


Artículo original

Cuestionario Revisado de Personalidad de Eysenck (versión reducida): análisis con la teoría de respuesta al ítem

Revised Eysenck Personality Questionnaire (Brief Version): Analysis with Item Response Theory

Facundo Juan Pablo Abal ¹  [ORCID](#); Damián Javier Ursino ²  [ORCID](#); Horacio Félix Attorresi ³  [ORCID](#)

¹ Doctor en Psicología. Universidad de Buenos Aires y Consejo Nacional de Investigaciones Científicas y técnicas (CONICET), Argentina.

² Becario de Doctorado Universidad de Buenos Aires.

³ Licenciado en Ciencias Matemáticas. Universidad de Buenos Aires, Argentina.

Fecha correspondencia:

Recibido: mayo 11 de 2020.

Aceptado: marzo 31 de 2021.

Forma de citar:

Abal, F.J. P., Ursino, D.J., & Attorresi, H.F. (2022).

Cuestionario revisado de personalidad de Eysenck (versión reducida): análisis con la teoría de respuesta al ítem. *Rev. CES Psico*, 15(1), 1-23.

<https://dx.doi.org/10.21615/cesp.5830>

Open access

© Derecho de autor

Licencia creative commons

Ética de publicaciones

Revisión por pares

Gestión por Open Journal System

DOI: 10.21615/cesp.5830

ISSN: 2011-3080

Publica con nosotros

Resumen

El objetivo de este trabajo es aportar nuevas evidencias de calidad psicométrica para la adaptación argentina de la versión reducida del Cuestionario de Personalidad de Eysenck (EPQ-RS). Participaron 1136 personas de población general (52.5% femenino, edad media = 29.6 años, DE = 11.9) residentes en Buenos Aires, Argentina. La adaptación argentina se compone de 42 ítems con formato de respuesta dicotómica. Se realizó un análisis factorial confirmatorio a partir de la matriz de correlaciones tetracóricas. Esto permitió replicar la estructura propuesta por Eysenck para el modelo PEN (Psicoticismo-Extraversión-Neuroticismo) y la escala Sinceridad. Posteriormente, se ajustó el modelo logístico de dos parámetros por separado para los ítems de cada escala. Los ítems no mostraron funcionamiento diferencial según género. La discriminación de los ítems resultó moderada-alta. Los parámetros b se localizaron en rangos acotados de cada uno de los rasgos medidos, lo que originó que la precisión de las escalas varíe en el recorrido de los continuos. La escala Neuroticismo aporta más información en niveles medios del rasgo, Psicoticismo en los medio-bajos y Extraversión en los medio-altos. La escala Sinceridad mostró una función de información relativamente plana en todo el recorrido del rasgo. Se brindan evidencias de validez basadas en la relación con otras pruebas que miden facetas

del neuroticismo y sintomatología. Las evidencias de validez y confiabilidad obtenidas ofrecen garantías de calidad suficientes para la aplicación de este instrumento en el contexto local y confirman la vigencia del modelo teórico que operacionaliza el EPQ-RS.

Palabras clave: neuroticismo; psicoticismo; extraversión; sinceridad; modelo logístico de dos parámetros; evaluación de la personalidad; EPQ-RS.

Abstract

The aim of this work is to provide new evidence of psychometric quality for the Argentinean adaptation of the brief version of the Eysenck Personality Questionnaire (EPQ-RS). 1136 people from the general population (52.5% female, mean age = 29.6 years, SD = 11.9) residing in Buenos Aires, Argentina participated. The Argentinean adaptation consists of 42 items with dichotomous response format. A confirmatory factor analysis was performed from the tetrachoric correlation matrix. This allowed replicating the structure proposed by Eysenck for the PEN model (Psychoticism - Extroversion - Neuroticism) and the Lie scale. Subsequently, the two-parameter logistic model was adjusted separately for the items of each scale. The items did not show differential functioning by gender. Items discrimination was moderate-high. Parameters b were located in narrow ranges of each one of the measured traits, which caused the precision of the scales to vary along the trait continuums. The Neuroticism scale provides more information at medium levels of the trait, Psychoticism in the medium-low and Extraversión in the medium-high. The Lie scale showed a relatively flat information function throughout the trait. Evidence of validity based on the relationship with other tests that measure facets of neuroticism and symptomatology is provided. The evidence of validity and reliability obtained offers sufficient quality guarantees for the application of this instrument in the local context and confirms topicality of the theoretical model that operationalizes the EPQ-RS.

Keywords: neuroticism; psychoticism; extraversion; lie; two-parameter logistic model; personality assessment; EPQ-RS.

