

# Evaluación adaptativa del Neuroticismo mediante el Cuestionario Revisado de Personalidad de Eysenck

## Adaptive Evaluation of Neuroticism through the Revised Eysenck Personality Questionnaire

Facundo Juan Pablo Abal<sup>1, 2\*</sup>, Gabriela Susana Lozzia<sup>2</sup>,  
Sofía Esmeralda Auné<sup>1, 2</sup>, Horacio Félix Attorresi<sup>2</sup>.

1 - Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Buenos Aires, Argentina.

2 - Instituto de Investigaciones de la Facultad de Psicología, Universidad de Buenos Aires, Argentina.

Recibido: 13/08/2018 Revisado: 14/01/2019 Aceptado: 28/01/2019

Introducción  
Objetivos  
Método  
Resultados  
Discusión  
Referencias

### Resumen

Se estudió la conveniencia de administrar adaptativamente los ítems de Neuroticismo pertenecientes a la adaptación argentina del Cuestionario Revisado de Personalidad de Eysenck. Participaron 594 personas de población general. Un 70% de los casos sirvió para calibrar los ítems según el modelo logístico de dos parámetros. El resto se usó para analizar evidencias de validez y confiabilidad del test adaptativo informatizado usando un criterio de parada mixto (error de estimación  $\leq .4$  o administración de 9 ítems). El 48.9% de los participantes alcanzó el error  $\leq .4$  luego de presentarles 6 ítems y el 69.1% requirió de 9 ítems o menos. La correlación entre el nivel estimado con el test completo y la versión adaptativa fue de .98. Tampoco varió sustancialmente la asociación con criterios externos. La discusión aborda las ventajas y limitaciones vinculadas a la falta de ítems que discriminen en los niveles bajos del rasgo.

**Palabras clave:** *Neuroticismo, Eysenck, test adaptativo informatizado, modelo logístico de dos parámetros, Teoría de Respuesta al Ítem*

### Abstract

This work analyses the convenience of adaptively administering the neuroticism items belonging to the Argentine adaptation of the Revised Eysenck Personality Questionnaire. The study included 594 people from the general population. Seventy percent of the cases were used to calibrate items according to the two-parameter logistic model. The rest of the cases were used to analyse evidence of validity and reliability of the computerized adaptive test using a mix stopping criterion (estimation error  $\leq .4$  or administration of 9 items). Results showed that 48.9% of the participants reached an  $\leq .4$  error after the administration of 6 items and that 69.1% required 9 items or less. The correlation between the level estimated with the full test and the adaptive version was of .98. The association with external criteria did not vary substantially either. The discussion addresses the advantages and limitations that stem from the lack of items which are discriminative at low-trait levels.

**Key words:** *Neuroticism, Eysenck, computerized adaptive test, two-parameter logistic model, Item Response Theory*

\* **Correspondencia a:** Facundo Juan Pablo Abal; Dirección postal: Zuviría 5691 (1439), Ciudad Autónoma de Buenos Aires. Teléfono: 54 11 46014331. E-mail: [fabal@psi.uba.ar](mailto:fabal@psi.uba.ar)

**Cómo citar este artículo:** Abal, F. J. P., Lozzia, G. S., Auné, S. E., & Attorresi, H. F. (2019). Evaluación adaptativa del Neuroticismo mediante el Cuestionario Revisado de Personalidad de Eysenck. *Revista Evaluar*, 19(1), 17-34. Recuperado de <https://revistas.unc.edu.ar/index.php/revaluar>

**Financiación:** Esta investigación fue financiada con subsidios de la Universidad de Buenos Aires UBACyT 2018 Código 20020170100200BA y UBACyT 2018 Código 20020170200001BA y de la Agencia Nacional de Promoción Científica y Tecnológica PICT-2017-3226.

## Introducción

El estudio conceptual e instrumental del *neuroticismo* (*N*) tiene una extensa tradición en el campo de la evaluación de la personalidad. Sin embargo, a pesar de que se registran algunas referencias más remotas (Thurstone & Thurstone, 1930; Woodworth, 1919), la intención de una medición psicométricamente válida del *N*, basada en una teoría sólida y con sustento empírico llegaría recién a partir de la década de 1940 con los desarrollos de Guilford (1940), Cattell (1957) y principalmente por la contribución de Eysenck (1947). Esto inauguró una concepción moderna del *N*, entendiéndolo como un continuo cuantitativo sobre el que es posible reconocer diferentes niveles de intensidad (Dumont, 2010; Jeronimus, Kotov, Riese, & Ormel, 2016).

El *N* fue la primera de las dimensiones identificadas por Eysenck en la construcción de su modelo de la personalidad, el cual incluye también las dimensiones *extraversión* y *psicoticismo*. La caracterización del *N* presenta en sus orígenes más lejanos la descripción de los cuatro temperamentos clásicos de la tradición hipocrática-galénica (Eysenck, 1947, 1952; Zuckerman, 2005). Siguiendo a Wundt, Eysenck vinculó el *N* con los temperamentos melancólico y colérico en tanto que para ambos se describe una tendencia a experimentar una reacción emocional de fuerte intensidad (como se citó en Bech, 2016). Los sujetos con niveles elevados de *N* son propensos a reaccionar de manera exagerada frente a todo tipo de estímulos y muestran dificultades para retornar al nivel emocional normal luego de estas experiencias. En el polo opuesto, la persona con bajo *N* es tranquila, controlada y mantiene un temperamento equilibrado en la mayoría de las circunstancias. Para Eysenck, el *N* se ubica en el nivel más alto de una estructura jerárquica y se conforma a partir de las intercorrelaciones de un

conjunto de rasgos primarios: depresión, miedo, ansiedad, preocupación, irracionalidad, tensión, timidez, vergüenza, sentimiento de culpa, baja autoestima, mal humor y emotividad (Eysenck & Eysenck, 1975). La combinación de estos aspectos en mayor o menor medida configura el nivel de estabilidad-inestabilidad emocional de la persona.

A lo largo de su obra Eysenck fue perfeccionando la medición del *N* y las otras variables de su modelo desarrollando una sucesión de cuestionarios basados en el análisis factorial (Furnham, Eysenck, & Saklofske, 2008). El Cuestionario de Personalidad Eysenck EPQ, (Eysenck & Eysenck 1975), su posterior versión revisada (EPQ-R) y su versión corta (EPQ-RS; Eysenck, Eysenck, & Barrett, 1985) podrían considerarse como el resultado final de un arduo trabajo en la construcción de una medida que operacionaliza las dimensiones básicas de personalidad que propuso y una escala de deseabilidad social (Bech, 2016). Mientras que el EPQ-R consta de 100 ítems, la versión corta emplea una selección de 48 elementos. Esta considerable reducción motivó a numerosos investigadores a efectuar adaptaciones locales de la versión corta (Aluja, García, & García, 2003; Francis, Lewis, & Ziebertz, 2006; Lewis, Francis, Shevlin, & Forrest, 2002; Squillace, Picón-Janeiro, & Schmidt, 2013; Tiwari, Singh, & Singh, 2009).

Aunque los cuestionarios de Eysenck no fueron diseñados para diagnosticar trastornos psiquiátricos, han resultado de gran utilidad en el ámbito clínico. Desde sus primeros estudios, Eysenck (1947) relacionó los niveles elevados de *N* con una predisposición a padecer lo que clásicamente se denominó trastornos neuróticos. En la actualidad el *N* es considerado como una dimensión de la personalidad normal prioritaria en virtud de su impacto en la salud pública (Vittengl,

2017; Widiger & Oltmanns, 2017). Su asociación con distintas psicopatologías ha llevado a concebirlo como un factor de vulnerabilidad común para el desarrollo y mantenimiento de diversos trastornos mentales (Jeronimus et al., 2016; Sauer-Zavala, Wilner, & Barlow, 2017). El *N* también es relevante por su vinculación con una gran variedad de enfermedades físicas, como cardiopatías, diabetes, asma o enfermedades inflamatorias intestinales (Lahey, 2009). Por esta demostrada contribución del *N* a tantas consecuencias negativas sobre la salud, en los últimos años se ha empezado a recomendar la detección de niveles altos de este rasgo en población general durante la asistencia clínica de rutina (Hengartner, Kawohl, Haker, Rössler, & Ajdacic-Gross, 2016; Tackett & Lahey, 2017; Widiger, 2009; Widiger & Oltmanns, 2017).

