

**APLICACIÓN DE LA TEORÍA DE RESPUESTA AL ÍTEM A LA ESCALA  
NEUROTICISMO DEL BIG FIVE INVENTORY**

**Facundo Juan Pablo Abal<sup>1</sup>, Javier Agustín Menéndez, Horacio Félix Attorresi**

Consejo Nacional de Investigaciones Científicas y Técnicas

Universidad de Buenos Aires

**APPLICATION OF THE ITEM RESPONSE THEORY TO THE NEUROTICISM  
SCALE OF THE BIG FIVE INVENTORY**

**APLICAÇÃO DA TEORIA DE RESPOSTA AO ITEM NA ESCALA DE  
NEUROTICISMO DOS BIG FIVE INVENTORY**

**RESUMEN.**

El impacto extendido que tiene el Neuroticismo en el sistema de salud ha motivado su medición para ser usada en tareas de prevención. Por su brevedad y su reconocimiento internacional, la escala de Neuroticismo del Big Five Inventory presenta características óptimas para estas aplicaciones. En este trabajo se propone profundizar el estudio de las propiedades psicométricas de esta escala utilizando Teoría de Respuesta al ítem (TRI). Colaboraron 742 adultos de Buenos Aires, Argentina (55.3% mujeres). Un primer análisis con TRI reveló la necesidad de eliminar uno de los ítems originales por presentar funcionamiento inadecuado. Para los 7 ítems restantes se verificaron los supuestos de unidimensionalidad e independencia local. Todos mostraron un ajuste aceptable al Modelo de Respuesta Graduada y una capacidad discriminativa media-alta. Tampoco se observó funcionamiento diferencial en función del género. Se hallaron correlaciones moderadas con las escalas del Inventario SCL-90-R. La Función de Información del Test mostró que la

---

<sup>1</sup> Correo electrónico: fabal@psi.uba.ar

precisión es aceptable aun para los valores extremos del rasgo, los cuales son de especial interés para las aplicaciones del ámbito clínico-epidemiológico.

**Palabras clave:** Neuroticismo – Estabilidad emocional – Modelo de los Cinco Factores - Teoría de Respuesta al ítem

## **ABSTRACT**

The widespread impact that Neuroticism has on the health system has motivated its measurement to be used in prevention tasks. Due to its brevity and international recognition, the Neuroticism scale of the Big Five Inventory presents optimal characteristics for these applications. In this work it is proposed to deepen the study of the psychometric properties of this scale using Item Response Theory (IRT). There were 742 adults from Buenos Aires, Argentina (55.3% women). At first IRT analysis revealed the need to eliminate one of the original items due to inadequate functioning. For the remaining 7 items, the assumptions of unidimensionality and local independence were verified. All showed an acceptable fit to the Graded Response Model and a medium-high discriminatory capacity. Gender differential functioning was not observed. Moderate correlations were found with the scales of the SCL-90-R Inventory. The Test Information Function showed that precision is acceptable even for extreme trait values, which are of special interest for clinical-epidemiological applications.

**Key words:** Neuroticism – Emotional stability – Five Factor Model - Item Response Theory

## **RESUMO**

O impacto generalizado que o Neuroticismo tem no sistema de saúde motivou a sua medição para ser utilizada em tarefas de prevenção. Devido à sua brevidade e reconhecimento internacional, a escala de Neuroticismo do Big Five Inventory apresenta características ótimas para estas aplicações. Neste trabalho propõe-se aprofundar o estudo das propriedades psicométricas desta escala utilizando a Teoria da Resposta ao item (TRI).

Havia 742 adultos de Buenos Aires, Argentina (55.3% mulheres). Uma primeira análise IRT revelou a necessidade de eliminar um dos itens originais devido ao funcionamento inadequado. Para os restantes 7 itens, foram verificados os pressupostos de unidimensionalidade e independência local. Todos apresentaram um ajuste aceitável ao Modelo de Resposta Graduada e uma capacidade discriminatória média-alta. O funcionamento diferencial de gênero também não foi observado. Correlações moderadas foram encontradas com as escalas do Inventário SCL-90-R. A Função de Informação do Teste mostrou que a acurácia é aceitável mesmo para valores extremos de traço, que são de especial interesse para aplicações clínico-epidemiológicas.

**Palavras-Chave:** Neuroticismo – Estabilidade emocional – Modelo dos Cinco Fatores - Teoria da Resposta ao item

## INTRODUCCIÓN.

El Neuroticismo (N) es considerado uno de los dominios fundamentales de la personalidad normal. Si bien es reconocida una extensa historia de antecedentes (e.g. Eysenck, 1947), su inclusión en el que actualmente se concibe como modelo hegemónico de la personalidad, el Modelo de los Cinco Factores (FFM, *Five Factor Model*), ha rubricado su relevancia como elemento indispensable en las teorías de los rasgos (Digman, 1990; Goldberg, 1993; McCrae y Costa, 2010). Este factor caracteriza la tendencia de las personas a percibir afectos negativos (tristeza, miedo, culpa, enojo, entre otros) de manera inusualmente intensa y/o recurrente. Dicho aspecto se asocia con la aparición de problemas de regulación emocional y manejo del estrés, preocupaciones constantes, sentimientos de insatisfacción y conflictos interpersonales (Goldberg, 1993; McCrae y Costa, 2010).

Aun cuando no se describe como un rasgo desadaptativo de la personalidad, en las últimas décadas se ha acumulado abundante evidencia empírica que demuestra la fuerte asociación de N con múltiples trastornos mentales y patologías orgánicas (Friedman, 2019; Hengartner, Kawohl, Haker, Rössler y Ajdacic-Gross, 2016; Lahey, 2009; Tackett y Lahey, 2017). Es entendido como un factor de vulnerabilidad subyacente para el desarrollo y mantenimiento de distintas afecciones (Sauer-Zavala, Wilner, y Barlow, 2017) con un

importante costo económico para los sistemas de salud (Lahey, 2009; Vittengl, 2017). Esto implica que las personas que puntúan alto en N no necesariamente revisten gravedad psicopatológica, pero sí cierta predisposición a padecer los trastornos psicológicos más comunes (Jeronimus, Kotov, Riese y Ormel, 2016), a exagerar la importancia de los síntomas físicos, a usar con mayor frecuencia los servicios de salud (Hajek, Bock, y König, 2017) y a un mayor consumo de medicación con o sin prescripción (Chapman y Goldberg, 2017).

