

## Invarianza por Sexo en la Escala de Detección del Trastorno de Ansiedad Generalizada (EDTAG)

### Gender Invariance of the Screening Scale for Generalized Anxiety Disorder (EDTAG)

Fabiola Gonzalez-Betanzos<sup>1</sup>

María Elena Rivera-Heredia<sup>2</sup>

Ferrán Padrós-Blázquez<sup>3</sup>

Universidad Michoacana de San Nicolás de Hidalgo, México

**Resumen.** En el presente trabajo se muestra la forma en la que se evalúa la invarianza utilizando los modelos de análisis factorial confirmatorio para medias y covarianzas (AFC-MACS) para datos categóricos y los modelos de Teoría de Respuesta al Ítem (TRI). Se ejemplifica el análisis de la invarianza en el estudio de la Escala de Detección del Trastorno de Ansiedad Generalizada (EDTAG) comparando hombres y mujeres. La EDTAG es una escala ampliamente utilizada en las instituciones de salud y por sus características (escala breve de 12 ítems dicotómicos) cualquier error de medida puede tener un impacto importante. En los resultados se muestra que la escala tiene la misma configuración, sin embargo, el ítem 9 muestra funcionamiento diferencial siendo los hombres quienes mayor probabilidad tienen de responder de manera afirmativa comparado con mujeres del mismo nivel de rasgo. Se discute la necesidad de éste tipo de análisis en las escalas.

**Palabras clave.** Invarianza, Análisis Factorial Confirmatorio, Teoría de la Respuesta al ítem, Funcionamiento Diferencial del ítem, Trastorno de ansiedad generalizada.

**Abstract.** The present paper shows how to evaluate invariance using confirmatory factor analysis models for means and covariances (CFA-MACS) for categorical data and models of Item Response Theory (IRT). Invariance analysis in the study of Detection Scale Generalized Anxiety Disorder comparing men and women is exemplified. The scale is widely used in health institutions and by its nature (a brief scale of 12 dichotomous items) any measurement error can have a major impact. Results have shown that the scale has the same configuration, but the item 9 shows differential item functioning, being males more likely to respond affirmatively compared to women in the same trait level. The need for this type of analysis on the scales is discussed.

**Keywords.** Invariance, Confirmatory Factor Analysis, Item Response Theory, Differential Item Functioning, Generalized Anxiety Disorder.

<sup>1</sup>Fabiola Gonzalez-Betanzos. Departamento Metodología y Psicometría, Universidad Michoacana de San Nicolás de Hidalgo, México. Dirección Postal: Universidad Michoacana de San Nicolás de Hidalgo. Gral. Francisco Villa #450. Edificio A, México. Email: [fabiolagonzalezbetanzos@hotmail.com](mailto:fabiolagonzalezbetanzos@hotmail.com)

<sup>2</sup>María Elena Rivera-Heredia. Facultad de Psicología, Universidad Michoacana de San Nicolás de Hidalgo, México. Email: [maelenarivera@gmail.com](mailto:maelenarivera@gmail.com)

<sup>3</sup>Ferrán Padrós-Blázquez. Facultad de Psicología, Universidad Michoacana de San Nicolás de Hidalgo, México. Email: [fpadros@uoc.com](mailto:fpadros@uoc.com)



## Introducción

Los resultados de la evaluación a través de los test pueden tener consecuencias adversas en la vida cotidiana de las personas. Basta pensar en un contexto clínico en el que una persona puede recibir tratamiento de una institución de salud en base a los resultados de una prueba. Estas circunstancias comprometen a los científicos a realizar estudios que garanticen la equidad en la medición y la validez de las inferencias que se hacen a partir de las puntuaciones en los test (Borsboom, Romeijn & Wicherts, 2008). Es difícil defender un test que resulte sesgado contra un grupo en función de su sexo, etnia, cultura u otra característica sociodemográfica. El término sesgado, en este contexto, implica que las puntuaciones tienen distinto significado para miembros de diferentes grupos. Así, en su desarrollo, los test deben pasar por procesos de análisis del sesgo para evitar la discriminación a las minorías étnicas (afroamericanos, hispanos, asiáticos) o entre sexos, por poner un ejemplo (González-Betanzos, 2011).

Una aproximación básica al estudio del sesgo es el estudio de la invarianza en la medida, la falta de invarianza se produce cuando el ítem (o el test) tiende a proporcionar puntuaciones distintas para personas que pertenecen a diferentes grupos y que, sin embargo, tienen el mismo nivel de rasgo (Shealy & Stout, 1993; Fidalgo, 1996). El término impacto se reserva para las diferencias reales entre los grupos (Camilli & Shepard, 1994).

En el presente trabajo esta propuesta teórica se aplica al problema psicométrico de análisis de la Escala de Detección del Trastorno de Ansiedad Generalizada—EDTAG (Carroll & Davidson, 2000). Este instrumento se ha propuesto como una prueba de cribado, una determinada puntuación en la EDTAG es uno de los criterios para determinar que la persona padece el trastorno, esta puntuación se ha determinado sin atender al sexo. Al mismo tiempo, se han encontrado puntuaciones mayores en las mujeres que en los hombres en dicho test. Por ello, el objetivo del presente trabajo es identificar si existe invarianza en la medida y/o si las diferencias entre hombres y mujeres son reales (i.e. existe impacto).

