

Un modelo bifactor para la Escala de Actitudes hacia la Estadística

José Moral de la Rubia¹, René Landero Hernández², Mónica Teresa González Ramírez³, Lucía del Carmen Quezada Berumen⁴
Universidad Autónoma de Nuevo León-México

Los objetivos del estudio fueron determinar las propiedades psicométricas de la Escala de Actitudes hacia la Estadística en México. En este estudio instrumental, se analizó una muestra no probabilística de 383 estudiantes de psicología mexicanos que respondieron a dos escalas actitudinales. El modelo bifactor definido por un factor general y tres factores específicos (seguridad, agrado-utilidad y motivación) mostró las mejores propiedades de ajuste a los datos y de validez convergente en sus modelos de medida. La actitud promedio fue neutra y más positiva en hombres que en mujeres. El factor general de actitud hacia la estadística tuvo correlación alta con actitud hacia la investigación. Se concluye el bifactor de tres factores representa mejor la relación entre los ítems que el modelo original de cinco factores correlacionados.

Palabras clave: Actitud, estadística, estudiantes, medición, validación.

A bifactor model for the Scale of Attitudes towards Statistics

The objectives of the study were to determine the psychometric properties of the Scale of Attitudes towards Statistics in Mexico. In this instrumental study, a non-probabilistic sample of 383 Mexican psychology students who responded to two attitude scales was analyzed. The bifactor model with three specific factors (safety, pleasure-usefulness, and motivation) showed the best properties of fit and convergent validity. The average attitude was neutral and more positive in men than in women. The general factor had a high correlation with

¹ Doctor en Psicología. Profesor-investigador en la Universidad Autónoma de Nuevo León, México. Contacto: jose.morald@uanl.edu.mx <https://orcid.org/0000-0003-1856-1458>

² Doctor en Trabajo Social y Políticas Comparadas de Bienestar Social. Profesor-investigador en la Universidad Autónoma de Nuevo León, México. Contacto: rene.landerohr@uanl.edu.mx <https://orcid.org/0000-0002-9694-8116>

³ Doctora en Psicología. Profesor-investigador en la Universidad Autónoma de Nuevo León, México. Contacto: jose_moral@hotmail.com <https://orcid.org/0000-0002-9058-9626>

⁴ Doctora en Psicología. Profesor-investigador en la Universidad Autónoma de Nuevo León, México. Contacto: luciaqb86@msn.com <https://orcid.org/0000-0003-4705-3225>

attitude towards research. It is concluded that the three-factor bifactor model represents the relationship between the items better than the original five-factor model.

Keywords: Attitude, statistics, measurement, students, validation.

Um modelo bifator para a escala de atitudes em relação à estatística

Os objetivos do estudo eram determinar as propriedades psicométricas da Escala de Atitudes em Relação à Estatística no México. Neste estudo instrumental, uma amostra não probabilística de 383 estudantes mexicanos de psicologia que responderam a duas escalas atitudinais foi analisada. O modelo bifator definido por um fator geral e três fatores específicos (segurança, prazer-utilidade e motivação) apresentou as melhores propriedades de ajuste aos dados e de validade convergente em seus modelos de mensuração. A atitude média foi neutra e mais positiva nos homens. O fator geral de atitude em relação à estatística teve alta correlação com a atitude em relação à pesquisa. Conclui-se que o bifator de três fatores representa melhor a relação entre os itens do que o modelo original de cinco fatores correlacionados.

Palavras-chave: Atitude, estatística, medição, alunos, validação.

Un modèle bifactoriel pour l'échelle d'attitudes envers la statistique

Les objectifs de cette étude étaient de déterminer les propriétés psychométriques de l'échelle d'attitudes envers les statistiques au Mexique. Dans cette étude instrumentale, un échantillon non probabiliste de 383 étudiants mexicains en psychologie ayant répondu à deux échelles d'attitudes a été analysé. Le modèle bifactoriel avec trois facteurs spécifiques (sécurité, plaisir-utilité, et motivation) a montré les meilleures propriétés d'ajustement et de validité convergente. L'attitude moyenne était neutre et plus positive chez les hommes que chez les femmes. Le facteur général avait une forte corrélation avec l'attitude envers la recherche. Il est conclu que le modèle bifactoriel à trois facteurs représente mieux la relation entre les items que le modèle original à cinq facteurs.

Mots clés: Attitude, statistiques, mesure, étudiants, validation.

El estudio de las actitudes hacia la estadística es de interés para los investigadores en el campo educativo (Peiró-Signes, Trull, Segarra-Oña, & García-Díaz, 2020). La estadística forma parte del currículo académico en la mayoría de las carreras universitarias (Cladera, Rejón-Guardia, Vich-i-Martorell, & Juaneda, 2019), incluida la psicología (Comas, Martins, Nascimento, & Estrada, 2017).

Cabe señalar que la psicología como ciencia empírica, desde sus inicios, ha hecho uso y ha contribuido al desarrollo de la ciencia estadística (Howard, 2019; Wright, 2009). En la carrera de psicología, la mayor proporción de estudiantes y egresados en cualquier país del mundo es femenina, aproximadamente ocho mujeres por cada dos hombres (Huang, Gates, Sinatra, & Barabási, 2020; Olos & Hoff, 2006; Rosli, Maat, & Rosli, 2017). Quizá este fuerte sesgo femenino y la presencia de cuerpos teórico-prácticos de corte filosófico y humanista en el campo disciplinar han generado gran resistencia a conceder mayor peso a los aspectos estadísticos y matemáticos en la organización curricular de la carrera (Ostertag & McNamara, 1991). No obstante, la consolidación de la psicología como una ciencia reputada y reconocida está ligada a su soporte lógico-matemático y estadístico (Howard, 2019; Wagenmakers et al., 2018).

En esta carrera, la enseñanza de la estadística presenta problemas didácticos debido a que los estudiantes acceden con bases heterogéneas (Comas et al., 2017; Vera & Díaz, 2013) y actitudes diversas formadas por las experiencias vividas durante el aprendizaje de las matemáticas y la estadística (Bautista, Morales, Dórame, & Peralta, 2016). Al influir la actitud en el rendimiento en las materias de estadísticas impartidas en la carrera (Rosli & Maat, 2017; Sesé, Jiménez, Montaña, & Palmer, 2015), la evaluación de la actitud hacia la estadística es un aspecto relevante (Ashaari, Judi, Tengku, & Wook, 2011; Mozafar, Akbar, & Hassan, 2017).

La Escala de Actitudes hacia la Estadística (EAE), creada en España por Auzmendi (1992), es uno de los instrumentos más usados en español para medir esta *actitud* definida como el posicionamiento en aspectos valorativos, emocionales y motivacionales ante la estadística como materia presente en el currículo de una ciencia (Ruiz de Miguel, 2015). La EAE está integrada por cinco factores: *utilidad* o valor que el estudiante otorga a la estadística para su futura vida profesional, *ansiedad* o temor que el estudiante manifiesta ante la asignatura, *confianza* o sentimiento de seguridad que provoca su habilidad en estadística, *agrado* o disfrute que provoca el trabajo con la estadística y *motivación* o inclinación hacia el estudio y utilización de la estadística. Esta estructura de cinco factores es la más usada en las investigaciones (Pérez, Aparicio, Bazán, & Abdounur, 2015; Torres, Aparicio, Bazán, & Abdounur, 2015), incluso en México (Bautista et al., 2016), que es el país en el que se hizo el presente estudio.

En estudiantes de economía españoles, Rojas, Escalera, Moreno y García (2017) confirmaron esta estructura de cinco factores y obtuvieron valores de confiabilidad por consistencia interna aceptables (de $\alpha = .72$ a $.79$). A través de un análisis de componentes principales de segundo orden, obtuvieron un único componente subyacente (de actitud hacia la estadística) que explicó el 55.8% de la varianza total.

No obstante, hay estudios divergentes con respecto al modelo de cinco factores. Unos estudios sustentan una estructura tetrafactorial. En esta línea, Darías (2000) exploró la estructura factorial de la EAE en una muestra de 188 estudiantes españoles de psicología y obtuvo cuatro factores; en este estudio, el ítem 25 se eliminó por saturar en varios factores con carga baja. También, Méndez y Macía (2007), al estudiar la estructura factorial de la escala en una muestra de 168 estudiantes universitarios chilenos, encontraron una estructura de cuatro factores; en esta validación se excluyó al ítem 20 por no saturar en ningún factor. Otros estudios muestran una estructura trifactorial (Bologna & Vaiman, 2013; Tejero & Castro, 2011). Incluso hay una propuesta de seis factores basada en análisis de conglomerados por variables (Vilá & Rubio, 2014).

Otro aspecto en que existe divergencia entre los estudios es con respecto a la diferencia en actitud promedio hacia la estadística entre ambos sexos. Por una parte, hay estudios que no reportan diferencias, ya sean hechos con la EAE (Pérez et al., 2015; Salinas & Mayén, 2016) u otras escalas (Comas et al., 2017; Sarikaya, Ok, Aydin, & Schau, 2018). Por otra parte, hay estudios que sí reportan diferencias. Auzmendi (1991) encontró una actitud más positiva con más confianza y menos ansiedad en hombres que en mujeres. Usando la Encuesta de Actitudes hacia la Estadística (SATS; Schau, Stevens, Dauphinee, & Del Vecchio, 1995), Mahmud y Zainol (2008) observaron que un promedio en la actitud hacia la estadística significativamente más positivo en mujeres que en hombres entre estudiantes malayos de diversos posgrados. También hallaron diferencia significativa en tres de sus seis factores: Afecto, Competencia y Valor. En Dificultad, Interés y Esfuerzo, los promedios fueron estadísticamente equivalentes. Usando el SATS, pero en una línea afín a Auzmendi (1991), Hommik y Luik (2017) en estudiantes de secundaria estonianos, Rejón-Guardia, Vich-I-Martorell, Juaneda y Cladera (2019) en estudiantes universitarios españoles de ciencias sociales y Opstad (2020) en estudiantes universitarios noruegos de economía reportaron un promedio en Valor más alto en hombres que en mujeres. También, en los hombres, los promedios fueron más altos en Competencia e Interés en el estudio de Hommik y Luik (2017); en Competencia y Afecto en el de Rejón-Guardia et al. (2019); y en Interés y Afecto en el de Opstad (2020). No obstante, el promedio en Esfuerzo fue más alto en mujeres que en hombres y no hubo diferencia en Dificultad en estos tres estudios.

A pesar del amplio uso de la EAE con sus cinco factores, su estructura factorial no está bien establecida y no está claro el perfil diferencial por sexos. Además, no ha sido validada en sus propiedades métricas en México, lo que es importante para su uso local, como en la población de estudiantes de psicología. Atendiendo a estas necesidades, este estudio instrumental tiene como objetivos: 1) establecer la estructura factorial de la EAE en una muestra extraída de una población de estudiantes mexicanos de psicología de una universidad pública, verificando que los

factores presenten confiabilidad por consistencia interna (precisión al medirse el constructo con distintos ítems), validez convergente (grado de certeza en que los ítems miden un mismo constructo) y validez discriminante (grado de certeza en que la varianza compartida entre los ítems de cada factor es mayor que la compartida entre los factores), y en caso de malas propiedades explorar un modelo nuevo; 2) describir las distribuciones de la puntuación total de la EAE y sus factores; 3) comparar los promedios entre mujeres y hombres; y 4) comprobar su validez de constructo concurrente en relación con la actitud hacia la investigación.

Se espera un buen ajuste del modelo de cinco factores correlacionados de Auzmendi (1992) con confiabilidad por consistencia interna al menos aceptable (Rojas et al., 2017), así como validez convergente y discriminante en sus factores. La expectativa con respecto a la distribución de las puntuaciones en la escala es que siga una distribución normal (Flores & Auzmendi, 2015). Se espera una actitud polarizada hacia la aceptación (Bautista et al., 2016; Comas et al., 2017). Las expectativas con respecto a las diferencias en los niveles actitudinales promedios entre ambos sexos no son claras, ya que los resultados de los estudios varían de no diferencia (Comas et al., 2017; Pérez et al., 2015; Salinas & Mayén, 2016; Sarikaya et al., 2018) a mayor promedio en hombres (Auzmendi; 1991; Hommik & Luik, 2017; Rejón-Guardia et al., 2019; Opstad, 2020) o mayor promedio en mujeres (Mahmud & Zainol, 2008). La hipótesis de validez concurrente es que exista una correlación positiva y alta entre la actitud hacia la estadística y la actitud hacia la investigación (Baños & Hurtado, 2016).