El N es un dominio disposicional con una base genética considerable pero que puede tener modificaciones causadas por factores ambientales (Sauer-Zavala et al., 2017; Widiger, 2009). Diversos estudios longitudinales permiten concluir que se trata de un rasgo relativamente más maleable de lo que se suponía anteriormente, impulsando el diseño de tratamientos con objetivos centrados en su disminución (e.g. Drake, Morris y Davis, 2017). En esta línea, en los últimos años cobró relevancia la detección de niveles elevados de N en entrevistas clínicas de rutina o en estudios epidemiológicos a gran escala (Widiger y Oltmanns, 2017) para la implementación de estrategias de intervención y prevención transdiagnósticas.

### *La medición del Neuroticismo*

Existen numerosos instrumentos desarrollados para la medición de N, siendo los derivados del FFM aquellos más utilizados en la actualidad (Raad y Mlačić, 2015). La mayoría constituyen escalas incluidas en cuestionarios extensos que evalúan todos los dominios del modelo y sus respectivas facetas. Entre estos, los inventarios NEO-PI-3 (McCrae y Costa, 2010) y NEO-IPIP (Goldberg et al., 2006) ocupan un lugar destacado. Otros tests, en cambio, se han propuesto exclusivamente para la medición del N y sus facetas (Taylor y DeBruin, 2006; Zanon y Hutz, 2009).

Cabe señalar, no obstante, que la cantidad de instrumentos disponibles para medir el N queda acotada cuando se piensa en las circunstancias que enmarcan las evaluaciones con fines de despistaje o estudios a gran escala. En estos casos, el contexto obliga a priorizar instrumentos que garanticen una medición válida y confiable con la menor inversión posible de tiempo (Donnellan, Oswald, Baird, y Lucas, 2006; Rammstedt y Beierlein,

2014). En este sentido, los autores de las pruebas tradicionales han propuesto versiones abreviadas como NEO-FFI-3 (McCrae y Costa, 2007) o escalas de 10 o 20 elementos con ítems del IPIP (Goldberg et al., 2006) para ahorrar tiempo de administración y disminuir el impacto de los errores de medida que podrían ocasionar la fatiga o aburrimiento del evaluado. También se han desarrollado pruebas cortas y ultra-cortas (Donnellan, et al., 2006; Gosling, Rentfrow y Swann, 2003; Rammstedt y John, 2007), que abren una controversia con respecto a si es posible ofrecer garantías mínimas de calidad psicométrica usando tan escasa cantidad de ítems (Ziegler, Kemper y Kruey, 2014).

Entre las variadas propuestas instrumentales, la escala N del Big Five Inventory (BFI, John, Donahue y Kentle, 1991) parece ser una de las opciones más razonable para realizar una medición rápida de este dominio. La escala se compone de ocho elementos con formato de respuesta tipo Likert, lo que resulta una cantidad de ítems razonable para captar las diferencias individuales con una precisión aceptable. A esto debe sumarse la evidencia que demuestra su utilidad en estudios transculturales y translingüísticos (Bleidorn et al., 2013; Gurven, von Rueden, Massenkoff, Kaplan, y Lero Vie, 2013; Schmitt et al., 2007). No solo cuenta con propiedades psicométricas adecuadas su versión original en inglés (John, Naumann, y Soto, 2008) sino que también en sus adaptaciones al chino (Carciofo Yang, Song, Du y Zhang, 2016), alemán (Rammstedt, 1997), holandés (Denissen, Geenen, van Aken, Gosling y Potter, 2008), francés (Plaisant, Courtois, Réveillère, Mendelsohn, y John, 2010), turco (Karaman, Dogan, y Coban, 2010), italiano (Ubbiali, Chiorri, Hampton y Donati, 2013) y árabe (Alansari, 2016).

Para el idioma castellano cobra relevancia la versión de España publicada por Benet-Martínez y John (1998), la cual ha sido aplicada sin modificaciones en Perú (Domínguez-Lara, Merino-Soto, Zamudio, y Guevara-Cordero, 2018) y México (Reyes Zamorano, Álvarez Carrillo, Peredo Silva, Miranda Sandoval y Rebolledo Pastrana, 2014). Los adaptadores de la versión colombiana (Salgado, Vargas-Trujillo, Schmutzler y Wills-Herrera, 2016) identificaron la necesidad de realizar ajustes léxicos y semánticos a la traducción para lograr una mejor adecuación a la propia cultura. Existe una adaptación Argentina que demostró poseer una propiedades psicométricas aceptables en población adolescente (Castro Solano y Casullo, 2001) y adulta (Castro Solano, 2002). No obstante,

los estudios de consistencia interna resultaron más bajos comparados con los resultados de otras culturas (alfas de .68 y .74 respectivamente).

A pesar de los avances alcanzados en la adaptación de esta escala, los estudios instrumentales se han generado principalmente en el marco de la Teoría Clásica de los Tests (TCT), presentando así las limitaciones de este modelo psicométrico (Muñiz, 2018). El análisis de ítems con TCT resulta de gran utilidad pero está lejos de ser exhaustivo, por lo que siempre es fructífero profundizar en el estudio del funcionamiento de cada elemento usando la perspectiva de la Teoría de Respuesta al Ítem (TRI). Mucho más valorado es este análisis en la medición de variables de la personalidad, donde la TRI comienza a tomar visibilidad (Abal, Lozzia, Aguerri, Galibert, y Attorresi, 2010; Reise y Revicki, 2015).