La EDTAG es un instrumento corto conformado por 12 ítems dicotómicos que detecta la presencia o ausencia de sintomatología del trastorno de ansiedad generalizado (TAG) según se observa en el DSM-IV. El instrumento se divide en 4 criterios: el criterio A que se refiere a la expectación aprensiva (ítems 1 y 2), el criterio B a la falta de control en la preocupación (ítems 1 y 2), el criterio C a los Síntomas fisiológicos (ítems del 5 al 10) y finalmente en el último criterio se recoge el aspecto temporal y su afectación en la vida cotidiana (ítems 11 y 12) que se relacionan con el criterio E.

La EDTAG fue adaptada al español por Bobes, García-Calvo, García-García y Rico-Villademoros (2006) en un estudio prospectivo multicéntrico, en una comparación entre pacientes y grupos sanos, la escala muestra una confiabilidad de 0.85, las pruebas señalan que una puntuación de 6 o más es indicativa de la presencia del TAG (sensibilidad = 84%, especificidad = 83%). Según estos autores se requiere en mínimo de una respuesta SI en los criterios A, B y E y tres respuestas afirmativas en el criterio C. En la aplicación se observó que las mujeres presentan puntuaciones significativamente mayores a los hombres, según los autores este hallazgo se corresponde con la prevalencia del trastorno en la población.

En el DSM-V (APA, 2013) se señala una prevalencia anual entre el 0.4 y el 3.6%, y una prevalencia a lo largo de la vida del 9% observándose el doble de casos en mujeres respecto a los hombres en los países que participan en la evaluación epidemiológica. En México, se ha estimado que el trastorno está presente alguna vez en la vida de la persona entre un 0.7%, y un 1,2% en los últimos 12 meses de la población general, observándose también una razón de 2 mujeres: 1 Hombre (Medina-Mora et al., 2003). Asegurar la invarianza en la medida en la EDTAG tiene especial relevancia pues sus resultados están siendo utilizados como criterio de referencia para diagnóstico, el tratamiento, la evaluación de intervenciones, para estudios epidemiológicos y como medidas de observación para introducir y mejorar políticas sanitarias (AERA, APA, NCME, 1999). Por ello, en el presente trabajo realizaremos dos tipos de análisis, en el primero se utilizarán modelos

de ecuaciones estructurales (MACS) y en el segundo se propone un análisis basado en los modelos de respuesta al ítem (TRI).

## Método

### *Participantes*

Participaron 1431 personas (véase características sociodemográficas en Tabla 1). La muestra estuvo formada por participantes de la población general, se buscó que la proporción entre rango de edades fuera similar. Los rangos de edad iban de 20 a 70 años con una media de 44.43 años y una desviación típica de 15.97.

### *Instrumentos*

*EDTAG: Escala de Detección del Trastorno de Ansiedad Generalizada de Carroll y Davidson (2000)*. Diseñada para detectar el trastorno de ansiedad generalizada consta de 12 ítems dicotómicos (“sí” o “no” respecto a la

Tabla 1

*Características sociodemográficas de la muestra.*

	(Población general)
n	1431
Sexo	
Femenino	841 (58.8%)
Masculino	590 (41.2%)
Edad	
Media	44.43
Desviación estándar	15.97
Rango	20-70
Años de escolaridad	
Media	10.30
Desviación estándar	5.12
Rango	0-29
Estado civil	
Soltero	441 (30.8%)
Casado/unión libre	829 (57.9%)
Separado/divorcio	80 (5.6%)
Viudo	77 (5.4%)

presencia de criterios del DSM IV) que sirven para determinar la presencia o ausencia de sintomatología del TAG de acuerdo con el DMS-IV. En esta investigación se utilizó la adaptación en español de Bobes et al. (2006).

*GADI (Generalized Anxiety Disorder Inventory, de Argyropoulos et al., 2007)*. Es un instrumento autoaplicado que consta de 18 ítems escala tipo Likert de 5 puntos (entre 0 “en absoluto” a 5 “el síntoma está presente en grado extremo”) útil para la evaluar la presencia o ausencia del trastorno por ansiedad generalizada así como su intensidad. La confiabilidad evaluada mediante el alfa fue de 0.948 para la subescala cognitiva, 0.84 para la subescala que evalúa los trastornos de sueño y 0.89 para la subescala de síntomas somáticos.

*BAI (Beck Anxiety Inventory)*. Consta de 21 ítems que se contestan en una escala de 4 puntos (de 0 a 3). Se utilizó la versión mexicana adaptada por Robles, Varela, Jurado y Páez (2001). Presentó una consistencia interna con valores de alfa de Cronbach de .84 en estudiantes universitarios y de .83 en población general.