Método

Participantes

Un total de 412 estudiantes de psicología de una universidad pública del noreste de México respondieron al cuestionario. Se eliminaron 29 casos por datos incompletos. La muestra analizada fue de

383 participantes con sus datos completos, lo que corresponde a una fracción de muestreo de 0.10. En la muestra analizada, 73.6% de los participantes fueron mujeres y 26.4% hombres. La muestra quedó integrada por estudiantes de los 10 semestres de la licenciatura, aunque predominaron los cuatro primeros semestres, representando el 84.6% de la función empírica de distribución. La media de edad fue 19.40 años con desviación estándar de 2.96.

Medición

El cuestionario en línea quedó integrado por la solicitud de consentimiento informado, las preguntas sobre sexo (hombre o mujer) y edad, la EAE y una escala de actitudes hacia la investigación.

La Escala de Actitudes hacia la Estadística (EAE; Auzmendi, 1992) consta de 25 ítems con rango de respuesta tipo Likert de cinco puntos (de 1 = “totalmente en desacuerdo” a 5 = “totalmente de acuerdo”). En el estudio de Auzmendi (1992), se encontró una confiabilidad por consistencia interna para los 25 ítems de $\alpha = .87$ a $.91$. Definido el número de factores por el criterio de Kaiser y extraídos por Componentes Principales con rotación Varimax, se obtuvo cinco factores con cinco ítems cada uno: utilidad (ítems 1, 6, 11, 20 y 21), ansiedad (ítems 2, 7, 12, 17 y 22), confianza (ítems 3, 8, 13, 18, 23), agrado (ítems 4, 9, 14, 19 y 24) y motivación (ítems 5, 10, 15, 16 y 25). Sus valores de confiabilidad por consistencia interna variaron de $\alpha = .61$ a $.84$. Una vez invertidas las puntuaciones de los cinco ítems de ansiedad y los cinco ítems de motivación (los 10 ítems inversos de la EAE), la puntuación en EAE se obtiene sumando los ítems; una mayor puntuación muestra una actitud más favorable hacia la estadística. Auzmendi (1992) comprobó la validez criterial de la EAE, usando como criterio la Encuesta de Actitud hacia la Estadística de Roberts y Bilderback (1980), y obtuvo una correlación muy alta ($r = .86$). A su vez, la validez concurrente la estableció con la Escala de Actitud hacia los Ordenadores (EAC) de Loyd y Gressard (1984) que cuenta con cuatro factores (ansiedad, confianza, agrado y utilidad). Las correlaciones variaron de moderadas a bajas en la correspondencia de factores

entre las dos escalas: $r = .37$ entre los factores de confianza, $.32$ entre los factores de agrado, $.27$ entre los factores de ansiedad y $.25$ entre los factores de utilidad. El factor de motivación de la EAE, que no tiene una correspondencia directa con ninguno de los cuatro factores de la EAC, tuvo su correlación más alta con el factor de utilidad de la EAC, $r = .23$ (Auzmendi, 1992).

Escala de Actitudes hacia la Investigación (EACIN; Aldana, Caraballo, & Babativa, 2016) con la adaptación a población mexicana de Quezada, Moral y Landero (2019). A la escala original de 34 ítems, Quezada et al. (2019) añadieron un nuevo ítem sobre un aspecto conductual de la actitud (ítem 35 “para ser sincero/a realmente lo que menos hago es leer) y eliminaron tres ítems por baja homogeneidad o consistencia interna (ítems 28, 30 y 32), por lo que se redujo a 32 ítems (EAI-32). Consta de tres factores: afectivo-conductual de ítems directos (ítems 2, 3, 6, 8, 10, 11, 13, 16, 17, 18, 21, 24 y 25; α ordinal = $.90$), cognitivo (ítems 7, 12, 15, 20, 22, 26, 29, 31 y 33; α ordinal = $.89$) y conductual-afectivo de ítems inversos (ítems 1, 4, 5, 9, 14, 19, 23, 27, 34 y 35; α ordinal = $.87$).

En la presente muestra, la confiabilidad por consistencia interna de los 32 ítems de la EAI-32 fue excelente (α ordinal = $.94$) y sus puntuaciones siguieron una distribución normal (prueba de D’Agostino-Pearson: $K^2 = 5.02$, $p = .081$). El factor afectivo-conductual de ítems directos tuvo confiabilidad por consistencia interna excelente (α ordinal = $.90$) y su distribución no se ajustó a la normalidad ($K^2 = 9.12$, $p = .010$), el factor cognitivo presentó confiabilidad por consistencia interna buena (α ordinal = $.89$) y tampoco se ajustó a la normalidad ($K^2 = 51.63$, $p < .001$) y el factor conductual-afectivo de ítems inversos mostró confiabilidad por consistencia interna buena (α ordinal = $.86$) y sí se ajustó a la normalidad ($K^2 = 0.25$, $p = .884$).

Procedimiento

En este estudio instrumental con diseño ex post facto de corte transversal, se usó un muestreo no probabilístico y se aplicó en línea un cuestionario de autorreporte. Seis maestros del área de investiga-

ción invitaron a sus alumnos a contestar un cuestionario en línea a través de la plataforma SurveyMonkey, además solicitaron a los estudiantes que difundieran el enlace con compañeros a través de sus redes sociales. El enlace estuvo disponible de octubre a diciembre del 2017. Siguiendo las normas de investigación de la Sociedad Mexicana de Psicología (2010), el cuestionario electrónico en su primera pantalla solicitaba el consentimiento informado. La participación fue voluntaria, no remunerada y anónima. Los responsables de la investigación estaban identificados, y había una dirección de correo electrónico para resolver dudas.

Análisis de datos

Los cálculos se hicieron con SPSS 24, módulo R versión 2.4 para SPSS 24, AMOS 16 y Excel 2013 con el complemento Real Statistics Resource Pack. Los contrastes fueron bilaterales con un nivel de significación de .05.

Debido al uso de un muestreo no probabilístico, se contrastó la aleatoriedad de la secuencia de datos (ordenados por el número secuencial asignado en el momento de su registro en línea) en los ítems de la EAE y las puntuaciones totales de EAE y EACIN, usando la prueba de rachas. A su vez, se comprobó la equivalencia entre la muestra y la población (estudiantes de la licenciatura en psicología de una universidad pública mexicana) en frecuencias de sexo y semestre cursado (prueba chi-cuadrado de bondad de ajuste para una muestra) y media de edad (prueba t de Student para una muestra). El tamaño del efecto para la prueba chi-cuadrado se calculó a través del coeficiente W de Cohen ($W < .10$ trivial, $.10$ a $.29$ pequeño, $.30$ a $.49$ medio, y $\geq .50$ grande) y para la prueba t de Student a través de la estadística d de Cohen ($d < 0.20$ trivial, 0.10 a $.49$ pequeño, 0.50 a 0.79 medio, y ≥ 0.80 grande) (Cohen, 1988). El contraste de la significación y estimación del tamaño del efecto de estas tres variables se hizo para ponderar la muestra en caso de un efecto medio o grande en relación con los objetivos de descripción de las distribuciones, comparación de medias y validez concurrente.

Para el primer objetivo sobre la estructura factorial, se contrastó el modelo de Auzmendi (1992) por análisis factorial confirmatorio (AFC), usando mínimos cuadrados no ponderados desde la matriz de correlación policórica, que es el procedimiento más recomendado para analizar variables ordinales, como son los ítems tipo Likert (Holgado, Morata, & Barbero, 2018). La estimación por intervalo de los parámetros se hizo por percentiles corregidos de sesgo con la simulación de 2,000 muestras. La confiabilidad por consistencia interna se determinó por el coeficiente omega (ω) y el coeficiente H de confiabilidad de constructo. Valores de ω y H entre .60 y .69 indican confiabilidad por consistencia interna cuestionable, entre .70 y .79 aceptable, entre .80 y .89 buena y $\geq .90$ excelente (Domínguez, 2016; Viladrich, Angulo, & Doval, 2017). La validez convergente de cada factor se evaluó por la varianza media extraída (*AVE*); valores de *AVE* $> .50$ reflejan validez convergente (Fornell & Larcker, 1981). Una varianza compartida entre dos factores menor que la *AVE* de cada factor fue el criterio de validez discriminante (Fornell & Larcker, 1981).

Ante problemas de validez discriminante, se especificó un modelo bifactor en el que la varianza de cada ítem es explicada por un factor general, un factor específico y un residuo; este tipo de modelo no solo da sostén teórico al cálculo de un puntaje total (factor general), sino que permite redefinir la participación de los factores específicos en la determinación de los ítems y, consecuentemente, replantea el problema de validez discriminante (Domínguez & Rodríguez, 2017). Otra opción hubiera sido especificar un modelo jerarquizado, el cual también justifica el cálculo de un puntaje total. No obstante, este tipo de modelo favorece la contribución del factor general en detrimento de los factores jerarquizados, y por ende incrementa el problema de validez discriminante. Además, usualmente arroja peor ajuste, ya que requiere la condición de tétradas o cocientes unitarios entre pares de productos de covarianzas, por lo que le afecta mucho la existencia de cargas cruzadas y correlaciones entre errores de medida (Mansolf & Reise, 2017). Debido a que el modelo jerarquizado, en la presente muestra, mos-

traba peor ajuste que el modelo bifactor e incrementaba el problema de validez discriminante, se optó por no incluirlo.

En el modelo bifactor, la confiabilidad de los ítems en su doble determinación por el factor general y el específico se estableció por los coeficientes omega ($\omega \geq .70$) y $H (\geq .70)$ y la validez convergente por la AVE ($> .50$). Se valoró la contribución de cada factor (específico y general) en cada dominio de ítems por diez índices: omega jerarquizada del factor específico = $\omega_{h(FE)}$ y del factor general = $\omega_{h(FG)}$, el coeficiente H del factor específico = H_{FE} y el factor general = H_{FG} , varianza común explicada por el factor específico = ECV_{FE} y el factor general = ECV_{FG} , y varianza media extraída por el factor específico = AVE_{FE} y el factor general = AVE_{FG} . Valores en el intervalo (.30, .70) en los índices de ω_h , H y ECV reflejan una contribución sustantiva; valores $\leq .30$ indican una contribución pobre y valores $\geq .70$ una contribución excesiva. Considerando una AVE mínima de .50 y una contribución de al menos 30%, el valor mínimo de AVE del factor específico o general debería ser .15 y el máximo de .70 (Domínguez & Rodríguez, 2017).

En el AFC, el ajuste de los modelos se valoró por ocho índices: χ^2/gl = chi-cuadrada relativa, GFI = índice de bondad de ajuste, $AGFI$ = índice de bondad de ajuste corregido, NFI = índice normado de ajuste, CFI = índice comparativo de ajuste, RFI = índice relativo de ajuste, $SRMR$ = residuo estandarizado cuadrático medio y $RMSEA$ = error de aproximación cuadrático medio. Valores de $\chi^2/gl \leq 2$, GFI , NFI , CFI y $RFI \geq .95$, $AGFI \geq .90$, y $SRMR$ y $RMSEA \leq .05$ reflejan buen ajuste; y valores de $\chi^2/gl \leq 3$, GFI , NFI , CFI y $RFI \geq .90$, $AGFI \geq .85$, $SRMR < .10$ y $RMSEA < .08$ reflejan un ajuste aceptable (Byrne, 2016). Se valoró la parsimonia por la razón de parsimonia (PR) o cociente entre los grados de libertad del modelo especificado y el independiente ($PR \geq .75$ alta y $\geq .50$ media) y tres índices parsimoniosos: $PNFI$ = índice normado de ajuste parsimonioso, $PCFI$ = índice comparativo de ajuste parsimonioso, y $PGFI$ = índice de bondad de ajuste parsimonioso; valores de $PNFI$ y $PCFI \geq .80$ y $PGFI \geq .70$ reflejan una relación ajuste-parsimonia buena, y $PNFI$ y $PCFI \geq .60$ y $PGFI \geq .50$ aceptable (Byrne, 2016).

