J. McGuire, R. P. Abelson, & J. W. Brehm (Eds.), *Attitude organization and change: An analysis of consistency among attitude components* (pp. 1–14). Yale University Press.

Rosseel, Y., Jorgensen, T. D., & De Wilde, L. (2025). *lavaan: Latent variable analysis* (Version 0.6-20) [R package]. https://doi.org/10.32614/CRAN.package.lavaan

Schepman, A., & Rodway, P. (2020). Initial validation of the General Attitudes towards Artificial Intelligence Scale. *Computers in Human Behavior Reports, 1*, 100014. https://doi.org/10.1016/j.chbr.2020.100014

Schepman, A., & Rodway, P. (2023). The General Attitudes towards Artificial Intelligence Scale (GAAIS): Confirmatory validation and associations with personality, corporate distrust, and general trust. *International Journal of Human–Computer Interaction, 39*(13), 2724–2741. https://doi.org/10.1080/10447318.2022.2085400

Society for Human Resource Management. (2023). *The use of artificial intelligence in HR* [Reporte de investigación]. SHRM. https://www.shrm.org/topics-tools/research

Society for Industrial and Organizational Psychology. (2023). *Principles for the validation and use of personnel selection procedures* (5.ª ed.). SIOP. https://www.siop.org/Research-Publications/SIOP-Principles

Sonderling, K. E., & Kelley, B. J. (2023). Filling the void: Artificial intelligence and private initiatives. *North Carolina Journal of Law & Technology, 24*(4), 153. https://scholarship.law.unc.edu/ncjolt/vol24/iss4/6

Spector, P. E. (1992). *Summated rating scale construction: An introduction* (Vol. 82). Sage Publications.

Starke, C., Baleis, J., Keller, B., & Marcinkowski, F. (2022). Fairness perceptions of algorithmic decision-making: A systematic review of the empirical literature. *Big Data & Society, 9*(2). https://doi.org/10.1177/20539517221115189

Strohmeier, S. (2020). Digital human resource management: A conceptual clarification. *German Journal of Human Resource Management, 34*(3), 345–365.

Subía-Arellano, A., Pérez-Vega, D., Guillen-García, S., & Cáceres-Fierro, N. (2023). Design of the attitudinal assessment scale towards artificial intelligence (EVAIA-1). En *2023 IEEE Seventh Ecuador Technical Chapters Meeting (ETCM)* (pp. 1–6). IEEE. https://doi.org/10.1109/ETCM58927.2023.10309175

Tristán-López, A. (2008). Modificación al modelo de Lawshe para el dictamen cuantitativo de la validez de contenido de un instrumento objetivo. *Avances en Medición, 6*(1), 37-48.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly, 27*(3), 425–478.

Weijters, B., Baumgartner, H., & Schillewaert, N. (2013). Reversed item bias: An integrative model. *Psychological Methods, 18*(3), 320–334. https://doi.org/10.1037/a0032121

Wilson, K., & Caliskan, A. (2024). Gender, race, and intersectional bias in resume screening via language model retrieval. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, 7*, 1578–1590.

Worthington, R. L., & Whittaker, T. A. (2006). Scale development research: A content analysis and recommendations for best practices. *The Counseling Psychologist, 34*(6), 806–838. https://doi.org/10.1177/0011000006288127

Xia, Y., & Yang, Y. (2019). RMSEA, CFI, and TLI in structural equation modeling with ordered categorical data: The story they tell depends on the estimation methods. *Behavior Research Methods, 51*(1), 409–428. https://doi.org/10.3758/s13428-018-1055-2

Zhang, B., Sun, T., Cao, M., & Drasgow, F. (2021). Using bifactor models to examine the predictive validity of hierarchical constructs: Pros, cons, and solutions. *Organizational Research Methods, 24*(3), 530–571.

Zhou, J., Verma, S., Mittal, M., & Chen, F. (2021). Understanding relations between perception of fairness and trust in algorithmic decision making. En *2021 IEEE International Conference on Behavioural and Social Computing (BESC)* (pp. 1–5). IEEE.

Conflicto de Intereses: Los/as autores/as declaran no tener conflictos de intereses asociados con la publicación.

Aprobación de la Junta de Revisión Institucional para la Protección de Seres Humanos en la Investigación: El presente estudio gestionó la aprobación y autorización de la Junta de Revisión Institucional (IRB) de la Universiad Albizu, San Juan.

Consentimiento Informado: Las personas participantes completaron el consentimiento informado de forma digital.

Financiamiento: Esta investigación no tuvo financiamiento.