Introducción

La teoría de la personalidad de Eysenck (1947, 1952, 1990) se ha consolidado durante el siglo XX como el modelo paradigmático de corte biológico-factorial con mayor evidencia empírica acumulada. Distinguiéndose de la mayoría de los teóricos de la personalidad de su época, Eysenck se propuso la integración de los aportes de la psicometría con reconocidas teorías de la psicología experimental (Revelle, 2016). De esta manera, llevó el estudio de los rasgos de la personalidad más allá de la descripción y buscó explicaciones causales mediante la identificación de sustratos neurofisiológicos, muchos de los cuales están siendo corroborados en la actualidad (Mitchell & Kumari, 2016). El planteo de Eysenck, conocido como el modelo PEN, reconoce que las diferencias individuales de la personalidad pueden reducirse a tres

superfactores esenciales: Psicoticismo (P), Extraversión (E) y Neuroticismo (N). Estos factores se ubican en el nivel más alto de una estructura jerárquica y se configuran a partir de las intercorrelaciones de rasgos primarios anidados (Eysenck & Eysenck, 1985).

El factor P se compone de un conjunto de rasgos primarios entre los que se destacan la impulsividad y el poco respeto por los convencionalismos. Los marcadores comportamentales más extremos de esta dimensión denotan problemas de socialización emparentados con conductas antisociales. Por el contrario, los individuos con bajos niveles de P se preocupan por los demás y se apegan al cumplimiento de las normas (Eysenck & Eysenck, 1985). Eysenck justificó la inclusión de este factor para describir la predisposición de los sujetos a la psicosis y al trastorno antisocial de personalidad. Sin embargo, esta idea es fuertemente cuestionada en la actualidad (Knežević et al, 2019).

La dimensión E puede caracterizarse por sus polos extremos: Extraversión-Introversión. Los niveles elevados de este rasgo describen a personas que se sienten altamente atraídas por las situaciones sociales. Son animadas y activas en la búsqueda de cambios y nuevas sensaciones. Por el contrario, las personas con niveles bajos de E tienden a ser reservadas e introspectivas. Prefieren mantenerse alejadas del contacto social y realizar actividades rutinarias (Eysenck & Eysenck, 1975).

Finalmente, la dimensión N describe un continuo que permite ubicar un nivel de estabilidad-inestabilidad emocional de la persona. Los sujetos con niveles elevados de N son propensos a experimentar frecuentemente ansiedad, miedo, vergüenza, sentimiento de culpa y mal humor. Tienden a reaccionar de forma exagerada frente a todos los estímulos y presentan dificultades para retornar al nivel emocional normal luego de estas experiencias (Eysenck, 1990). En el otro extremo de esta dimensión, la persona con bajo N es tranquila, controlada y mantiene un temperamento equilibrado en la mayoría de las circunstancias (Eysenck & Eysenck, 1975).

Eysenck y sus colaboradores desarrollaron una sucesión de cuestionarios para la medición de las variables de su modelo. Estos se fueron perfeccionando a medida que engrosaba su teoría psicobiológica y usaba análisis factorial para estudiar psicométricamente los ítems (Furnham, Eysenck, & Saklofske, 2008). Los cuestionarios permitieron aportar mayor evidencia empírica para sostener la universalidad del modelo PEN en al menos dos líneas de investigación: la comparación entre distintos grupos etarios (Eysenck & Eysenck, 1975) y la invariancia cultural (e.g. Bowden, Saklofske, van de Vijver, Sudarshan, & Eysenck, 2016; Eysenck & Barrett, 2013). Además, en el devenir histórico de la psicometría se han refinado los procedimientos estadísticos recomendados para los análisis factoriales con datos categóricos (Lloret-Segura, Ferreres-Traver, Hernández-Baeza, & Tomás-Marco, 2014) y, sin embargo, la estructura interna del modelo PEN se ha mantenido estable.

El Cuestionario de Personalidad Eysenck Revisado (EPQ-R) y su forma breve (EPQ-RS) (Eysenck, Eysenck, & Barrett, 1985) son considerados el resultado más acabado de un arduo trabajo en la construcción de una medida para operacionalizar las dimensiones básicas de personalidad que propuso Eysenck (Bech, 2016). Existe una versión posterior, el Cuestionario *Eysenck Personality Profiler* (Eysenck, Wilson, & Jackson, 1996) que reavivó el interés teórico por la estructura interna definiendo subdimensiones para los factores, pero su extensión (la versión breve consta de 200 ítems) ha complicado su implementación práctica.

El EPQ-R también incluye una escala de Sinceridad para evaluar la tendencia a dar una imagen positiva de sí mismo, aunque su interpretación no es sencilla y ha sido objeto de controversias (Katz & Francis, 1991; Massey 1980). Los ítems están redactados de tal manera que buscan registrar o bien el acuerdo con comportamientos normativos/morales deseables, pero poco probables, o bien el desacuerdo con conductas contrarias a la norma, pero bastante comunes (comportamientos deshonestos o transgresiones menores). Sin embargo, es difícil determinar si la Escala realmente mide sinceridad en las respuestas o un rasgo de la personalidad que describe una estricta aceptación de las normas sociales.