Al no confirmarse el modelo de cinco factores, ya sea de factores correlacionados o bifactor, se exploró uno nuevo, usando análisis factorial exploratorio (AFE). Se siguió un enfoque hipotético deductivo, explorando y contrastando los modelos en la misma muestra, por lo que finalmente se trata de un contraste en sentido débil; un contraste en sentido fuerte requeriría que se hiciera en una nueva muestra extraída de la misma población (Flora & Flake, 2017). El número de factores se determinó por el análisis paralelo de Horn, coordenadas óptimas y media mínima de las correlaciones parciales al cuadrado. Son los métodos que proporcionan mejor resultado para averiguar el verdadero número de factores en los estudios metodológicos de simulación, especialmente cuando convergen en un mismo número de factores (Courtney, 2013). La extracción de factores se hizo por mínimos cuadrados no ponderados desde la matriz de correlación policórica y la rotación de la matriz de cargas factoriales por Promax, siguiendo las sugerencias de Jöreskog, Olsson y Wallentin (2016).

Para el segundo objetivo de describir las distribuciones, se calcularon estadísticos descriptivos y se contrastó el ajuste de las puntuaciones en la puntuación total y los factores a una distribución normal por la prueba K^2 de D'Agostino-Pearson. Ya que el semestre sesgó la composición de la muestra con un tamaño del efecto grande, los cálculos se hicieron tanto en la muestra ponderada como en la muestra sin ponderar. Se compararon las medias de las puntuaciones en la escala y sus factores entre las dos estimaciones por la prueba t de Student para dos muestras independientes y las varianzas por la prueba F de Snedecor.

Para el tercer objetivo de comparar las medias entre ambos sexos se usó la prueba t de Student para dos muestras independientes y el tamaño del efecto se estimó por la g de Hedges. Valores $g < 0.20$ reflejan un tamaño del efecto trivial, 0.10 a .49 pequeño, 0.50 a 0.79 medio, y ≥ 0.80 grande (Hedges & Olkin, 1985). Los cálculos se ejecutaron en la muestra ponderada y sin ponderar. En las variables en las que hubo diferencia por sexo, adicionalmente, sus medias se compararon por análisis de varianza de dos factores de grupos independientes, introduciendo el sexo y el semestre como factores; asimismo, las medias

de las variables diferenciales por sexo se compararon por semestre en la muestra de mujeres y en la de hombres por medio de un análisis de varianza de un factor de grupos independientes. El tamaño del efecto se midió por el coeficiente eta parcial al cuadrado en el ANOVA de dos factores (η_p^2) y eta al cuadrado (η^2) en el ANOVA de un factor. Valores de η_p^2 y $\eta^2 < .01$ reflejan un tamaño del efecto trivial, de .01 a .05 pequeño, de .06 a .13 mediano y $\geq .14$ grande (Cohen, 1988).

Para el cuarto objetivo de validez concurrente, las correlaciones se calcularon por el coeficiente de Pearson (r); valores entre .30 y .49 se interpretaron como una fuerza de asociación media, entre .50 y .69 fuerte, entre .70 y .89 muy fuerte, y $\geq .90$ perfecta (Byrne, 2016; Cohen, 1988). Los cálculos también se hicieron tanto en la muestra ponderada como en la muestra sin ponderar. La diferencia entre las correlaciones de las dos estimaciones se contrastó por la prueba Z de Fisher para dos muestras independientes.

Resultados

Contraste de la aleatoriedad de los datos y equivalencia entre la muestra y la población en sexo, edad y semestre cursado

Al contrastar la aleatoriedad de la secuencia de 383 datos para cada uno de los 25 ítems que integran la escala EAE, así como para las puntuaciones en EAE ($n_o = 184$, $n_i = 199$, $R = 194$, $Z = 0.18$, $p = .854$) y EACIN ($n_o = 208$, $n_i = 175$, $R = 199$, $Z = 0.82$, $p = .414$), se mantuvo la hipótesis nula de secuencia aleatoria en todos los casos.

Los porcentajes de mujeres y hombres en la muestra (73.6% y 26.4%, respectivamente) fueron estadísticamente equivalentes (prueba chi-cuadrado de bondad de ajuste: $\chi^2[1, N = 383] = 3.84$, $p = .536$) a los porcentajes dentro de la población de estudiantes de licenciatura (75% y 25%). No obstante, hubo una sobrerrepresentación de alumnos de primer y cuarto semestre en la muestra (prueba chi-cuadrado de bondad de ajuste: $\chi^2[9, N = 383] = 370.38$, $p < .001$), siendo el tamaño del efecto sobre la composición muestral grande (W

de Cohen = $0.98 > 0.50$). Consecuentemente, el promedio de edad en la muestra ($M = 19.40$ años, IC 95% [19.10, 19.70]) fue significativamente menor (prueba t de Student para una muestra: $t[382] = -2.65$, $p = .008$;) que el de la población ($M = 19.80$, IC 95% [19.76, 19.84]), aunque el tamaño del efecto de la edad sobre la composición muestral fue trivial (d de Cohen = $0.14 < 0.20$).

Los cálculos para la descripción de la distribución de las puntuaciones en la escala y sus factores (segundo objetivo), comparación de medias en la escala y sus factores entre ambos sexos (tercer objetivo) y validez concurrente en relación con actitud hacia la investigación (cuarto objetivo), se hicieron tanto con la muestra ponderada por los porcentajes por semestre como en la muestra sin ponderar, debido al tamaño del efecto grande del semestre sobre la composición muestral. No obstante, la prueba de validez estructural o análisis sobre los ítems (primer objetivo) se hicieron solo en la muestra sin ponderar por la dificultad de cálculo con variables ponderadas.

Contraste del modelo hipotético de cinco factores correlacionados

Todos los parámetros fueron significativos. El ajuste del modelo fue bueno por siete de los ocho índices y aceptable por *SRMR* (Tabla 1). Los factores de ansiedad, confianza y agrado mostraron confiabilidad por consistencia interna buena, $.80 \leq \omega$ y $H < .90$, y validez convergente, $AVE > .50$. El factor de utilidad tuvo confiabilidad por consistencia interna buena y una *AVE* de $.49$. La confiabilidad por consistencia interna del factor de motivación fue aceptable, $.70 \leq \omega$ y $H < .80$, pero no tuvo validez convergente, $AVE = .34 \leq .50$ (Tabla 1). Dos correlaciones entre factores fueron mayores que $.80$ y no hubo validez discriminante en cinco de los diez emparejamientos entre factores.

Tabla 1*Índices de ajuste y parsimonia*

Índice de ajuste	Interpretación		Modelos					
	Bueno	Acep.	5Fc-25	BF-5FE-23	4Fc-25	BF-4FE-23	3Fc-25	BF-3FE-21
χ^2			473.482	620.505	689.582	461.511	729	262.824
gl			265	207	269	207	272	168
χ^2/gl	≤ 3	≤ 2	1.787	2.998	2.562	2.230	2.680	1.564
GFI	$\geq .95$	$\geq .90$	.975	.964	.964	.970	.962	.981
$AGFI$	$\geq .90$	$\geq .85$	.970	.951	.956	.960	.954	.974
NFI	$\geq .95$	$\geq .90$	.968	.953	.954	.960	.951	.975
CFI	$\geq .95$	$\geq .90$	.986	.968	.971	.977	.946	.968
RFI	$\geq .95$	$\geq .90$	.964	.942	.948	.951	.969	.991
$RMSEA$	$\leq .05$	$< .08$	.045	.072	.064	.057	.066	.038
$SRMR$	$\leq .05$	$< .10$	.067	.082	.080	.066	.084	.058
PR	$A \geq .75$	$M \geq .50$	.833	.818	.897	.818	.907	.800
$PNFI$	$\geq .80$	$\geq .70$	.855	.779	.855	.785	.862	.780
$PCFI$	$\geq .80$	$\geq .70$	.870	.792	.871	.799	.879	.793
$PGFI$	$\geq .70$	$\geq .60$	.795	.723	.798	.727	.805	.714

Nota. Índices: χ^2 = valor mínimo de la función de discrepancia o chi-cuadrado, gl = grados de libertad, χ^2/gl = chi-cuadrada relativa, GFI = índice de bondad de ajuste, $AGFI$ = índice de bondad de ajuste corregido, NFI = índice normado de ajuste, CFI = índice comparativo de ajuste, RFI = índice relativo de ajuste, $RMSEA$ = error de aproximación cuadrático medio, $SRMR$ = residuo estandarizado cuadrático medio, PR = razón de parsimonia (A = alta y M = media), $PNFI$ = índice normado de ajuste parsimonioso, $PCFI$ = índice comparativo de ajuste parsimonioso, y $PGFI$ = índice de bondad de ajuste parsimonioso. Modelos: 5Fc-25 = Modelo de Auzmendi et al. (1992) de cinco factores correlacionados para los 25 ítems de la EAE, BF-5FE-23 = Modelo bifactor con cinco factores específicos definidos desde Auzmendi et al. (1992) sin los ítems 3 y 23 del factor de confianza, 4Fc-25 = modelo de cuatro factores correlacionados especificados a partir del AFE con los 25 ítems de la EAE, BF-4FE-23 = modelo bifactor con cuatro factores específicos definidos desde el AFE sin los ítems 4 y 14 del factor de agrado-importancia, 3Fc-25 = modelo de tres factores correlacionados con los 25 ítems de la EAE y BF-3FE-21 = modelo bifactor con tres factores específicos y 21 ítems (sin ítems 4, 11, 14 y 23). Datos de entrada: matriz de correlación policórica. Método para optimizar la función de discrepancia: Mínimos cuadrados no ponderados. Análisis unigrupo: $N = 383$. Fuente: elaborada por los autores.

Ante la falta de validez discriminante, se definió un modelo bifactor. Este tipo de modelo puede no solo dar sostenimiento teórico al cálculo de un puntaje total (factor general), sino que permite redefinir la participación de los factores específicos que determinan los ítems. Los pesos de medida del factor específico de confianza no fueron significativos sobre los ítems 3 ($\beta = .15$, 95% IC $[-.01, .24]$, $p = .053$) y 23 ($\beta = .082$, IC 95% $[-.13, .23]$, $p = .443$), por lo que estos dos ítems se eliminaron. Al estimar el modelo bifactor sin los ítems 3 y 23, todos los pesos de medida fueron significativos. El ajuste fue bueno por *GFI*, *AGFI*, *NFI* y *CFI* y aceptable por los cuatro índices restantes (Tabla 1). Los 23 ítems mostraron confiabilidad por consistencia interna excelente, ω y $H \geq .90$, validez convergente, $AVE > .50$, y el factor general y el conjunto de los cinco factores específicos tuvieron contribuciones sustantivas. Los cinco ítems de utilidad y los cinco de ansiedad presentaron confiabilidad por consistencia interna buena, $.80 \leq \omega$ y $H < .90$, validez convergente, $AVE > .50$, y el factor general y el específico tuvieron contribuciones sustantivas. Los tres ítems de confianza y los cinco ítems de agrado también mostraron confiabilidad por consistencia interna buena y validez convergente, pero el factor general tuvo una contribución excesiva en detrimento de estos dos factores específicos. La confiabilidad por consistencia interna del factor de motivación fue aceptable, $.70 \leq \omega$ y $H < .80$, pero no tuvo validez convergente, $AVE = .37 \leq .50$, y la contribución del factor general fue pobre (Tabla 2).

Aún eliminando un ítem o dos del factor de motivación, el problema de validez convergente persistía, por lo que se optó por explorar un nuevo modelo.

