

Actualización psicométrica y funcionamiento diferencial de los ítems en el State Trait Anxiety Inventory (STAI)

Alejandro Guillén-Riquelme y Gualberto Buela-Casal
Universidad de Granada

La ansiedad es uno de los problemas psicológicos con mayor prevalencia. Entre los instrumentos para medirla se encuentra el State Trait Anxiety Inventory. Este cuestionario mide ansiedad rasgo (factor de personalidad que predispone a sufrir o no ansiedad) y ansiedad estado (los factores ambientales que protegen o generan ansiedad). La adaptación española del cuestionario se realizó en 1982, por ello el objetivo del estudio es realizar una revisión del STAI. Para ello, se reunió una muestra de 1.036 adultos. Se realizó un análisis de fiabilidad mediante alfa de Cronbach (0,90 para ansiedad rasgo y 0,94 para ansiedad estado). También se realizó una reducción factorial con unos resultados similares a los de la adaptación original. Además, se comprobó si existía funcionamiento diferencial de ítem por sexo y únicamente una de las 40 preguntas mostró problemas. Por último se realizó comparaciones t-Student con los valores de la adaptación; mientras el rasgo varía en 1 punto, el estado tiene diferencias de hasta 6 puntos. En estos resultados de forma general se observa que el STAI mantiene unas adecuadas propiedades métricas y que, además, ha sido sensible al aumento de estímulos ambientales que producen estrés.

Psychometric revision and differential item functioning in the State Trait Anxiety Inventory (STAI). One of the psychological problems with highest prevalence is anxiety. The State Trait Anxiety Inventory is one of the instruments to measure it. This questionnaire assesses Trait Anxiety (understood as a personality factor that predisposes one to suffer from anxiety) and State Anxiety (refers to environment factors that protect from or generate anxiety). The questionnaire was adapted in Spain in 1982. Therefore, the goal of the study is to review the current psychometric properties of the STAI. A total of 1036 adults took part in the study. Cronbach's alpha reliability was .90 for Trait and .94 for State Anxiety. Factor analysis showed similar results compared with the original data. Moreover, differential item functioning (DIF) was carried out to explore sex bias. Only one of the 40 items showed DIF problems. Lastly, a t-Test was run, comparing the original and current values; whereas Trait Anxiety varied in 1 point, State Anxiety had differences of up to 6 points. In general, this result shows that the STAI has maintained adequate psychometric properties and has also been sensitive to increased environmental stimuli that produce stress.

En la actualidad, la ansiedad es uno de los problemas psicológicos con mayor prevalencia. Alonso et al. (2004b) indican que, en una muestra representativa de seis países europeos, un 13,6% había tenido algún problema de ansiedad a lo largo de su vida y un 6,4% en el último año; resultado similar al obtenido por Andlín-Sobocki y Wittchen (2005), nuevamente en europeos, donde un 12% presentaba trastornos de ansiedad. De la misma forma Wang et al. (2007) hallaron que un 6,8% de los encuestados españoles había tenido problemas por ansiedad y trastornos emocionales en el último año. Además, un 26,1% de las consultas psiquiátricas formales se deben a trastornos de ansiedad (Alonso et al., 2004c) y un 28,3% del total de casos de

ansiedad muestran comorbilidad con otros trastornos (Alonso et al., 2004a).

Existen numerosos cuestionarios para medir la ansiedad (Inventario de Ansiedad de Beck, Inventario de Situación y Respuestas Ansiosas, Escala de Ansiedad Manifiesta de Taylor...). Entre los cuestionarios para evaluar este trastorno se encuentra el State-Trait Anxiety Inventory (STAI), desarrollado por Spielberger, Gorsuch y Lushene (1970). Este instrumento fue creado para evaluar la ansiedad según el modelo del propio Spielberger, que postulaba que dicho trastorno está constituido por dos componentes: un factor de personalidad que comprendería las diferencias individuales, relativamente estables, para responder ante situaciones percibidas como amenazantes con una subida en la ansiedad. También se define como una tendencia a percibir las situaciones como más amenazantes (ansiedad rasgo). El segundo factor (ansiedad estado) hace referencia a un período transitorio caracterizado por un sentimiento de tensión, aprensión y un aumento de la actividad del sistema nervioso autónomo, pudiendo variar tanto en el tiempo como en la intensidad (Spielberger et al., 1970). La elección de

este cuestionario sobre el resto de pruebas para evaluar ansiedad se debe a que el STAI es uno de los más utilizados; de hecho Muñiz y Fernández-Hermida (2010) encontraron que es el 7º cuestionario más utilizado en España.