Si bien el EPQ-R presenta una extensión considerable de 100 ítems, la versión breve está conformada por una selección de sólo 48 elementos (Eysenck et al., 1985). La posibilidad de obtener una medición rápida de la personalidad impulsó una gran cantidad de adaptaciones a distintos idiomas y culturas (Alexopoulos & Kalaitzidis, 2004; Almiro, Moura, & Simões, 2016; Aluja, García, & García, 2003; Dominguez, Villegas, Yauri, Aravena, & Ramírez, 2013; Francis, Lewis, & Ziebertz, 2006; Lewis, Francis, Shevlin, & Forrest, 2002; Lewis & Musharraf, 2014; Squillace, Picón Janeiro, & Schmidt, 2013; Tiwari, Singh, & Singh, 2009). Recientemente Colledani, Anselmi y Robusto (2018, 2019) decidieron analizar los ítems de la EPQ-R desde la perspectiva de la Teoría de Respuesta al Ítem (TRI) y efectuaron una nueva selección de los ítems. Los autores generaron versiones reducidas de 48 y 24 ítems como alternativas a las propuestas desarrolladas por Eysenck et al. (1985) y Francis et al. (1992). Sin embargo, sus resultados no mejoran la debilidad psicométrica que suele presentar la escala de Psicoticismo y sacrifican la comparabilidad frente a posibles estudios transculturales.

El EPQ-R y su versión reducida son instrumentos reconocidos a nivel internacional y ampliamente usados en diferentes contextos. El objetivo de este trabajo es aportar nuevas evidencias de validez y confiabilidad para la adaptación argentina del EPQ-RS basadas en métodos psicométricos avanzados. Específicamente, considerando lineamientos metodológicos actuales definidos para la realización de estudios factoriales (Lloret-Segura et al., 2014) y el análisis de la calidad de los reactivos a partir de la aplicación de la TRI (Muñiz, 2018; Reise & Revicki, 2015).

Método

Participantes

Colaboraron 1136 personas adultas residentes en el área metropolitana de Buenos Aires, Argentina. Los participantes se seleccionaron a partir de un muestreo no probabilístico por accesibilidad. El 52.5% pertenece al género femenino y la edad promedio resultó de 29.6 años ($DE=11.9$; $Mín=18$, $Máx=82$). El 81.4% consignó estar empleado con un trabajo de tiempo completo o medio tiempo. En cuanto al nivel educativo, la mayoría de los participantes habían completado estudios secundarios (56.9%), seguido por universitarios (22.3%) y estudios de nivel superior no universitarios tales como tecnicaturas (estudios técnicos) y profesorado (15.5%). Sólo un 5.3% no alcanzó a concluir el nivel secundario. Un porcentaje mayoritario de participantes señaló tener un nivel socioeconómico medio (80.1%) mientras que se registraron menores proporciones de nivel medio-bajo (10.4%) y medio-alto (9.5%).

Instrumentos

Eysenck Personality Questionnaire Revised - short version, EPQ-RS (Eysenck et al., 1985). Si bien el cuestionario original se compone de 48 ítems, en la adaptación argentina quedó reducido a 42 (Squillace et al., 2013). Los ítems tienen un formato de respuesta dicotómico y sólo ocho de ellos están redactados en sentido inverso. Los adaptadores obtuvieron la estructura propuesta por Eysenck mediante un análisis factorial exploratorio. Registraron índices de consistencia interna adecuados para Neuroticismo (Alfa = .84), Extraversión (Alfa = .77) y Sinceridad (Alfa = .71). Para Psicoticismo, en cambio, resultó mínimamente aceptable (Alfa = .66) según el criterio de DeVellis (2016).

Banco de ítems de Neuroticismo (BIN, Abal, Auné, & Attorresi, 2019). Consta de 54 ítems con formato de respuesta politómica (*En desacuerdo, Ligeramente en desacuerdo, Ligeramente de acuerdo y De acuerdo*). El instrumento mide las seis facetas de Neuroticismo propuestas por McCrae y Costa (2010). Cuenta con evidencias de validez basadas en la estructura interna y los ítems están calibrados con el modelo de respuesta graduada de la TRI. Los coeficientes de consistencia interna Alfa de Cronbach oscilaron entre .70 (Hostilidad y Vulnerabilidad) y .83 (Depresión).