La demanda de una evaluación a gran escala y/o con fines de rastreo conlleva la necesidad práctica de seleccionar instrumentos cortos que permitan la detección rápida de personas que presentan un riesgo potencial (Baldasaro, Shanahan, & Bauer, 2013). Sin embargo, optimizar la medición de los rasgos no es una tarea sencilla. Son numerosos los estudios instrumentales que en los últimos años han perseguido la construcción de una medida breve para *N* así como también para otros dominios de la personalidad (Donnellan, Oswald, Baird, & Lucas, 2006; Gosling, Rentfrow, & Swann, 2003; Soto & John, 2016, 2017). Pero las estrategias para reducir la cantidad de ítems en el marco de la Teoría Clásica de Test (TCT) tienen un costo en la calidad psicométrica que no siempre es considerado (Credé, Harms, Niehorster, & Gaye-Valentine, 2012). Frente a la validación de las escalas breves, los investigadores se enfrentan a encontrar un equilibrio razonable entre resguardar validez de contenido (Milojev, Osborne, Greaves, Barlow, & Sibley, 2013; Ziegler, Kemper, & Krueger, 2014) o precisión en la medida (Baldasaro

et al., 2013; Sibley, 2012).

Más recientemente, y al amparo del crecimiento de la Teoría de Respuesta al Ítem, se han desarrollado Test Adaptativos Informatizados (TAIs) para la medición de rasgos de la personalidad (Reise & Revicki, 2015). La implementación de una estrategia adaptativa para la medición del *N* puede resultar particularmente eficiente en la detección de niveles elevados del rasgo dado que podría ofrecer un marco de referencia psicométrico que permita acortar el cuestionario sin comprometer la precisión o validez (Chang, 2015).

## Objetivos

La motivación para construir una escala breve que mida *N* no es novedosa. En efecto, Eysenck (1958) elaboró tempranamente una primera escala con apenas seis ítems. Posteriormente, algunos autores se propusieron acortar el EPQ-RS desde la TCT y ensayaron versiones abreviadas (EPQ-RA) que también utilizan seis ítems (Francis, Brown, & Philipchalk, 1992; Sandín, Valiente, Chorot, Olmedo, & Santed, 2002). Pero la reducción en la cantidad de reactivos tuvo como consecuencia una irremediable disminución en la confiabilidad de la escala (Lewis et al., 2002).

La medición de *N* mediante EPQ-RS acumula sólidas propiedades psicométricas tanto a nivel local como internacional. Considerando las ventajas que brindaría la evaluación adaptativa de este constructo en un ámbito clínico-epidemiológico y a la luz de los avances metodológicos mencionados, se plantean los siguientes objetivos: a) evaluar si los 12 ítems de *N* de la adaptación argentina de la EPQ-RS cumplen con los requisitos psicométricos necesarios para ser administrados adaptativamente y b) examinar si la versión adaptativa de esta escala podría mostrar ventajas por sobre la administración

tradicional.

## Método

### *Participantes*

Se contó con la colaboración de 594 personas residentes en el área metropolitana de Buenos Aires, Argentina. El 52% pertenecían al género femenino y la edad osciló entre 18 y 82 años alcanzando un valor promedio de 32.9 años (DE = 12.1). El 79% consignó estar empleado y el 31% se encontraba cursando una carrera terciaria o universitaria al momento de ser evaluado. El 54.1% reportó tener un nivel de estudios secundarios completo y sólo un 6.2% no alcanzó este nivel. Un 18.5% presentó estudios terciarios (completos o incompletos) y un 21.2% registró un nivel universitario (completo o incompleto).

### *Instrumentos*

***Escala Neuroticismo del Eysenck Personality Questionnaire revised short version, EPQ-RS (Eysenck et al., 1985; adaptación de Squillace et al., 2013).*** El cuestionario original consta de 48 ítems, pero en la adaptación argentina quedó reducido a 42. De estos, sólo 12 son utilizados para medir el *N*. Los ítems están redactados en forma de pregunta y el formato de respuesta es dicotómico (*Sí-No*). Los adaptadores del instrumento replicaron la estructura factorial propuesta por Eysenck y registraron un índice de consistencia interna satisfactorio (KR-20 = .84) para la escala *N*. En el presente estudio se halló un índice, ligeramente más alto, de .85 para la muestra de calibración.

***Inventario de síntomas SCL-90-R (Derogatis, 1994).*** Es un autoinforme compuesto por 90 ítems que evalúan patrones de síntomas psicológicos con el fin de detectar personas en

riesgo. Los elementos presentan un formato de respuesta politómica en la que el sujeto responde si el síntoma le ha preocupado o molestado *nada, muy poco, poco, bastante* o *mucho* durante los últimos siete días. Los ítems se agrupan para posibilitar la medición de nueve dimensiones clínicas (*somatización, obsesiones/compulsiones, sensibilidad interpersonal, depresión, ansiedad, hostilidad, ansiedad fóbica, ideación paranoide y psicoticismo*) y tres índices globales (*índice de severidad global, total de síntomas positivos e índice de malestar positivo*). Los estudios de la adaptación argentina de este inventario muestran suficientes evidencias de validez y confiabilidad tanto en población clínica (Sánchez & Ledesma, 2009) y no clínica (Casullo, 2004). En cuanto al presente estudio, el análisis de la consistencia interna de todos los ítems del inventario mostró un coeficiente alfa de Cronbach de .96 mientras que para las dimensiones clínicas este coeficiente osciló entre .77 y .86.

### *Procedimiento*

Los individuos respondieron el protocolo en formato lápiz y papel sin tiempo límite. Antes de comenzar la administración se les explicó que la prueba tenía como objetivo la medición de aspectos de su personalidad y que no había respuestas correctas o incorrectas. Se les comunicó sobre las garantías de anonimato y confidencialidad de sus respuestas. También se les informó sobre el carácter voluntario de su participación y la posibilidad de abandonar la evaluación en cualquier momento de la actividad. Todas estas consideraciones fueron precisadas por escrito y formaron parte de la redacción del consentimiento que debieron firmar los sujetos para participar. El protocolo administrado incluyó otras escalas no consideradas para los fines de esta investigación.

### *Análisis de datos*

Se dividió a los participantes en dos grupos de manera aleatoria. Las respuestas de un 70% de los sujetos ( $n = 416$ ) fueron empleadas para calibrar los ítems según el modelo logístico de dos parámetros (ML2p) mientras que el resto de los individuos ( $n = 178$ ) fueron considerados exclusivamente para analizar el TAI.

### *Calibración de los ítems*

**Supuestos del ML2p.** Se realizó un análisis factorial confirmatorio (AFC) con el programa Mplus (Muthén & Muthén, 2010) a fin de verificar el supuesto de unidimensionalidad. Se estimaron los parámetros con el método robusto de mínimos cuadrados ponderados sobre la base de la matriz de correlaciones tetracóricas. Para analizar el ajuste se aplicaron los criterios recomendados por Byrne (2012) para los índices de ajuste comparativo ( $CFI \geq .90$ ) y de Tucker-Lewis ( $TLI \geq .90$ ) y el error medio cuadrático de aproximación ( $RMSEA \leq .08$ ).

Si bien es habitual dar por satisfecho el supuesto de independencia local de los ítems a partir de verificar la unidimensionalidad del constructo, los tests de personalidad construidos originalmente desde la perspectiva de la teoría clásica suelen incluir ítems redundantes que ocasionan problemas de dependencia local (Reise & Revicki, 2015). Es por esta razón que, habiendo procedimientos específicos para su comprobación, se decidió estudiar el supuesto de independencia local. Su análisis se implementó mediante los estadísticos  $\chi^2$  que suministra el programa MODFIT (Stark, 2001) para contrastar la independencia a partir de pares y tríos de ítems. Para hacer manejable el número de combinatorias de pares

$\binom{n}{2}\binom{n}{2}$  y tríos  $\binom{n}{3}\binom{n}{3}$  de ítems, el programa analizó la dependencia dividiendo a los 12

ítems de la prueba en cuatro conjuntos de tres ítems (Drasgow, Levine, Tsien, Williams, & Mead, 1995). Posteriormente, se calcularon los estadísticos  $\chi^2$  para cada ítem (simple), para todas las combinaciones de pares posibles al interior de cada conjunto de ítems (12 pares) y para los cuatros tríos. Los valores del ratio  $\chi^2/gl$  superiores a 3 reflejan problemas de desajuste del modelo a los datos ocasionados por la dependencia local entre los ítems (Abad, Olea, Ponsoda, & García, 2011).

**Estimación y Ajuste del ML2p.** El ML2p supone que la probabilidad  $P_i(\theta)$  de que un individuo con un nivel de rasgo  $\theta$  brinde una determinada respuesta a un ítem dicotómico está en función de los parámetros de localización  $b_i$  y de discriminación  $a_i$  del ítem. Su formulación viene dada por la expresión:

$$P_i(\theta) = \frac{1}{1 + e^{-1.7 a_i(\theta - b_i)}}$$

La aplicación del ML2p se implementó con MULTILOG (Thissen, 2003). La estimación de los parámetros se efectuó por máxima verosimilitud marginal (MML) con un criterio de convergencia de .0001. Se estimaron 12 parámetros de localización y 12 parámetros de discriminación. Así también se estimaron valores de  $\theta$  para cada sujeto sobre la base de sus patrones de respuesta a la prueba completa. Se examinó el funcionamiento diferencial de los ítems (DIF) en función del género a partir del test de Wald modificado, realizado con el software IRTPRO (Cai, Thissen, & Du Toit, 2011).