Cada una de las escalas del STAI (ansiedad rasgo y ansiedad estado) está compuesta por 20 ítems; una parte de los mismos redactada de forma positiva y otra de forma negativa. Diversos estudios mostraron reducciones factoriales muy robustas con cuatro factores. Estos factores estaban basados en la inversión de parte de las preguntas y se definían como ansiedad rasgo afirmativa y negativa y ansiedad estado afirmativa y negativa (Hishinuma, Miyamoto, Nishimura y Nahulu, 2000; Suzuki, Tsukamoto y Abe, 2000; Vigneau y Cormier, 2008). En la adaptación española (Spielberger, Gorsuch y Lushene, 1982) se encuentran resultados muy similares. El número de factores que subyacen en el STAI han sido objeto de numerosos análisis, existiendo defensores del modelo bifactorial, tetrafactorial y, en menor medida, un modelo pentafactorial, compuesto por los factores presentados anteriormente más un factor, compuesto por dos ítems, llamado *happiness* (Suzuki et al., 2000).

Este instrumento ha sido traducido a más de cuarenta lenguas (Spielberger y Díaz-Guerrero, 1976, 1983, 1986; Spielberger, Díaz-Guerrero y Strelau, 1990) y se han realizado más de 3.000 investigaciones que lo utilizan (Spielberger, 1989). En el 2002, Barnes, Harp y Jung realizaron una revisión del STAI y encontraron 909 artículos (publicados desde el año 1990 hasta el 2000) que lo citaban. Así pues, hasta el año 2000 se realizaron, al menos, 3.909 investigaciones con el cuestionario.

El STAI se adaptó al español a partir de los trabajos de Bermúdez (1978a; 1978b) y su versión comercial se realizó por la sección de estudios de TEA Ediciones (Spielberger et al., 1982). En estos artículos se comprueba que las medias y la fiabilidad (alfa de Cronbach y correlación test-retest) resultan similares a la versión original. Respecto a la dimensionalidad de la escala, se mantiene el modelo de cuatro factores presentado anteriormente.

El objetivo del presente estudio es analizar las propiedades psicométricas del STAI, puesto que su adaptación española se realizó hace casi treinta años, y comprobar su fiabilidad. Entre los análisis se incluirá una prueba del funcionamiento diferencial del ítem para analizar posibles sesgos, producidos por el efecto del sexo. Este análisis es pertinente puesto que no se ha realizado anteriormente en la adaptación española. Se espera que los índices psicométricos del cuestionario sean adecuados; pese a ello también es posible que los valores normativos hayan variado respecto al original por el paso del tiempo.

Método

Participantes

La muestra está compuesta por 1.036 participantes (352 hombres y 667 mujeres). Las edades oscilan desde 18 hasta los 69 años. La edad media es de 29,64 (DT= 10,51). Las mujeres tienen una edad media de 29,83 años (DT= 10,31) y los hombres de 29,30 años (DT= 10,91).

Instrumentos

Para el estudio se utilizó el State-Trait Anxiety Inventory (STAI) (Spielberger et al., 1982). La adaptación española fue realizada por la sección de estudios de TEA.

Este inventario se compone de veinte ítems para cada una de las subescalas (ansiedad estado y ansiedad rasgo). La escala de respuesta es de tipo Likert, puntuando desde 0 (nada) hasta 3 (mucho). Los totales se obtienen sumando los valores de los ítems (tras las inversiones de las puntuaciones en los ítems negativos). Por ello, los totales de ansiedad rasgo y de ansiedad estado abarcan desde 0 hasta 60, correspondiéndose una mayor puntuación con mayor ansiedad detectada. No cuenta con un criterio clínico en su versión española, aunque sí lo tiene en la versión original, así como en muchas adaptaciones.

Procedimiento

La muestra se recogió en Granada, Córdoba, Jaén, Madrid, Barcelona, Valencia, Ourense, Santiago de Compostela y Bilbao. El tipo de muestreo fue no probabilístico por cuotas (ser mayor de edad y no estar cursando una carrera universitaria). Para el muestreo se contó con la colaboración de organizaciones públicas y privadas. También se administró el cuestionario en lugares con una gran afluencia de personas, como aeropuertos, estaciones, centros comerciales, etc. El cuestionario se administró por un único investigador. Se daba el test a los sujetos y era completado en presencia del administrador para posibles dudas o problemas. En todos los casos se informó a los participantes de la confidencialidad del estudio.