### *Objetivo*

Tomando en consideración el reconocimiento internacional que presenta el BFI y rescatando la potencial utilidad de su escala de Neuroticismo para aplicaciones del ámbito clínico-epidemiológico, en este trabajo se propone un análisis detallado de los ítems que miden este dominio bajo la óptica de la TRI. Si bien la TRI ofrece una amplia gama de modelos para analizar formatos tipo Likert (politómico de respuesta de ordenada), numerosos estudios encontraron evidencia empírica de que el Modelo de Respuesta Graduada (MRG) de Samejima (2010) brinda una adecuada estructura paramétrica para describir las respuestas a ítems de personalidad (e.g. Maples-Keller, et al., 2017). En este sentido, se busca enriquecer los estudios de validez y confiabilidad de la escala a nivel de los ítems que la componen.

## **MÉTODO**

### *Participantes*

Colaboraron 742 adultos residentes en área metropolitana de Buenos Aires, Argentina, seleccionados a partir de un muestreo no probabilístico por accesibilidad. La edad de los participantes registró un promedio de 38 años (DE = 13.9, Mín = 18, Máx = 81). Del total

un 55.3% pertenecían al género femenino y un 79.5% manifestó tener un empleo regular de medio tiempo o tiempo completo. En cuanto al estado civil, el 40.7% era solteros/as y un 47.7% casado/as. El 69.6% se autopercibió con un nivel socioeconómico medio. Finalmente, el máximo nivel de estudios completados se distribuyó de la siguiente manera: 7.8% inferior a secundario, 52.2% secundario, 12.6% terciario y 27.4% universitario.

### *Instrumentos*

Cuestionario sociodemográfico. Se construyó ad-hoc para esta investigación con el objetivo de recopilar datos tales como género, edad, situación laboral, estado civil, entre otras características.

Escala Neuroticismo del Big Five Inventory (John et al., 1991). El cuestionario completo consta de 44 ítems que permiten evaluar los dominios de la personalidad según el FFM. Para esta investigación sólo fueron utilizados los ocho ítems diseñados para la medición del N. Los enunciados de los ítems tienen una base común (Yo me veo a mí mismo/a como alguien...) que se combina con afirmaciones específicas de cada ítem (e.g. que puede ser tenso/a). Tienen un formato respuesta de cinco opciones (desde 1 - Completo desacuerdo a 5 - Completo acuerdo). Se optó por la adaptación de Castro Solano y Casullo (2001) por considerarla más ajustada a las características lingüísticas locales, comparada con la realizada por Benet-Martínez y John (1998). Desde la perspectiva clásica, el análisis de la consistencia interna de los ítems reveló coeficientes elevados para los datos del presente estudio: Alfa de Cronbach = .82, 95% ICs método bootstrap [.80, .84], Alfa ordinal = .85, 95% ICs método bootstrap [0.83, 0.86], y Omega = .82, 95% ICs método bootstrap [.80, .84].

Inventario de síntomas SCL-90-R (Derogatis, 1994). Inventario de síntomas SCL-90-R (Derogatis, 1994). Es un autoinforme de 90 ítems diseñado para detectar personas en riesgo psicológico por experimentar (en la última semana) sintomatología asociada a nueve dimensiones clínicas (Somatización, Obsesiones/compulsiones, Depresión, Sensitividad Interpersonal, Ansiedad, Hostilidad, Ansiedad Fóbica, Ideación Paranoide y Psicoticismo). El instrumento brinda además tres índices globales (Índice de Severidad Global, Total de

Síntomas Positivos e Índice de Malestar Positivo). Su adaptación argentina presenta evidencias de validez y confiabilidad en población clínica (Sánchez y Ledesma, 2009).

### *Procedimiento*

Los participantes respondieron el protocolo en formato lápiz y papel sin tiempo límite. Se explicó que la evaluación sería de carácter voluntario, anónimo y confidencial y que no obtendrían una devolución de los resultados. Estas consideraciones fueron precisadas por escrito en un consentimiento informado que los participantes debieron firmar. Este estudio cuenta con el aval de la Comisión de Evaluación de Conductas Responsables en Investigación de la Facultad de Psicología de la Universidad de Buenos Aires.

### *Análisis de datos*

Se realizó un Análisis Factorial Confirmatorio (AFC) con Mplus (Muthén y Muthén, 2010) para verificar el supuesto de unidimensionalidad del MRG. Se estimaron los parámetros con el método robusto de mínimos cuadrados ponderados (Weighted Least Squares Mean and Variance Adjusted, WLSMV) sobre la base de la matriz de correlaciones policóricas. Para verificar el ajuste se adoptaron los criterios habitualmente recomendados en la literatura psicométrica (e.g. Byrne, 2012): CFI y TLI  $\geq$  .90 y RMSEA  $\leq$  .08. Se recurrió al estadístico  $\chi^2_{LD}$  (Chen y Thissen, 1997) para corroborar el supuesto de independencia local de los ítems. Siguiendo Reise y Rodríguez (2016), se consideró que los pares de ítems con  $\chi^2_{LD} > 10$  señalan dependencia local entre los elementos.

Con respecto a la calibración del MRG, se estimaron los parámetros de los ítems y de los participantes con el método de Máxima Verosimilitud Marginal usando IRTPRO 4.2 (Cai, Thissen, y du Toit, 2011). Por las características que presenta el MRG, cada uno de los ítems de N del BFI demanda la estimación de un parámetro de pendiente (a) y cuatro parámetros de localización ( $b_1$ ,  $b_2$ ,  $b_3$  y  $b_4$ ). Los parámetros b estiman, en la misma escala del rasgo  $\theta$ , las posiciones de los umbrales que separan las opciones de respuesta en la escala Likert. De esta manera, en los ítems redactados en sentido directo, el parámetro  $b_1$



describe el nivel de rasgo que demanda el contenido del elemento para tener una probabilidad igual o mayor a .50 de escoger la categoría 2 (o una categoría superior) en lugar de optar por la categoría 1. De forma análoga se interpretan el resto de los parámetros b, considerando una segmentación acumulativa de la respuesta politómica. Para examinar el ajuste de los ítems al MRG se estudió analizó el índice S- $\chi^2$ . Se adoptó el criterio de Kang y Chen (2011), en el que un  $p > .05$  se interpreta como indicativo de ajuste aceptable.