El ajuste se analizó a partir de los métodos gráficos que brinda MODFIT. El programa proporciona una comparación de las probabilidades observadas y esperadas para 25 niveles del rasgo permitiendo establecer si el modelo predice adecuadamente las curvas

empíricas. Adicionalmente, se estudiaron otros indicadores indirectos de ajuste recomendados por Rubio, Aguado, Hontangas y Hernández (2007): a) una convergencia en la estimación con una cantidad razonable de iteraciones, b) valores estimados acordes a lo esperable, c) errores estándar relativamente bajos.

#### ***Determinación del algoritmo adaptativo.***

Los ítems calibrados fueron usados como insumos para el estudio del TAI. Se efectuó una simulación sobre datos reales (simulación post hoc) operando el programa Firestar-D (Choi, 2009). A partir de la matriz de datos recabados el software selecciona progresivamente aquellas respuestas correspondientes a los ítems que presentaría en el caso de estar respondiendo un TAI.

Se consideró la media del rasgo como una estimación inicial razonable de  $\theta$  para dar comienzo a la administración (Embretson & Reise, 2000). Luego de la respuesta a cada ítem, la estimación provisional de  $\theta$  se realizó con el método bayesiano de estimación esperada a posteriori (EAP) utilizando a la normal estándar como distribución a priori. En la selección progresiva de los ítems se aplicó el método de máxima información de Fisher consistente en presentar al individuo el ítem que resulte más informativo para el nivel provisional de  $\theta$  estimado hasta ese momento. En este punto se añadió la especificación de que la selección se realizó eligiendo al azar uno de los tres ítems con máxima función de información. Se esperó con esto alcanzar una mejor representatividad del contenido en el muestreo de los ítems. Se definió un criterio de parada mixto por el cual la administración de ítems finalizaría al conseguirse un nivel de error en la estimación del rasgo igual o inferior a .4 o al presentarse 9 ítems (25% menos de ítems que los utilizados en la versión

convencional del test).

Para determinar si este algoritmo adaptativo brinda un nivel de rasgo aproximado al que obtendría la persona al responder a todos los ítems, se correlacionaron los niveles  $\theta$  estimados a partir de este TAI con los estimados al responder todos los ítems. Además, se analizó si este tipo de administración podría afectar las evidencias de validez convergente con las escalas del SCL-90-R. Por ello, se aplicaron pruebas de diferencia de  $r$  de Pearson para muestras relacionadas (Steiger, 1980) con el fin de comparar si los coeficientes de correlación con estas escalas diferían al estimar las puntuaciones de  $N$  con el TAI o con el test completo.

## **Resultados**

### *Fase de calibración de los ítems*

***Supuestos del ML2p.*** Los índices de ajuste comparativos evidenciaron valores adecuados (TLI = .94 y CFI = .95). Si se considera el índice de ajuste absoluto RMSEA = .075 (IC 90% .063 - .087), el límite superior del IC podría suponer cierto desajuste. No obstante, en términos generales, es posible aceptar razonablemente la unidimensionalidad del constructo a partir de estos resultados. Todas las cargas factoriales para la solución unidimensional fueron estadísticamente significativas ( $p < .05$ ) con valores que variaron entre .51 (ítem 3) y .89 (ítem 11). Con respecto al supuesto de independencia local, en la Tabla 1 se reproduce la distribución de frecuencias y resúmenes estadísticos de los valores de las ratio  $\chi^2/gl$  obtenidos con MODFIT. Como se puede observar, la contrastación de la independencia local de los ítems por pares y tríos mostró que todas las ratios resultaron menores a 3, apuntando a la verificación de este supuesto.

**Tabla 1**

Frecuencias y estadísticos descriptivos de la ratio  $\chi^2/gl$  para contrastar la independencia local de los ítems.

	Tabla de frecuencias			Media	DE
	< 1	1 < 2	2 < 3		
Simple	12	0	0	0.003	0.006
Pares de ítems	8	4	0	0.622	0.523
Tríos de ítems	1	3	0	1.066	0.612

**Aplicación del ML2p.** El proceso de estimación requirió de 26 iteraciones para alcanzar el criterio de convergencia de .0001, lo cual resulta en una cantidad de ciclos razonable. Los parámetros estimados para los ítems presentaron valores dentro de un rango esperable y los errores de estimación resultaron relativamente bajos (Tabla 2).

El valor promedio de los parámetros de localización de los 12 ítems fue de 0.24 (DE = 0.27) y sus valores mínimo y máximo fueron de -0.04 (*harto*) y 0.88 (*sufriente*). Estos resultados indican que los ítems se localizaron con una fuerte concentración en los niveles medio-altos del rasgo. Se constató una significativa falta de ítems que permitan registrar las diferencias individuales de personas con niveles bajos del rasgo. A nivel descriptivo es posible señalar que los contenidos de los ítems demandan un nivel de  $N$  relativamente elevado como para tener más chances de responder en el sentido del rasgo. En efecto, la distribución de los  $\theta$  estimados para los sujetos osciló entre -1.31 y 0.72 con una media de -0.27 (DE = 0.63) lo que denota una tendencia de los evaluados a puntuar en niveles medio-bajos de  $N$ .

**Tabla 2**

Parámetros de los ítems, estudio de DIF y uso en el Test Adaptativo Informatizado (TAI).

Ítems	Parámetros		Test de Wald			% de uso en el TAI
	$b$ ( $se$ )	$a$ ( $se$ )	$\chi^2$	$gl$	$p$	
1) Altibajos en ánimo	-0.02 (.07)	1.58 (.32)	0.9	2	.6232	82
2) Desdichado sin motivo	0.21 (.09)	1.15 (.25)	0.9	2	.6332	83
3) Irritable	0.33 (.15)	0.58 (.17)	0.2	2	.9166	20
4) Harto	-0.04 (.09)	1.19 (.26)	2.2	2	.3276	74
5) Culpa	0.11 (.10)	0.94 (.22)	0.3	2	.8682	61
6) Nervioso	0.14 (.12)	0.73 (.19)	0.3	2	.8619	27
7) Sufriente	0.88 (.08)	1.82 (.48)	2.0	2	.3626	45
8) Tensa	0.59 (.12)	0.90 (.23)	1.3	2	.5113	35
9) Desgano	0.09 (.08)	1.22 (.26)	3.9	2	.1432	90
10) Vida monótona	-0.04 (.11)	0.80 (.19)	0.2	2	.9015	35
11) Inestable	0.28 (.07)	1.72 (.37)	1.8	2	.4039	89
12) Sentimiento de soledad	0.33 (.09)	1.17 (.27)	0.5	2	.7739	79

**Nota.** Ítem: Expresión que sintetiza el indicador indagado por cada ítem;  $se$ : Error de estimación.

Los parámetros de discriminación mostraron un promedio de 1.15 (DE=0.39), lo que supone que los ítems presentan una potencia discriminativa