Diseño

Según la clasificación propuesta por Montero y León (2007), el presente estudio sería instrumental, puesto que evalúa las propiedades psicométricas de un cuestionario. Se siguieron las recomendaciones metodológicas de Carretero-Dios y Pérez (2007).

Análisis de datos

En el análisis de la fiabilidad se empleó el alfa de Cronbach. La comprobación de la estructura de la escala se realizó mediante análisis factorial. El método de extracción ha sido máxima verosimilitud y la rotación varimax. Para establecer el número de factores se realizó un análisis paralelo al 95% de confianza.

Para comparar los datos del estudio con los de la adaptación española se utilizó la prueba t-Student para una sola muestra, complementados con el tamaño del efecto mediante la *d* de Cohen.

Se comprobó si había funcionamiento diferencial del ítem (DIF en adelante). El DIF se define como la diferencia en la probabilidad que tienen los participantes de dos o más grupos, en este caso sexo, de responder de forma similar, tras haber sido igualados en habilidad (Millsap y Meredith, 1992). Según Mellenbergh (1995) se distingue entre DIF uniforme (la probabilidad de responder correctamente al ítem es mayor para uno de los grupos uniformemente para todos los niveles de la variable) y DIF no uniforme (la diferencia de probabilidad para responder correctamente al ítem no es igual a lo largo de los diferentes niveles del atributo).

Para la detección del DIF existen varios procedimientos. En el presente artículo se usará el análisis de regresión logística, propuesto por Miller y Spray (1993). Esta técnica permite detectar el DIF uniforme y no uniforme en ítems politómicos (Hidalgo y López-Pina, 2004). Los análisis, tal como se presentan en Hidalgo y Gómez (2006), consisten en un modelo de regresión logística binaria donde la variable dependiente es el grupo (en este caso

sexo) y las variables independientes son la puntuación total en el test, la respuesta al ítem y la interacción entre ambas. La estrategia para detectar DIF, siguiendo las indicaciones de Hidalgo, Gómez y Padilla (2005), consiste en ajustar tres modelos en distintas etapas. En la primera etapa se introduce en la ecuación la puntuación total del sujeto en el test, ajustando así el Modelo 1 (M1) en base a la ausencia de DIF. En la segunda etapa se añade la variable de agrupamiento (Modelo 2, M2). Si la explicación de este modelo respecto al M1 fuese significativa sería indicio de DIF uniforme. En la etapa 3 se introduce la interacción entre el grupo y la puntuación total (Modelo 3, M3). Si la explicación que añade este modelo, respecto a los anteriores, fuese significativa habría DIF no uniforme.

Una vez detectados los ítems con DIF se eliminan sus puntuaciones del total del cuestionario. Tras la purificación se repiten el análisis en todos los ítems con el total corregido. Si aparecen nuevos ítems con DIF se puede proceder a una segunda depuración del total.

Por último, en los ítems con DIF cabe averiguar si se mantiene en todas las escalas de respuesta. Para ello se utiliza el análisis de categorías acumuladas propuesto por Mellenbergh (1995). El análisis consiste en varias dicotomizaciones de las categorías de respuesta. Se repite el análisis de DIF con las diferentes agrupaciones realizadas. En este caso se compara el 0 con el 1, 2 y 3 agrupados, el 0 y 1 contra el 2 y 3, y el 0, 1 y 2 contra el 3.

Para la realización de todos los análisis descriptos se utilizó el programa estadístico SPSS 15.0, excepto para el cálculo del tamaño del efecto mediante la *d* de Cohen, realizado con el G*Power 3.1.2.

Resultados

Primeramente se realizó un análisis de las puntuaciones medias y desviación típica de los ítems para detectar puntuaciones extremas. Ninguna de las medias supera 1,78, siendo la menor de 0,45. Este valor, pese a ser bajo, no es de extrañar, puesto que no se trata de una muestra clínica, no esperando por ello puntuaciones elevadas. Las desviaciones típicas se sitúan entre 0,71 y 1,11. Esto indica que los sujetos no tienen un patrón de respuesta con altas frecuencias en los valores centrales o en los extremos.