Tabla 2
Índices de confiabilidad y validez convergente

Modelo	Factor	No. de ítems	ω_h	Confiabilidad				Validez convergente					
				<i>H</i>				<i>ECV</i>					
				T	FE	FG	T	FE	FG	T	FE	FG	FE
5Fc_25	U	5	.824				.852						.490
	An	5	.874				.917						.591
	C	5	.861				.887						.559
	A	5	.888				.912						.618
	M	5	.718				.734						.342
BF_5FE_23	U	5	.834	.384		.450	.855	.615	.668	.461	.539	.510	.235
	An	5	.883	.365		.519	.912	.642	.769	.410	.590	.613	.251
	C	3	.881	.162		.719	.883	.312	.807	.184	.816	.711	.131
	A	5	.890	.229		.662	.909	.499	.833	.258	.742	.626	.161
	M	5	.731	.467		.264	.752	.634	.433	.645	.355	.365	.236
4Fc_25	FG	23	.964	.372		.591	.974	.868	.939	.379	.621	.553	.209
	S	9	.919				.935					.564	
	A	7	.888				.914					.537	
	U	4	.681				.700					.352	
BF_4FE_23	M	5	.794				.838					.445	
	S	9	.923	.494		.429	.938	.820	.782	.535	.465	.581	.311
	A	5	.838	.189		.650	.860	.508	.769	.266	.734	.526	.140
	U	5	.832	.208		.624	.908	.702	.763	.331	.669	.535	.177
	M	4	.698	.426		.271	.736	.600	.402	.628	.372	.381	.239
	FG	23	.958	.396		.562	.971	.904	.915	.443	.557	.524	.232

Modelo	Factor	No. de ítems	Confiabilidad			Validez convergente					
			H			ECV			AVE		
			T	FE	FG	T	FE	FG	T	FE	FG
3Fc_25	S	9	.919						.564		
	AU	7	.906						.498		
	M	5	.727						.310		
BF_3FE_21	S	9	.923	.390	.533	.940	.776	.832	.431	.569	.252
	AU	7	.885	.347	.539	.896	.696	.782	.408	.592	.220
	M	5	.735	.398	.337	.757	.599	.506	.556	.444	.372
	FG	21	.955	.416	.539	.965	.879	.905	.444	.556	.231

Nota. M = Modelos factoriales; 5Fc_25 = cinco factores correlacionados y 25 ítems de Auzmendi (1992); U = utilidad: ítems 1, 6, 11, 20, 21, An = ansiedad: ítems 2, 7, 12, 17, 22, C = confianza: ítems 3, 8, 13, 18, 23, A = agrado: ítems 4, 9, 14, 19, 24 y M = motivación: ítems 5, 10, 15, 16, 25. BF_5FE_23 = bifactor con cinco específicos sin los ítems 3 y 23 (U, An, C sin los ítems 3 y 23, A y M) y un factor general (FG). 4Fc = cuatro factores correlacionados: S = seguridad (ítems 2, 3, 7, 8, 12, 13, 17, 18 y 22), A = agrado-importancia (ítems 4, 9, 14, 19, 20, 21 y 24), U = utilidad (ítems 1, 5, 6, 11 y 23) y M = motivación (ítems 10, 15, 16 y 25). BF_4FE_23 = bifactor con cuatro específicos sin los ítems 3 y 23 (S, A sin ítems 4 y 14, U y M). 3Fc_25 = tres factores correlacionados con los 25 ítems: S = seguridad (ítems 2, 3, 7, 8, 12, 13, 17, 18 y 22), AU = agrado-utilidad (ítems 1, 4, 6, 9, 14, 19, 20, 21, 23 y 24) y M = motivación (ítems 5, 10, 11, 15, 16 y 25). BF_3FE_21 = bifactor con tres específicos con 21 ítems (S, AU sin ítems 4, 14 y 23 y M sin ítems 11). Estadísticos: ω = coeficiente omega (T = omega total o en la doble determinación por el factor específico y general, FE = omega jerarquizada del factor específico, $\omega_{h(fg)}$ = omega jerarquizada del factor general), H = índice H (T = total o en la doble determinación por el factor específico y general, FE = del factor específico, FG = del factor general), ECV = varianza compartida explicada (FE = por factor específico y FG = por factor general), AVE = varianza media extraída (T = total o por la doble determinación del factor específico y general, FE = por factor específico, FG = por factor general). En negrilla valores muy bajos o excesivos. Fuente: elaborada por los autores.

Exploración de un nuevo modelo: una estructura de cuatro factores

El número de factores fue cuatro por el análisis paralelo de Horn, coordenadas óptimas y media mínima de las correlaciones parciales al cuadrado. Al extraerlos, se explicó el 56.2% de la varianza total. Al rotar la matriz de cargas factoriales e interpretar los factores desde su contenido, se configuró un primer factor con nueve indicadores (ítems 2, 3, 7, 8, 12, 13, 17, 18 y 22) con confiabilidad por consistencia interna excelente, $\omega = .92$ y $H = .93 \geq .90$, y validez convergente, $AVE = .56 > .50$ (índices calculados desde la matriz estructural). A este primer factor se le denominó seguridad. El segundo factor quedó definido por siete indicadores (ítems 4, 9, 14, 19, 20, 21 y 24) con confiabilidad por consistencia interna buena por $\omega = .89$ y excelente por $H = .91$, así como validez convergente, $AVE = .58 > .50$. A este segundo factor se le denominó agrado-importancia. El tercer factor quedó conformado por cuatro indicadores (ítems 10, 15, 16 y 25) con confiabilidad por consistencia interna aceptable, $.70 \geq \omega = .79$ y $H > .80$, pero sin validez convergente, $AVE = .36 \leq .50$. Este tercer factor corresponde al factor de motivación de Auzmendi (1992), faltándole el ítem 5. El cuarto factor quedó constituido por cinco indicadores (ítems 1, 5, 6, 11 y 23) con confiabilidad por consistencia interna buena, $.80 \geq \omega = .83$ y $H = .82 < .90$, aunque con validez convergente algo débil, $AVE = .45 \leq .50$. A este cuarto factor se le denominó utilidad. Los cuatro factores mostraron validez discriminante.

Cabe señalar que, al forzar la solución a cinco factores, se configuró un primer factor que mezcla los ítems de confianza y ansiedad, salvo el ítem 2 de ansiedad que conformó como indicador único el quinto factor. Por otra parte, los factores, de utilidad, motivación y agrado presentaron algunos ítems mezclados. Claramente, un modelo de cinco factores es inadecuado para los presentes datos.

Contraste del modelo de cuatro factores basado en el AFE

Se especificó un modelo de cuatro factores correlacionados en línea con el modelo original y por ser un modelo más parsimonioso

que el modelo bifactor. Además, permite comprender mejor entre qué factores específicos existen problemas de validez discriminante y si en realidad constituyen un único factor. En el modelo de cuatro factores correlacionados, todos los parámetros fueron significativos. Su ajuste fue bueno por *GFI*, *AGFI*, *NFI* y *CFI* y aceptable por los otros cuatro índices (Tabla 1). No obstante, los valores de motivación y utilidad presentaron validez convergente débil y no hubo validez discriminante entre los factores de agrado-importancia y utilidad, $r^2 = .63 > AVE$ de ambos factores (Tabla 2).

Ante este problema de validez discriminante y no cumplirse con los criterios de validez convergente en dos factores, se volvió a especificar el modelo de cuatro factores desde su configuración en el AFE como un modelo bifactor, como se hizo con el modelo de cinco factores para ver si se resolvían estos problemas. El peso del factor específico de agrado no fue significativo sobre los ítems 4, 9 y 14. Se eliminaron los ítems 4 y 14, pero se conservó el ítem 9, pues pasó a ser significativo. En este modelo, todos los ítems fueron significativos. Su ajuste fue bueno por seis índices y aceptable por χ^2/df y *SRMR* (Tabla 1).

Los 23 ítems en su determinación directa por el factor general y uno de los cuatro factores específicos tuvieron una confiabilidad por consistencia interna excelente, ω y $H \geq .90$, y validez convergente, $AVE = .52 > .50$. Las contribuciones del factor general y el conjunto de los cuatro factores específicos fueron sustantivas. Los nueve ítems de seguridad tuvieron confiabilidad por consistencia interna excelente y los cinco ítems de agrado buena, $.80 \leq \omega$ y $H < .90$. En ambos conjuntos de ítems hubo validez convergente, $AVE > .50$, y las contribuciones del factor general y específico fueron sustantivas. Los cinco ítems de utilidad tuvieron una confiabilidad por consistencia interna buena, $.80 \leq \omega$ y $H < .90$; no obstante, su AVE fue ligeramente menor que .50 ($AVE = .49$) y la contribución del factor específico fue pobre. Los cuatro ítems de motivación tuvieron una confiabilidad por consistencia interna aceptable, $.70 \leq \omega$ y $H < .80$, pero no tuvieron validez convergente, $AVE = .38 < .50$. La contribución del factor específico fue sustantiva, pero la del factor general fue pobre (Tabla 2). Si se elimina

el ítem 25, se fortalece la validez convergente del factor de motivación, $AVE = .42$, pero se debilita la confiabilidad por consistencia interna, $\omega = .68$ y $H = .70$, y persiste la contribución pobre del factor general (Tabla 2).

El replanteamiento del problema de validez discriminante a través de un modelo bifactor no lo resolvió, ya que la contribución del factor general fue excesiva y la del factor específico fue muy pobre en los ítems de agrado-importancia. Por el contrario, el factor general tuvo una contribución muy pobre en los factores de motivación y utilidad, los cuales a su vez mostraron problemas de validez convergente, especialmente el primero.

Exploración y contraste de un modelo de tres factores

Con la intención de juntar los factores de agrado-importancia y utilidad en uno, retomando la información sobre el problema de validez discriminante entre estos dos factores proporcionada por el modelo de factores correlacionados, se exploró una configuración con tres factores. El factor de seguridad se reprodujo con sus nueve ítems. Apareció un factor de agrado-utilidad con 10 ítems (ítems 1, 4, 6, 9, 14, 19, 20, 21, 23 y 24). En este factor quedaron los siete ítems de agrado (ítems 4, 9, 14, 19, 20, 21 y 24) y tres ítems de utilidad, cuando el ítem 21 ya estaba incluido. El ítem 11 de utilidad pasó al tercer factor de motivación que ahora cuenta con sus cinco indicadores originales (ítems 5, 10, 11, 15, 16 y 25).

Como en los modelos de cinco y cuatro factores, se inició especificando un modelo de factores correlacionados. Al contrastar este modelo de tres factores correlacionados, el ajuste fue bueno por GFI , $AGFI$, NFI y CFI y aceptable por los otros cuatro índices (Tabla 1). Persistió la validez convergente pobre del factor de motivación, $AVE = .31$, con una confiabilidad por consistencia interna aceptable, $\omega = .73$ y $H = .74$. Además, hubo falta de validez discriminante entre agrado-utilidad y motivación, $r^2 = .35 > AVE$ de motivación (Tabla 2).

Debido a los problemas de validez discriminante entre dos factores y validez convergente en un factor, se reespecificó como un modelo bifactor. Se halló que el factor específico de agrado-utilidad no tenía pesos de medida significativos sobre los ítems 4, 14 y 23. Al eliminar estos tres ítems, el peso de medida del factor de motivación fue muy pobre sobre el ítem 11, $\lambda = -.15$, IC 95% $[-.27, -.02]$, por lo que también se eliminó. De este modo, se obtuvo el modelo más satisfactorio (Figura 1). Todos sus parámetros fueron significativos. El ajuste a los datos del modelo fue bueno por siete de los ocho índices y aceptable por uno, $SRMR = .06$. Su parsimonia fue alta, $PR = .80$. La relación entre ajuste y parsimonia fue buena por $PGFI$ y aceptable por $PNFI$ y $PCFI$ (Tabla 1).

En su determinación por el factor general y uno de los tres factores específicos, los 21 ítems tuvieron una confiabilidad excelente, ω y $H = .97 \geq .90$, y validez convergente ($AVE > .50$). Además, las contribuciones del factor general y el conjunto de los cuatro factores específicos fueron sustantivas. Los nueve ítems de seguridad tuvieron confiabilidad por consistencia interna excelente y los siete ítems de agrado-utilidad buena, $.80 \leq \omega$ y $H < .90$. Ambos conjuntos de ítems tuvieron validez convergente, $AVE > .50$, y las contribuciones de los factores general y específico fueron sustantivas. La confiabilidad por consistencia interna de los cinco ítems de motivación fue aceptable, $.70 \leq \omega$ y $H < .80$ y el factor general y el específico tuvieron contribuciones sustantivas, pero no tuvieron validez convergente, $AVE = .37 \leq .50$ (Tabla 2).