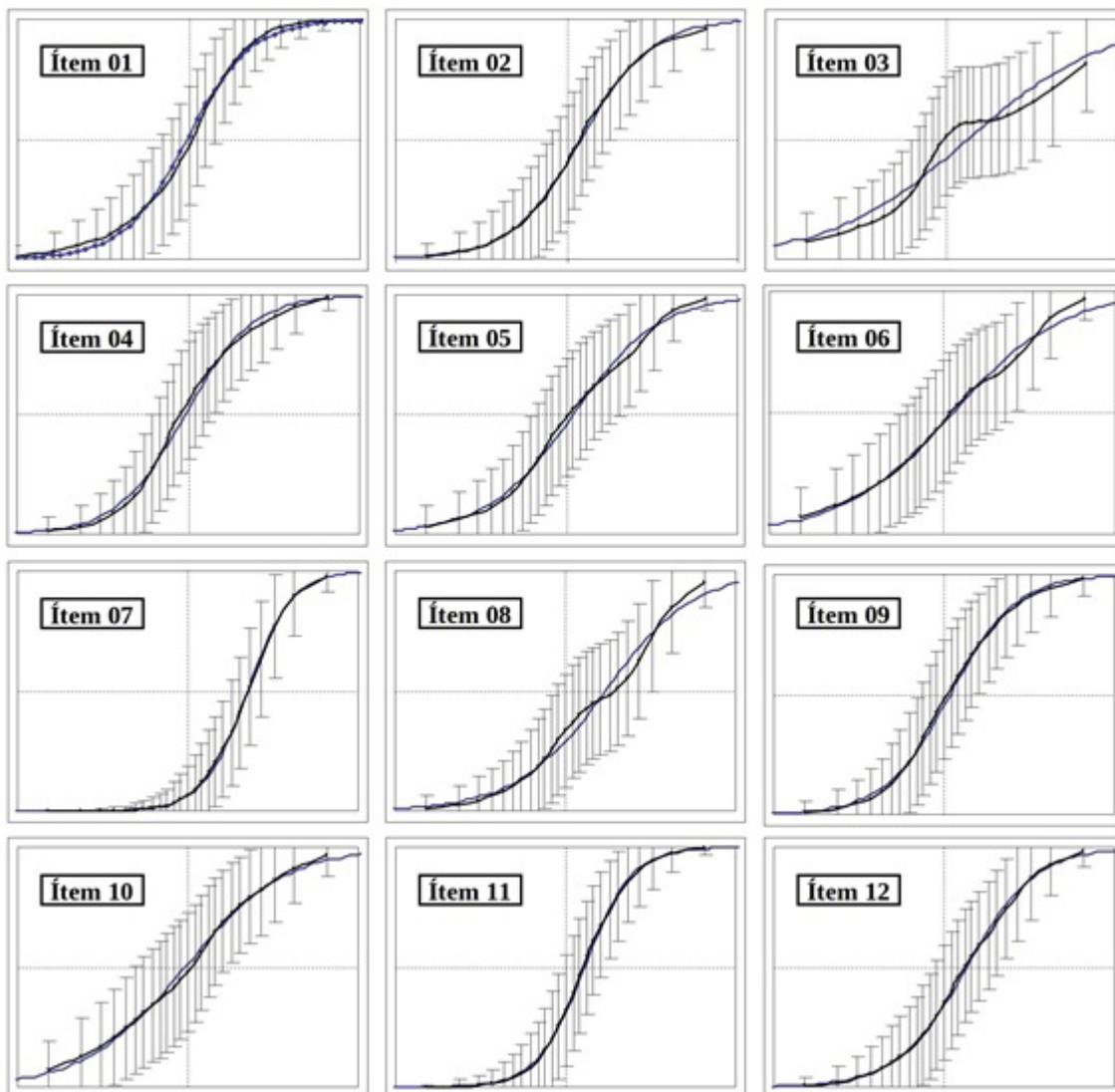
moderada en torno a sus respectivos valores de  $b$ . El único ítem que presentó un valor bajo de discriminación indaga sobre la tendencia del individuo a enojarse fácilmente ( $a = .58$ ). Entre

los ítems con mayores valores de  $a$  se destacan aquellos que preguntan por cambios en el estado del ánimo y por percibirse como una persona que sufre mucho.

En la Tabla 2 también se aporta información sobre el estudio del DIF de los ítems en función del género. Los valores  $p$  asociados al estadístico  $\chi^2$  del test de Wald señalan que no se observan diferencias estadísticamente significativas en los

parámetros de los ítems estimados para mujeres y varones.

En la Figura 1 se muestran los gráficos de ajuste que proporciona MODFIT para cada ítem. Para todos los elementos se verifica que la curva característica del ítem se encuentre dentro del intervalo de confianza asociado a la probabilidad observada de los distintos niveles del rasgo que son contrastados.



**Figura 1**

Gráficos de ajuste para los ítems de la Escala de Neuroticismo al Modelo Logístico de dos Parámetros.

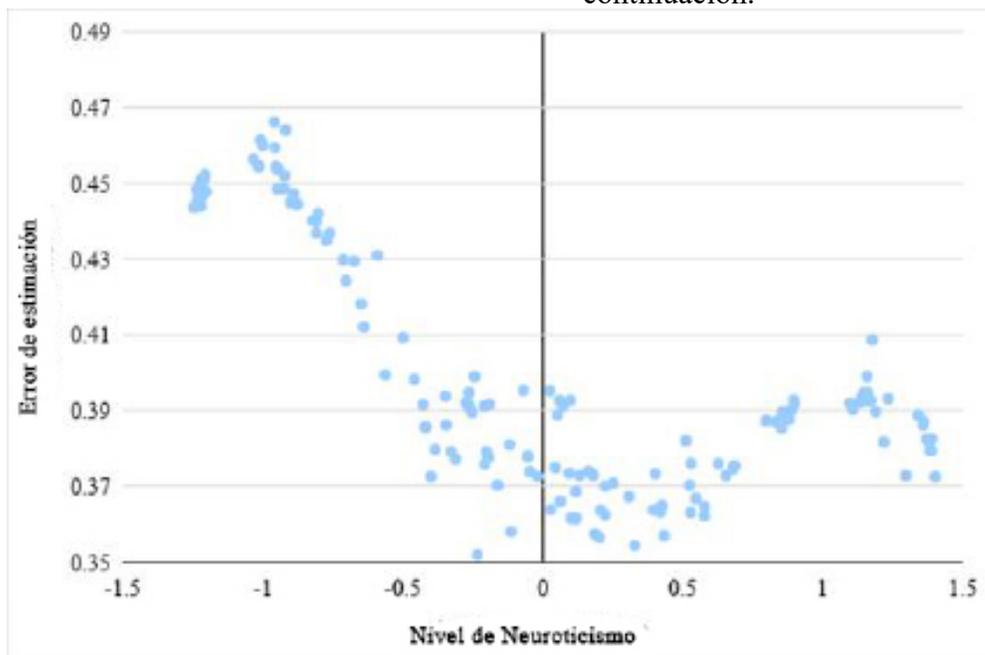
### Fase de simulación del TAI

Al analizar los resultados obtenidos tras la simulación del TAI con criterio de parada mixto se encontró una correlación alta y positiva entre los valores  $\theta$  estimados con el test completo y su versión adaptativa ( $r = .98$ ;  $p < .001$ ) y también entre los errores de estimación respectivos ( $r = .77$ ;  $p < .001$ ). La asociación resultó también muy elevada al correlacionar el TAI y el puntaje total obtenido desde la perspectiva de la clásica ( $r = .96$ ;  $p < .001$ ). La diferencia media entre los niveles de rasgo estimados para los evaluados en el TAI y sus correspondientes niveles estimados a partir del test completo (sesgo empírico) resultó de .018.

En las condiciones que establece el criterio de parada adoptado se requirió administrar un promedio de 7.2 ítems por sujeto. El 48.9% de los participantes alcanzó un error de  $\leq .4$  luego de presentarles 6 ítems y el 69.1% requirieron de 9 ítems o menos. El 30.9% restante ( $n = 55$ )

no alcanzó el error prefijado y se interrumpió su evaluación por alcanzar el tope de cantidad de ítems.

Los errores de estimación variaron entre .35 y .47 con una media de .39 (DE = .032). Esto implica que se alcanzó un nivel de precisión adecuado considerando la reducción en la cantidad de ítems administrados. El error máximo (.47) en la estimación del  $\theta$  del TAI se corresponde con una confiabilidad clásica de .78. Esto significa que ningún individuo fue evaluado con confiabilidad menor que .78. El error promedio es equivalente a una confiabilidad de .85. Esto supone que para la mayoría de los sujetos evaluados la confiabilidad del TAI, en términos clásicos, fue similar a la que se obtuvo con el test en su formato convencional, pero requirió administrar entre un 50% y 75% menos de ítems. Asimismo, los casos en los que la estimación se realizó con mayor error no son mejor evaluables con la versión completa de la escala ya que no hay ítems que aporten máxima información en esos niveles, como se verá a continuación.



**Figura 2**

Diagrama de dispersión de los errores de estimación en función de los niveles de rasgo estimados a partir del TAI.

Se obtuvo una correlación de  $-.72$  ( $p < .001$ ) entre el nivel de rasgo estimado y el error, indicando que a menor  $\theta$  estimado se verificó un mayor error de estimación. Este hallazgo era esperable ya que los ítems de la prueba se localizan en un rango restringido del rasgo por encima de la media de  $\theta$ . La Figura 2 exhibe el diagrama de dispersión de los errores de estimación en función del  $\theta$  estimado con el TAI. Allí se observa que la mayoría de los sujetos con niveles de  $\theta \geq -.5$  presentan un error de  $\leq .4$ . En cambio, los sujetos que no alcanzaron el error de estimación prefijado al finalizar la administración adaptativa mostraron un  $\theta$  promedio de  $-.96$  (DE =  $.36$ ) y un error promedio de  $.44$  (DE =  $.0012$ ).