Seguidamente se realizó un análisis de fiabilidad, mediante el alfa de Cronbach, obteniendo, para los ítems de ansiedad rasgo, un alfa de 0,90. Solo uno de los veinte ítems de la escala mejoraba el alfa (en una milésima) si era eliminado. Por su parte, la ansiedad estado alcanzó un alfa de 0,94, con un solo ítem cuya eliminación mejoraba el alfa en cuatro milésimas.

Una vez establecido el funcionamiento adecuado de los ítems se pasó a la reducción factorial de los mismos para ver la estructura factorial subyacente. En primer lugar, la prueba Kaiser-Meyer-Olkin (KMO= 0,96), así como la prueba de esfericidad de Bartlett ($\chi^2_{:780} = 19,37$; $p < 0,001$), mostraron la adecuación de la muestra para realizar el análisis factorial. Para la elección del número de factores se utilizó un análisis paralelo. Se extrajeron cuatro factores. Sus autovalores son 14,40, 2,60, 2,48 y 1,33, y la varianza total explicada conjuntamente es de 52,04%. Las saturaciones de cada ítem en los diferentes factores se presentan en la tabla 1.

En las saturaciones se observa que el factor 3 corresponde a los ítems de la ansiedad rasgo redactados de forma positiva, mientras que los ítems que saturan en el factor 2 pertenecen a la ansiedad estado redactados de forma positiva. Los otros dos factores, espe-

cialmente el primero, aglutinan ítems de ambas subescalas cuya redacción original es negativa y sus puntuaciones se han invertido. Esta estructura es similar a la hallada en la adaptación original de cuatro factores (ansiedad estado y rasgo, positiva y negativa, respectivamente). El hecho de establecer cuatro factores en función de la presencia o ausencia de ansiedad estado y rasgo se puede considerar un artefacto estadístico, por lo que se interpretará el resto de análisis en función de ansiedad estado y rasgo.

En siguiente lugar se analizaron las medias del total de las subescalas. Se calcularon pruebas t-Student sobre una media, utilizando para el contraste los valores del manual de la adaptación

Tabla 1
Matriz de saturaciones de los ítems de STAI en los factores rotados

	Ítem	Factor 1	Factor 2	Factor 3	Factor 4
Rasgo	1	0,67	0,14	0,31	0,11
	2	0,28	0,09	0,32	-0,01
	3	0,29	0,24	0,47	0,08
	4	0,32	0,15	0,42	0,07
	5	0,13	0,12	0,40	0,17
	6	0,44	0,15	0,22	0,06
	7	0,23	0,25	0,25	0,10
	8	0,30	0,25	0,45	0,16
	9	0,11	0,15	0,64	0,19
	10	0,70	0,11	0,23	0,09
	11	-0,04	0,11	0,46	-0,02
	12	0,20	0,15	0,59	0,51
	13	0,37	0,09	0,38	0,61
	14	0,08	0,15	0,34	0,11
	15	0,40	0,20	0,51	0,06
	16	0,68	0,07	0,30	0,20
	17	0,13	0,20	0,59	0,11
	18	0,19	0,12	0,60	0,06
	19	0,39	0,17	0,30	0,28
	20	0,19	0,26	0,55	0,07
Estado	1	0,40	0,59	0,10	0,24
	2	0,54	0,35	0,16	0,56
	3	0,22	0,75	0,19	0,08
	4	0,28	0,60	0,20	0,07
	5	0,63	0,44	0,07	0,13
	6	0,15	0,79	0,17	0,13
	7	0,20	0,38	0,37	-0,04
	8	0,48	0,37	0,14	0,08
	9	0,30	0,60	0,26	0,08
	10	0,63	0,33	0,09	0,10
	11	0,43	0,20	0,25	0,65
	12	0,12	0,74	0,24	0,13
	13	0,01	0,33	0,15	0,00
	14	0,26	0,49	0,28	0,01
	15	0,52	0,54	0,09	0,12
	16	0,72	0,21	0,14	0,24
	17	0,34	0,58	0,30	0,06
	18	0,13	0,70	0,20	0,11
	19	0,74	0,19	0,17	0,14
	20	0,67	0,37	0,13	0,16

Nota: se ha empleado el método de *maxima verosimilitud* y la rotación *varimax*

Tabla 2
Diferencias de medias por escala y sexo respecto a la adaptación original¹ mediante *t-Student*