Se empleó el test de Wald modificado para detectar el posible Funcionamiento Diferencial de los Ítems (Differential Item Functioning, DIF) en función del género. También se analizó la relación de N con las variables que mide el Inventario SCL-90-R con el coeficiente rho de Spearman. En lo que respecta a la confiabilidad, se obtuvo el coeficiente de fiabilidad marginal (Cai et al., 2011) de la TRI. Adicionalmente, aprovechando las ventajas que proporciona la TRI con medidas locales de precisión, se analizaron las Funciones de Información y de Error Estándar del test.

## **RESULTADOS**

### *Depuración*

El AFC realizado a partir de los ocho ítems que componen la escala mostró un ajuste aceptable al modelo unidimensional requerido como supuesto para la aplicación del MRG (CFI = .95, TLI = .92; RMSEA = .072; IC 90% .059 - .086). En cuanto a la independencia local, la mayoría de los pares de ítems mostraron resultados favorables ( $\chi^2_{LD} < 10$ ) y sólo dos pares de ítems excedieron el límite aceptable para satisfacer este supuesto. El máximo valor de  $\chi^2_{LD}$  fue registrado para el par de ítems 5 y 8 ( $\chi^2_{LD} = 12.2$ ) en tanto que el par de ítems 2 y 7 obtuvieron un  $\chi^2_{LD}$  de 10.2. Estos resultados son esperables en la medida en que se trata de pares de ítems relativamente redundantes (por similitud o por su negación) en el contenido muestreado.

Para alcanzar un criterio de convergencia de .00001 se requirieron de 45 iteraciones del proceso de estimación de parámetros. Con respecto al ajuste de los datos al MRG se registraron valores del estadístico S- $\chi^2$  aceptables para seis ítems ( $p > .05$ ). Sin embargo,

los ítems 6 y 8 no cumplieron con este criterio, obteniendo valores p de .02 y .03 respectivamente.

El análisis de estos resultados sugiere la necesidad de descartar elementos que componen la versión original de la escala por presentar un funcionamiento inadecuado. Se adoptó una postura conservadora y, consecuentemente, se priorizó la eliminación del ítem 8 (...que se pone nervioso/a fácilmente), el cual demostró indicadores inaceptables tanto en el estudio de la independencia local como en el ajuste. Esta decisión arrojó resultados favorables en los análisis subsiguientes, por lo que no resultó necesario descartar otros ítems.

### *Propiedades psicométricas*

La eliminación del ítem 8 no modificó sustantivamente los índices de ajuste al modelo unidimensional registrados en el AFC (CFI = .95, TLI = .93; RMSEA = .069; IC 90% .052-.086). Todas las cargas factoriales (tabla 1) fueron estadísticamente significativas ( $p < .05$ ). La depuración sí trajo consecuencias positivas en el estudio de independencia local de los ítems dado que los valores del estadístico  $\chi^2_{LD}$  oscilaron entre 1.1 y 9.8. Aunque resultan próximos al criterio definido por Reise y Rodríguez (2016), alcanzan para aceptar que los ítems son localmente independientes pese a la similitud observada en el contenido de algunos pares de ítems.

La calibración de los ítems con el MRG resultó exitosa luego de eliminar el ítem 8. Se reportó una disminución en el total de ciclos necesarios durante el proceso de estimación de parámetros para alcanzar el criterio de convergencia de .0001, requiriendo en este caso un total de 38. A su vez, según lo arrojado por los estadísticos  $S-\chi^2$ , todos los ítems demostraron un ajuste satisfactorio al MRG (tabla 1).

Tabla 1. *Pesos factoriales, aplicación del MRG y funcionamiento diferencial*

Ítem	Peso	Parámetros del MRG					Bondad de Ajuste			Test de Wald	
		$a(se)$	$b_1(se)$	$b_2(se)$	$b_3(se)$	$b_4(se)$	$\chi^2$	$gl$	$p$	$\chi^2(gl=5)$	$p$

1	.61	1.30	0.09	0.80	1.79	2.82	64.4	73	.75	7.0	.22
		(0.12)	(0.07)	(0.09)	(0.15)	(0.23)					
2(R)	.66	1.55	-1.34	-0.22	1.06	2.02	46.1	70	.98	3.4	.64
		(0.13)	(0.10)	(0.06)	(0.09)	(0.14)					
3	.59	1.27	-1.59	-0.75	0.39	1.61	66.9	74	.70	3.7	.59
		(0.11)	(0.13)	(0.09)	(0.08)	(0.13)					
4	.48	0.92	-3.14	-1.75	-0.44	0.80	84.7	78	.28	2.1	.84
		(0.10)	(0.32)	(0.18)	(0.10)	(0.12)					
5 (R)	.75	2.07	-1.01	-0.10	0.78	1.43	55.5	63	.73	3.8	.58
		(0.17)	(0.08)	(0.06)	(0.07)	(0.10)					
6	.67	1.58	-1.56	-0.81	0.17	1.13	50.0	68	.95	3.4	.65
		(0.13)	(0.12)	(0.08)	(0.06)	(0.09)					
7 (R)	.65	1.51	-0.96	0.20	1.10	1.97	80.6	73	.25	4.4	.49
		(0.12)	(0.09)	(0.07)	(0.10)	(0.15)					

Los parámetros de discriminación revelaron que la escala presenta valores entre moderados y altos de pendiente según los criterios de Baker (2001) siendo el promedio de  $a = 1.47$  ( $DE = 0.03$ ). El parámetro de pendiente más elevado se encontró en el ítem 5 ( $a = 2.07$ ) y el parámetro más bajo en el ítem 4 ( $a = 0.92$ ). Los parámetros de umbral se localizaron en niveles del rasgo que van desde  $-3.14$  ( $b_1$  ítem 4) hasta  $2.82$  ( $b_4$  ítem 1). Los  $\theta$  estimados para los 742 participantes presentaron valores entre  $-2.29$  y  $2.59$  con una media de  $-0.0001$  ( $DE = 0.9$ ).