### *Procedimiento*

En el presente estudio se aplicó la traducción española disponible Bobes et al. (2006) esta versión fue revisada por tres jueces con experiencia en el ámbito de la psicometría y con conocimientos sobre el trastorno de ansiedad generalizada, ninguno de los jueces consideró pertinente la modificación de los ítems. Los datos fueron recogidos en población general a través de la solicitud de participación anónima y voluntaria a transeúntes.

### *Metodologías para el análisis*

La invarianza en la medida (IM) significa que las propiedades de medida del test o de los ítems de un test deberían ser independientes de las características de las personas, excepto por la característica específica que se está midiendo con el test. Formalmente se considera que la distribución de las puntuaciones observadas en las respuestas a un test debería depender únicamente del espacio de la dimensión latente que se evalúa (Mellenbergh, 1989; Meredith, 1993). A saber,

$$f(\mathbf{X}|\eta, g) = f(\mathbf{X}|\eta) \quad (1)$$

Donde:

$f(\bullet)$  es la función de distribución de probabilidad de la variable observada

$\eta$  es el vector d- dimensional de puntuaciones factoriales

$g$  es la variable de agrupación

Según la ecuación 1 los valores de las variables observables (que son las respuestas a los ítems) deben depender únicamente de los valores de las variables latentes (también llamados factores o dimensiones), lo que significaría que existe una independencia entre  $\mathbf{X}$  y  $g$ . Para estudiar la IM es necesario especificar una relación funcional entre las variables observables y la dimensión que representan, dicha relación puede ser lineal o no-lineal. El primer caso corresponde al análisis factorial confirmatorio definido a partir de los modelos de ecuaciones estructurales de medias y covarianzas (AFC-MACS), el segundo se relaciona con los modelos de teoría de la respuesta al ítem (TRI, Elosua, 2005).

*El Análisis Factorial Confirmatorio (AFC) para datos categóricos y la Invarianza.*

El Modelo de AFC es un modelo de regresión lineal en el cual un número de variables observadas, las respuestas a los ítems, se explica a partir de un número de variables latentes subyacentes llamadas factores. El modelo lineal parte de la ecuación que define la puntuación observada (o vector de puntuaciones -  $\mathbf{X}$ , a través de un modelo de regresión en el que la variable independiente se define como:

$$\mathbf{X} = \boldsymbol{\tau} + \Lambda\boldsymbol{\eta} + \boldsymbol{\varepsilon} \quad (2)$$

Donde:

$\mathbf{x}$  Es el valor de la variable aleatoria n-dimensional  $\mathbf{X}$

$\boldsymbol{\tau}$  Es el vector n-dimensional de interceptos

$\Lambda$  Es la matriz nxd de pesos factoriales de los  $d$  factores

$\boldsymbol{\eta}$  Es el vector d-dimensional de puntuaciones factoriales

$\boldsymbol{\varepsilon}$  Es el vector de residuos n-dimensionales de la regresión de  $\mathbf{X}$  sobre  $\boldsymbol{\eta}$

Este modelo asume que las variables observadas son continuas y tienen una distribución normal, la existencia de una relación lineal entre variables observadas y latentes, que los residuos se distribuyen de manera normal, que son independientes entre ellos y que son independientes de las variables latentes (Jöreskog, 1971).

Sin embargo, se sabe que la mayoría de las medidas obtenidas en los ítems que componen las escalas o los cuestionarios en la investigación en las ciencias sociales no dan como resultado valores continuos en escala de razón, sino más bien valores discretos, en el caso de la EDTAG la variable observada es dicotómica.

Si se trata a estas variables como variables continuas en el AFC se viola el supuesto de normalidad multivariada y, en el contexto del análisis multigrupo, se podría distorsionar la estructura factorial de los diferentes grupos, haciendo problemática la investigación sobre equivalencia (Lubke & Muthén, 2004). Para evitar este problema, Millsap y Yun-Tein (2004) han propuesto recientemente el empleo de un modelo de factor común multigrupo para medidas categóricas. Los autores señalan que éste modelo ha recibido poca atención en la investigación de la invarianza.

La definición de invarianza propuesta por Mellenbergh (1989) se aplica también al modelo para variables categóricas, en el caso de los modelos unidimensionales (unifactoriales), y para una adecuada comparación con modelos de TRI, se realizan contrastes estadísticos sobre la igualdad de pesos factoriales (invarianza métrica) e interceptos (invarianza escalar) a través de los grupos. Pero se añade una restricción sobre los umbrales latentes que hacen que la variable latente continua se torne discreta en las respuestas observadas.