Sexo	Factor	Media en la adaptación original	Media en el presente estudio	Diferencia de medias	t	g.l. ²	p ³	d de Cohen
Hombres	Rasgo	20,19 (8,89)	18,96 (10,00)	1,23	-2,24	329	0,02	0,20
	Estado	20,54 (10,56)	15,87 (9,92)	6,67	-8,48	322	0,00	0,56
Mujeres	Rasgo	24,99 (10,05)	23,35 (10,60)	1,64	-3,85	617	0,00	0,10
	Estado	23,30 (11,93)	18,20 (11,62)	5,10	-10,58	604	0,00	0,45

¹ Datos obtenidos del manual de la adaptación española. Entre paréntesis se incluye la desviación típica
² Grados de libertad
³ Probabilidad

española. Este análisis se realizó para ambos sexos, ya que son los valores que ofrece la adaptación.

En la tabla 2 se puede observar, además de que todos los valores son significativos, que en las puntuaciones rasgo las diferencias son cercanas a 1,5 puntos, mientras que en la ansiedad estado las diferencias están en torno a los 5 puntos. Los tamaños del efecto son muy bajos en rasgo y elevados en la ansiedad estado. Parece ser que el rasgo no ha cambiado apenas respecto a los valores originales, mientras que la ansiedad estado sí ha sufrido fuertes modificaciones.

En la adaptación española se encontraron diferencias por sexo (Spielberger et al., 1982). Para comprobar si estas diferencias se mantienen en ambas subescalas se utilizó la prueba *t-Student*. En la ansiedad rasgo (asumiendo varianzas iguales; $F= 2,08$, $p= 0,150$) hay diferencias estadísticamente significativas ($t_{(946)} = -6,20$; $p<0,001$). Para la ansiedad estado (asumiendo varianzas distintas; $F= 17,60$, $p<0,001$) también se encuentran diferencias por sexo ($t_{(764,21)} = -3,19$; $p= 0,001$).

Para aportar evidencias de que las diferencias encontradas anteriormente no se deben a un problema de DIF se realizó un análisis del mismo. Los resultados se muestran en la tabla 3.

En la tabla 3 se observa que no se detecta DIF en ninguno de los ítems de ansiedad estado (significación menor a 0,01). Sin embargo, en la escala de ansiedad rasgo hay cuatro ítems (3, 5, 7 y 19) cuyo modelo 2 es significativo al 0,01, por lo que es un primer indicador de la presencia de DIF uniforme. Ningún ítem de ambas subescalas muestra DIF no uniforme. El siguiente paso es comprobar el tamaño del efecto mediante las diferencias en el R^2 del modelo 2 respecto al 1. Siguiendo los criterios propuestos por Jodoin y Gierl (2001), un valor inferior a 0,035 informaría de un DIF inapreciable; de 0,035 a 0,07 un DIF moderado, y mayor a 0,07 correspondería a un DIF elevado. En este caso el ítem 19 mostraría DIF moderado, mientras que, en los otros tres ítems, sería inapreciable. Seguidamente se depuró el total, eliminando estos cuatro ítems para repetir los análisis.

En la segunda fase se realizó de nuevo el procedimiento con el total depurado. Nuevamente los ítems 3, 5, 7 y 19 fueron significativos al 0,01, sin embargo los tamaños del efecto (excluyendo el del ítem 3, con un 0,061) son muy bajos, siendo el mayor de ellos de 0,026. Además de estos ítems, el 13, 14 y el 20 también fueron significativos, no obstante los tamaños del efecto son de 0,012, 0,010 y 0,010, respectivamente (inapreciables). Por todo ello, se puede concluir que únicamente el ítem 3 muestra un DIF medio.

Una vez se detecta el DIF, es necesario analizar si se encuentra en todas las categorías de respuesta o solo en algunas. Para ello, la escala del ítem 3 (único con DIF) se dicotomizó mediante las agrupaciones de Mellenbergh (1995). Con las categorías dicotomizadas se realizan los mismos análisis que anteriormente con los ítems. En ninguna de las alternativas de respuesta se detecta DIF no uniforme. Sin embargo, se observa DIF uniforme al comparar la opción cero contra la uno, dos y tres de forma conjunta. Por tanto, en el ítem 3 únicamente existe DIF al comparar la primera opción («casi nunca») con las otras tres. Es decir, que hombres y mujeres, igualados en nivel de ansiedad, responden con diferente frecuencia en esta opción.