Al explorar el número de factores sin los ítems 4, 11, 14 y 23, los tres criterios empíricos convergieron en tres y los factores especificados en el AFC se conformaron en la matriz de configuraciones. Consecuentemente, la configuración de los tres factores se reproduce al explorar la matriz de correlaciones.

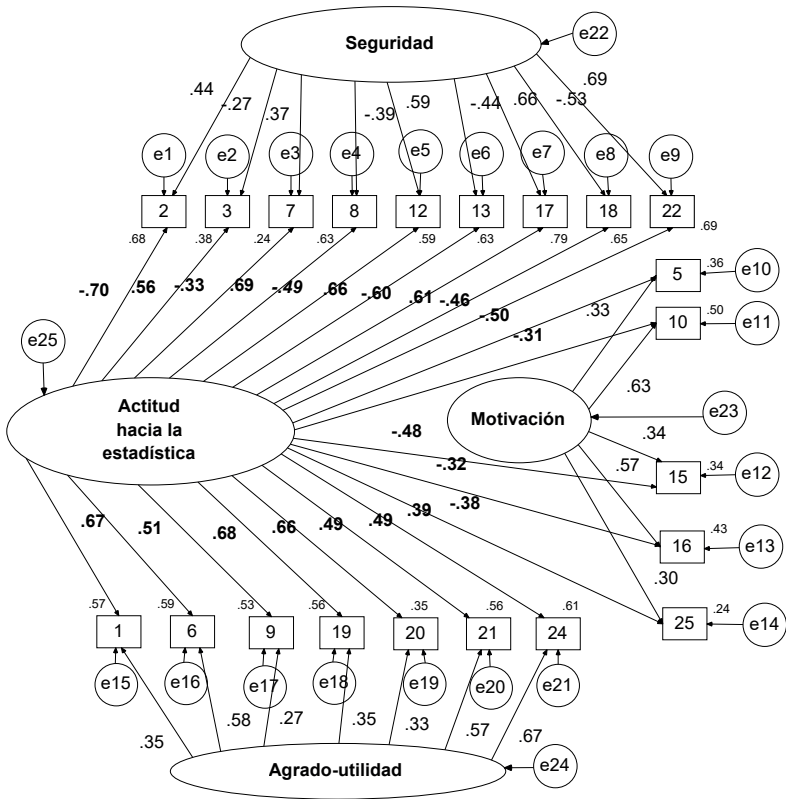


Figura 1. Modelo bifactor con tres factores específicos y 21 ítems estimado por mínimos cuadrados no ponderados. Fuente: elaborada por los autores.

Distribuciones de las puntuaciones en la EAE-21 y sus tres factores

Usando el modelo bifactor de tres factores específicos con 21 ítems (EAE-21), las puntuaciones se obtuvieron a través del cociente entre la suma de los ítems puntuados en sentido de aceptación y el número de ítems sumados. Se interpretaron en sentido absoluto, al dividir el continuo de puntuaciones de 1 a 5 en cinco intervalos en correspondencia

Tabla 3

Estadísticos descriptivos y contraste de la normalidad de EAE-21 en la muestra sin ponderar y ponderada por el semestre

Estadísticos	Sin ponderar				Ponderado por semestre			
	Factor general	Seguridad	Agrado-utilidad	Motivación	Factor general	Seguridad	Agrado-utilidad	Motivación
<i>Mín</i>	1.52	1	1	1	1.52	1	1	1
<i>Máx</i>	4.95	5	5	5	4.95	5	5	5
<i>M</i>	3.22	3.32	3.12	3.17	3.25	3.30	3.19	3.25
<i>Mo</i>	3	3	3.29	3	3.62	3.22	3	3
<i>DE</i>	0.57	0.73	0.71	0.63	0.65	0.81	0.82	0.65
<i>Z_{sk}</i>	1.54	-0.86	1.45	1.58	0.31	-1.05	1.14	0.85
<i>Z_{K3}</i>	1.72	1.68	0.49	3.12	0.35	0.54	-1.79	0.76
<i>P10</i>	2.54	2.44	2.29	2.40	2.48	2.11	2.14	2.40
<i>P20</i>	2.81	2.78	2.57	2.76	2.76	2.78	2.57	2.80
<i>P25</i>	2.90	3	2.57	2.80	2.88	3	2.57	2.80
<i>P30</i>	2.95	3	2.71	2.80	2.95	3	2.71	3
<i>P40</i>	3.05	3.11	3	3	3.05	3.11	3.00	3
<i>P50</i>	3.19	3.33	3.14	3.20	3.19	3.22	3.14	3.20
<i>P60</i>	3.29	3.44	3.29	3.20	3.38	3.44	3.29	3.40
<i>P70</i>	3.48	3.67	3.43	3.40	3.62	3.67	3.43	3.60
<i>P75</i>	3.57	3.78	3.43	3.60	3.62	3.78	3.71	3.60
<i>P80</i>	3.67	3.89	3.71	3.60	3.76	3.89	4.00	3.80
<i>P90</i>	3.95	4.22	4.14	4	4.24	4.22	4.43	4.20
<i>K²</i>	5.35	3.56	2.33	12.25	0.22	1.39	4.49	1.30
<i>p</i>	.069	.169	.311	.002	.897	.498	.106	.522

Nota. *N* = 383. Modelo bifactor con 3 factores específicos y 21 ítems: Seguridad = ((6 - i2) + i3 + (6 - i7) + i8 + (6 - i12) + i13 + (6 - i17) + i18 + (6 - i22))/9, Agrado-utilidad = (i1 + i6 + i9 + i19 + i20 + i21 + i24)/7, y Motivación = ((6 - i5) + (6 - i10) + (6 - i15) + (6 - i16) + (6 - i25))/5. *Mín* = valor mínimo, *Máx* = valor máximo, *M* = media aritmética, *Mo* = moda, *DE* = desviación estándar, *Z_{sk}* = valor estandarizado del coeficiente de asimetría basado en el tercer momento, *Z_{K3}* = valor estandarizado del exceso de curtosis, *P10* a *P90* = percentiles. Contraste de la normalidad por la prueba *K²* de D'Agostino-Pearson: *K²* = estadístico de contraste, *p* = probabilidad a la cola derecha. Los cálculos se hicieron con la muestra sin ponderar y con la muestra ponderada por el semestre (0.35 para el primer semestre, 1.27 para el segundo, 1.76 para el tercero, 0.45 para el cuarto, 1.62 para el quinto, 4.76 para el sexto, 2.90 para el séptimo, 3.89 para el octavo, 7.50 para el noveno y 7.50 para el décimo). Fuente: elaborada por los autores.

Comparaciones de medias entre ambos sexos

Al comparar las medias de las puntuaciones en EAE-21 (factor general) y sus tres factores específicos entre ambos sexos, no hubo diferencias significativas en la muestra sin ponderar. En la muestra ponderada por el semestre, la media de los hombres fue significativamente más alta que la de las mujeres en el factor general (puntuación total de EAE-21) con un tamaño del efecto pequeño y en el factor específico de seguridad con un tamaño del efecto medio (Tabla 4).

Tabla 4

Diferencia de medias por sexo en la muestra ponderada por semestre

EAE-21	Sexo	Descriptivos		Levene		Student			<i>g</i> de Hedges
		<i>M</i>	<i>SD</i>	<i>F</i>	<i>p</i>	<i>t</i>	<i>gl</i>	<i>p</i>	
Factor general	Mujer	3.20	0.65	0.01	.948	-2.63	381	.009	-0.32
	Hombre	3.41	0.63						[-0.56, -0.09]
Seguridad	Mujer	3.20	0.79	0.31	.575	-4.46	381	< .001	-0.53
	Hombre	3.62	0.78						[-0.77, -0.29]
Agrado-utilidad	Mujer	3.16	0.81	0.53	.469	-1.10	381	.273	-0.13
	Hombre	3.27	0.86						[-0.37, 0.10]
Motivación	Mujer	3.27	0.65	0.49	.483	0.72	381	.473	0.09
	Hombre	3.21	0.65						[-0.14, 0.33]

Nota. Descriptivos: *M* = media aritmética y *DE* = desviación estándar. Prueba de homogeneidad de varianzas de Levene: *F* = estadístico de contraste, *p* = valor de probabilidad en la cola derecha. Prueba *t* de Student, asumiendo homogeneidad de varianzas: *t* = estadístico de contraste, *gl* = grados de libertad, *p* = probabilidad en un contraste a dos colas. Tamaño del efecto: *g* de Hedges en su estimación puntual y con un intervalo de confianza al 95%. Fuente: elaborada por los autores.

En la muestra ponderada, hubo diferencia significativa de medias en EAE-21 por semestre ($F[8, 92] = 9.81, p < .001$), y el tamaño del

efecto del semestre sobre la actitud hacia la estadística fue muy grande ($\eta^2 = .49$). La tendencia fue a incrementarse el promedio según el semestre es más avanzado ($R^2 = .25$ en el modelo general lineal; $r_s = .48$, $p < .001$). En la muestra ponderada de mujeres, no hubo diferencia significativa de medias en EAE-21 por semestre ($F[9, 281] = 1.58$, $p = .123$; $\eta^2 = .05$). En la muestra sin ponderar, estas diferencias no fueron significativas ni en hombres ($F[8, 82] = 1.91$, $p = .067$) ni en mujeres ($F[9, 281] = 1.06$, $p = .382$).

En el factor de seguridad, la diferencia de medias fue significativa en la muestra ponderada de ambos sexos, aunque el efecto fue mayor en hombres ($F[8, 96] = 8.01$, $p < .001$; $\eta^2 = .44$) que en mujeres ($F[9, 281] = 3.34$, $p = .001$; $\eta^2 = .10$). Además, en hombres, hubo una tendencia lineal, es decir, la seguridad fue mayor en los últimos semestres que en los primeros ($R^2 = .21$; $r_s = .55$, $p < .001$), cuando no hubo ninguna tendencia lineal o monótona creciente en mujeres ($R^2 = .01$; $r_s = -.02$, $p = .795$). En la muestra sin ponderar, estas diferencias no fueron significativas ni en hombres ($F[8, 96] = 1.76$, $p = .096$) ni en mujeres ($F[9, 281] = 1.61$, $p = .110$).

Al realizar un análisis de varianza con la puntuación total del EAE-21 como variable comparada y el sexo y semestre como factores fijos en la muestra ponderada, el efecto de la interacción entre el semestre y el sexo fue significativo ($F[6, 280] = 6.12$, $p < .001$) y de tamaño medio ($\eta_p^2 = .12$), al igual que el efecto del semestre ($F[7, 280] = 3.60$, $p = .001$; $\eta_p^2 = .08$). El efecto del sexo fue significativo ($F[1, 280] = 7.79$, $p = .006$), pero de tamaño pequeño ($\eta_p^2 = .03$). En la muestra sin ponderada, ni el semestre, ni el sexo ni su interacción tuvieron un efecto significativo.

Al hacer este análisis con el factor de seguridad como variable comparada, los efectos del semestre, el sexo y la interacción entre ambos fueron significativos en la muestra ponderada. Por orden de tamaño del efecto, como en la puntuación total, primero quedó el efecto de la interacción y su tamaño fue grande ($F[6, 280] = 7.30$, $p < .001$; $\eta_p^2 = .14$), luego se ubicó el efecto del semestre y su tamaño fue medio ($F[7, 280] = 4.32$, $p < .001$; $\eta_p^2 = .10$) y, por último, el efecto del sexo y su

tamaño también fue medio ($F[1, 280] = 26.61, p < .001; \eta_p^2 = .09$). En la muestra sin ponderada, ni el semestre, ni el sexo ni su interacción tuvieron un efecto significativo.

Validez de constructo convergente

En la muestra sin ponderar por el semestre, el factor general y los tres factores específicos correlacionaron significativa y positivamente con la EAI-32 y sus tres factores con valores de .53 a .19 con media de .37. Entre los factores, el de agrado-utilidad tuvo las correlaciones más altas, salvo con el factor conductual afectivo de ítems inversos de la EAI-32, cuya correlación más alta fue con motivación de la EAE-21 (Tabla 4).