En nuestro caso para la comparación por sexo, llamaremos grupo de referencia a las mujeres y grupo focal a los hombres. En el análisis se especifica primero un modelo de línea base compacto donde, en cada caso, los pesos y los interceptos se restringen a ser iguales en

los dos grupos (se iguala  $\lambda_r = \lambda_f$  y  $\tau_r = \tau_f$  donde  $r =$  grupo de referencia y  $f =$  grupo focal, ver ecuación 3). El siguiente paso se especifica un modelo aumentado en el que se liberan los parámetros del ítem que se prueba para saber si es invariante (ver ecuación 4).

i) Modelo compacto -modelo de línea base

$$\lambda_{ir} = \lambda_{if}, \tau_{ir} = \tau_{if}, \gamma_{ir} = \gamma_{if} \quad (3)$$

ii) Modelo aumentado

$$\lambda_{ir} \neq \lambda_{if}, \tau_{ir} \neq \tau_{if}, \gamma_{ir} \neq \gamma_{if} \quad (4)$$

Donde  $i$  es el ítem que se prueba para invarianza

Donde  $\gamma_{ig}$  es el umbral de la variable latente en el ítem  $i$  para el grupo  $g$  (ya sea  $r =$  referencia o  $f =$  focal).

Una vez que se han estimado tanto el modelo compacto como el modelo aumentado se obtiene el estadístico de bondad de ajuste Chi-cuadrado, debido a que el modelo compacto está anidado en el aumentado, se puede comparar el cambio en el Chi-cuadrado con grados de libertad igual a la diferencia de los grados de libertad de los modelos respectivos. En cada comparación un resultado significativo es evidencia de que se viola la equivalencia entre los parámetros y que por lo tanto no hay Invarianza en la medida, a esta prueba se le conoce como Razón de verosimilitudes.

*La Teoría de la Respuesta al Ítem (TRI) y la Invarianza.*

Tanto los modelos de AFC como los de TRI relacionan la respuesta en un ítem con el nivel de rasgo de un examinado. Sin embargo, mientras que los primeros representan una relación lineal, en la TRI el modelo de respuesta al ítem es una función matemática logística, denominada curva característica del ítem (CCI). En esta función se incluyen dos tipos de parámetros, el de los sujetos y los de los ítems. El parámetro de los sujetos se denomina nivel del rasgo ( $\theta$ ). Los parámetros de los ítems dependen del modelo. En el caso del modelo logístico de 2 parámetros (2PL) los parámetros para los ítems son el parámetro de discriminación ( $a$ ) que representa la pendiente de la función y el parámetro de dificultad ( $b$ ), que indica

el valor de  $\theta$  correspondiente al punto de máxima pendiente de la CCI. El Modelo 2PL se escribe como:

$$P(X_{ij} = 1 | \theta_j) = \frac{1}{1 + \exp(-1.7 * a_i(\theta_j - b_i))} \quad (5)$$

Donde

$P(X_{ij} = 1 | \theta_j)$  significa que la probabilidad de responder 1 = Si ( $X_{ij} = 1$ ) a un ítem en el EDTAG depende del nivel del rasgo que se evalúa, en nuestro caso es el Trastorno de ansiedad generalizado, esta probabilidad se explica mediante una función logística con dos parámetros y el nivel de rasgo de la persona.

En el marco de la TRI se llama Funcionamiento Diferencial del Ítem (DIF) a la Invarianza en la medida (IM) y dice que existe DIF cuando las curvas características de los ítems (CCI) son diferentes para los grupos focal y de referencia; es decir cuando para el mismo nivel de rasgo los grupos no tienen la misma probabilidad de elegir la opción de un ítem.

Existen dos tipos básicos de DIF, el DIF uniforme y el DIF no-uniforme. En el primero, uno de los grupos, casi siempre el de referencia, tiene mayor probabilidad de dar una respuesta en todos los niveles del rasgo (ver figura 1 -izquierda) mientras que en el DIF no-uniforme la probabilidad de emitir una respuesta es mayor en el grupo de referencia en algunos niveles del rasgo, mientras que en otros niveles la probabilidad es mayor para el grupo focal (ver figura 1-derecha) (Finch & French, 2007; Martínez-Arias, Hernández-Lloreda & Hernández-Lloreda, 2006).

Por lo tanto, se considera que existe invarianza (o que no existe DIF) si los parámetros del ítem permanecen invariantes a través de los grupos. Existen diversos procedimientos basados en TRI para detectar DIF: chi-cuadrado, la medida del área, la prueba de razón de verosimilitud (LR). Entre ellas, esta última se considera la más general. En esta prueba, descrita por Thissen, Steinberg y Wainer (1993), se compara el ajuste de los modelos anidados, estimados mediante procedimientos de máxima verosimilitud marginal. Este procedimiento

es muy similar al que se emplea en MACS. Para ello se ajusta un modelo en el que se asume la igualdad de los parámetros estimados a través de los grupos de referencia y focal (modelo compacto); posteriormente, este modelo se compara con un modelo donde los parámetros del ítem estudiado se liberan (modelo aumentado),

i) Modelo compacto

$$a_{ir} = a_{if}, b_{ir} = b_{if} \quad (6)$$

ii) Modelo aumentado

$$a_{ir} \neq a_{if}, b_{ir} \neq b_{if} \quad (7)$$

donde  $i$  es el ítem que se prueba para DIF

y se aplica la prueba de razón de verosimilitud (LR) que se define como:

$$LR_i = \frac{L_c}{L_{A_i}} \quad (8)$$

Donde  $L_c$  = probabilidad de los datos según los parámetros del modelo compacto y  $L_{A_i}$  = probabilidad de los datos según los parámetros del modelo aumentado. La transformación del logaritmo natural de esta función se toma como una prueba estadística distribuida como  $\chi^2$  bajo la hipótesis nula:

$$\chi^2(M) = -2\ln(LR) = -2\ln L_c + 2\ln L_{A_i} \quad (9)$$

Para la Ecuación 9,  $M$  es el número de grados de libertad y es igual a la diferencia en el número de parámetros estimados en el modelo compacto respecto del modelo aumentado por lo que, en caso de un modelo de dicotómico de dos parámetros, dado que se liberan los parámetros  $a_i$  y  $b_i$ , el estadístico sigue una distribución chi-cuadrado con 2 grados de libertad. Si el contraste toma un valor estadísticamente significativo, esto indica que el modelo compacto se ajusta peor que el modelo aumentado.

Para usar la prueba de LR, un valor de chi-cuadrado se calcula para cada ítem en el test. Aquellos ítems con valor significativo de  $\chi^2$  se dice

que presentan DIF. En la tradición de TRI, si esto ocurre, se calculan dos test adicionales. En el primero se permite variar el parámetro de discriminación y, posteriormente, se hace una prueba sobre el parámetro de dificultad, por lo que se puede saber si existe DIF uniforme o no uniforme.

Tanto el modelo lineal AFC-MACS como el modelo no lineal de TRI se utilizarán para analizar la IM de los datos del EDTAG. En el primer caso se utiliza el programa MPLUS 7.0 (Muthen & Muthen, 2012) que permite trabajar con variables categóricas y en el segundo caso el análisis se hace con el programa llamado IRTLRDIF v. 2 de Thissen (2001).

Para evaluar el ajuste (Akaike, 1974) de los modelos se utilizará el valor de Chi-cuadrado ( $\chi^2$ ), así como la información de otros indicadores de bondad de ajuste (Bollen y Long, 1993, Elosua, 2005). Entre ellos se utiliza el índice de ajuste general (GFI; Jöreskog & Sorbom, 1993), la raíz media cuadrática del error de aproximación (RMSEA; Hu & Bentler, 1999), el criterio de información de Akaike (AIC; Akaike, 1974) y el índice de ajuste comparativo (CFI; Bentler, 1990). Además se utilizará el criterio de Cheung y Resnold (2002), para ello se observará la diferencia entre los valores del índice comparativo de Bentler (CFI); si el valor de la diferencia entre dos modelos anidados es superior a .01 en favor del modelo con menos restricciones, deberá rechazarse el modelo con más restricciones.

## Resultados

Las medias aritméticas y las desviaciones típicas obtenidas con la versión original de la EDTAG son 4.66 (DT= 3.55) para las mujeres y 4.01 (DT= 3.37) para los hombres. La diferencia de medias entre ambas muestras es estadísticamente significativa [ $t = 2.80, p < .05$ ]. El coeficiente alfa de Cronbach es de .85.

### *Invarianza factorial con AFC-MACS*

La invarianza factorial con AFC-MACS requiere una serie de pasos que inician con la determinación del modelo de línea base que se establece para cada

grupo de manera separada. Este modelo representa el mejor ajuste a los datos dado que no se impone ninguna restricción en la estimación de los parámetros. La estimación progresiva de la invarianza comienza con el modelo de invarianza configural (Tabla 2). Los índices de ajuste obtenidos (Tabla 3) permiten aceptar la equivalencia de la configuración entre hombres y mujeres aunque el valor de Chi-cuadrado excede el criterio para aceptar la hipótesis de invarianza, el resto de los índices contradicen dicha conclusión. El índice de ajuste (GFI = .95) y la raíz media cuadrática (RMSEA = .06) nos permiten aceptar que ambos grupos comparten la misma configuración.

Cuando añadimos las restricciones sobre los pesos factoriales se prueba el modelo de invarianza métrica. Los valores de la Tabla 3 señalan una pérdida significativa en el ajuste del modelo. El índice de ajuste general

(GFI = .1) y la raíz media cuadrática (RMSEA = .074). Si hacemos uso del criterio anidado, la diferencia entre CFI's entre el modelo de configuración y el de invarianza métrica ( $\Delta\text{CFI} = .018$ ) nos conduce a señalar una posible falta de invarianza métrica entre los grupos.

Para evaluar el (los) pesos específicos que podrían causar una falta en la invarianza métrica se compara el modelo compacto y se van liberando uno a uno los pesos factoriales de los ítems, en la Tabla 3 se muestran los resultados para la comparación. En negritas (ver tablas 2 y 3) se muestra el único ítem en el que se encuentra un ajuste significativamente mejor entre el modelo en el que se estiman los pesos para cada grupo respecto del modelo en el que los pesos son los mismos para hombres y mujeres ( $\Delta\chi^2 = 8.32$  y un incremento en CFI = .015).