Discusión

En los resultados se ha comprobado que las medias y desviaciones típicas de los ítems son adecuadas (Carretero-Dios y Pérez, 2007). Si bien algunas de las medias son bajas, esto se debe a que es un cuestionario de ansiedad (en su versión original y en muchas adaptaciones tiene baremos clínicos) administrado en población normal. Por ello, las medias algo bajas son esperables. Respecto a la fiabilidad los valores alfa de Cronbach son adecuados dando indicios de una adecuada consistencia interna.

El análisis factorial muestra un porcentaje de varianza explicada superior al 50%. Además, el análisis paralelo indica que se deben extraer cuatro factores, al igual que en la adaptación española (Spielberger et al., 1982). Vera-Villaruel, Buéla-Casal y Spielberger (2007) también encuentran cuatro factores al realizar dos extracciones separadamente (una para rasgo y otra para estado), con una versión en español del cuestionario. Cabe indicar que estos factores no replican exactamente los esperados teóricamente, encontrando problemas en los factores que incluyen los ítems invertidos de ambas subescalas. Si bien estos datos muestran cierta confusión, son muy similares a los obtenidos por Bermúdez (1978a, 1978b), así como a los del manual de la adaptación española (Spielberger et al., 1982). Pese a que los datos no son tan claros como en otras revisiones de la versión en inglés, muestran una estructura tetrafactorial similar a las anteriormente halladas, aportando un nuevo indicio de que el test conserva sus propiedades métricas.

Respecto a la comparación de medias respecto al original los resultados son muy interesantes. En primer lugar cabe indicar que los baremos y medias de la adaptación española se realizó con 318 varones y 365 mujeres trabajadoras. El manual no ofrece más da-

Tabla 3
Funcionamiento diferencial del ítem: probabilidades (p) asociadas a los modelos (M) de la regresión y tamaño del efecto

	Ítem	χ^2 (M2-M1)	p	χ^2 (M3-M2)	p	R Nagelkerke (M3-M2)	R Nagelkerke (M2-M1)
Rasgo	1	4,07	0,131	0,09	0,955	0,000	0,006
	2	0,57	0,753	1,24	0,538	0,002	0,001
	3	45,10	0,000	6,30	0,043	0,009	0,061
	4	0,04	0,982	0,90	0,638	0,001	0,000
	5	20,95	0,000	0,07	0,967	0,000	0,029
	6	0,00	1,000	1,79	0,408	0,003	0,000
	7	14,14	0,001	0,05	0,977	0,000	0,020
	8	1,70	0,428	0,29	0,867	0,001	0,002
	9	2,42	0,298	0,03	0,985	0,000	0,003
	10	5,46	0,065	0,42	0,809	0,000	0,008
	11	3,01	0,222	0,74	0,692	0,001	0,004
	12	1,76	0,415	4,69	0,096	0,007	0,002
	13	8,29	0,016	0,05	0,978	0,000	0,012
	14	6,87	0,032	5,14	0,077	0,007	0,010
	15	4,85	0,089	2,58	0,275	0,003	0,007
	16	5,80	0,055	0,02	0,993	0,000	0,008
	17	0,16	0,925	1,19	0,552	0,002	0,000
	18	1,07	0,587	2,51	0,285	0,000	0,001
	19	12,40	0,002	0,05	0,974	0,000	0,017
	20	6,86	0,032	1,78	0,411	0,002	0,010
Estado	1	1,14	0,565	1,74	0,420	0,003	0,001
	2	1,08	0,582	4,94	0,085	0,008	0,001
	3	0,18	0,914	5,08	0,079	0,007	0,000
	4	6,57	0,037	2,00	0,368	0,003	0,009
	5	8,88	0,012	1,68	0,433	0,002	0,013
	6	1,40	0,497	2,00	0,368	0,003	0,002
	7	0,05	0,973	1,35	0,510	0,002	0,000
	8	2,88	0,237	0,83	0,661	0,001	0,004
	9	0,39	0,823	2,23	0,327	0,004	0,000
	10	0,02	0,988	1,23	0,541	0,002	0,000
	11	7,53	0,023	2,10	0,350	0,003	0,011
	12	3,77	0,152	3,90	0,142	0,006	0,005
	13	1,82	0,404	3,14	0,208	0,005	0,002
	14	0,18	0,912	2,54	0,282	0,004	0,000
	15	4,29	0,117	2,47	0,290	0,004	0,006
	16	1,38	0,501	0,00	1,000	0,000	0,002
	17	0,85	0,652	0,76	0,685	0,001	0,001
	18	1,14	0,565	4,10	0,129	0,006	0,001
	19	0,19	0,911	0,70	0,703	0,001	0,000
	20	3,10	0,212	4,24	0,120	0,007	0,004