Los evaluados con menores niveles de rasgo también fueron los que necesitaron mayor cantidad de ítems y, sin embargo, no alcanzaron el nivel de precisión esperado. En este sentido, la finalización de la administración en 9 ítems parece justificada para efectuar un uso más eficiente del tiempo y del esfuerzo del evaluado. Los tres ítems adicionales que le permitirían al sujeto completar el instrumento aportaron muy poca información para la medición dado que se encontraban alejados de la localización del individuo en la escala del rasgo. En la Tabla 2 se observa el porcentaje de casos en los que fue administrado cada ítem. Los ítems menos usados son los que presentaron menor capacidad discriminativa como el 3 (*irritable*) y el 6 (*nerviosismo*). Entre los más aplicados aparecen aquellos que tienen parámetros  $a$  entre moderado y alto y cuyos parámetros  $b$  de localización fueron próximos a la media del rasgo. El ítem 7 (*sufriente*) presentó el valor más elevado de discriminación, pero solo fue usado en el 35% de los casos porque se localiza en un nivel medio-alto de la escala. Debido a que el TAI se iniciaba con un ítem seleccionado al azar entre los tres más informativos para el nivel de rasgo

medio, los ítems 9 (*desgano*), 11 (*inestable*) y 1 (*altibajos en el ánimo*) fueron los utilizados para arrancar la administración.

Finalmente, el estudio de la asociación entre el  $N$  y las escalas del SCL-90-R mostró resultados acordes a lo esperable desde una perspectiva teórica (Tabla 3). Aun remitiendo a la personalidad normal, el  $N$  es un rasgo que describe una predisposición a padecer trastornos psicopatológicos. En este sentido, es coherente que los individuos tiendan a manifestar mayor intensidad y/o variedad de sintomatología psicológica cuanto mayor nivel de  $N$  experimenten. Si bien los resultados van en este sentido, el hallazgo más destacable que se observa en la Tabla 3 es que las evidencias de validez convergente con las escalas del SCL-90-R no registraron diferencias estadísticamente significativas al ser estudiadas con las estimaciones de  $\theta$  surgidas del TAI o del test completo. Esto significa que la reducción en la cantidad de los ítems que presenta la versión adaptativa no impactó sustantivamente en estas evidencias de validez.

## Discusión

Si bien pareciera que el modelo de personalidad de Eysenck ha sido opacado por la preponderancia actual del modelo de los Cinco Factores (Digman, 1990; Goldberg, 1999; McCrae & Costa, 2010), el uso de sus cuestionarios no ha perdido vigencia ni en su versión para población adulta (Almiro, Moura, & Simões, 2016; Colledani, Anselmi, & Robusto, 2018; Moeller et al., 2015; Picconi, Jackson, Balsamo, Tommasi, & Saggino, 2018; Zambrano-Cruz, 2011) ni en su versión para adolescentes (Alcázar-Córcoles, Verdejo-García, Bouso-Sáiz,

**Tabla 3**

Correlación de las estimaciones del TAI y el test completo con criterios externos.

Escalas del SCL90-R	TAI	Test completo	Prueba de diferencia de $r$ de Pearson
Somatizaciones	.40**	.40**	$t(175) = .36$ ; ns
Obsesiones y Compulsiones	.51**	.52**	$t(175) = .54$ ; ns
Sensitividad Interpersonal	.50**	.51**	$t(175) = .46$ ; ns
Depresión	.65**	.66**	$t(175) = .56$ ; ns
Ansiedad	.56**	.57**	$t(175) = .64$ ; ns
Hostilidad	.43**	.43**	$t(175) = .37$ ; ns
Ansiedad Fóbica	.38**	.38**	$t(175) = .21$ ; ns
Ideación Paranoide	.36**	.36**	$t(175) = .28$ ; ns
Psicoticismo	.45**	.45**	$t(175) = .45$ ; ns
Índice de Severidad Global	.52**	.53**	$t(175) = .55$ ; ns
Total de Síntomas Positivos	.49**	.50**	$t(175) = .53$ ; ns
Índice de Malestar Positivo	.52**	.53**	$t(175) = .55$ ; ns

**Nota.** \*\* $p < .001$ .

Reuelta-Menéndez, & Ramírez-Lira, 2017; Colledani, Robusto, & Anselmi, 2018). Esto se debe fundamentalmente a que da como resultado un modelo paradigmático causal de corte biologicista que cuenta con un importante volumen de evidencia experimental acumulada (Flores-Mendoza et al., 2016; Revelle, 2016). Incluso los estudios recientes sobre una fundamentación genética del  $N$  (Calboli et al., 2010; Luciano et al., 2018; Mitchell & Kumari, 2016; Smith et al., 2016) y la replicación del modelo en diversas culturas (Bowden, Saklofske, van de Vijver, Sudarshan, & Eysenck, 2016; Eysenck & Barrett, 2013) muestran cómo se sostiene la preocupación de Eysenck por encontrar bases empíricas que complementen los hallazgos obtenidos con el análisis factorial de los cuestionarios.

Dentro del modelo psicobiológico de Eysenck, la dimensión  $N$  presenta atributos relevantes que justifican su selección para los fines de este trabajo. En el plano teórico,  $N$  se

destaca por haber sido caracterizado con relativa concordancia por las diversas teorías factorialistas de la personalidad contemporáneas a Eysenck y por ser reconocido por desarrollos posteriores (De Raad & Mlačić, 2015; Zuckerman, 2005). Existen pocos modelos empíricos o teóricos de la estructura general de la personalidad que no incluyan un dominio como  $N$  (Digman, 1990). A su vez, en el plano aplicado, la vinculación de  $N$  con distintos trastornos psicopatológicos y enfermedades físicas ofrece como ventaja la posibilidad de pensar estrategias transdiagnósticas de detección y prevención (Hajek, Bock, & König, 2017; Lahey, 2009; Vittengl, 2017; Widiger & Oltmanns, 2017).

La metodología TAI aquí implementada ha demostrado proporcionar estimaciones de los niveles de  $N$  de los evaluados con un grado de precisión aceptable aun cuando se administra sólo una parte de los ítems que componen la escala  $N$  del EPQ-RS. Similares resultados se

han registrado al estudiar evidencias de validez de versiones adaptativas de tests reconocidos como el MMPI-2 (Forbey & Ben-Porath, 2007; Salazar-Schneiderman, Tonetti, Sommers, & Greene, 2016) y el NEO-PI-R (Reise & Henson, 2000) y se encontraron idénticos estándares de calidad de la medida administrando solo la mitad de reactivos que la equivalente convencional. La disminución de la cantidad de ítems de  $N$  del EPQ-RS administrados puede variar entre un 50% y 75% dependiendo de la localización del sujeto en la dimensión del rasgo.

Un aspecto que conviene discutir es la posibilidad de que, al administrar una menor cantidad de ítems con el TAI, se vea afectada la representatividad del contenido muestreado. La validez de contenido del TAI está supeditada a los ítems con los que cuenta el EPQ-RS. El algoritmo adaptativo selecciona el ítem de entre todos los disponibles basándose en un criterio de calidad psicométrica. Su administración dependerá de la información que pueda aportar el ítem por su cercanía al nivel del rasgo estimado para ese sujeto independientemente del contenido al que alude el indicador. Con respecto a la implementación adaptativa del EPQ-RS, la correlación elevada observada entre el  $\theta$  estimado con el TAI y el  $\theta$  obtenido con el test completo demuestra que las variaciones en el contenido no afectaron sustantivamente la medición del constructo. Tampoco se alteró la relación del  $N$  con las dimensiones sintomáticas evaluadas mediante la SCL-90-R. En consecuencia, la infra o supra representación de un rasgo primario durante la medición adaptativa de una persona no tuvo consecuencias en la estimación de su rasgo.

Si se quisiera alcanzar un mayor control sobre el contenido de los ítems deberían identificarse con más precisión los rasgos primarios del  $N$  e incorporarse más elementos para la medición

de estos aspectos. Estudios posteriores a la construcción del EPQ-R llevaron a Eysenck y sus colaboradores a definir siete rasgos primarios: inferioridad, infelicidad, ansiedad, dependencia, hipocondría, culpa y obsesión (Eysenck, Barrett, Wilson, & Jackson, 1992), pero luego se consideró que los primeros tres eran los más descriptivos de la dimensión (Eysenck, Wilson, & Jackson, 1996). Obsesión se encontró débilmente asociado al resto de los rasgos mientras que dependencia, hipocondría y culpa ofrecían información redundante (Jackson, Furnham, Forde, & Cotter, 2000). La delimitación conceptual precisa del  $N$  y la identificación de sus rasgos primarios son objeto de debate actual al interior de las teorías factorialistas de la personalidad (Tackett & Lahey, 2017; Widiger, 2009).