Al observar la relación de los parámetros de los ítems con sus respectivos contenidos se puede apreciar que todos los  $b$  del ítem 1 (... que es depresivo/a o triste) resultaron positivos. Esto señala que se trata de un ítem cuya capacidad discriminativa resulta eficiente para los niveles más elevados de  $N$ . En cambio, el contenido del ítem 4 (... que se preocupa mucho por todo) discrimina mejor en los niveles medios y bajos. Por último, es posible apreciar que los ítems redactados en sentido inverso tendieron a registrar mayores parámetros de pendiente.

El estudio del DIF en función del género reveló que no se registran diferencias estadísticamente significativas en los parámetros de los ítems estimados para los grupos de mujeres y varones. Como se observa en la tabla 1, todos los p-valores asociados al estadístico  $\chi^2$  del test de Wald fueron superiores a .05, permitiendo descartar la presencia de DIF.

#### *Relación de N con dimensiones sintomatológicas*

Todas las dimensiones clínicas del SCL-90-R mostraron asociaciones significativas y moderadas con N (tabla 2). La intensidad de los coeficientes rho de Spearman aumenta si se consideran los índices globales de severidad sintomática que brinda la prueba.

Tabla 2. *Asociación entre Neuroticismo y escalas del SCL-90-R*

	Som	Obs	Dep	SI	Ans	Hos	AF	IP	Psi	ISG	TSP	IMP
N	.42**	.43**	.49**	.45**	.54**	.38**	.38**	.36**	.47**	.55**	.57**	.58**

*Nota.* N = Neuroticismo; Som = Somatización; Obs = Obsesiones/Compulsiones; Dep = Depresión; SI = Sensitividad Interpersonal; Ans = Ansiedad; Hos = Hostilidad; AF = Ansiedad Fóbica; IP = Ideación Paranoide; Psi = Psicoticismo; ISG = Índice de Severidad Global; TSP = Total de Síntomas Positivos; IMP = Índice de Malestar Positivo.

\*\*  $p < .001$

#### *Análisis de la confiabilidad*

El instrumento depurado reveló un coeficiente de fiabilidad marginal de .81 demostrando que, globalmente, obtiene mediciones precisas según este indicador de la TRI. La FIT muestra una relativa simetría centrada en los niveles medio-altos del rasgo (figura 1). En efecto, alcanza sus niveles máximos de información aproximadamente a menos de un desvío por encima de la media. En este mismo rango de  $\theta$ , el error de estimación registró un valor mínimo de 0.42, aumentando hacia los niveles extremos del rasgo. Para  $\theta$  iguales a -2.4 el error de estimación mínimo resultó de 0.60 (equivalente a una confiabilidad clásica

de .64). En el otro polo del rasgo, incluso con un  $\theta$  de 2.8 el error fue de 0.57, lo que implica una confiabilidad clásica aceptable de .68.

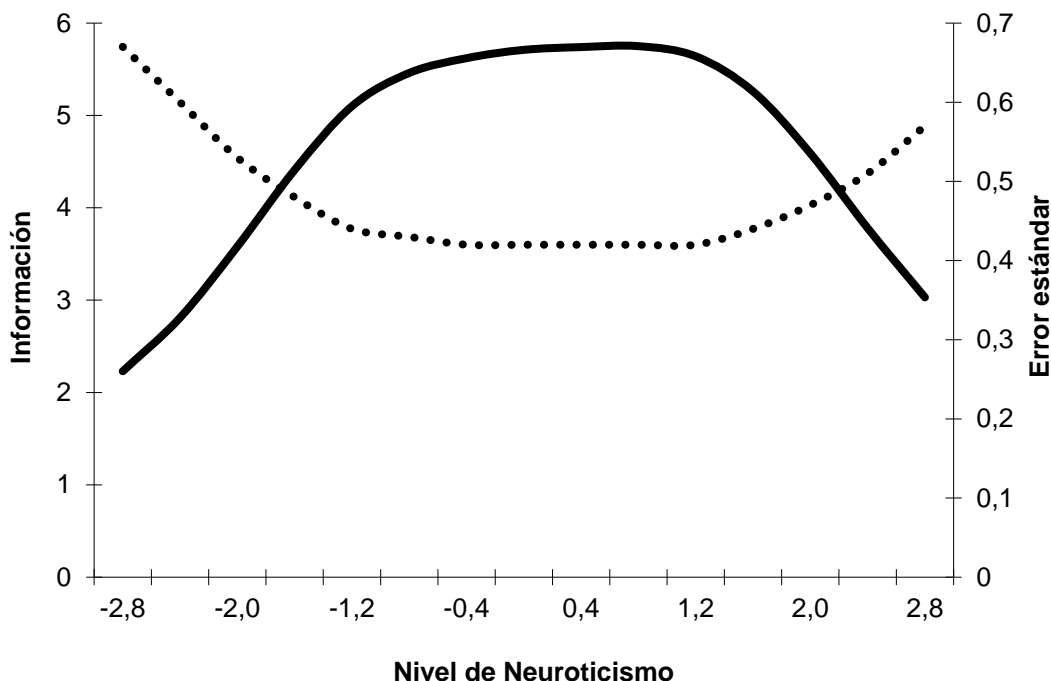


Figura 1. Funciones de Información y Error estándar del test

## DISCUSIÓN

La aplicación del MRG a la escala N brindó una comprensión más acabada de la relación entre el constructo y los contenidos de cada ítem, identificando la capacidad de las opciones de respuesta para discriminar en distintos niveles de la variable. Sólo debió descartarse uno de los ítems de la escala original mientras que el resto mostró un ajuste adecuado al MRG. La eliminación de este ítem no repercutió en la validez de contenido dado que el estudio de independencia local dejó en evidencia que podría ser considerado redundante (aunque su redacción está en sentido inverso) con respecto al ítem 5 (... que es emocionalmente estable y no se altera fácilmente). El uso de ítems redundantes es una práctica frecuente en el marco de la TCT porque eleva la consistencia interna, pero implica

