

EDUCATIONIS MOMENTUM

vol. 4, n.º 1, 2018, pp. 55-77. ISSN (impr.): 2414-1364; (online): 2517-9853

Inventario de Ansiedad ante Exámenes
(versión en español): diferencias según sexo
y valores normativos en universitarios de Lima

Test Anxiety Inventory–Spanish version:
Sex Differences and Normative Data
in College Students from Lima

Sergio DOMINGUEZ-LARA
Universidad de San Martín de Porres
sdominguezl@usmp.pe

Recibido: 2018.08.20

Aprobado: 2018.10.08

Resumen

El objetivo del presente trabajo fue analizar la invarianza de medición del Test Anxiety Inventory (versión en español: TAI-E) entre varones y mujeres, así como la obtención de valores normativos (baremos) en estudiantes universitarios de Lima. Participaron 866 estudiantes (71.1 % mujeres) de tres universidades privadas ubicadas en Lima Metropolitana, con edades comprendidas entre 17 y 52 años ($M = 20.38$). La invarianza de medición fue evaluada mediante un análisis factorial confirmatorio de grupo múltiple, el que indica que el grado de invarianza obtenido es suficiente para afirmar que la medición de la ansiedad ante exámenes es similar entre varones y mujeres. Del mismo modo, los coeficientes de confiabilidad utilizados (α , ω y H) presentan magnitudes elevadas; finalmente, los baremos fueron construidos con percentiles debido a que los puntajes del TAI-E no se aproximan a la normalidad. Se discuten los resultados y se sugieren guías para utilizar las normas.

Palabras clave: Inventario de Ansiedad, confiabilidad, validez, invarianza factorial, universitarios, baremos

Abstract

The aim of this paper was analyze the measurement invariance of Test Anxiety Inventory (Spanish Version: TAI-S) between men and women, and the obtention of normative data in college students from Lima. Were evaluated 866 college students (71.1 % women) with aged between 17 and 52 years ($M = 20.38$). The measurement invariance was performed with multigroup confirmatory factor analysis, which shows that degree of measurement invariance is enough for claim that the measurement is the same among men and women. Same way, the reliability coefficients (α , ω y H) have high magnitude. The normative data were elaborated with percentiles because the TAI-S scores don't have normal distribution. The results are discussed and practical guidelines are given for the use of the norms.

Keywords: Test Anxiety, reliability, validity, factorial invariance, university students, normative data

El *Test Anxiety Inventory* [TAI] (Spielberger, 1980; Spielberger, González, Taylor, Algaze, & Anton, 1978) es uno de los instrumentos más usados para la valoración de la Ansiedad ante Exámenes [AE] (Chapell et al., 2005) tanto en investigación como en entornos clínicos (Oktedalen & Hagtvet, 2011; Szafranski, Barrera, & Norton, 2012), y cuenta con estudios instrumentales en diversas partes del mundo, incluida una versión en español (Bauermeister, Collazos, & Spielberger, 1983).

El TAI se basa en el modelo dual (Liebert & Morris, 1967; Spielberger, 1980), es decir, considera que la AE se manifiesta de dos formas: fisiológica (emocionalidad) y cognitiva (preocupación); sin embargo, un análisis minucioso de los resultados proporcionados por diversos estudios instrumentales en diferentes contextos culturales permite concluir que el TAI realiza una evaluación unidimensional de la AE, debido a que existe poca diferenciación empírica entre las dos dimensiones evaluadas, es decir, un pronunciado solapamiento entre ellas (Dominguez-Lara & Cruz-Contreras, 2017).

La AE es uno de los tópicos más estudiados desde su formalización como constructo hace más de sesenta años (Mandler & Sarason, 1952) y sus aplicaciones son vigentes en distintos ámbitos de la psicología (Stöber & Pekrun, 2004). Así mismo, sus múltiples correlatos lo erigen como un constructo de trascendental importancia (Dominguez-Lara & Cruz-Contreras, 2017), por lo que es necesario su estudio a fin de evitar consecuencias negativas en el estudiante a corto y largo plazo.

A pesar de ello, no existen estudios nacionales que informen sobre la magnitud del problema en estudiantes universitarios, es decir, la prevalencia de AE, quizá el principal inconveniente es la ausencia de investigaciones que brinden datos normativos (baremos) y puntos de corte apropiados. A la fecha, el único trabajo que brinda baremos del TAI-E está centrado en población escolar (Aliaga, Ponce, Bernaola, & Pecho, 2001) y su uso no sería apropiado en población universitaria.

En ese estudio (Aliaga et al., 2001), la construcción de normas fue realizada por separado: varones y mujeres. Esta decisión se sustentó en las diferencias estadísticas entre las puntuaciones de varones y mujeres, a favor de estas últimas. Las puntuaciones más elevadas en AE en mujeres se observan consistentemente hasta la fecha en diversos estudios empíricos e instrumentales

(Chapell et al., 2005; Di María & Di Nuovo, 1990; Farooqui, Ghani, & Spielberger, 2012; Kavakci, Semiz, Kartal, Dikici, & Kugu, 2014; Oktedalen & Hagtvet, 2011; Onyeizugbo, 2010), incluso en aquellas investigaciones que no usan el TAI-E como medida de AE (Heredia, Piemontesi, Furlán, & Hodapp, 2008). La decisión para construir baremos separados debe fundamentarse en las desigualdades existentes en el proceso de medición del constructo, es decir, con base en los resultados de procedimientos que evalúen la posible presencia de un sesgo en la medición (Gómez-Benito, Hidalgo, & Guilera, 2010) y no solo en las diferencias observadas entre las puntuaciones directas de ambos grupos, como se sugiere habitualmente (Szafranski et al., 2012).