La concentración de los parámetros de localización de los ítems en los niveles medio-altos del rasgo latente podría revelar una fortaleza o una debilidad de la escala, pero esto dependerá de los objetivos para los que se la pretenda usar. Si la finalidad de la evaluación es estudiar la distribución del rasgo en personalidad normal se hallarán dificultades en la especificación del extremo inferior en la bipolaridad del constructo. A nivel conceptual este polo se define como la tendencia a permanecer estable y tranquilo incluso al ser confrontado por estresores importantes (Furnham et al., 2008). La falta de ítems localizados en los niveles bajos de la escala imposibilita describir los rasgos distintivos que caracterizan ese rango de la variable. Esto significa que a nivel operacional la especificación de la estabilidad emocional se circunscribe a la ausencia de manifestaciones relacionadas con un nivel elevado de  $N$ . Como consecuencia, la estimación del nivel del rasgo de los sujetos próximos al polo de la estabilidad emocional resulta más imprecisa. Esta limitación es un problema de la escala de partida y no es una

consecuencia del TAI.

Los ítems del instrumento resultan más efectivos para detectar niveles elevados de  $N$ , que son justamente las puntuaciones que presentan más relevancia en el contexto clínico-epidemiológico. Esto significa que la localización de los ítems en niveles medio-altos del rasgo beneficia especialmente a la aplicación de la prueba para tareas de evaluación con fines de rastillaje. En esta línea, la implementación mediante un formato de administración adaptativo contribuye a realizar esta evaluación de manera más eficiente en dos sentidos. En primer lugar, la reducción en la cantidad de ítems posibilita que el tiempo ahorrado pueda ser aprovechado para la medición de otras variables o una simplificación de la administración que evite la fatiga innecesaria del evaluado. El ahorro de tiempo podría ser considerado despreciable en la práctica cotidiana de una evaluación individual o en grupos reducidos. Sin embargo, la economía de los tiempos de administración se vuelve significativa si la finalidad es evaluar a un conjunto numeroso de personas o en estudios a gran escala (Baldasaro et al., 2013).

En segundo lugar, el procedimiento adaptativo es más eficiente en comparación con una prueba de igual cantidad de ítems pero con formato fijo. La evaluación adaptativa selecciona los ítems para ofrecer garantías de una medición más precisa incluso en los valores extremos del rasgo. Estos niveles del rasgo son los que presentan mayor error si se emplean las pruebas convencionales (Olea & Ponsoda, 2013; Reise & Revicki, 2015; Rubio et al., 2007).

El algoritmo adaptativo ensayado en el presente estudio propone una manera de reducir la cantidad de ítems de la EPQ-RS desde una aproximación alternativa a la aplicada en estudios que se basan en la Teoría Clásica de Tests. La

selección de los ítems que componen los tests con formato fijo se realiza en función de identificar los elementos con mayor capacidad discriminativa general (Francis et al., 1992; Sandín et al., 2002). Sin embargo, en el marco de la TRI, la discriminación de un ítem se considera en torno al valor de su parámetro  $b$ . El ítem que se localiza en un nivel del rasgo en el que se hayan posicionado pocos sujetos será escasamente utilizado en el TAI a pesar de tener una elevada capacidad discriminativa (como sucedió con el ítem 7). También podría considerarse una diferencia con respecto a la validez de contenido. En las versiones breves con un formato fijo se acota el muestreo de indicadores evaluados, pero se corre el peligro de que conlleve a restringir posibles ramificaciones del constructo. En cambio, en la versión adaptativa existe la posibilidad de que cualquiera de los ítems de la prueba original sea administrado siempre que el nivel del individuo se halle cerca de su localización.

Aun cuando los resultados son alentadores es fundamental rescatar el carácter preliminar de estos. El escaso número de elementos que componen la escala limitan considerablemente la posibilidad de alcanzar una medida más precisa y eficiente. Es habitual que las mediciones adaptativas se realicen a partir de un banco que incluya una mayor cantidad de ítems (Reise & Revicki, 2015). En el presente estudio se ha recurrido a un conjunto de ítems pertenecientes a un instrumento corto pero con fuertes evidencias de validez y estudios de adaptación local. Se ha demostrado que a pesar de su escaso número de ítems es posible contar con una versión adaptativa de esta escala.

En futuros estudios se buscará incorporar nuevos ítems tendientes a identificar indicadores del constructo que detecten niveles bajos de  $N$ . Esto no resulta una tarea sencilla. Eysenck y Eysenck

(1969) concluyeron que es desaconsejable emplear ítems redactados en sentido negativo al rasgo para medir el *N*. Además, el estudio de la unidimensionalidad requiere la inversión en la codificación de las respuestas de los ítems negativos, por lo que no es seguro que queden localizados en los niveles bajos del rasgo. Parece recomendable apelar a ítems cuyos indicadores sean plasmados en los enunciados con una frecuencia de aparición más esporádica o con una manifestación de menor intensidad. Es complejo poder predecir en qué nivel de rasgo se localizará el ítem, pero se debería recurrir a matices más sutiles en la redacción de los enunciados.

## Referencias

- Abad, F. J., Olea, J., Ponsoda, V., & García, C. (2011). *Medición en Ciencias Sociales y de la Salud*. Madrid, España: Síntesis.
- Alcázar-Córcoles, M. A., Verdejo-García, A. J., Bouso-Sáiz, J. C., Revuelta-Menéndez, J., & Ramírez-Lira, E. (2017). Propiedades psicométricas del cuestionario de personalidad EPQ-A en una muestra de adolescentes hispanohablantes. *Anuario de Psicología Jurídica*, 27(1), 51-56. doi: [10.1016/j.apj.2017.02.002](https://doi.org/10.1016/j.apj.2017.02.002)
- Almiro, P. A., Moura, O., & Simões, M. R. (2016). Psychometric properties of the European Portuguese version of the Eysenck Personality Questionnaire - Revised (EPQ-R). *Personality and Individual Differences*, 88, 88-93. doi: [10.1016/j.paid.2015.08.050](https://doi.org/10.1016/j.paid.2015.08.050)
- Aluja, A., García, O., & García, L. F. (2003). A psychometric analysis of the revised Eysenck Personality Questionnaire short scale. *Personality and Individual Differences*, 35(2), 449-460. doi: [10.1016/S0191-8869\(02\)00206-4](https://doi.org/10.1016/S0191-8869(02)00206-4)
- Baldasaro, R. E., Shanahan, M. J., & Bauer, D. J. (2013). Psychometric properties of the Mini-IPIP in a large, nationally representative sample of young adults. *Journal of Personality Assessment*, 95(1), 74-84. doi: [10.1080/00223891.2012.700466](https://doi.org/10.1080/00223891.2012.700466)
- Bech, P. (2016). Neuroticism (Eysenck's Theory). En V. Zeigler-Hill & T. K. Shackelford (Eds.), *Encyclopedia of Personality and Individual Differences* (pp. 1-4). doi: [10.1007/978-3-319-28099-8\\_1094-1](https://doi.org/10.1007/978-3-319-28099-8_1094-1)
- Bowden, S. C., Saklofske, D. H., van de Vijver, F. J. R., Sudarshan, N. J., & Eysenck, S. B. G. (2016). Cross-cultural measurement invariance of the Eysenck Personality Questionnaire across 33 countries. *Personality and Individual Differences*, 103, 53-60. doi: [10.1016/j.paid.2016.04.028](https://doi.org/10.1016/j.paid.2016.04.028)
- Byrne, B. M. (2012). *Structural equation modeling with Mplus: Basics, concepts, applications, and programming*. New York, NY: Routledge. doi: [10.4324/9780203807644](https://doi.org/10.4324/9780203807644)
- Cai, L., Thissen, D., & du Toit, S. H. C. (2011). IRTPRO for Windows [software de cómputo]. Lincolnwood, IL: Scientific Software International.
- Calboli, F. C. F., Tozzi, F., Galwey, N. W., Antoniadis, A., Mooser, V., Preisig, M., ... Balding, D. J. (2010). A genome-wide association study of neuroticism in a population-based sample. *PLoS One*, 5(7), 1-7. doi: [10.1371/journal.pone.0011504](https://doi.org/10.1371/journal.pone.0011504)
- Casullo, M. (2004). Síntomas psicopatológicos en adultos urbanos. *Psicología y Ciencia Social*, 6(1), 49-57. Recuperado de <https://biblat.unam.mx/es/revista/psicologia-y-ciencia-social>
- Cattell, R. B. (1957). *Personality and Motivation Structure and Measurement*. Oxford, Inglaterra: World Book Co.
- Chang, H. H. (2015). Psychometrics behind computerized adaptive testing. *Psychometrika*, 80(1), 1-20. doi: [10.1007/s11336-014-9401-5](https://doi.org/10.1007/s11336-014-9401-5)
- Choi, S. W. (2009). Firestar: Computerized adaptive testing simulation program for Polytomous Item Response Theory models. *Applied Psychological Measurement*, 33(8), 644-645. doi: [10.1177/0146621608329892](https://doi.org/10.1177/0146621608329892)
- Colledani, D., Anselmi, P., & Robusto, E. (2018). Using Item Response Theory for the development of a new