Al hacer estos mismos cálculos en la muestra ponderada por el semestre, 12 de las 16 correlaciones se incrementaron (promedio de .42, con un mínimo de .13 y máximo de .55) y se cumplió el supuesto de normalidad bivariada en todos los casos. Al hacer las comparaciones por la prueba Z para dos muestras independientes, hubo cuatro diferencias significativas en un contraste a dos colas. Las correlaciones del factor de motivación de la EAE-21 con la puntuación total del EAI-32 ($Z = 2, p = .045$) y sus factores afectivo-conductual de ítems directos ($Z = 2.29, p = .022$) y cognitivo ($Z = 2.33, p = .020$) fueron más altas en la muestra ponderada que en la muestra sin ponderar. También la correlación entre agrado-utilidad de EAE-21 y el factor conductual-afectivo de ítems inversos de EAI-32 fue significativamente más alta ($Z = 2.45, p = .014$) en la muestra ponderada que en la muestra sin ponderar. No obstante, si se aplicase la corrección de Bonferrini al nivel de significación, ninguna diferencia sería estadísticamente significativa: $\alpha = .05/16 = .003$ (Tabla 5).

Tabla 5

Correlaciones de la EAE-21 con la escala de actitud a la investigación en la muestra sin ponderar y ponderada por el semestre y comparación de ambas estimaciones

EAE-21	Actitud hacia la investigación			
	EAI-32	AC-ID	COG	CA-II
Sin ponderar				
Factor general	.514 [.430, .595] ^b	.425 [.329, .515] ^b	.343 [.254, .432] ^b	.500*** [.413, .587]
Seguridad	.371 [.270, .464] ^b	.284 [.179, .385] ^b	.256 [.159, .345] ^b	.384*** [.291, .477]
Agrado-utilidad	.494 [.397, .582] ^b	.483 [.376, .575] ^b	.350 [.252, .437] ^b	.366 [.255, .466] ^b
Motivación	.380 [.274, .471] ^b	.242 [.136, .346] ^b	.205 [.090, .312] ^b	.502*** [.415, .589]
Ponderado por semestre				
Factor general	.522*** [.445, .591]	.469*** [.387, .544]	.293*** [.199, .382]	.549*** [.475, .615]
Seguridad	.337*** [.245, .423]	.312*** [.219, .400]	.125* [.025, .222]	.389*** [.301, .471]
Agrado-utilidad	.533*** [.457, .601]	.497*** [.418, .569]	.334*** [.242, .420]	.509*** [.431, .580]
Motivación	.497*** [.418, .569]	.391*** [.303, .473]	.360*** [.270, .444]	.536*** [.461, .604]
Sin ponderar - Ponderada				
Factor general	-0.151 ^{ns}	-0.758 ^{ns}	0.767 ^{ns}	-0.932 ^{ns}
Seguridad	0.536 ^{ns}	-0.424 ^{ns}	1.877 ^{ns}	-0.081 ^{ns}
Agrado-utilidad	-0.730 ^{ns}	-0.254 ^{ns}	0.250 ^{ns}	-2.448 [*]
Motivación	-2.002 [*]	-2.289 [*]	-2.329 [*]	-0.642 ^{ns}

Nota. Significación contrastada bilateralmente por la prueba t con 381 grados de libertad: * $p < .05$, ** $p < .01$, *** $p < .001$ en un contraste a dos colas. Se comprobó la normalidad bivariada por las pruebas basadas en la asimetría (U-test) y curtosis (W-tets) ^b Estimación por intervalo con un nivel de confianza al 95% usando el método de muestreo repetitivo de percentiles con una simulación de 2,000 muestras, al no mantenerse el supuesto de normalidad bivariada. En todos los casos el sesgo fue menor que .001. EAI-32 = puntuación total del EAI-32, AC-ID = factor afectivo-conductual de ítems directos, COG = factor cognitivo, CA-II = factor conductual-afectivo de ítems inversos. Los cálculos se hicieron con la muestra sin ponderar y con la muestra ponderada por el semestre (0.35 para el primer semestre, 1.27 para el segundo, 1.76 para el tercero, 0.45 para el cuarto, 1.62 para el quinto, 4.76 para el sexto, 2.90 para el séptimo, 3.89 para el octavo, 7.50 para el noveno y 7.50 para el décimo). Comparación entre las correlaciones estimadas con la muestra sin ponderar y ponderada por el semestre: Z = estadístico basado en la transformación de Fisher para dos muestras independientes; probabilidad bajo la hipótesis nula de diferencia nula en un contraste a dos colas: ns = $p \geq .05$ y * $p < 0.05$. Fuente: elaborada por los autores.

Discusión

El primer objetivo del estudio se enfocó a determinar la estructura factorial de la EAE. Se partió del modelo de cinco factores correlacionados de Auzmendi (1992), ya que es el modelo original y más usado en los estudios empíricos (Bautista et al., 2016; Pérez et al., 2015; Torres et al., 2015). Su ajuste a los datos y la relación entre ajuste y parsimonia fueron buenos. Como debilidades, el factor de motivación careció de validez convergente, lo que ocurrió también en los otros modelos contrastados, por lo que no sería una debilidad tan crítica. No obstante, en cinco de las diez relaciones entre pares de factores no hubo validez discriminante y, al explorar nuevos modelos, los criterios empíricos para determinar el número de factores indicaron menos de cinco factores. Precisamente, al extraer cinco factores con los 25 ítems, se evidenció que los ítems de confianza y ansiedad cargaban en un solo factor, configurándose el quinto factor con un solo ítem de ansiedad. Si se considera su contenido, “la estadística no se me da”, más que ansiedad corresponde a autoeficacia o confianza en la propia capacidad para el desempeño en estadística. Por tanto, confianza (seguridad, no ansiedad) y ansiedad (desconfianza, inseguridad) son los polos de un mismo factor. Además, los factores de utilidad y agrado presentaron algunos ítems mezclados, por lo que el modelo de cinco factores de Auzmendi (1992) no se reprodujo en la exploración factorial de los datos.

Sin abandonar el modelo de cinco factores de Auzmendi (1992), se optó por dos variantes de este que sustentan con más fuerza el cálculo de un puntaje total. Por una parte, se definió un modelo bifactor, en el cual la varianza de cada ítem es explicada por un factor general, un factor específico y un residuo; en este modelo se atribuye la varianza compartida entre los factores al factor general, por lo que los factores específicos aparecen como independientes. Además, se replantea la validez discriminante en términos de contribución sustantiva del factor general y específico en cada dominio de contenido (Domínguez & Rodríguez, 2017). Por otra parte, se especificó un modelo jerarquizado,

en el cual un factor de orden superior determina a los cinco factores de orden inferior, pero este último se omitió debido a su peor ajuste en comparación con los otros modelos, aparte de mostrar problemas de validez convergente en tres de sus factores.

En el modelo bifactor, se logró una estimación con parámetros significativos, buen ajuste a los datos, parsimonia alta y una relación entre ajuste y parsimonia de aceptable a buena, al eliminar dos ítems del factor específico de confianza. Cabe señalar que los dos reactivos eliminados (ítems 3 y 23) no habían sido reportados como problemáticos en otros estudios de validación con modelos de factores correlacionados (Bologna & Vaiman, 2013; Darías, 2000; Méndez & Macía, 2007; Tejero & Castro, 2011). Esto indica que estos dos ítems evalúan la actitud hacia la estadística de una forma general sin referencia específica a los contenidos de confianza. Como debilidades, su ajuste, parsimonia y relación entre ajuste y parsimonia fueron menores que en el modelo de factores correlacionados, el cual conserva los 25 ítems originales de la escala. El factor general tuvo una contribución excesiva en detrimento del factor específico en los tres ítems de confianza y los cinco ítems de agrado. Precisamente, estos fueron los factores afectados por el problema de validez discriminante, por lo que el replanteamiento de la correlación excesiva entre factores no se resolvió con la especificación del modelo bifactor. Además, el factor de motivación no tuvo validez convergente y la contribución del factor general fue pobre en el dominio de contenido de este factor. Por tanto, el modelo bifactor con cinco factores específicos no constituye una mejor opción que el modelo de cinco factores correlacionados. Consecuentemente, se exploró un nuevo modelo.

Al explorar el número de factores, este no fue cinco, como también reportan otros estudios (Darías, 2000; Méndez & Macía, 2007). La convergencia de criterios indicó cuatro factores sustantivos. Al extraer cuatro factores, el primer factor de nueve ítems se asemeja al factor de seguridad definido por Darías (2000) e incluye los cinco ítems del factor de ansiedad de Auzmendi (1992). Se le denominó seguridad/calma. El segundo factor de siete ítems incluye los cinco ítems del

factor de agrado de Auzmendi (1992), pero es más amplio como en la solución de Darías (2000), quien denominó a este factor agrado-importancia, por lo que se le dio este nombre. El tercer factor de cuatro ítems se asemeja al factor de motivación de Auzmendi (1992), al tener cuatro de sus cinco ítems y consecuentemente se le denominó motivación. El cuarto factor con cinco ítems contiene tres de los cinco ítems del factor de utilidad de Auzmendi (1992). Se le ha añadido el ítem 5 que también hace referencia a la utilidad práctica de la estadística y el ítem 23 de confianza en poder dominar la estadística. Sus ítems están incluidos en el factor de aspectos motivacionales de Méndez y Macía (2007) y se le denominó utilidad.

Al contrastar este modelo de cuatro factores correlacionados por AFC, no se requirió eliminar ningún ítem. El ajuste fue de bueno a aceptable con una parsimonia alta. Sus debilidades fueron la confiabilidad por consistencia interna cuestionable y la falta de validez convergente del factor de motivación, como en el contraste del modelo de cinco factores, y la falta de validez discriminante entre agrado-importancia y utilidad. Cabe señalar que el factor de aspectos motivacionales de Méndez y Macía (2007) con ocho ítems no solo incluía los cinco ítems del presente factor de utilidad, sino también tres de los cuatro ítems del factor de motivación. No obstante, en el presente trabajo estos dos factores fueron claramente diferenciables.

Ante la falta de validez discriminante entre agrado-importancia y utilidad se consideró que se podían juntar en un factor. Al explorar la configuración con tres factores, estos 12 ítems sí aparecen juntos en un factor, aunque sin los ítems 5 y 11 que pasan al factor de motivación. Ahora este tercer factor de 10 ítems se asemeja más al de aspectos motivacionales de Méndez y Macía (2007); se le denominó agrado-utilidad, aunque también se le podría nombrar como valoración positiva de la estadística. El factor de motivación con seis indicadores recupera el ítem 5 que aparece en el modelo de Auzmendi (1992) y se le añade el ítem 11 de utilidad. El factor de seguridad conserva sus nueve indicadores. Al contrastar este modelo de tres factores correlacionados, sigue persistiendo la falta de validez convergente del factor de motivación y aparece

falta de validez discriminante entre el factor de agrado-utilidad y el de motivación. Este modelo de tres factores se asemeja mucho al modelo simplificado de calma/ansiedad, predisposición activa o motivación y percepción que se tienen de la utilidad-importancia de la estadística de Tejero y Castro (2011), así como al de Bologna y Vaiman (2013).

Al especificar un modelo bifactor con tres factores específicos como solución al problema de validez discriminante, los ítems 4, 14 y 23 del factor de valoración positiva de la estadística se eliminaron, ya que el factor específico careció de peso significativo sobre los mismos. A su vez, se eliminó el ítem 11 al tener el factor específico de motivación un peso bajo sobre el mismo y no ser congruente con su configuración original. Aun moviendo el ítem 11 con carga simple o cruzada al factor de agrado-utilidad, el resultado era malo, siendo la mejor opción eliminarlo. De este modo, se alcanzó el mejor modelo. Su ajuste fue bueno, su parsimonia alta, su relación ajuste-parsimonia varió de aceptable a buena, las contribuciones del factor general y el específico fueron sustantivas en cada dominio, lo que supera el problema de validez discriminante y los tres factores contaron con confiabilidad. Salvo el factor de motivación, el factor general y los otros dos específicos contaron con validez convergente. Como en los anteriores modelos, el factor de motivación presentó una validez convergente pobre, explicando más de un tercio de varianza extraída de los cinco ítems, aunque menos de un medio.