Algunos autores argumentan que las mujeres puntúan más alto en las pruebas de AE a causa de las elevadas exigencias sociales hacia su desempeño (Huang, 2013; Putwain & Daly, 2014; Rosário et al., 2008), lo que les produciría más temor en comparación con los varones (Benson, Bandalos, & Hutchinson, 1994). Además de la ansiedad propia de la evaluación durante el examen, estarían involucrados procesos cognitivos asociados al temor al fracaso, desaprobación social u otros afines. Estas creencias son ajenas al constructo AE y, por ende, irrelevantes para su correcta medición. En tal sentido, de ser verdadera esa afirmación y a pesar de que es inevitable la presencia del error de medición, los instrumentos utilizados probablemente están sesgados a favor de los varones, dichas diferencias serían a causa de otros constructos y no necesariamente de la AE.

En ese sentido, cuando el objetivo de la investigación es comparar grupos de forma apropiada, el procedimiento preliminar de relevancia en la actualidad es el análisis de la invarianza de medición (Byrne, 2008; Dimitrov, 2010). Este consiste en una serie de procedimientos que permiten saber si el constructo es medido de manera similar entre los grupos estudiados, y de ese modo conocer la presencia (o ausencia) de diferencias que podrían ser atribuidas a algún tipo de sesgo a favor de uno de los grupos.

Algunas investigaciones han examinado el grado de invarianza de medición del TAI entre varones y mujeres; mientras que un estudio lo hizo de forma completa y concluyó a favor de una estructura invariante (Everson, Millsap, & Rodriguez, 1991), otro lo hizo al interior de un modelo explicativo y evaluó únicamente la invarianza métrica de cargas factoriales (Benson et al., 1994); finalmente, un estudio solo analizó el nivel configuracional, es decir, si la

estructura hipotetizada de dos factores era similar entre varones y mujeres (Ware, Galassi, & Dew, 1990). No obstante, el único criterio para evaluar los distintos niveles de invarianza fue el cambio en el estadístico χ^2 , lo que en la actualidad es poco recomendado debido a su sensibilidad tanto al tamaño muestral como a la falta de normalidad de los datos (Chen, 2007; Dimitrov, 2010), como es el caso de las puntuaciones del TAI, predominantemente asimétricas. Hace falta implementar estudios que superen esas limitaciones metodológicas.

Una vez comprobada la invarianza de medición podrían elaborarse baremos que permitan un acercamiento más preciso a la prevalencia de AE en población universitaria, o un uso en un contexto terapéutico (Ergene, 2003). Por el contrario, si el grado de invarianza presentado no fuera el suficiente (al menos invarianza fuerte), los resultados no serían comparables entre sí ya que las diferencias no estarían en función de la AE y, en consecuencia, tendrían que elaborarse normas por separado.

En ese sentido, los objetivos del presente estudio fueron: a) analizar la invarianza de medición entre estudiantes varones y mujeres pertenecientes a diversas carreras respecto a la AE evaluada con el TAI-E; b) elaborar datos normativos del TAI-E en universitarios; y c) analizar la confiabilidad de los puntos de corte elegidos para determinar la presencia de elevada AE.

Método

Diseño

Se trata de un estudio empírico de tipo comparativo (Ato, López, & Benavente, 2013) porque busca analizar las diferencias existentes entre varones y mujeres respecto a la AE mediante métodos multivariados (análisis factorial de grupo múltiple, [AFGM]). A su vez, es un estudio instrumental (Ato et al., 2013), cuyo objetivo es el análisis de las propiedades psicométricas del TAI-E y elaborar baremos.

Participantes

La muestra intencional estuvo formada por 866 estudiantes (71.1 % mujeres) cuyas edades estuvieron comprendidas entre los 17 y 52 años ($M = 20.38$; $DE = 4.134$; 91.4 % entre 17 y 25 años). Todos ellos pertenecían a tres universidades privadas ubicadas en Lima Metropolitana ($n_1 = 579$; $n_2 = 230$; $n_3 = 57$) y cursan del primero al séptimo semestre de tres carreras profesionales ($n_{\text{psicología}} = 606$; $n_{\text{contabilidad}} = 203$; $n_{\text{enfermería}} = 57$). No se encontraron discrepancias estadísticamente significativas en la distribución de varones y mujeres según la universidad ($\chi^2_{[2]} = 9.310$; $p = 0.01$; $V = 0.104$), ni en cuanto a la edad ($t_{(864)} = 0.856$; $p = 0.392$; $d = 0.07$). Aquellos casos con valores perdidos ($n = 10$) fueron eliminados del análisis.

Instrumento: Inventario de Ansiedad ante Exámenes

Fue utilizada la versión adaptada a universitarios limeños (Dominguez-Lara & Cruz-Contreras, 2017; Villegas, Dominguez, Sotelo, & Sotelo, 2015) que evalúa de manera unidimensional la AE por medio de 19 ítems con cuatro alternativas de respuesta (*casi nunca*, *algunas veces*, *frecuentemente* y *casi siempre*). Los ítems se presentan en el mismo orden original (Spielberger, 1980), excepto por el primer ítem que fue eliminado debido a una nula relación con los demás ítems (Villegas et al., 2015).

Procedimiento

El TAI-E fue aplicado en los distintos claustros universitarios en el horario habitual de clase, previa solicitud de permiso a las autoridades y en coordinación con los docentes. Antes de la evaluación se solicitó la colaboración a los estudiantes, se les explicó que dicha actividad formaba parte de una investigación y que la participación era voluntaria y no serían compensados económica ni académicamente. Todos los estudiantes participaron, cuando finalizó la evaluación se les agradeció por su colaboración.