- short form of the Eysenck Personality Questionnaire-Revised. *Frontiers in Psychology*, 9, 1-13. doi: [10.3389/fpsyg.2018.01834](https://doi.org/10.3389/fpsyg.2018.01834)
- Colledani, D., Robusto, E., & Anselmi, P. (2018). Development of a new abbreviated form of the Junior Eysenck Personality Questionnaire-Revised. *Personality and Individual Differences*, 120, 159-165. doi: [10.1016/j.paid.2017.08.037](https://doi.org/10.1016/j.paid.2017.08.037)
- Credé, M., Harms, P., Niehorster, S., & Gaye-Valentine, A. (2012). An evaluation of the consequences of using short measures of the Big Five personality traits. *Journal of Personality and Social Psychology*, 102(4), 874-888. doi: [10.1037/a0027403](https://doi.org/10.1037/a0027403)
- De Raad, B., & Mlačić, B. (2015). Big Five Factor model, theory and structure. En J. D. Wright (Ed.), *International Encyclopedia of the Social & Behavioral Sciences* (2<sup>a</sup>. ed., Vol. II, pp. 559-566). Oxford, Reino Unido: Elsevier. doi: [10.1016/b978-0-08-097086-8.25066-6](https://doi.org/10.1016/b978-0-08-097086-8.25066-6)
- Derogatis, L. (1994). *SCL-90-R. Symptom Checklist-90-R. Administration, Scoring and Procedures Manual*. Minneapolis, MN: National Computer System.
- Digman, J. M. (1990). Personality structure: Emergence of the Five-Factor model. *Annual Review of Psychology*, 41(1), 417-440. doi: [10.1146/annurev.psych.41.1.417](https://doi.org/10.1146/annurev.psych.41.1.417)
- Donnellan, M. B., Oswald, F. L., Baird, B. M., & 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](https://doi.org/10.1037/1040-3590.18.2.192)
- Drasgow, F., Levine, M. V., Tsien, S., Williams, B. A., & Mead, A. D. (1995). Fitting Polytomous Item Response Theory models to multiple-choice tests. *Applied Psychological Measurement*, 19(2), 143-165. doi: [10.1177/014662169501900203](https://doi.org/10.1177/014662169501900203)
- Dumont, F. (2010). *A History of Personality Psychology: Theory, Science, and Research from Hellenism to the Twenty-first Century*. New York, NY: Cambridge University. doi: [10.1017/cbo9780511676093.002](https://doi.org/10.1017/cbo9780511676093.002)
- Embretson, S. E., & Reise, S. P. (2000). *Item Response Theory for Psychologists*. Mahwah, NJ: Erlbaum.
- Eysenck, H. J. (1947). *Dimensions of Personality*. Londres, Reino Unido: Routledge & Kegan Paul.
- Eysenck, H. J. (1952). *The scientific study of personality*. Londres, Reino Unido: Routledge & Kegan Paul.
- Eysenck, H. J. (1958). A short questionnaire for the measurement of two dimensions of personality. *Journal of Applied Psychology*, 42(1), 14-17. doi: [10.1037/h0041738](https://doi.org/10.1037/h0041738)
- Eysenck, H. J., & Eysenck, S. B. G. (1969). *Personality structure and measurement*. Londres, Reino Unido: Routledge & Kegan Paul.
- Eysenck, H. J., & Eysenck, S. B. G. (1975). *Manual of the Eysenck Personality Questionnaire*. Londres, Reino Unido: Hodder and Stoughton.
- Eysenck, H. J., Barret, P., Wilson, G. D., & Jackson, C. J. (1992). Primary trait measurement of the 21 components of the P-E-N System. *European Journal of Psychological Assessment*, 8(2), 109-117.
- Eysenck, H. J., Wilson, G. D., & Jackson, C. J. (1996). *Manual of the Eysenck Personality Profiler (Short)*. Guilford, Inglaterra: Psi-Press.
- Eysenck, S. B. G., & Barrett, P. (2013). Re-introduction to cross-cultural studies of the EPQ. *Personality and Individual Differences*, 54(4), 485-489. doi: [10.1016/j.paid.2012.09.022](https://doi.org/10.1016/j.paid.2012.09.022)
- Eysenck, S. B. G., Eysenck, H. J., & Barrett, P. (1985). A revised version of the Psychoticism Scale. *Personality and Individual Differences*, 6(1), 21-29. doi: [10.1016/0191-8869\(85\)90026-1](https://doi.org/10.1016/0191-8869(85)90026-1)
- Flores-Mendoza, C., Ardila, R., Gallegos, M., Sampaio-Braga, L., Carvalhais-Santiago, B. M., & Andrade, D. M. (2016). Hans Eysenck in Latin America: His influence in the psychology, the study of personality and individual differences. *Personality and Individual Differences*, 103, 68-73. doi: [10.1016/j.paid.2016.04.049](https://doi.org/10.1016/j.paid.2016.04.049)
- Forbey, J. D., & Ben-Porath, Y. S. (2007). Computerized adaptive personality testing: A review and illustration

- with the MMPI-2 computerized adaptive version. *Psychological Assessment*, 19(1), 14-24. doi: [10.1037/1040-3590.19.1.14](https://doi.org/10.1037/1040-3590.19.1.14)
- Francis, L. J., Lewis, C. A., & Ziebertz, H. G. (2006). The Short-Form Revised Eysenck Personality Questionnaire (EPQR-S): A German edition. *Social Behavior Personality: An International Journal*, 34(2), 197-204. doi: [10.2224/sbp.2006.34.2.197](https://doi.org/10.2224/sbp.2006.34.2.197)
- Francis, L. J., Brown, L. B., & Philipchalk, R. (1992). The development of an abbreviated form of the Revised Eysenck Personality Questionnaire (EPQR-A): Its use among students in England, Canada, the U.S.A. and Australia. *Personality and Individual Differences*, 13(4), 443-449. doi: [10.1016/0191-8869\(92\)90073-X](https://doi.org/10.1016/0191-8869(92)90073-X)
- Furnham, A., Eysenck, S. B. G., & Saklofske, D. H. (2008). The Eysenck personality measures: Fifty years of scale development. En G. J. Boyle, G. Matthews & D. H. Saklofske (Eds.), *The Sage handbook of personality theory and assessment* (Vol. II, pp. 199-218). Los Angeles, CA: Sage. doi: [10.4135/9781849200479.n10](https://doi.org/10.4135/9781849200479.n10)
- Goldberg, L. R. (1999). A broad-bandwidth, public-domain, personality inventory measuring the lower-level facets of several five-factor models. En I. Mervielde, I. Deary, F. De Fruyt & F. Ostendorf (Eds.), *Personality Psychology in Europe*, (Vol. VII, pp. 7-28). Tilburg, The Netherlands: Tilburg University.
- Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, 37(6), 504-528. doi: [10.1016/S0092-6566\(03\)00046-1](https://doi.org/10.1016/S0092-6566(03)00046-1)
- Guilford, J. P. (1940). *An inventory of factors STDCR*. California: Sheridan Supply.
- Hajek, A., Bock, J. O., & 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), 1-15. doi: [10.1371/journal.pone.0181716](https://doi.org/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](https://doi.org/10.1016/j.jpsychores.2016.03.012)
- Jackson, C. J., Furnham, A., Forde, L., & Cotter, T. (2000). The structure of the Eysenck Personality Profiler. *British Journal of Psychology*, 91(2), 223-239. doi: [10.1348/000712600161808](https://doi.org/10.1348/000712600161808)
- Jeronimus, B. F., Kotov, R., Riese, H., & 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](https://doi.org/10.1017/S0033291716001653)
- Lahey, B. B. (2009). Public health significance of neuroticism. *American Psychologist*, 64(4), 241-256. doi: [10.1037/a0015309](https://doi.org/10.1037/a0015309)
- Lewis, C. A., Francis, L. J., Shevlin, N., & Forrest, S. (2002). Confirmatory factor analysis of the French translation of the Abbreviated Form of the Revised Eysenck Personality Questionnaire (EPQR-A). *European Journal of Psychological Assessment*, 18(2), 179-185. doi: [10.1027//1015-5759.18.2.179](https://doi.org/10.1027//1015-5759.18.2.179)
- Luciano, M., Hagenaars, S. P., Davies, G., Hill, W. D., Clarke, T. K., Shirali, M. ... Deary, I. J. (2018). Association analysis in over 329,000 individuals identifies 116 independent variants influencing neuroticism. *Nature Genetics*, 50(1), 6-11. doi: [10.1038/s41588-017-0013-8](https://doi.org/10.1038/s41588-017-0013-8)
- McCrae, R. R., & Costa, P. T. Jr. (2010). *NEO Inventories professional manual*. Odessa, FL: Psychological Assessment Resources.
- Milojev, P., Osborne, D., Greaves, L. M., Barlow, F. K., & Sibley, C. G. (2013). The Mini-IPIP6: Tiny yet highly stable markers of Big Six personality. *Journal of Research in Personality*, 47(6), 936-944. doi: [10.1016/j.jrp.2013.09.004](https://doi.org/10.1016/j.jrp.2013.09.004)
- Mitchell, R. L. C., & Kumari, V. (2016). Hans Eysenck's