una violación del supuesto de independencia local de los ítems al modelizar el instrumento con TRI (Abal et al, 2010; Reise y Rodríguez, 2016).

Se han evidenciado dos resultados para destacar con respecto a la calidad psicométrica de la escala al ser estudiada con TRI y que complementan las conclusiones derivadas de la calibración de los ítems. El primero remite a que los siete ítems no mostraron DIF en función del género, lo que significa que el instrumento brinda mediciones invariantes para mujeres y varones. El segundo resultado relevante se desprende del análisis de la FIT. La fiabilidad clásica se determinada mediante un coeficiente global y que se presume constante todo el espectro del rasgo. En cambio, la TRI ofrece la ventaja de analizar la confiabilidad de la prueba para los diferentes niveles (Muñiz, 2018). La elevación observada en la FIT en torno a los niveles medios del rasgo es esperable en virtud de que se trata de una escala construida para población general. No obstante, se ha podido verificar que, aun para valores extremos, la precisión de la medida resulta aceptable. Esto es particularmente útil si se recuerda que los puntajes altos son de interés para la detección de personas con niveles riesgosos de N.

Al hilo de la potencial aplicación de N en el contexto clínico-epidemiológico, las correlaciones con las nueve dimensiones del SCL-90-R están en línea con otros estudios en los que se hallaron asociaciones de este dominio con sintomatología internalizante y, en menor medida, también con sintomatología externalizante (Jeronimus et al, 2016; Kotov, Gámez, Schmidt, y Watson, 2010; Ormel, et al., 2013). La variación en la intensidad de estas asociaciones en función del tipo de sintomatología se explica en parte por el muestreo de contenido de los ítems de N. Aspectos vinculados a la labilidad emocional, hostilidad e impulsividad tienen escasa o nula representación en los ítems del BFI, condición que también se observa en otras operacionalizaciones breves de este constructo (Goldberg et al., 2006; McCrae y Costa, 2007). En este sentido, parece más relevante destacar las asociaciones con los índices globales de SCL-90-R dado que presentaron correlaciones de mayor intensidad con N.

En cuanto a las limitaciones del presente estudio, se reconoce la necesidad de avanzar sobre la determinación de un punto de corte para facilitar la identificación de personas con niveles clínicamente atendibles de N. Cabría la posibilidad de definir provisoriamente un punto de corte basado en percentiles, pero resulta más pertinente validar un puntaje que

maximice la sensibilidad de la escala para disminuir la proporción de falsos negativos. Los próximos estudios estarán encaminados hacia este objetivo. En el futuro, la información derivada de la medición de este dominio podría formar parte de protocolos para que los profesionales de distintas áreas de la salud sepan de qué manera orientar a los pacientes para la búsqueda de un tratamiento precoz.

## REFERENCIAS

- Abal, F.J.P., Lozzia, G.S, Aguerri, M.E., Galibert, M.S. y Attorresi, H.F. (2010). La escasa aplicación de la Teoría de Respuesta al Ítem en Tests de Ejecución Típica. *Revista Colombiana de Psicología*, 19 (1) 111-122.
- Alansari, B. (2016). The Big Five Inventory (BFI): Reliability and validity of its Arabic translation in non-clinical sample. *European Psychiatry*, 33, 209-210. doi: 10.1016/j.eurpsy.2016.01.500
- Baker, F. B. (2001). *The basics of item response theory*. Washington: ERIC Publications.
- Benet-Martínez, V. y John, O. P. (1998). Los Cinco Grandes across cultures and ethnic groups: Multitrait multimethod analyses of the Big Five in Spanish and English. *Journal of Personality and Social Psychology*, 75 (3), 729-750. doi: 10.1037/0022-3514.75.3.729
- Bleidorn, W., Klimstra, T. A., Denissen, J. J. A., Rentfrow, P. J., Potter, J., & Gosling, S. (2013). Personality maturation around the world: A cross-cultural examination of social-investment theory. *Psychological Science*, 24 (12), 2530 - 2540. doi: 10.1177/0956797613498396.
- Byrne, B.M. (2012). *Structural equation modeling with Mplus: Basics, concepts, applications, and programming*. Nueva York: Routledge.
- Cai, L., Thissen, D., y du Toit, S. (2011). *IRTPRO users guide*. Lincolnwood, IL: Scientific Software International.
- Carciofo, R., Yang, J., Song, N., Du, F. y Zhang, K. (2016). Psychometric evaluation of Chinese-language 44-item and 10-item Big Five Personality Inventories, including correlations with chronotype, mindfulness and mind wandering. *PLOS ONE*, 11 (2), doi: 10.1371/journal.pone.0149963