Invarianza entre grupos

Luego de un análisis descriptivo preliminar de los ítems, la invarianza de medición de la AE entre varones y mujeres fue explorada mediante un AFGM

(Brown, 2006; Byrne, 2008; Little, 1997). Este procedimiento se ejecutó a fin de garantizar que la potencial comparación entre varones y mujeres tenga como base una medición equivalente de la AE en ambos grupos (Dimitrov, 2010) y, de corroborarse, posibilitaría la construcción de un único baremo para la muestra total. Por el contrario, si la medición de la AE difiere según los grupos, es necesario elaborar baremos por separado.

El AFGM es utilizado para evaluar consecutiva y acumulativamente la invarianza de medición, centrada en variables observadas, a través de la igualdad de configuración total del instrumento en ambos grupos (invarianza configuracional), de los coeficientes de configuración o cargas factoriales (invarianza métrica), de los interceptos (invarianza fuerte) y residuales (invarianza estricta) (Byrne, 2008; Dimitrov, 2010).

Para dicho procedimiento se usó el programa EQS 6.2 (Bentler & Wu, 2012) bajo las siguientes condiciones: método de máxima verosimilitud (Beauducel & Herzberg, 2006), matrices de covarianzas y estadísticos robustos. Si bien correspondería el uso de una matriz de correlaciones policóricas dada la naturaleza ordinal de los ítems (Dominguez-Lara, 2014), se optó por las matrices de covarianzas debido a que se implementó el modelamiento de los interceptos, es decir, de la media de cada ítem.

Para valorar el ajuste de los modelos estudiados fueron usados diversos índices como el SRMR (≤ 0.05), RMSEA (≤ 0.05), CFI (≥ 0.95), y la prueba general χ^2 , ajustada para atenuar el efecto de la falta de normalidad de las variables [$SB-\chi^2$] (Satorra & Bentler, 1994). Así mismo, los criterios para determinar la presencia de cada grado de invarianza fueron las variaciones en el CFI [$\Delta_{CFI} < 0.01$] (Cheung & Rensvold, 2002), el RMSEA [$\Delta_{RMSEA} > -0.015$] (Chen, 2007), y SRMR [$\Delta_{SRMR} \geq 0.030$] (Chen, 2007). Finalmente, se calcularon en cada paso los índices modificadores [IM] (Sörbom, 1989) con el objetivo de evaluar individualmente la equivalencia de los parámetros para cada ítem (cargas factoriales e interceptos) durante todo el proceso, ya que si bien los índices de ajuste son importantes también lo son otros indicadores relevantes en la toma de decisiones (Dominguez-Lara, 2016a).

De este modo, una vez realizadas las etapas del análisis de invarianza de medición, alcanzar la invarianza fuerte permitiría concluir que la medición

es similar entre los dos grupos, lo que haría posible la comparación entre las medias latentes de varones y mujeres.

Confiabilidad

Preliminarmente fue evaluado el modelo tau-equivalente (Dunn, Baguley, & Brunnsden, 2014; Graham, 2006) con el objetivo de justificar el uso posterior del coeficiente α (Cronbach, 1951) como estimador de la confiabilidad de las puntuaciones. Las variaciones en el CFI [$\Delta_{CFI} < - 0.01$] (Cheung & Rensvold, 2002), el RMSEA [$\Delta_{RMSEA} > - 0.015$] (Chen, 2007) y SRMR [$\Delta_{SRMR} \geq 0.030$] (Chen, 2007) informan sobre la pertinencia de considerar como estadísticamente equivalentes las cargas factoriales de los ítems del TAI-E. Posteriormente, se reporta el coeficiente α con intervalos de confianza [IC] (Dominguez, 2016b) utilizando el módulo *ICalfa* (Dominguez-Lara & Merino-Soto, 2015), coeficiente ω (McDonald, 1999) y el coeficiente H (Hancock & Mueller, 2001). Por último, en vista de la potencial presencia de puntuaciones extremas, fueron calculadas versiones robustas de los coeficientes α y ω (Zhang & Yuan, 2016).

Elaboración de baremos

Las características distribucionales de los datos fue analizada con la prueba de Shapiro y Wilk (1965) dado que presenta mejores resultados en comparación a otras pruebas de normalidad (Ghasemi & Zahediasl, 2012), a fin de elegir el tipo de puntuación equivalente para la construcción del baremo, es decir, T o percentiles (P_c), en el caso del cumplimiento de la normalidad univariada; sin embargo, cabe mencionar que para constructos como la AE es más probable encontrar una distribución no normal, porque evalúa comportamientos que no están simétricamente distribuidos en la población de estudio (Micceri, 1989; Norton, Cosco, Doyle, Done, & Sacker, 2013; Ho & Yu, 2015). Del mismo modo, fue construido un baremo general a fin de poder captar con mayor facilidad las desviaciones respecto a los patrones generales de conducta (Seisdedos, 2000).

Luego de ello fue calculado el coeficiente $K2$ (Livingston, 1972). Este coeficiente es útil para conocer la confiabilidad de cada punto de corte elegido y valorar qué tan precisa es la clasificación realizada (Fernández & Merino,

2014). A su vez, no es sensible a las características distribucionales de las puntuaciones, por ejemplo, cuando no se cumple el supuesto de normalidad (Gempp & Saiz, 2014), lo que facilita su aplicación aún en distribuciones asimétricas.

Resultados

Invarianza de medición

El modelo unidimensional fue analizado en varones y mujeres por separado, previamente se evidenció características descriptivas y distribucionales adecuadas. En ambos grupos la asimetría ($g_1 < 2$) y curtosis ($g_2 < 7$) de los ítems alcanzaron magnitudes moderadas (Finney & DiStefano, 2013), y el supuesto de normalidad multivariada no fue alcanzado en vista de que el coeficiente de multivariado de Mardia (1970) obtuvo valores elevados [> 70] (Rodríguez & Ruiz, 2008). A pesar de ello, los procedimientos robustos ejecutados pueden superar estas limitaciones distribucionales.