Tabla 2

*Estructura factorial de la Escala de Detección del Trastorno de Ansiedad Generalizado (EDTAG), pesos factoriales para las mujeres, hombres y el modelo invariante*

Item	mujeres	hombres	invarianza	estricta
	$\lambda$	$\lambda$	$\lambda$	$\tau$
1.La mayoría de los días me siento nervioso/a	.89	.82	.86	.453
2.La mayoría de los días me preocupo por muchas cosas	.74	.76	.75	.121
3.La mayoría de los días no puedo parar de preocuparme	.84	.82	.83	.672
4.La mayoría de los días me resulta difícil controlar mis preocupaciones	.77	.85	.80	.733
5.Me siento inquieto/a, intranquilo/a, o con los nervios de punta	.82	.81	.81	.555
6.Me siento cansado/a fácilmente	.54	.65	.59	.188
7.Tengo problemas para concentrarme	.51	.50	.50	.243
8.Me enfado o irritado fácilmente	.65	.64	.64	.263
9.Mis músculos están tensos y agarrotados	<b>.69</b>	<b>.52</b>	.62	.549
10.Tengo problemas de sueño	.58	.49	.54	.339
11.Las cosas que ha señalado anteriormente, ¿afectaron a su vida diaria...?	.86	.83	.85	.452
12.¿fueron suficiente molestas como para que pensara en buscar ayuda?	.68	.77	.72	.727

Tabla 3  
 Índices de bondad de ajuste para los modelos de invarianza para el AFC-MACS

	$\chi^2$	GFI	RMSEA	CFI	$\Delta$	$\chi^2$	$p$	$\Delta$ CFI
Configuración	443.47	0.95	0.060	0.940				
Métrica	621.45	0.91	0.07	0.925				
Restringido	237.80	0.97	0.070	0.951	0.00	1.000		.001
Item_1	236.05	0.97	0.073	0.951	1.75	0.186		.0001
Item_2	235.54	0.97	0.074	0.949	2.26	0.133		.002
Item_3	236.04	0.97	0.072	0.952	1.76	0.185		.001
Item_4	234.25	0.97	0.073	0.951	3.55	0.060		.0001
Item_5	235.25	0.97	0.074	0.950	2.55	0.110		.001
Item_6	235.06	0.97	0.073	0.951	2.74	0.098		.0001
Item_7	237.20	0.97	0.074	0.948	0.60	0.439		.003
Item_8	235.98	0.97	0.074	0.949	1.82	0.177		.002
<b>Item_9</b>	<b>229.48</b>	<b>0.97</b>	<b>0.071</b>	<b>0.954</b>	<b>8.32</b>	<b>0.004</b>		<b>.015</b>
Item_10	236.10	0.97	0.074	0.949	1.70	0.192		.002
Item_11	236.46	0.97	0.073	0.951	1.34	0.248		.0001
Item_12	235.65	0.97	0.073	0.951	2.15	0.142		.0001

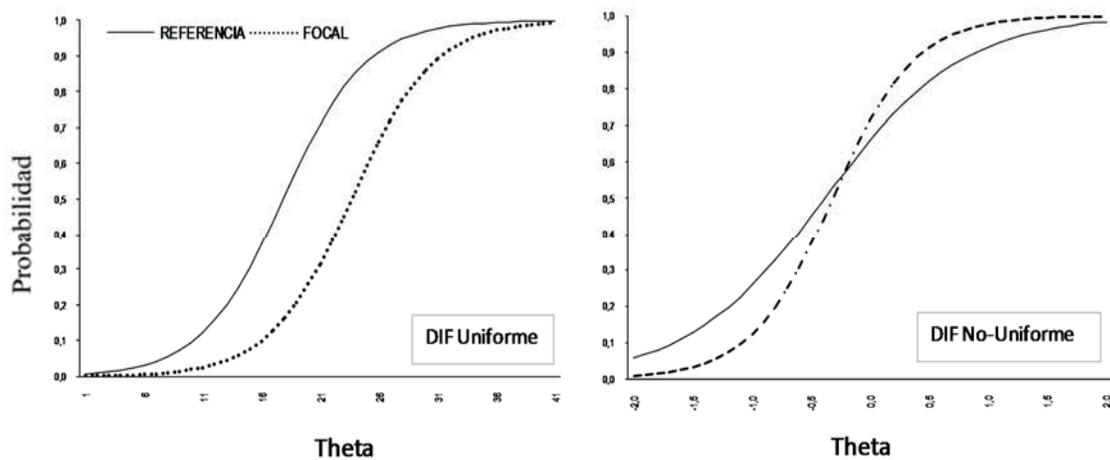


Figura 1. Representa las CCI de los grupos de referencia y focal para el caso en el que existe DIF uniforme (izquierda) o DIF no uniforme (derecha)

*Invarianza factorial con modelos de TRI*

La Tabla 4 presenta los resultados del análisis FDI con el método LR-DIF (loglikelihood –Differential Item Functioning, o prueba de Razón de similitudes para el funcionamiento diferencial del Ítem). Se presenta el estadístico G2 que se distribuye como Chi-cuadrado con grados de libertad igual al número de parámetros que se restringen en el modelo. Dado que trabajamos con un modelo de dos parámetros los grados de libertad son 2. Un valor de G2 de 3.84 es indicativo de que el ítem no es invariante.