- interface between the brain and personality: Modern evidence on the cognitive neuroscience of personality. *Personality and Individual Differences*, 103, 74-81. doi: [10.1016/j.paid.2016.04.009](https://doi.org/10.1016/j.paid.2016.04.009)
- Moeller, S. B., Bech, P., Kessing, L., Mortensen, E. L., Austin, S. F., & Bukh, J. D. (2015). A psychometric validation analysis of Eysenck's neuroticism and extraversion scales in a sample of first time depressed patients. *Journal of Depression and Anxiety*, 4(4), 1-6. doi: [10.4172/2167-1044.1000202](https://doi.org/10.4172/2167-1044.1000202)
- Muthén, L., & Muthén, B. (2010). *Mplus User's Guide*. (6aed.). Los Angeles, CA: Muthén & Muthén.
- Olea, J., & Ponsoda, V. (2013). *Tests adaptativos informatizados*. Madrid, España: UNED.
- Picconi, L., Jackson, C. J., Balsamo, M., Tommasi, M., & Saggino, A. (2018). Factor structure and measurement invariance across groups of the Italian Eysenck Personality Questionnaire - Short form (EPP-S). *Personality and Individual Differences*, 123, 76-80. doi: [10.1016/j.paid.2017.11.013](https://doi.org/10.1016/j.paid.2017.11.013)
- Reise, S. P., & Henson, J. M. (2000). Computerization and adaptive administration of the NEO PI-R. *Assessment*, 7(4), 347-364. doi: [10.1177/107319110000700404](https://doi.org/10.1177/107319110000700404)
- Reise, S. P., & Revicki, D. A. (2015). *Handbook of Item Response Theory Modeling Applications to Typical Performance Assessment*. Nueva York, NY: Routledge.
- Revelle, W. (2016). Hans Eysenck: Personality theorist. *Personality and Individual Differences*, 103, 32-39. doi: [10.1016/j.paid.2016.04.007](https://doi.org/10.1016/j.paid.2016.04.007)
- Rubio, V. J., Aguado, D., Hontangas, P. M., & Hernández, J. M. (2007). Psychometric properties of an emotional adjustment measure. An application of the Graded Response Model. *European Journal of Psychological Assessment*, 23(1), 39-46. doi: [10.1027/1015-5759.23.1.39](https://doi.org/10.1027/1015-5759.23.1.39)
- Salazar-Schneiderman, Tonetti, C. S., Sommers, L. M., & Greene, R. L. (2016). Advances in the use and interpretation of the MMPI-2. En U. Kumar (Ed.), *The Wiley Handbook of Personality Assessment* (pp. 158-172). Nueva York, NY: John Wiley & Sons. doi: [10.1002/9781119173489.ch12](https://doi.org/10.1002/9781119173489.ch12)
- Sánchez, R. O., & Ledesma, R. D. (2009). Análisis psicométrico del Inventario de Síntomas Revisado (SCL-90-r) en población clínica. *Revista Argentina de Clínica Psicológica*, 18(3), 265-274. Recuperado de <http://www.clinicapsicologica.org.ar>
- Sandín, B., Valiente, R. M., Chorot, P., Olmedo, M., & Santed, M. A. (2002). Versión española del cuestionario EPQR-Abreviado (EPQR-A) (I): Análisis exploratorio de la estructura factorial. *Revista de Psicopatología y Psicología Clínica*, 7(3), 195-205. doi: [10.5944/rppc.vol.7.num.3.2002.3933](https://doi.org/10.5944/rppc.vol.7.num.3.2002.3933)
- Sauer-Zavala, S., Wilner, J. G., & Barlow, D. H. (2017). Addressing neuroticism in psychological treatment. *Personality Disorders: Theory, Research, and Treatment*, 8(3), 191-198. doi: [10.1037/per0000224](https://doi.org/10.1037/per0000224)
- Sibley, C. G. (2012). The Mini-IPIP6: Item Response Theory analysis of a short measure of the Big-Six Factors of personality in New Zealand. *New Zealand Journal of Psychology*, 41(3), 20-31. Recuperado de <http://www.psychology.org.nz/publications-media/new-zealand-journal-of-psychology/#.W2mQkdVKjIV>
- Smith, D. J., Escott-Price, V., Davies, G., Bailey, M. E. S., Colodro-Conde, L., Ward, J., ... O'Donovan, M. C. (2016). Genome-wide analysis of over 106000 individuals identifies neuroticism-associated loci. *Molecular Psychiatry*, 21, 749-757. doi: [10.1038/mp.2016.49](https://doi.org/10.1038/mp.2016.49)
- Soto, C. J., & John, O. P. (2016). The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social Psychology*, 113(1), 117-143. doi: [10.1037/pspp0000096](https://doi.org/10.1037/pspp0000096)
- Soto, C. J., & John, O. P. (2017). Short and extra-short forms of the Big Five Inventory-2: The BFI-2-S and BFI-2-XS. *Journal of Research in Personality*, 68, 69-81. doi: [10.1016/j.jrp.2017.02.004](https://doi.org/10.1016/j.jrp.2017.02.004)
- Squillace, M., Picón-Janeiro, J., & Schmidt, V. (2013).

- Adaptación local del Cuestionario Revisado de Personalidad de Eysenck (Versión abreviada). *Revista Evaluar*, 13(1), 19-37. Recuperado de <https://revistas.unc.edu.ar/index.php/revaluar>
- Stark, S. (2001). MODFIT: A computer program for model-data fit. Manuscrito inédito, University of Illinois, Urbana-Champaign, IL, U.S.A.
- Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, 87(2), 245-251. doi: [10.1037/0033-2909.87.2.245](https://doi.org/10.1037/0033-2909.87.2.245)
- Tackett, J. L., & Lahey, B. B. (2017). Neuroticism. En T. A. Widiger (Ed.), *The Oxford Handbook of the Five Factor Model*. New York, NY: Oxford University. doi: [10.1093/oxfordhb/9780199352487.013.14](https://doi.org/10.1093/oxfordhb/9780199352487.013.14)
- Thissen, D. (2003). MULTILOG [software]. Chicago: Scientific Software International.
- Thurstone, L. L., & Thurstone, T. G. (1930). A neurotic inventory. *The Journal of Social Psychology*, 1(1), 3-30. doi: [10.1080/00224545.1930.9714128](https://doi.org/10.1080/00224545.1930.9714128)
- Tiwari, T., Singh, A. L., & Singh, I. L. (2009). The Short-Form Revised Eysenck Personality Questionnaire: A Hindi edition (EPQRS-H). *Industrial Psychiatry Journal*, 18(1), 27-31. doi: [10.4103/0972-6748.57854](https://doi.org/10.4103/0972-6748.57854)
- Vittengl, J. R. (2017). Who pays the price for high neuroticism? Moderators of longitudinal risks for depression and anxiety. *Psychological Medicine*, 47(10), 1794-1805. doi: [10.1017/S0033291717000253](https://doi.org/10.1017/S0033291717000253)
- Widiger, T. A. (2009). Neuroticism. En M. R. Leary & R. H. Hoyle (Eds.), *Handbook of individual differences in social behavior* (pp. 129-146). New York, NY: Guilford.
- Widiger, T. A., & 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](https://doi.org/10.1002/wps.20411)
- Woodworth, R. S. (1919). Personal Data Sheet. Chicago, IL: C. H. Stoelting.
- Zambrano-Cruz, R. (2011). Revisión sistemática del Cuestionario de Personalidad de Eysenck (Eysenck Personality Questionnaire - EPQ). *Liberabit*, 17(2), 147-155. Recuperado de <http://revistaliberabit.com/es>
- Ziegler, M., Kemper, C. J., & 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](https://doi.org/10.1027/1614-0001/a000148)
- Zuckerman, M. (2005). *Psychobiology of personality* (2<sup>a</sup> ed.). New York, NY: Cambridge University. doi: [10.1017/cbo9780511813733](https://doi.org/10.1017/cbo9780511813733)