El ajuste del modelo unidimensional se mostró satisfactorio tanto en mujeres (SB - $\chi^2_{[152]} = 340.022$ ($p < 0.01$), CFI = 0.945, RMSEA (IC 90 %) = 0.045 [0.038, 0.051], SRMR = 0.039); como en varones (SB - $\chi^2_{[152]} = 224.272$ ($p < 0.01$), CFI = 0.948, RMSEA (IC 90 %) = 0.044 [0.031, 0.055], SRMR = 0.050).

En cuanto al AFGM, la invarianza configuracional obtuvo índices de ajuste cercanos a los hallados en cada grupo por separado: SB - $\chi^2_{(304)} = 560.122$ ($p < 0.01$), CFI = 0.946, RMSEA (IC 90 %) = 0.044 (0.038, 0.050), SRMR = 0.045. Estos indicadores sirven como línea base para imponer las restricciones gradualmente en la evaluación de las siguientes etapas del análisis de invarianza. Posteriormente, se fijó la igualdad de coeficientes de configuración (invarianza métrica), mostrando buen ajuste (SB - $\chi^2_{(323)} = 585.476$ ($p < 0.01$), CFI = 0.945, RMSEA (IC 90 %) = 0.043 [0.038, 0.049], SRMR = 0.055), así como cambios no significativos en el CFI ($\Delta_{CFI} = -0.001$), RMSEA ($\Delta_{RMSEA} = -0.001$) y SRMR ($\Delta_{SRMR} = 0.010$). Una exploración de los IM indica que el ítem 3 difiere entre los grupos ($\chi^2 = 5.753$, $p = 0.016$), pero la magnitud de sus estadísticos no fue elevada y dicha

variación puede ser a causa del error de muestreo; de este modo, se puede concluir que existe invarianza métrica.

Por último, se evaluó la invarianza fuerte y se obtuvo un ajuste similar a los anteriores: SB - $\chi^2_{(342)} = 653.051$ ($p < 0.01$), CFI = 0.944, RMSEA (IC 90 %) = 0.045 (0.039, 0.050), SRMR = 0.056. De este modo, las variaciones fueron mínimas en el CFI ($\Delta_{CFI} = -0.001$), RMSEA ($\Delta_{RMSEA} = -0.002$) y SRMR ($\Delta_{SRMR} = 0.001$). De igual manera que el procedimiento anterior, algunos IM presentaron significancia estadística, aunque su magnitud no fue lo suficientemente grande como para relajar tales restricciones; a pesar de todo, el del intercepto del ítem 13 fue elevado ($\chi^2 = 12.844$, $p < 0.001$).

En vista de ese escenario, fue relajada la restricción y evaluado nuevamente el ajuste del modelo. No existieron variaciones significativas en los índices de ajuste ($\Delta_{CFI} = 0.000$; $\Delta_{RMSEA} = 0.000$; $\Delta_{SRMR} = 0.000$). Por tal motivo, se decide mantener la restricción de igualdad del intercepto del ítem 13 entre los grupos, se concluye entonces que existe invarianza fuerte entre varones y mujeres.

Análisis comparativo

Ante el cumplimiento de la invarianza fuerte fueron comparadas las medias latentes de varones y mujeres. Preliminarmente, la media latente del primer grupo (mujeres) se fijó en cero (Brown, 2006). El ajuste fue estadísticamente similar al modelo de invarianza fuerte ($\Delta_{CFI} = 0.000$; $\Delta_{RMSEA} = 0.000$; $\Delta_{SRMR} = 0.001$), y existe diferencia estadísticamente significativa en cuanto a la AE a favor de las mujeres ($\Delta = 2.065$, $p < 0.001$). En cuanto a la comparación de las puntuaciones observadas, las mujeres puntuaron de forma estadísticamente significativa por encima de los varones ($t_{[864]} = 2.390$; $p = 0.017$), pero sin la suficiente significancia práctica ($d = 0.18$) (Dominguez-Lara, 2016c) que justifique la construcción de baremos separados según el sexo.

Confiabilidad

Con la muestra total fue analizado nuevamente el ajuste del modelo unidimensional, se obtuvieron indicadores aceptables: SB - $\chi^2_{(35)} = 137.051$ ($p < 0.001$), CFI = 0.992, RMSEA (IC 90 %) = 0.057 (0.047, 0.068), SRMR = 0.051;

cargas factoriales moderadas ($\lambda > 0.60$) y estadísticamente significativas ($p < 0.001$).

Con el objetivo de evaluar el modelo tau-equivalente se restringió la igualdad de cargas factoriales, se tomó como base el modelo congénérico evaluado previamente. Bajo esta condición los índices de ajuste alcanzaron magnitudes meritorias: SB - $\chi^2_{(35)} = 137.051$ ($p < 0.001$), CFI = 0.992, RMSEA (IC 90 %) = 0.057 (0.047, 0.068), SRMR = 0.051; y los cambios no fueron significativos en el CFI ($\Delta_{CFI} = -0.002$), RMSEA ($\Delta_{RMSEA} = 0.003$), y SRMR ($\Delta_{SRMR} = 0.018$) respecto al modelo unidimensional (congénérico). Ello indica que se cumple la tau-equivalencia entre los ítems del TAI-E.

Tanto el coeficiente α (0.917; IC95 % .906, 0.927) como los coeficientes ω (0.919) y H (0.922) tuvieron magnitudes elevadas (Hancock & Mueller, 2001; Merino, Navarro, & García, 2014; Ponterotto & Charter, 2009; Raykov & Hancock, 2005). Del mismo modo, en control de la influencia de las puntuaciones extremas, los coeficientes robustos son similares a los anteriores ($\alpha_{robusto} = 0.922$; $\omega_{robusto} = 0.927$).