- Castro Solano, A. y Casullo, M. M. (2001). Rasgos de personalidad, bienestar psicológico y rendimiento académico en adolescentes argentinos. *Interdisciplinaria*, 18 (1), 65-85.
- Castro Solano, A. (2002). *Técnicas de evaluación psicológica en los ámbitos militares*. Buenos Aires: Paidós
- Chapman, B.P. y Goldberg, L.R. (2017). Act-Frequency Signatures of the Big Five. *Personality and Individual Differences*, 116, 201-205. doi: 10.1016/j.paid.2017.04.049.
- Chen, W., y Thissen, D. (1997). Local dependence indices for item pairs using item response theory. *Journal of Educational and Behavioral Statistics*, 22, 265-289. doi: 10.2307/1165285
- Denissen, J. J. A., Geenen, R., van Aken, M. A. G., Gosling, S. D. y Potter, J. (2008). Development and validation of a Dutch translation of the Big Five Inventory (BFI). *Journal of Personality Assessment*, 90, 152-157. doi: 10.1037/t62652-000
- Derogatis, L. (1994). *SCL-90-R. Symptom Checklist-90-R. Administration, Scoring and Procedures Manual*. Minneapolis: National Computer System.
- Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology*, 41, 417 - 440. doi: 10.1146/annurev.ps.41.020190.002221
- Drake, M. M., Morris, D. y Davis, T. J. (2017). *Neuroticism's susceptibility to distress: Moderated with mindfulness*. *Personality and Individual Differences*, 106, 248-252. doi: 10.1016/j.paid.2016.10.060
- Domínguez-Lara, S., Merino-Soto, C., Zamudio, B., y Guevara-Cordero, C. (2018). Big Five Inventory en Universitarios Peruanos: Resultados Preliminares de su Validación. *Psyche*, 27 (2), 1 - 12. doi: 10.7764/psyche.27.2.1052
- Donnellan, M. B., Oswald, F. L., Baird, B. M., y Lucas, R. E. (2006). The Mini-IPIP scales: Tiny-yet-effective measures of the big five factors of personality. *Psychological Assessment*, 18 (2), 192-203. doi: 10.1037/1040-3590.18.2.192
- Eysenck, H. J. (1947). *Dimensions of Personality*. Londres: Routledge y Kegan Paul.
- Friedman, H. S. (2019). Neuroticism and health as individuals age. *Personality Disorders: Theory, Research, and Treatment*, 10 (1), 25 - 32. doi: 10.1037/per0000274
- Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., y Gough, H.C. (2006). The Personality Item Pool and the future of public-domain



- personality measures. *Journal of Research in Personality*, 40, 84–96.  
doi:10.1016/j.jrp.2005.08.007.
- Goldberg, L.R. (1993). The structure of phenotypic personality traits. *American Psychologist*, 48 (1), 26-34. doi: 10.1037/0003-066X.48.1.26
- Gosling, S. D., Rentfrow, P. J., y Swann, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, 37, 504–528. doi: 10.1016/S0092-6566(03)00046-1
- Gurven, M., von Rueden, C., Massenkoff, M., Kaplan, H., y Lero Vie, M. (2013). How universal is the Big Five? Testing the five-factor model of personality variation among forager–farmers in the Bolivian Amazon. *Journal of Personality and Social Psychology*, 104 (2), 354–370. doi:10.1037/a0030841
- Hajek, A; Bock, J.O. y König, H.H. (2017). The role of personality in health care use: Results of a population-based longitudinal study in Germany. *PLoS One*, 12 (7):e0181716. doi: 10.1371/journal.pone.0181716.
- Hengartner, M. P., Kawohl, W., Haker, H., Rössler, W., & Ajdacic-Gross, V. (2016). Big Five personality traits may inform public health policy and preventive medicine: Evidence from a cross-sectional and a prospective longitudinal epidemiologic study in a Swiss community. *Journal of Psychosomatic Research*, 84, 44 - 51. doi: 10.1016/j.jpsychores.2016.03.012.
- Jeronimus, B.F., Kotov, R., Riese, H. y Ormel, J. (2016). Neuroticism’s prospective association with mental disorders halves after adjustment for baseline symptoms and psychiatric history, but the adjusted association hardly decays with time: a meta-analysis on 59 longitudinal/prospective studies with 443 313 participants. *Psychological Medicine*, 46 (14), 2883-2906. doi:10.1017/S0033291716001653.
- John, O. P., Naumann, L. P., y Soto, C. J. (2008). Paradigm shift to the integrative Big Five trait taxonomy: History, measurement, and conceptual issues. En O. P. John, R. W. Robins, y L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (3rd ed., pp. 114–158). Nueva York: Guilford.
- John, O. P., Donahue, E. M. y Kentle, R. L. (1991). *The Big Five Inventory — versions 4a and 54*. Berkeley: Institute of Personality and Social Research.

- Kang, T., y Chen, T. T. (2011). Performance of the generalized S-X2 item fit index for the graded response model. *Asia Pacific Education Review*, 12 (1), 89 – 96. doi: 10.1007/s12564-010-9082-4.
- Karaman, N. G., Dogan, T., y Coban, A. E. (2010). A study to adapt the big five inventory to Turkish. *Procedia - Social and Behavioral Sciences*, 2 (2), 2357-2359. doi: 10.1016/j.sbspro.2010.03.336.
- Kotov, R., Gamez, W., Schmidt, F., y Watson, D. (2010). Linking ‘‘big’’ personality traits to anxiety, depressive, and substance use disorders: A meta-analysis. *Psychological Bulletin*, 136, 768–821. doi: 10.1037/a0020327.
- Lahey, B. B. (2009). Public health significance of neuroticism. *American Psychologist*, 64, 241–256. doi: 10.1037/a0015309
- Maples-Keller, J. L., Bunnell, B. E., Kim, S. J., y Rothbaum, B. O. (2017). The use of Virtual Reality technology in the treatment of anxiety and other psychiatric disorders. *Harv. Rev. Psychiatry*, 25, 103–113. doi: 10.1097/HRP.0000000000000138.
- McCrae, R. y Costa, P. (2007). Brief versions of the NEO-PI-3. *Journal of Individual Differences*, 28 (3), 116-128. doi: 10.1027/1614-0001.28.3.116
- McCrae, R. y Costa P. (2010). *NEO Inventories professional manual*. Odessa, FL: Psychological Assessment Resources.
- Muñiz, J. (2018). *Introducción a la Psicometría: teoría clásica y TRI*. Madrid: Pirámide.
- Muthén, L. y Muthén, B. (2010). *Mplus User's Guide*, 6th Edn. Los Angeles: Muthén & Muthén.
- Ormel, J., Bastiaansen, A., Riese, H., Bos, E. H., Servaas, M., Ellenbogen, M., Rosmalen, J. G., Aleman, A. (2013). The biological and psychological basis of neuroticism: Current status and future directions. *Neuroscience and Biobehavioral Reviews*, 37, 59–72. doi: 10.1016/j.neubiorev.2012.09.004
- Plaisant, O., Courtois, R. Réveillère, C., Mendelsohn, G.A. y John, O. P. (2010). Validation par analyse factorielle du Big Five Inventory français (BFI-Fr). Analyse convergente avec le NEO-PI-R. *Annales Medico-Psychologiques*, 168, 97–106. doi: 10.1016/j.amp.2009.09.003.
- Raad, B. y Mlačić, B. (2015). Big Five Factor Model, Theory and Structure. En J. D. Wright (Ed), *International Encyclopedia of the Social y Behavioral Sciences* (pp.