En el análisis el ítem 9 tiene un valor de  $G^2 = 6.1$  lo que significa que este ítem presenta funcionamiento diferencial. En la figura 2 se presentan los valores de los parámetros para cada grupo, como puede observarse, en el ítem 9 los hombres tienen mayor probabilidad de responder que “Tienen los músculos agarrotados” que las mujeres, aun cuando tengan el mismo nivel de ansiedad.

Tabla 4

*Estadísticos para el análisis con el método de TRI*

	G2	gl	a	b
Item_1	0	2	2.91	.48
Item_2	0	2	1.67	.35
Item_3	2.7	2	2.03	.86
Item_4	0	2	1.56	.88
Item_5	0	2	2.18	.6
Item_6	0	2	0.96	.74
Item_7	0	2	0.82	.82
Item_8	0	2	1.22	.64
Item_9	6.1	2	1.41	.73
Item_10	0	2	0.93	.65
Item_11	0	2	2.57	.44
Item_12	1.7	2	1.68	.85

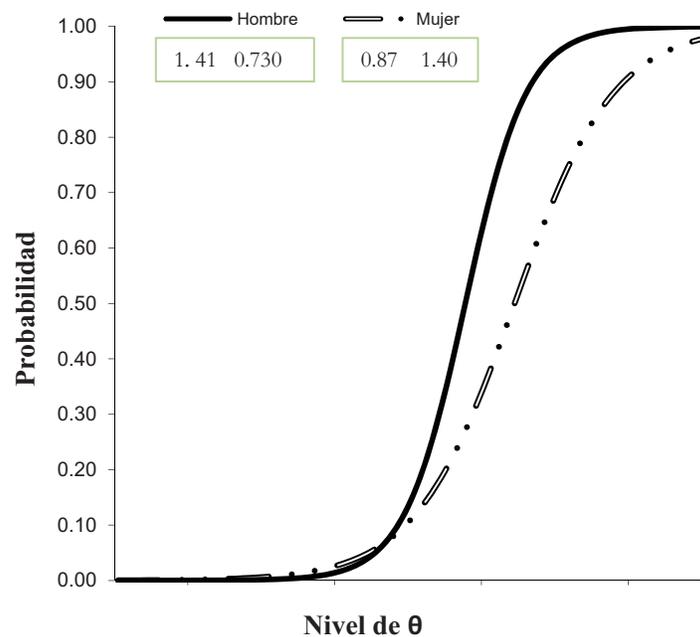


Figura 2. Representa las CCI del ítem 9 para los hombres y las mujeres

En los resultados tanto el análisis con los modelos de ecuaciones estructurales para el análisis factorial confirmatorio con modelos de medias y covarianzas (AFC-MACS) como con los análisis de la Teoría de la Respuesta al Ítem (TRI) coinciden en señalar que el ítem 9 no es invariante entre los grupos. Mientras que en la puntuación observada la diferencia entre las medias es de 0.66 en el análisis con la variable latente la diferencia de medias es de 0.28 mayor para las mujeres que para los hombres.

### Discusión

El objetivo del presente estudio fue realizar un análisis de invarianza entre hombres y mujeres en la Escala de Detección del Trastorno de Ansiedad Generalizado (EDTAG) partiendo de dos tipos de relaciones funcionales entre el espacio latente y las variables observadas: i) Un modelo lineal y un modelo logarítmico.

En el modelo lineal se ha mostrado la evaluación progresiva de la invarianza estableciendo primero que en ambos sexos el test se comporta de manera unidimensional, es decir todos los ítems representan la medida del Trastorno de Ansiedad Generalizado (TAG) como variable latente (invarianza configural). Sin embargo, el análisis señala que no existe invarianza métrica en la medida a nivel global, para conocer específicamente en donde se produce la falta de equivalencia se llevaron a cabo análisis por ítem en el que se probó la equivalencia en los parámetros del modelo: i.e. los pesos factoriales y los umbrales, mediante un procedimiento que descubre la falta de ajuste en modelos anidados denominada prueba de Razón de verosimilitudes. Este análisis mostró que el ítem 9 no representa a la variable latente de igual forma en ambos grupos.

Como señalan diversos autores, el estudio de la invarianza desde la perspectiva lineal se acerca a la evaluación del funcionamiento diferencial del ítem que describe la relación de manera logística entre los ítems y las dimensiones latentes del rasgo.

Estos análisis han sido diseñados para su uso en escalas unidimensionales. El presente estudio pone en

evidencia la necesidad de realizar análisis de invarianza en las escalas de medición, especialmente en aquellas escalas cortas donde una diferencia de un punto podría significar un falso positivo.

Sin embargo, estos resultados no proveen respuestas por sí solos, se sabe que un segundo paso en el análisis, se relaciona con la revisión e interpretación de contenidos que permitan conocer porque existe un comportamiento diferencial entre hombres y mujeres en estos ítems. Además es importante tener en cuenta a las variables que pueden estar influyendo y para ello es necesario profundizar en el estudio sobre sesgo en el marco de la validez de constructo.