Datos normativos

A partir de la existencia de invarianza de medición entre varones y mujeres y de que las diferencias no poseen significancia práctica, puede elaborarse un único baremo en vista que la medición de la AE no es distinta según el sexo. La puntuación total del TAI-E no se aproxima a una distribución normal ($SW_{(866)} = 0.967$, $p < 0.001$). Las respuestas tienden hacia el espectro más bajo de puntuaciones ($M = 36.289$; $DE = 10.703$) y poseen asimetría y curtosis de baja magnitud ($g_1 = 0.568$; $g_2 = -0.118$); en tal sentido, los baremos serán construidos en base a Pc.

Fueron calculados diversos percentiles (Dominguez-Lara, 2016d; Glaesmer et al., 2012; Nordin & Nordin, 2013; Vázquez, Duque, & Hervás, 2013), se tomó como referencia los cuartiles uno (Pc 25) y tres (Pc 75) para establecer los puntos de corte de los niveles alto ($> Pc 75$) y bajo ($< Pc 25$) en la evaluación de la AE. En cuanto a la delimitación de niveles, puntuaciones mayores que 44 ($> Pc 75$) indican elevada AE y menores que 27 ($< Pc 25$) se asocian con niveles bajos de AE (tabla 1).

Tabla 1*Datos normativos del TAI-E en universitarios*

PD	Pc	K2*	PD	Pc	K2*
19 – 21	5	.977 - .973	36 – 37	55	.917
22 – 23	10	.970 - .967	38 – 39	60	.919 - .922
24 – 25	15	.964 - .961	40	65	.926
26	20	.957	41	70	.931
27 – 28	25	.953 - .948	42 – 44	75	.935 - .945
29	30	.943	45 – 46	80	.950 - .951
30 – 31	35	.938 - .933	47 – 48	85	.959 - .962
32	40	.928	49 – 51	90	.966 - .971
33 – 34	45	.924 - .921	52 – 56	95	.974 - .981
35	50	.918	> 57	99	.983

PD: Puntuación directa. Pc: Percentil. K2: Coeficiente K2.

* Calculado con el puntaje más bajo y más alto del punto de corte.

En negrita: zona de elevada ansiedad ante exámenes.

Discusión

El objetivo principal del presente trabajo fue la obtención de valores normativos del TAI-E en estudiantes universitarios de Lima, previo análisis de invarianza de medición entre varones y mujeres. Si bien la diferencia estadísticamente significativa de puntuaciones en AE a favor de las mujeres parece corresponder con la literatura previa (Chapell et al., 2005; Di María & Di Nuovo, 1990; Farooqui et al., 2012; Kavakci et al., 2014; Oktedalen & Hagtvvet, 2011; Onyeizugbo, 2010), el aporte de la presente investigación es el análisis de la invarianza de medición del TAI-E entre varones y mujeres.

Hasta la fecha, los estudios que reportaban invarianza entre varones y mujeres con el TAI en muestras anglohablantes solo brindaban reportes parciales o con procedimientos metodológicos que actualmente no son considerados fiables (Benson et al., 1994; Everson et al., 1991; Ware et al., 1990). Esto daba la

sensación de que, efectivamente, la AE era invariante según el sexo y que las diferencias observadas podrían ser explicadas por dicho constructo; sin embargo, un análisis pormenorizado de esos reportes permite indicar que los hallazgos no fueron concluyentes. Esto quiere decir que las comparaciones pudieron estar sesgadas y las conclusiones derivadas carecerían de sustento ya que no fueron analizados todos los niveles posibles de invarianza, sobre todo la invarianza fuerte que, en última instancia, brinda evidencias para comparar puntajes (Dimitrov, 2010).

De este modo, con los métodos implementados se trata de superar esos inconvenientes y los resultados permitirían concluir, al menos con la muestra de estudio, que la medición de la AE realizada con el TAI en varones y mujeres es similar y, por tanto, las desigualdades observadas son susceptibles de ser interpretadas; sin embargo, y pese a existir diferencias estadísticamente significativas, la significancia práctica, es decir, la magnitud de las diferencias (Ferguson, 2009; Fritz, Morriz, & Richler, 2012) no fue elevada. En tal sentido, si llega a concluirse que existen diferencias significativas entre varones y mujeres respecto a la AE, probablemente los resultados estén sobreestimándose al basarse únicamente en la significancia estadística, que suele estar influida por el tamaño muestral, dejando de lado la significancia práctica (Dar, Serlin, & Omer, 1994; Frías, Pascual, & García, 2002; Monterde-Bort, Pascual, & Frías, 2006).

En tal sentido, ya que los puntajes podrían considerarse como estadísticamente equivalentes, parece ser que los estudiantes evaluados, independientemente del sexo, experimentan similar AE. Lamentablemente, al no existir estudios preliminares sobre las potenciales diferencias de sexo en población peruana, no podría concluirse que dicha equivalencia se debe a un cambio social en cuanto a las exigencias hacia la mujer como potencial causa de la elevada AE, como se plantea en otros manuscritos (Huang, 2013; Putwain & Daly, 2014; Rosário et al., 2008), o simplemente, que dichas diferencias nunca existieron en este contexto, en contraposición a otras culturas donde priman fuertes estereotipos de género respecto a las mujeres (Farooqui et al., 2012).

Otro aspecto que vale resaltar fueron los elevados coeficientes de confiabilidad (> 0.90) tanto de sus puntuaciones observadas (α) como de la variable latente (ω y H). Además de su dimensionalidad, estos indicadores resaltan aún más cualidades psicométricas del TAI-E, ya que se espera que el instrumento sea

usado para fines de apoyo al diagnóstico, o al menos para su inclusión en evaluaciones masivas y, posteriormente, como parte del perfil del educando.