- 559–566), 2nd edition, Vol 2. Oxford: Elsevier. doi:10.1016/b978-0-08-097086-8.25066-6
- Rammstedt, B. y Beierlein, C. (2014). Can't we make it any shorter? *Journal of Individual Differences*, 35 (4), 212–220. doi: 10.1027/1614-0001/a000141
- Rammstedt, B. y John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, 41, 203-212. doi: 10.1016/j.jrp.2006.02.001.
- Rammstedt, B. (1997). *Die deutsche Version des Big Five Inventory (BFI): Übersetzung und Validierung eines Fragebogens zur Erfassung des Fünf-Faktoren-Modells der Persönlichkeit*. (Tesis doctoral inédita). Bielefeld: Universität Bielefeld.
- Reise, S. P. y Rodriguez, A. (2016). Item response theory and the measurement of psychiatric constructs: some empirical and conceptual issues and challenges. *Psychol Med*, 46 (10), 2025-2039. doi: 10.1017/ S0033291716000520.
- Reise, S. P. y Revicki, D. A. (2015). *Handbook of Item Response Theory Modeling Applications to Typical Performance Assessment*. Nueva York: Routledge.
- Reyes Zamorano, E., Álvarez Carrillo, C., Peredo Silva, A., Miranda Sandoval, A. y Rebolledo Pastrana, I. M. (2014). Psychometric properties of the Big Five Inventory in a Mexican sample. *Salud Mental*, 37, 491-497. doi: 10.17711/sm.0185-3325.2014.059
- Salgado, E., Vargas-Trujillo, E., Schmutzler, J., y Wills-Herrera, E. (2016). Uso del Inventario de los Cinco Grandes en una muestra colombiana. *Avances en Psicología Latinoamericana*, 34(2), 365-382.
- Samejima, F. (2010). Graded Response Model. En W. J. van der Linden (Ed.). *Handbook of Item Response Theory*, Volume 1: Models (pp. 95-108). Boca Raton: Chapman y Hall/CRC.
- Sánchez, R., & Ledesma, R. (2009) Análisis Psicométrico del Inventario de Síntomas Revisado (SCL-90-R) en Población Clínica. *Revista Argentina en Clínica Psicológica*, XVIII (3), 265-274.
- Sauer-Zavala, S., Wilner, J. y Barlow, D. H. (2017). Addressing neuroticism in psychological treatment. *Personality Disorders: Theory, Research, and Treatment*, 8 (3), 191-198. doi: 10.1037/per0000224.

- Schmitt, D. P., Allik, J., McCrae, R. R., Benet-Martínez, V., et al. (2007). The geographic distribution of Big Five personality traits: Patterns and profiles of human self-description across 56 nations. *Journal of Cross-Cultural Psychology*, 38, 173-212. doi: 10.1177/0022022106297299
- Tackett, J. L. y Lahey, B. B. (2017). Neuroticism. En T. A. Widiger (Ed). *The Oxford handbook of the five factor model* (pp. 39-56). Nueva York: Oxford University Press.
- Taylor, N., y De Bruin, G.P. (2006). BTI. *Manual of the Basic Traits Inventory*. Johannesburgo, Sudáfrica: JvR.
- Ubbiali, A., Chiorri, C., Hampton, P. y Donati, D. (2013). Psychometric properties of the Italian adaptation of the Big Five Inventory (BFI). *Bollettino di Psicologia Applicata*, 266, 37 - 46.
- Vittengl, J. R. (2017). Who pays the price for high neuroticism? Moderators of longitudinal risks for depression and anxiety. *Psychological Medicine*, 1-12. doi:10.1017/S0033291717000253.
- Widiger, T. A. y Oltmanns, J. R. (2017). Neuroticism is a fundamental domain of personality with enormous public health implications. *World Psychiatry*, 16 (2), 144–145. doi: 10.1002/wps.20411.
- Widiger, T. A. (2009). Neuroticism. En M. R. Leary y R. H. Hoyle (Ed), *Handbook of individual differences in social behavior* (pp. 129-146). Nueva York: Guilford Press.
- Zanon, C. y Hutz, C. S. (2009). Propriedades psicométricas da Escala Fatorial de Neuroticismo e do Questionário de Ruminação e Reflexão. *Aval. Psicol*, 8 (2), 279-281.
- Ziegler, M., Kemper, C. J., y Kruey, P. (2014). Short scales – Five misunderstandings and ways to overcome them. *Journal of Individual Differences*, 35 (4), 185-189. doi: 10.1027/1614-0001/a000148.

Recibido: 6 de junio del 2019

Aceptado: 26 de junio del 2019

SIN CONFLICTO DE INTERESES