Fortalezas adicionales de este modelo son que el factor general justifica la obtención de una puntuación total y los factores específicos justifican la obtención de las puntuaciones en tres dominios de contenido: seguridad/calma, predisposición activa o motivación y valoración positiva de la estadística (agrado/utilidad/importancia). Además, los tres criterios empíricos para determinar el número de factores convergieron en tres con 21 ítems y se reprodujo el modelo especificado al extraer los factores y rotar oblicuamente la matriz factorial.

El modelo bifactor de tres factores con 21 ítems resultó la mejor representación de la estructura subyacente a los ítems con clara ventaja teórica. La segunda opción sería el modelo original de Auzmendi

(1992). No obstante, los problemas de validez discriminante entre factores y la no reproducción del modelo a nivel exploratorio, cuestionan este modelo, a pesar de su buen ajuste a los datos.

¿Qué aporta este modelo bifactor con tres factores específicos frente al modelo original de Auzmendi (1992)? Esencialmente una mejor justificación de la selección de ítems para medir dominios específicos de la actitud hacia la estadística y el cálculo de un puntaje total. Estos tres dominios específicos de contenido son una simplificación de los cinco dominios que Auzmendi consideró que se repetían con más frecuencia en el conjunto de escalas de las cuales fue tomando reactivos para conformar inicialmente la suya. Precisamente, los métodos de análisis usados son más sofisticados y adecuados para variables ordinales que los implementados por Auzmendi (1992), basados en el procedimiento Little Jiffy, actualmente cuestionado (Jöreskog et al., 2016). Los resultados no implican una nueva teoría para comprender la actitud, sino una simplificación de la ya existente, e incluso es una propuesta muy afin a dos anteriores también formuladas desde resultados empíricos (Tejero & Castro, 2011; Bologna & Vaiman, 2013).

El segundo objetivo enunciado fue describir la distribución de la escala. Conforme a las expectativas (Flores & Auzmendi, 2015), la distribución del factor general se ajustó a la normalidad, así como las distribuciones de los tres factores específicos. Cabe señalar que la distribución del factor de motivación mostró un histograma con un decaimiento más abrupto hacia las colas que el de una distribución normal en la muestra sin ponderar; no obstante, esta leptocurtosis fue leve y en la muestra ponderada el perfil fue simétrico y mesocúrtico con un buen ajuste a la normalidad. Por tanto, la escala y sus factores pueden ser baremados por puntuaciones T con media 50 y desviación estándar de 10. Se ha señalado que las variables cuantitativas con distribución normal son propias de fenómenos naturales y sociales sometidos a selección por presión adaptativa al entorno (Smerlak & Youssef, 2015). En este caso sería el entorno académico que demanda habilidades en estadística para la comprensión de los reportes de investigación y el análisis de datos.

Las medidas de tendencia central (media aritmética y mediana) resultaron equivalentes al ser estimadas en la muestra ponderada y en la muestra sin ponderar, lo que indica que son estables. Estas medidas pueden ser fácilmente interpretadas en términos absolutos desde las etiquetas de respuesta, cuando se usan puntuaciones promedio por ítem en lugar de puntuaciones suma (Ávila et al., 2015). De acuerdo a esta interpretación, la actitud fue neutra en el factor general de actitud hacia la estadística y los específicos, como en los estudios de Rodríguez (2011) en estudiantes argentinos de humanidades y Baños y Hurtado (2016) en estudiantes de pedagogía españoles. Estos estudios también usaron interpretaciones en sentido absoluto. Aunque domina una actitud neutra que representa a casi tres quintos de los estudiantes en el factor general, un tercio muestra aceptación frente a un décimo rechazo. Por tanto, el balance final se inclina más hacia la aceptación, como en los estudios de Comas et al. (2017) y Walker y Brakke (2017). En el estudio de Bautista et al. (2016) hecho en México, las actitudes en estudiantes de ciencias sociales se polarizaron más hacia la aceptación que en el presente estudio, ya que 15% mostraron rechazo, 40% neutralidad y 45% aceptación.

El tercer objetivo fue comparar los promedios entre ambos sexos sin una expectativa clara. En la muestra sin ponderar, los datos apoyan la hipótesis de no diferencia entre sexos (Comas et al., 2017; Pérez et al., 2015; Salinas & Mayén, 2016; Sarikaya et al., 2018). Sin embargo, los datos de la muestra ponderada por semestre apoyan la hipótesis de diferencia con mayor seguridad y una actitud más positiva en hombres que en mujeres (Auzmendi, 1991; Hommik & Luik, 2017; Opstad, 2020; Rejón-Guardia et al. 2019), siendo el tamaño del efecto del sexo medio sobre la seguridad y pequeño sobre la actitud general. La muestra sin ponderar está sesgada por una mayor participación de los alumnos de los primeros semestres. Al observar la tendencia de las medias en las muestras ponderadas, esta es ascendente en hombres, con una mayor seguridad y una actitud más positiva en los últimos semestres que en los primeros semestres, cuando no hay una tendencia lineal ni monótona creciente en la muestra ponderada de mujeres, ni en la muestra sin ponderar de

ambos sexos. Consecuentemente, la diferencia entre mujeres y hombres es efecto, sobre todo, de una seguridad creciente y una mayor positividad hacia la estadística en los hombres en la medida que van avanzando o acabando la carrera, cuando no es así en mujeres. Cabe señalar que, aunque el efecto de la interacción entre el semestre y el sexo tuvo el mayor peso, el factor sexo fue en sí mismo significativo y con un tamaño del efecto no trivial en la muestra ponderada, cuando se incluyó el sexo y el semestre como factores independientes en un análisis de varianza.

El último objetivo fue comprobar la validez de constructo convergente de la EAE. Para este fin se consideró la actitud hacia la investigación. La expectativa era de correlación directa y alta. Se confirmó la expectativa con correlaciones muy semejantes en la muestra ponderada y sin ponderar. A semejanza del estudio de Baños y Hurtado (2016), el factor de agrado-utilidad o valoración positiva de la estadística tuvo la correlación más alta y el de motivación la más baja.

Los modelos bifactor mostraron que los cinco ítems de motivación son bastante específicos con una relación con la actitud general hacia la estadística menor que la de los otros dominios de contenido. De aquí se podría conjeturar que este factor específico esté más relacionado con la motivación hacia la carrera o con el motivo de logro (Lee, Hayes, Seitz, DiStefano, & O'Connor, 2016), pero los presentes datos no lo confirman, ya que motivación fue independiente de la calificación en la materia de investigación del semestre anterior. No obstante, fue el factor más correlacionado con el factor conductual-afectivo de ítems inversos del ECIN-32. Por tanto, parece estar midiendo un aspecto de predisposición hacia la acción sin implicar resultados efectivos.

Como limitación del estudio, se tiene el uso de un muestreo no probabilístico, por lo que las inferencias deben tomarse con la debida cautela dentro de la población de estudiantes de psicología en la que se colectó la muestra. Su aplicación a otras poblaciones debe realizarse como hipótesis a contrastar y como datos comparativos. Cabe señalar que la aleatoriedad de las series de datos se mantuvo por la prueba de rachas. Además, la muestra presenta equivalencia en porcentajes de sexo a la de su población de estudiantes de licenciatura de la cual fue

extraída, lo que habla de la calidad de los datos obtenidos. No obstante, la muestra presenta un sesgo en cuanto a la composición por semestres; de ahí que se hicieron los análisis con la muestra ponderada y sin ponderar por semestre. En la muestra sin ponderar, están sobre-representados el primer y cuarto semestre en detrimento de los últimos semestres. Por tanto, la muestra representa sobre todo a los alumnos que no han elegido su área de acentuación y que llevan las materias de estadística y metodología más generales.

Se concluye que, en la presente muestra extraída de una población de estudiantes de psicología de una universidad pública mexicana, el modelo de cinco factores correlacionados originariamente propuesto para la EAE es forzado. El número de factores es menor. Tras eliminar cuatro ítems, el mejor modelo es el bifactor con tres factores específicos: seguridad, agrado-utilidad y motivación. Las puntuaciones en el factor general y los tres factores específicos presentan distribución normal, por lo que la EAE-21 puede ser baremada por puntuaciones T. Los hombres presentan mayor seguridad y una actitud más positiva hacia la estadística que las mujeres y esta diferencia es, sobre todo, atribuible a una interacción entre la carrera y el sexo. La seguridad y positividad aumenta en los hombres en la medida que van acabando la carrera cuando no existe esta tendencia en las mujeres. La actitud promedio es neutra; no obstante, un tercio de los estudiantes poseen una actitud polarizada hacia la aceptación contra un décimo polarizada hacia el rechazo, inclinando la balanza hacia la aceptación. A su vez, la EAE-21 muestra validez de constructo convergente en relación con la actitud hacia la investigación.

Se requiere contrastar, en nuevas muestras (de la misma población u otras poblaciones), este modelo frente a otros modelos propuestos, incluido el de Auzmendi (1992), para comprobar que no es solo un artificio de los métodos de análisis usados o una singularidad de la muestra analizada. Además, se sugiere probar la estabilidad temporal del modelo bifactor de tres factores específicos. Finalmente, se recomienda seguir usando este instrumento en la investigación aplicada en el campo educativo.

Referencias

- Aldana, G. M., Caraballo, G. J., & Babativa, D. A. (2016). Escala para medir Actitudes hacia la Investigación (EACIN): validación de contenido y confiabilidad. *Revista Aletheia*, 8(2), 104-121. <https://doi.org/10.11600/21450366.8.2aletheia.104.121>
- Ashaari, N. S., Judi, H. M., Tengku, H. M., & Wook, M. T. (2011). Student's attitude towards statistics course. *Procedia - Social and Behavioral Sciences*, 18, 287-294. <https://doi.org/10.1016/j.sbspro.2011.05.041>
- Auzmendi, E. (1991). Factors related to statistics: a study with a Spanish sample. En AERA (Eds.), *Proceedings of the annual meeting of the American Educational Research Association* (pp. 3-7). Chicago, IL: AERA Press.
- Auzmendi, E. (1992). *Las actitudes hacia la matemática-estadística en las enseñanzas medias y universitarias*. Bilbao: Mensajero.
- Ávila, M. L., Stinson, J., Kiss, A., Brandão, L. R., Uleryk, E., & Feldman, B. M. (2015). A critical review of scoring options for clinical measurement tools. *BMC Research Notes*, 8, (612): 1-11. <https://doi.org/10.1186/s13104-015-1561-6>
- Bailey, T. H., & Phillips, L. J. (2016). The influence of motivation and adaptation on students' subjective well-being, meaning in life and academic performance. *Higher Education Research & Development*, 35(2), 201-216. <https://doi.org/10.1080/07294360.2015.1087474>
- Baños, R. V., & Hurtado, M. R. (2016). Actitudes hacia la estadística en el alumnado del grado de pedagogía de la Universidad de Barcelona. *Revista de Docencia Universitaria*, 14(1), 131-150. <https://doi.org/10.4995/redu.2016.5766>
- Bautista, J., Morales, M., Dórame, L., & Peralta, G. (2016). Un estudio sobre las actitudes hacia la estadística en estudiantes universitarios. *Epistemus*, 20(10), 48-54. <https://doi.org/10.36790/epistemus.v10i20.22>
- Bologna, E. L., & Vaiman, M. (2013). Actitudes, experiencia previa y nivel de logro en estadística en la carrera de psicología. En J. M.