En vista de estos hallazgos y de las características distribucionales de las puntuaciones del TAI-E, fueron elegidos los percentiles para elaborar datos normativos generales. Si bien pudo optarse por la construcción de normas para varones y para mujeres, la medición de la AE con el TAI-E es invariante, y las diferencias observadas parecen diluirse cuando se considera la significancia práctica en lugar de la estadística, por lo que no se justificaría tal decisión. Inclusive en otros estudios los puntos de corte difieren por un punto entre varones y mujeres (Kavakci et al., 2014), lo que los hace estadísticamente equivalentes a juzgar por la probable superposición de sus respectivos intervalos de confianza.

Como era de esperarse, existió una mayor cantidad de personas que obtuvieron puntuaciones bajas, produciendo una distribución asimétrica. A pesar de ello, fueron obtenidos sendos puntos de corte con bajo error de medición alrededor de ellos ($K2 > 0.90$), lo que posibilita una clasificación (p. ej. persona con alta AA) confiable.

Esto trae dos importantes implicancias prácticas: en primer lugar, contar con baremos posibilita el desarrollo de estudios preliminares de prevalencia en población universitaria; y en segundo lugar, en vista de las bondades psicométricas del TAI-E, podría ser utilizado como evaluación de entrada y salida en las intervenciones destinadas a la reducción de la AE.

Finalmente, existieron algunas limitaciones que podrían ser minimizadas en próximas investigaciones. Respecto a las características de la muestra, si bien el tamaño muestral es elevado, quizá un punto débil sea la proporción entre varones y mujeres, aunque en otros estudios de estandarización llevados a cabo con constructos similares este aspecto no fue percibido como una debilidad en el estudio (Furlán, Pérez, Moyano, & Cassady, 2010) ya que refleja la realidad que involucra las carreras estudiadas. Así mismo, aunque la cantidad de encuestados es similar a otros estudios de validación de instrumentos en población universitaria (Gempp & Avendaño, 2008; Furlán et al., 2010), hace falta incorporar estudiantes de más carreras profesionales para lograr una muestra completamente representativa, debido a que probablemente la AE sea experimentada de forma distinta entre las distintas profesiones.

Por último, la falta de contrastes adicionales del punto de corte con criterios empíricos (grupos definidos o constructos teóricamente relevantes) podría sugerir un uso mesurado del TAI-E para fines diagnósticos en entornos educativos, siempre acompañado de una entrevista clínica a fin de descartar otros aspectos asociados (p. ej. ansiedad) (Cardona-Arias, Pérez-Restrepo, Rivera-Ocampo, Gómez-Martínez, & Reyes, 2015).

Se recomienda para futuras investigaciones comparar los grupos con baja y elevada AE respecto a constructos asociados a manifestaciones negativas de la afectividad (estrategias cognitivas de regulación emocional, inteligencia emocional, burnout académico, etc.), inclusive estudios longitudinales para evaluar su impacto en la conducta académica del estudiante (procrastinación académica) o en la satisfacción con sus estudios.

Concluimos que la evaluación de la AE mediante el TAI-E se muestra invariante respecto al sexo, los indicadores de confiabilidad son adecuados y los puntos de corte parecen ser útiles para la realización de evaluaciones individuales.

Referencias

- Aliaga, J., Ponce, C., Bernaola, E., & Pecho, J. (2001). Características psicométricas del inventario de autoevaluación de la ansiedad ante exámenes (IDASE). *Paradigmas, Revista de Actualización Profesional*, 2(3), 11-29.
- Ato, M., López, J., & Benavente, A. (2013). Un sistema de clasificación de los diseños de investigación en psicología. *Anales de Psicología*, 29(3), 1038-1059.
- Bauermeister, J., Collazos, J., & Spielberger, C. (1983). The construction and validation of the Spanish form of the Test Anxiety Inventory: Inventario de Autoevaluación sobre Exámenes (IDASE). En C. Spielberger & Diaz-Guerrero (Eds.), *Cross-cultural Anxiety* (pp. 67-85). Washington: McGraw-Hill.

- Beauducel, A., & Herzberg, P. Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Structural Equation Modeling, 13*, 186-203.
- Bentler, P. M., & Wu, E. J. C. (2012). *EQS 6.2 for windows* [Statistical Program]. Encino, CA: Multivariate Software, Inc.
- Benson, J., Bandalos, D., & Hutchinson, S. (1994). Modeling test anxiety among men and women. *Anxiety, Stress, and Coping, 7*(2), 131-148.
- Brown, T. (2006). *Confirmatory Factor Analysis for Applied Research*. New York: The Guilford Press.
- Byrne, B. (2008). Testing for multigroup equivalence of a measuring instrument: A walk through the process. *Psicothema, 20*(4), 872-882.
- Cardona-Arias, J., Pérez-Restrepo, D., Rivera-Ocampo, S., Gómez-Martínez, J., & Reyes, A. (2015). Prevalencia de ansiedad en estudiantes universitarios. *Diversitas: Perspectivas en Psicología, 11*(1), 79-89.
- Chapell, M. S., Blanding, Z. B., Silverstein, M. E., Takahashi, M., Newman, B., Gubi, A., & McCann, N. (2005). Test anxiety and academic performance in undergraduate and graduate students. *Journal of Educational Psychology, 97*(2), 268-274.
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14*, 464-504.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*(2), 233-255.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297-334.
- Dar, R., Serlin, R. C., & Omer, H. (1994). Misuse of statistical tests in three decades of psychotherapy research. *Journal of Consulting and Clinical Psychology, 62*, 75-82.