tos al respecto. En base a ello, se observa que los tamaños muestrales son similares y no hay características diferentes en la presente muestra. Al comparar puntuaciones entre ambas muestras las medias difieren un máximo de 1,64 puntos en la ansiedad rasgo, mientras que alcanzan entre 5 y 6 puntos de diferencia en la ansiedad estado. Este hecho puede ser indicio de que el cuestionario conserva unas buenas propiedades métricas pese al paso del tiempo, ya que la escala de ansiedad rasgo mantiene las puntuaciones (los factores de personalidad son relativamente estables). Al contrario, en los factores ambientales la variación ha sido notable ya que éstos cambian con más facilidad a lo largo del tiempo (Eysenck, 1995).

En el análisis de DIF se encuentra un nuevo indicio sobre la validez del cuestionario. Se observa que únicamente un ítem mues-

tra funcionamiento diferencial por sexo. Para comprender mejor la causa del sesgo en la respuesta de hombres y mujeres cabe realizar un análisis del contenido del ítem. La cuestión planteada es «Siento ganas de llorar», en este caso parece lógico pensar que, pese a que se hayan igualado hombres y mujeres en nivel de ansiedad, los hombres serán más propensos a puntuar que casi nunca sienten ganas de llorar. Mientras, las mujeres tendrán una respuesta más equilibrada en las diferentes alternativas, lo que explica el DIF hallado.

En conclusión, que la escala se mantenga estable en la medida de la ansiedad rasgo pero haya variado en la ansiedad estado, además de mantener las medidas psicométricas y no verse afectado por DIF, son buenos indicios para considerar adecuada la utiliza-

ción de la escala. No obstante, cabe indicar que el estudio cuenta con la limitación de ser un muestreo incidental, lo que puede influir en las medidas. Para investigaciones futuras se debería replicar el

experimento en varias muestras e incluir medidas de la validez de contenido, criterio y predictiva. Además, habría que rehacer los baremos, especialmente para la subescala de ansiedad estado.