- Contreras, G. R. Cañadas, M. M. Gea & P. Arteaga (Eds.), *Actas de las jornadas virtuales en didáctica de la estadística, probabilidad y combinatoria* (pp. 91-103). Granada: Universidad de Granada.
- Byrne, B. (2016). *Structural equation modeling with AMOS* (3ª ed.). New York, NY: Routledge. <https://doi.org/10.4324/9781315757421>
- Cladera, M., Rejón-Guardia, F., Vich-i-Martorell, G. A., & Juaneda, C. (2019). Tourism students' attitudes toward statistics. *Journal of Hospitality, Leisure, Sport & Tourism Education*, 24(1), 202-210. <https://doi.org/10.1016/j.jhlste.2019.03.002>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2ª ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.
- Comas, C., Martins, J. A., Nascimento, M. M., & Estrada, M. A. (2017). Estudio de las actitudes hacia la estadística en estudiantes de psicología. *Boletim de Educação Matemática*, 31(57), 479-496. <https://doi.org/10.1590/1980-4415v31n57a23>
- Courtney, M. G. R. (2013). Determining the number of factors to retain in EFA: Using the SPSS R-Menu v2.0 to make more judicious estimations. *Practical Assessment, Research & Evaluation*, 18(8), 1-14. <https://doi.org/10.2147/JHL.S35483>
- Darías, E. J. (2000). Escala de actitudes hacia la estadística. *Psicothema*, 12(2), 175-178.
- Domínguez, S. A. (2016). Evaluación de la confiabilidad del constructo mediante el coeficiente H: breve revisión conceptual y aplicaciones. *Psychologia*, 10(2), 87- 94. <https://doi.org/10.21500/19002386.2134>
- Domínguez, S. A., & Rodríguez, A. (2017). Índices estadísticos de modelos bifactor. *Interacciones. Revista de Avances en Psicología*, 3(2), 59-65. <https://doi.org/10.24016/2017.v3n2.51>
- Flora, D. B., & Flake, J. K. (2017). The purpose and practice of exploratory and confirmatory factor analysis in psychological research: Decisions for scale development and validation. *Canadian Journal of Behavioural Science*, 49(2), 78-88. <https://doi.org/10.1037/cbs0000069>

- Flores, L. W. O., & Auzmendi, E. E. (2015). Análisis de la estructura factorial de una escala de actitud hacia las matemáticas. *Aula de Encuentro*, 1(17), 45-77. Recuperado de <https://revistaselectronicas.ujaen.es/index.php/ADE/article/view/2256>
- Fornell, C., & Larcker, D. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18(1), 39-50. <https://doi.org/10.2307/3151312>
- Hedges, L. V., & Olkin, I (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.
- Holgado, F. P., Morata, M. A., & Barbero, M. I. (2018). Confirmatory factor analysis of ordinal variables: A simulation study comparing the main estimation methods. *Avances en Psicología Latinoamericana*, 36(3), 601-617. <https://doi.org/10.12804/revistas.urosario.edu.co/apl/a.4932>
- Hommik, C., & Luik, P. (2017). Adapting the Survey of Attitudes Towards Statistics (SATS-36) for Estonian secondary school students. *Statistics Education Research Journal*, 16(1), 228-239.
- Howard, G. S. (2019). The present and future of methodology and statistics in psychology. *The Humanistic Psychologist*, 47(1), 26-51. <https://doi.org/10.1037/hum0000111>
- Huang, J., Gates, A. J., Sinatra, R., & Barabási, A. L. (2020). Historical comparison of gender inequality in scientific careers across countries and disciplines. *Proceedings of the National Academy of Sciences*, 117(9), 4609-4616; <https://doi.org/10.1073/pnas.1914221117>
- Jöreskog, K. G., Olsson, U. H., & Wallentin, F. Y. (2016). Exploratory factor analysis (EFA). En *Multivariate analysis with LISREL* (pp. 257-282). Cham, Suiza: Springer Series in Statistics. https://doi.org/10.1007/978-3-319-33153-9_6
- Lee, C. S., Hayes, K. N., Seitz, J., DiStefano, R., & O'Connor, D. (2016). Understanding motivational structures that differentially predict engagement and achievement in middle school science.

- International Journal of Science Education*, 38(2), 192-215.
<https://doi.org/10.1080/09500693.2015.1136452>
- Loyd, B. H., & Gressard, C. (1984). Reliability and factorial validity of computer attitude scales. *Educational and Psychological Measurement*, 44, 501-505. <https://doi.org/10.1177/0013164484442033>
- Mahmud, Z., & Zainol, M. S. (2008). Examining postgraduate students' perceived competency in statistical data analysis and their attitudes toward statistics. *International Journal of Education and Information Technologies*, 2(1), 79-86.
- Mansolf, M., & Reise, S. P. (2017). When and why the second-order and bifactor models are distinguishable. *Intelligence*, 61(2), 120-129. <https://doi.org/10.1016/j.intell.2017.01.012>
- Méndez, D., & Macía, F. (2007). Análisis factorial confirmatorio de la escaladeactitudes hacia la estadística. *Cuadernos de Neuropsicología*, 3(1), 174-371. Recuperado de http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S0718-41232007000300017
- Mozafar, D., Akbar, S. A., & Hassan, A. (2017). Effects of mastery learning on academic achievement and attitudes toward statistics in psychology students. *Educational Development of Jundishapur*, 8(1), 126-135.
- Olos, L., & Hoff, E.-H. (2006). Gender ratios in European psychology. *European Psychologist*, 11(1), 1-11. <https://doi.org/10.1027/1016-9040.11.1.1>
- Opstad, L. (2020). Attitudes towards statistics among business students: do gender, mathematical skills and personal traits matter? *Sustainability*, 12(15), 6104. <https://doi.org/10.3390/su12156104>
- Ostertag, P. A., & McNamara, J. R. (1991). "Feminization" of psychology: the changing sex ratio and its implications for the profession. *Psychology of Women Quarterly*, 15, 349-369. <https://doi.org/10.1111/j.1471-6402.1991.tb00413.x>
- Peiró-Signes, Á., Trull, Ó., Segarra-Oña, M., & García-Díaz, J. C. (2020). Attitudes towards statistics in secondary education:

- findings from fsQCA. *Mathematics*, 8(5), article 804. <https://doi.org/10.3390/math8050804>
- Pérez, L. E., Aparicio, A. S., Bazán, J. L., & Abdounur, O. J. (2015). Actitudes hacia la estadística de estudiantes universitarios de Colombia. *Educación Matemática*, 27(3), 111-149.
- Quezada, L. C., Moral, J., & Landero, R. (2019). Estudio de validación de la Escala de Actitud hacia la Investigación en estudiantes mexicanos de psicología. *Evaluar*, 19(1), 1-16. <https://doi.org/10.35670/1667-4545.v19.n1.23874>
- Rejón-Guardia, F., Vich-I-Martorell, G.A., Juaneda, C., & Cladera, M. (2019). Gender differenced in attitudes towards statistics in social science degrees. En L. Gómez-Chova, A. López-Martínez, & I. Candel-Torres (Eds.), *EDULEARN19 Proceedings of the 11th International Conference on Education and New Learning Technologies* (pp. 2933-2941). Palma de Mallorca: IATED Academy. <https://doi.org/10.21125/edulearn.2019.0784>
- Roberts, D. M., & Bilderback, E. W. (1980). Reliability and validity of a "Statistics Attitude Survey". *Educational and Psychological Measurement*, 40, 235-238. <https://doi.org/10.1177/001316448004000138>
- Rodríguez, N. (2011). Actitudes de los estudiantes universitarios hacia la estadística. *Interdisciplinaria*, 28(2), 199-205.
- Rojas, C. A., Escalera, M. E., Moreno, E., & García, A. (2017). Motivación, ansiedad, confianza, agrado y utilidad. Los factores que explican la actitud hacia las matemáticas en los estudiantes de economía. *International Journal of Developmental and Educational Psychology*, 2(1), 527-539. <https://doi.org/10.17060/ijodaep.2017.n1.v2.875>
- Rosli, M. K., & Maat, S. M. (2017). Attitude towards statistics and performance among post-graduate students. *AIP Conference Proceedings*, 1847, article 030004. doi:10.1063/1.4983881
- Rosli, M. K., Maat, S. M., & Rosli, R. (2017). Students' attitude and anxiety towards statistics: A descriptive analysis. *Research on Education and Psychology*, 1(1), 47-56.

- Ruiz de Miguel, C. (2015). Actitudes hacia la estadística de los alumnos del Grado en Pedagogía, Educación Social y Maestro de Educación Primaria en la UCM. *Educación XXI*, 18(2), 351-374. <https://doi.org/10.5944/educXX1.12158>
- Salazar, L. R. (2018). Examining the relationship between math self-efficacy and statistics motivation in the introductory business statistics course: self-compassion as a mediator. *Decision Sciences. Journal of Innovative Education*, 16(2), 140-160. <https://doi.org/10.1111/dsji.12149>
- Salinas, J., & Mayén, S. (2016). Estudio exploratorio de las actitudes hacia la estadística en estudiantes mexicanos de bachillerato. *Avances de Investigación en Educación Matemática*, 10, 73-90. <https://doi.org/10.35763/aiem.v0i10.130>
- Sarikaya, E. E., Ok, A., Aydin, Y. C., & Schau, C. (2018). Turkish version of the Survey of Attitudes Toward Statistics: factorial structure invariance by gender. *International Journal of Higher Education*, 7(2), 121-127. <https://doi.org/10.5430/ijhe.v7n2p121>
- Schau, C., Stevens, J., Dauphinee, T. L., & Del Vecchio, A. (1995). The development and validation of the Survey of Attitudes Toward Statistics. *Educational and Psychological Measurement*, 55(5), 868-875. <https://doi.org/10.1177/0013164495055005022>
- Sesé, A., Jiménez, R., Montaña, J., & Palmer, A. (2015). ¿Pueden las actitudes hacia la estadística y la ansiedad estadística explicar el rendimiento de los estudiantes? *Revista de Psicodidáctica*, 20(2), 285-304. <https://doi.org/10.1387/RevPsicodidact.13080>
- Smerlak, M., & Youssef, A. (2017). Limiting fitness distributions in evolutionary dynamics. *Journal of Theoretical Biology*, 416, 68-80. <https://doi.org/10.1016/j.jtbi.2017.01.005>
- Sociedad Mexicana de Psicología (2010). *Código ético del psicólogo* (5a ed). Ciudad de México: Trillas.
- Tejero, C., & Castro, M. (2011). Validación de la escala de actitudes hacia la estadística en estudiantes españoles de ciencias de la actividad física y del deporte. *Revista Colombiana de Estadística*, 34(1), 1-14.

- Torres, F., Aparicio, A. S., Bazán, J. L., & Abdounur, O. J. (2015). Actitudes hacia la estadística en universitarios del área de las ciencias de una universidad pública de Chile. *Educação Matemática Pesquisa*, 17(1), 45-73.
- Vera, O. Y., & Díaz, C. (2013). Dificultades de estudiantes de psicología en relación al contraste de hipótesis. En J. M. Contreras, G. R. Cañadas, M. M. Gea, & P. Arteaga (Eds.), *Actas de las jornadas virtuales en didáctica de la estadística, probabilidad y combinatoria* (pp. 197-203). Granada, España: Departamento de Didáctica de la Matemática de la Universidad de Granada.
- Vilá, R., & Rubio, M. J. (2016). Actitudes hacia la estadística en el alumnado del grado de pedagogía de la Universidad de Barcelona. *Revista de Docencia Universitaria*, 14(1), 131-149. <https://doi.org/10.4995/redu.2016.5766>
- Viladrich, C., Angulo-Brunet, A., & Doval, E. (2017). A journey around alpha and omega to estimate internal consistency reliability. *Annals of Psychology*, 33(3), 755-782. <https://doi.org/10.6018/analesps.33.3.268401>
- Wagenmakers, E., Marsman, M., Jamil, T., Ly, A. Verhagen, J., Love, J., ... Morey, R. D. (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review*, 25, 35-57. <https://doi.org/10.3758/s13423-017-1343-3>
- Walker, E. R., & Brakke, K. E. (2017). Undergraduate psychology students' efficacy and attitudes across introductory and advanced statistics courses. *Scholarship of Teaching and Learning in Psychology*, 3(2), 132-140. <https://doi.org/10.1037/stl0000088>
- Wright, D. B. (2009). Ten statisticians and their impacts for psychologists. *Perspectives on Psychological Sciences*, 4(6):587-597. <https://doi.org/10.1111/j.1745-6924.2009.01167.x>

Recibido: 2019-08-19

Revisado: 2020-08-16

Aceptado: 2021-03-21