- Di Maria, F., & Di Nuovo, S. (1990). Gender differences in social and test anxiety. *Personality and Individual Differences*, 11(5), 525-530.
- Dimitrov, D. M. (2010). Testing for factorial invariance in the context of construct validation. *Measurement and Evaluation in Counseling and Development*, 43(2), 121-149.
- Dominguez-Lara, S. (2014). ¿Matrices Policóricas/Tetracóricas o Matrices Pearson? Un estudio metodológico. *Revista Argentina de Ciencias del Comportamiento*, 6(1), 39-48.
- Dominguez-Lara, S. (2016a). Evaluación de modelos estructurales, más allá de los índices de ajuste. *Enfermería Intensiva*, 27, 84-85.
- Dominguez-Lara, S. (2016b). Intervalos de confianza en el reporte de la fiabilidad: un análisis necesario. *Anales del Sistema Sanitario de Navarra*, 39(1), 169-170.
- Dominguez-Lara, S. (2016c). Diferencias entre grupos y magnitud del efecto: un análisis complementario. *Archivos Argentinos de Pediatría*, 114 (4), e300-e301.
- Dominguez-Lara, S. (2016d). Valores normativos de una escala de autoeficacia académica en estudiantes universitarios de Lima. *Interacciones*, 2(2), 91-98.
- Dominguez-Lara, S., & De la Cruz-Contreras, F. (2017). Análisis estructural y desarrollo de una versión breve de la versión en español del Inventario de Ansiedad ante Exámenes (TAI-E) en universitarios de Lima. *Interacciones*, 3(1), 7-17.
- Dominguez-Lara, S., & Merino-Soto, C. (2015) ¿Por qué es importante reportar los intervalos de confianza del coeficiente alfa de Cronbach? *Revista Latinoamericana de Ciencias Sociales, Niñez y Juventud*, 13(2), 1326-1328.
- Dunn, T. J., Baguley, T., & Brunnsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, 105(3), 399-412.

- Ergene, T. (2003). Effective interventions on test anxiety reduction a meta-analysis. *School Psychology International*, 24(3), 313-328.
- Everson, H. T., Millsap, R. E., & Rodriguez, C. M. (1991). Isolating gender differences in test anxiety: A confirmatory factor analysis of the Test Anxiety Inventory. *Educational and Psychological Measurement*, 51(1), 243-251.
- Farooqui, Y. N., Ghani, R., & Spielberger, C. D. (2012). Gender differences in test anxiety and academic performance of medical students. *International Journal of Psychology and Behavioral Sciences*, 2(2), 38-43.
- Ferguson, C. J. (2009). An effect size primer: a guide for clinicians and researchers. *Professional Psychology: Research and Practice*, 40(5), 532-538.
- Fernández, M., & Merino, C. (2014). Error de medición alrededor de los puntos de corte en el MBI-GS. *Liberabit*, 20(2), 209-218.
- Finney S. J., y DiStefano, C. (2013). Nonnormal and categorical data in structural equation modeling. En G. R. Hancock y R. O. Mueller (Eds.), *Structural Equation Modeling: A second course* (pp. 269-492). Greenwich, CT: Information Age Publishing Inc.
- Frías, M. D., Pascual, J., & García, J. F. (2002). La hipótesis nula y la significación práctica. *Metodología de las Ciencias del Comportamiento (Volumen especial)*, 181-185.
- Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, 141(1), 2-18.
- Furlán, L., Pérez, E., Moyano, M., & Cassady, J. (2010). Propiedades psicométricas y estandarización de la Escala de Ansiedad Cognitiva frente a los Exámenes a la población universitaria argentina. *Evaluar*, 10, 22-31.

- Gempp, R., & Avendaño, C. (2008). Datos normativos y propiedades psicométricas del SCL-90-R en estudiantes universitarios chilenos. *Terapia Psicológica, 26*(1), 39-58.
- Gempp, R. & Saiz, J. L. (2014). El coeficiente K2 de Livingston y la fiabilidad de una decisión dicotómica en un test psicológico. *Universitas Psychologica, 13*(1), 217-226.
- Ghasemi, A., & Zahediasl, S. (2012). Normality Tests for Statistical Analysis: A Guide for Non-Statisticians. *International Journal of Endocrinology and Metabolism, 10*(2), 486-489.
- Glaesmer, H., Rief, W., Martin, A., Mewes, R., Brahler, E., Zenger, M., & Hinz, A. (2012). Psychometric properties and population-based norms of the Life Orientation Test Revised (LOT-R). *British Journal of Health Psychology, 17*, 432-445.
- Gómez-Benito, J., Hidalgo, M. D., & Guilera, G. (2010). El sesgo de los instrumentos de medición. Tests justos. *Papeles del Psicólogo, 31*(1), 75-84.
- Graham, J. M. (2006). Congeneric and essentially tau-equivalent estimates of score reliability: What they are and how to use them. *Educational and Psychological Measurement, 66*, 930-944.
- Hancock, G. R., & Mueller, R. O. (2001). Rethinking construct reliability within latent variable systems. En R. Cudeck, S. H. C. du Toit & D. Sörbom (Eds.), *Structural equation modeling: Past and present. A Festschrift in honor of Karl G. Jöreskog* (pp. 195-261). Chicago: Scientific Software International.
- Heredia, D., Piemontesi, S., Furlán, L., & Hodapp, V. (2008). Adaptación del Inventario Alemán de Ansiedad frente a los Exámenes: GTAI-A. *Evaluar, 8*, 46-60.