Referencias

- Alonso, J., Angermeyer, M.C., Bernert, S., Bruffaerts, R., Brugha, T.S., Bryson, H., et al. (2004a). 12-Month comorbidity patterns and associated factors in Europe: Results from the European Study of the Epidemiology of Mental Disorders (ESEMeD) Project. *Acta Psychiatrica Scandinavica*, 109 (Suppl. 420), 28-37.
- Alonso, J., Angermeyer, M.C., Bernert, S., Bruffaerts, R., Brugha, T.S., Bryson, H., et al. (2004b). Prevalence of mental disorders in Europe: Results from the European Study of the Epidemiology of Mental Disorders (ESEMeD) Project. *Acta Psychiatrica Scandinavica*, 109 (Suppl. 420), 21-27.
- Alonso, J., Angermeyer, M.C., Bernert, S., Bruffaerts, R., Brugha, T.S., Bryson, H., et al. (2004c). Use of mental health services in Europe: Results from the European Study of the Epidemiology of Mental Disorders (ESEMeD) Project. *Acta Psychiatrica Scandinavica*, 109 (Suppl. 420), 47-54.
- Andlin-Sobocki, P., y Wittchen, H.U. (2005). Cost of anxiety disorders in Europe. *European Journal of Neurology*, 12 (Suppl. 1), 39-44.
- Barnes, L., Harp, D., y Jung, W.S. (2002). Reliability generalization of scores on the Spielberger State-Trait Anxiety Inventory. *Educational and Psychological Measurement*, 62, 603-618.
- Bermúdez, J. (1978a). Anxiety and performance. *Revista de Psicología General y Aplicada*, 151, 183-207.
- Bermúdez, J. (1978b). Functional analysis of anxiety. *Revista de Psicología General y Aplicada*, 153, 617-634.
- Carretero-Dios, H., y Pérez, C. (2007). Standards for the development and review of instrumental studies: Considerations about test selection in psychological research. *International Journal of Clinical and Health Psychology*, 7, 863-882.
- Eysenck, H.J. (1995). Un modelo de personalidad: rasgos generales. En M.D. Avia y M.L. Sánchez-Bernados (Comps.), *Personalidad: aspectos cognitivos y sociales*. Madrid: Pirámide.
- Hidalgo, M.D., y Gómez, J. (2006). Nonuniform DIF detection using discriminant logistic analysis and multinomial logistic regression: A comparison for polytomous items. *Quality & Quantity*, 40, 805-823.
- Hidalgo, M.D., Gómez, J., y Padilla, J.L. (2005). Regresión logística: alternativas de análisis en la detección del funcionamiento diferencial del ítem. *Psicothema*, 17, 509-515.
- Hidalgo, M.D., y López-Pina, J.A. (2004). Differential item functioning detection and effect size: A comparison between logistic regression and Mantel-Haenszel procedures. *Educational and Psychological Measurement*, 64, 903-915.
- Hishinuma, E.S., Miyamoto, R.H., Nishimura, S.T., y Nahulu, L.B. (2000). Differences in State-Trait Anxiety Inventory Scores for ethnically diverse adolescents in Hawaii. *Cultural Diversity and Ethnic Minority*, 6, 73-83.
- Jodoin, M.G., y Gierl, M.J. (2001). Evaluating type I Error and power rates using an effect size measure with the logistic regression procedure for DIF detection. *Applied Measurement in Education*, 14, 329-349.
- Mellenbergh, G.J. (1995). Conceptual notes on models for discrete polytomous item responses. *Applied Psychological Measurement*, 19, 91-100.
- Miller, T.R., y Spray, J.A. (1993). Logistic discriminant function analysis for DIF identification of polytomously scored items. *Journal of Educational Measurement*, 30, 107-122.
- Millsap, R.E., y Meredith, W. (1992). Inferential conditions in the statistical detection of measurement bias. *Applied Psychological Measurement*, 16, 389-402.
- Montero, I., y León, O.G. (2007). A guide for naming research studies in Psychology. *International Journal of Clinical and Health Psychology*, 7, 847-862.
- Muñiz, J., y Fernández-Hermida, J.R. (2010). La opinión de los psicólogos españoles sobre el uso de los test. *Papeles del Psicólogo*, 31, 108-121.
- Spielberger, C.D. (1989). *State-Trait Anxiety Inventory: A comprehensive bibliography* (2ª ed.). Palo Alto, CA: Consulting Psychologists Press.
- Spielberger, C.D., y Díaz-Guerrero, R. (Eds.) (1976). *Cross-cultural anxiety*. Washington, DC: Hemisphere/Wiley.
- Spielberger, C.D., y Díaz-Guerrero, R. (Eds.) (1983). *Cross-cultural anxiety* (Vol. 2). Washington, DC: Hemisphere/Wiley.
- Spielberger, C.D., y Díaz-Guerrero, R. (Eds.) (1986). *Cross-cultural anxiety* (Vol. 3). Washington, DC: Hemisphere/Wiley.
- Spielberger, C.D., Díaz-Guerrero, R., y Strelau, J. (Eds.) (1990). *Cross-cultural anxiety* (Vol. 4). Washington, DC: Hemisphere/Wiley.
- Spielberger, C.D., Gorsuch, R., y Lushene, R. (1970). *Manual for the State-Trait Anxiety Inventory*. Palo Alto, California: Consulting Psychologist Press.
- Spielberger, C.D., Gorsuch, R.L., y Lushene, R. (1982). *Manual del Cuestionario de Ansiedad Estado/Rasgo (STAI)*. Madrid, España: TEA Ediciones.
- Suzuki, T., Tsukamoto, K., y Abe, K. (2000). Characteristic factor structures of the Japanese version of the State-Trait Anxiety Inventory: Coexistence of positive-negative and State-Trait Factor Structures. *Journal of Personality Assessment*, 74, 447-458.
- Vera-Villarreal, P., Buela-Casal, G., y Spielberger, C.D. (2007). Preliminary analysis and normative data of the State-Trait Anxiety Inventory (STAI) in adolescent and adults of Santiago, Chile. *Terapia Psicológica*, 25, 155-162.
- Vigneau, F., y Cormier, S. (2008). The factor structure of the State-Trait Anxiety Inventory: An alternative view. *Journal of Personality Assessment*, 90, 280-285.
- Wang, P.S., Aguilar-Gaxiola, S., Alonso, J., Angermeyer, M.C., Borges, G., Bromet, E.J., et al. (2007). Use of mental health services for anxiety, mood and substance disorders in 17 countries in the WHO world mental health surveys. *The Lancet*, 370, 841-850.