El presente estudio demuestra que en el EDTAG el ítem 9 que señala “Mis músculos están tensos y agarrotados” se comporta distinto para hombres que tiene mayor probabilidad de responder afirmativamente, esto puede indicar que la sintomatología en hombres y en mujeres es diferente en los procesos ansiosos.

### Referencias

- AERA-APA-NCME. (1999). *Standards for educational and psychological testing*. Washington, DC.: American Psychological Association.
- Akaike, H. (1974). *A new look at statistical model identification*. *Transactions on Automatic Control*, AC-19.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed. ed.). Washington, DC: Author.
- Argyropoulos, S., Ploubidis, G., Wright, T., Palm, M., Hood, S., Nash, J., . . . Potokar, J. (2007). Development and validation of the generalized anxiety disorder inventory (GADI). *Journal of Psychopharmacology*, 21, 145-152.
- Bobes, J., García-Calvo, C., García-García, M., & Rico-Villademoros, F. (2006). Propiedades psicométricas de la versión española de la Escala de Detección del Trastorno de Ansiedad Generalizada según DSM-IV de Carroll y Davidson. *Actas Españolas de Psiquiatría*, 34, 83-93.

- Bollen, K., & Long, J. (1993). Introduction. En K. Bollen, & J. Long, *Testing structural equation models* (págs. 1-9). Newbury Park, CA.: SAGE.
- Borsboom, D., Romeijn, J., & Wicherts, J. (2008). Measurement invariance versus selection invariance: Is fair selection possible? *Psychological Methods, 13*, 75-97.
- Camilli, G., & Shepard, L. A. (1994). *Methods for identifying biased test items*. Thousand Oaks, CA: Sage publications.
- Carroll, B. J., & Davidson, J. R. (2000). *Screening Scale for DSM-IV GAD*. Copyright.
- Cheung, G., & Rensvold, R. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*, 233-255.
- Elosua, P. (2005). Evaluación progresiva de la invarianza factorial entre las versiones original y adaptada de una escala de autoconcepto. *Psicothema, 17*(2), 356-362.
- Fidalgo, A. (1996). Funcionamiento diferencial de los ítems. En J. Muñiz, *Psicometría* (p. 370-455). Madrid: Universitas.
- Finch, W., & French, B. (2007). Detection of Crossing Differential Item Functioning: a comparison of four methods. *Educational and psychological Measurement, 67*(4), 565-582.
- Gonzalez-Betanzos, F. (2011). *Funcionamiento Diferencial del Ítem en Test Adaptativos Informatizados*. Tesis del doctorado. Madrid: Universidad Autónoma de Madrid.
- Hu, L., & Bentler, P. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1-55.
- Jöreskog, K. (1971). Simultaneous factor analysis en several populations. *Psychometrika, 36*, 409-426.
- Jöreskog, K., & Sörbom, D. (1993). *LISREL 8: user's guide*. Chicago: Scientific Software International.
- Lubke, G., & Muthen, B. (2004). Applying multigroup confirmatory factor models for continuous outcomes to Likert scale data complicates meaningful group comparison. *Structural Equation Modeling, 11*, 514-534.
- Martínez-Arias, R., Hernández-Lloreda, M., & Hernández-Lloreda, M. (2006). *Psicometría*. Madrid: Alianza Editorial.
- Medina-Mora, M. E., Borges, G., Lara-Muñoz, C., Benjet, C., Blanco-Jaimes, J., Bautista, C., . . . Aguilar-Gaxiola, S. (2003). Prevalencia de trastornos mentales y uso de servicios: Resultado de la encuesta nacional de epidemiología psiquiátrica en México. *Salud Mental, 26*(4), 1-16.
- Mellenbergh, G. (1989). Item bias and item response theory. *International Journal of Mathematical and Statistical Psychology, 24*, 49-82.
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*(4), 525-543.
- Millsap, R., & Yun-Tein, J. (2004). Assessing Factorial Invariance in Ordered-Categorical Measures. *Applied Psychological Measurement, 39*(3), 479-515.
- Muthen, L., & Muthen, B. (2012). *Mplus user's guide*. Los Angeles, CA.: Muthen & Muthen.
- Robles, R., Varela, R., Jurado, S., & Páez, C. (2001). Versión mexicana del Inventario de Ansiedad de Beck: Propiedades psicométricas. *Revista Mexicana de Psicología, 18* (2), 211-218.
- Shealy, R., & Stout, W. F. (1993). A model-based standardization approach that separates true bias/DIF from group ability differences and detects test bias/DIF as well as item bias/DIF. *Psychometrika, 58*, 159-194.
- Thissen, D. (2001). *IRTLRTDIF (version 2.02b)*. Chapel Hill, NC: L.L. Thustone Psychometric Laboratory.

Recibido: 13 de abril de 2015  
 Aceptado: 18 de setiembre de 2015