- Ho, A. D., & Yu, C. Y. (2015). Descriptive Statistics for Modern Test Score Distributions Skewness, Kurtosis, Discreteness, and Ceiling Effects. *Educational and Psychological Measurement, 75*(3), 365-388.
- Huang, C. (2013). Gender differences in academic self-efficacy: a meta-analysis. *European Journal of Psychology of Education, 28*(1), 1-35.
- Kavakci, O., Semiz, M., Kartal, A., Dikici, A., & Kugu, N. (2014). Test anxiety prevalence and related variables in the students who are going to take the university entrance examination. *Dusunen Adam, 27*(4), 301.
- Liebert, R. M., & Morris, L. W. (1967). Cognitive and emotional components of test anxiety: A distinction and some initial data. *Psychological Reports, 20*(3), 975-978.
- Little, T. D. (1997). Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioral Research, 32*(1), 53-76.
- Livingston, S. A. (1972). Criterion-referenced applications of classical test theory. *Journal of Educational Measurement, 9*(1), 13-26.
- Mandler, G., & Sarason, S. B. (1952). A study of anxiety and learning. *Journal of Abnormal and Social Psychology, 47*, 166-713.
- Mardia, K. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika, 57*, 519-530.
- McDonald R. P. (1999). *Test theory: A unified treatment*. Mahwah, N. J.: L. Erlbaum Associates.
- Merino, C., Navarro, J., & García, W. (2014). Revisión de la consistencia interna del Inventario de Inteligencia Emocional de Bar-On, EQ-I: YV. *Revista Peruana de Psicología y Trabajo Social, 3*(1), 141-154.
- Micceri, T. (1989). The unicorn, the normal curve and other improbable creatures. *Psychological Bulletin, 105*(1), 156-166.

- Monterde-Bort, H., Pascual, J., & Frías, M. D. (2006). Errores de interpretación de los métodos estadísticos: importancia y recomendaciones. *Psicothema, 18*(4), 848-856.
- Nordin, M., & Nordin, S. (2013). Psychometric evaluation and normative data of the Swedish versión of the 10-ítem perceived stress scale. *Scandinavian Journal of Psychology, 54*, 502-507.
- Norton, S., Cosco, T., Doyle, F., Done, J., & Sacker, A. (2013). The Hospital Anxiety and Depression Scale: a meta confirmatory factor analysis. *Journal of Psychosomatic Research, 74*(1), 74-81.
- Oktedalen, T., & Hagtvet, K. A. (2011). A revised version of the Norwegian adaptation of the Test Anxiety Inventory in a heterogeneous population. *Scandinavian Journal of Educational Research, 55*(5), 475-487.
- Onyeizugbo, E. U. (2010). Self-efficacy, gender and trait anxiety as moderators of test anxiety. *Electronic Journal of Research in Educational Psychology, 8*(1), 299-312.
- Ponterotto, J. G., & Charter R. A. (2009). Statistical extensions of Ponterotto and Ruckdeschel's (2007) reliability matriz for estimating the adequacy of internal consistency coefficients. *Perceptual and Motor Skills, 108*, 878-886.
- Putwain, D., & Daly, A. L. (2014). Test anxiety prevalence and gender differences in a sample of English secondary school students. *Educational Studies, 40*(5), 554-570.
- Raykov, T., & Hancock, G. R. (2005). Examining change in maximal reliability for multiple-component measuring instruments. *British Journal of Mathematical and Statistical Psychology, 58*(1), 65-82.
- Rodriguez, M., & Ruiz, M. (2008). Atenuación de la asimetría y de la curtosis de las puntuaciones observadas mediante transformaciones de variables: Incidencia sobre la estructura factorial. *Psicológica, 29*, 205-227.

- Rosário, P., Núñez, J. C., Salgado, A., González-Pienda, J. A., Valle, A., Joly, C., & Bernardo, A. (2008). Ansiedad ante los exámenes: relación con variables personales y familiares. *Psicothema*, 20(4), 563-570.
- Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. En A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399-419). Thousand Oaks, CA: Sage.
- Seisdedos, N. (2000). Adaptación y tipificación: unos lujos indispensables (Estudio transcultural de los rasgos de personalidad). *Revista Iberoamericana de Diagnóstico y Evaluación Psicológica*, 10(2), 37-50.
- Shapiro, S. S. & Wilk, M. B. (1965). An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, 52(3/4), 591-611.
- Sörbom, D. (1989). Model modification. *Psychometrika*, 54(3), 371-384.
- Spielberger, C. D. (1980). *Test Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.
- Spielberger, C. D., González, H., Taylor, C., Algaze, B., & Anton, W. (1978). Examination stress and test anxiety. En C. Spielberger & I. Sarason (Eds.), *Stress and Anxiety*. Washintong: Hemisphere.
- Stöber, J., & Pekrun, R. (2004). Advances in test anxiety research. *Anxiety, Stress & Coping*, 17(3), 205-211.
- Szafranski, D. D., Barrera, T. L., & Norton, P. J. (2012). Test anxiety inventory: 30 years later. *Anxiety, Stress, & Coping: An international Journal*, 25(6), 667-677.
- Vázquez, C., Duque, A., & Hervás, G. (2013). Satisfaction with Life Scale in a representative sample of Spanish adults: validation and normative data. *Spanish Journal of Psychology*, 16, 1-15.
- Villegas, G., Dominguez, S., Sotelo, N., & Sotelo, L. (2015). Propiedades psicométricas del Inventario de Autoevaluación de la Ansiedad

ante Exámenes (IDASE) en universitarios de Lima. *Revista Mexicana de Psicología Educativa*, 3(1), 15-21.

Ware, W. B., Galassi, J. P., & Dew, K. M. H. (1990). The test anxiety inventory: A confirmatory factor analysis. *Anxiety Research*, 3(3), 205-212.

Zhang, Z., & Yuan, K. H. (2016). Robust coefficients alpha and omega and confidence intervals with outlying observations and missing data: Methods and software. *Educational and Psychological Measurement*, 76, 387-411.