Inventario de síntomas SCL-90-R (Derogatis, 1994). Se compone de 90 ítems que examinan patrones de síntomas psicológicos. Los ítems se agrupan en nueve dimensiones clínicas (Somatización, Obsesiones/Compulsiones, Sensitividad 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 ítems indagan sobre la frecuencia con que cada síntoma le ha preocupado o molestado al sujeto en los últimos siete días y se responde con un formato de cinco opciones (de *nada* a *mucho*). El inventario cuenta con una adaptación argentina con suficientes evidencias de validez y confiabilidad (Casullo, 2004; Sanchez & Ledesma, 2009). En cuanto al presente estudio, el

análisis de la consistencia interna de las nueve dimensiones sintomatológicas registró valores de Alfa de Cronbach desde 0.77 (Ansiedad Fóbica) hasta 0.86 (Depresión).

Procedimiento

El diseño de investigación fue aprobado por el Comité de Conductas Responsables de la Facultad de Psicología de la Universidad de Buenos Aires, sede en donde se lleva adelante el presente proyecto de investigación. Para la recolección de los datos se entrenó en tareas de evaluación psicológica a un grupo compuesto por estudiantes avanzados de la carrera de grado de Psicología y graduados que cursaban estudio de posgrado. Estos colaboradores se encargaron de contactar a los participantes de la investigación y de administrar los instrumentos incluidos en el protocolo. Su labor fue supervisada por los autores del artículo. Como criterios de inclusión de los participantes del estudio se consideraron: edad de 18 años o más, nacionalidad argentina y residencia en el área metropolitana de Buenos Aires. Adicionalmente, se fijaron pautas en la selección de los casos para alcanzar un balance en cuanto al sexo y edad de los participantes. Se excluyeron personas no dispuestas o incapacitadas para firmar el consentimiento informado e individuos con algún tipo de relación estrecha con los evaluadores.

A las personas evaluadas se les informó que el objetivo de la prueba consistía en la evaluación de distintos aspectos de la personalidad y que los ítems no poseían respuestas correctas o incorrectas. Además, se les comunicó que su participación era de carácter voluntaria, anónima y se garantizaba la confidencialidad de sus respuestas, y que podían abandonar la prueba en cualquier momento. Estas consideraciones fueron precisadas por escrito en el consentimiento informado. Una vez firmado, se procedió con la administración del protocolo en formato de lápiz y papel sin tiempo límite. Al finalizar la participación los evaluados recibieron los correos electrónicos del evaluador y del director del proyecto por si necesitaban resolver alguna duda.

Análisis de datos

A fin de obtener evidencias de validez basadas en la estructura interna se aplicó un Análisis Factorial Confirmatorio usando el programa Mplus (Muthén & Muthén, 2010). El modelo propuesto busca replicar la versión alcanzada en el estudio exploratorio realizado por Squilace et al. (2013). Se estimaron los parámetros con el método de mínimos cuadrados ponderados robustos (*Weighted Least Squares Mean and Variance Adjusted*, WLSMV) con base en la matriz de correlaciones tetracóricas. El ajuste del modelo a los datos se analizó con los criterios recomendados por Byrne (2012): índices de ajuste comparativo $CFI \geq .90$ y de Tucker-Lewis $TLI \geq .90$ y el error medio cuadrático de aproximación $RMSEA \leq .08$.

El estudio con la TRI se llevó adelante aplicando el Modelo Logístico de dos parámetros (ML2p) de Birnbaum (1968). La decisión de aplicar este modelo responde, en primer lugar, a la naturaleza dicotómica de los datos. Por otro lado, dentro de los modelos dicotómicos, el ML2p es el que se elige con más frecuencia para el análisis de ítems de dos alternativas de tests de

personalidad (Abal, Lozzia, Aguerri, Galibert, & Attorresi, 2010; Reise & Waller, 1990, 2009) y, particularmente, para la modelización de las escalas del EPQ-RS (Bauermeister & Gallacher, 2020; Colledani, Anselmi et al., 2018; Colledani, Robusto, & Anselmi, 2018; Ferrando, 2003). La posibilidad de estimar libremente el parámetro de discriminación supone una ventaja para el ajuste del ML2p por sobre el Modelo de Rasch, dado que los ítems de personalidad difícilmente presentan discriminación homogénea. A su vez, la aplicabilidad del parámetro c (denominado habitualmente de *pseudo-azar*) del Modelo logístico de tres parámetros no ha arrojado aún una interpretación clara en el contexto de los tests de comportamiento típico (Abal, Auné, & Attorresi, 2015; Min, Zickar, & Yankov, 2018).

La calibración de los ítems según el ML2p se llevó adelante con IRTPRO 4.2 (Cai, Thissen, & du Toit, 2011). Este modelo postula que la probabilidad $P_i(\vartheta)$ de que un individuo con un nivel de rasgo ϑ elija la opción clave en 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:

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

Previa aplicación de la TRI se verificó el supuesto de unidimensionalidad requerido por el modelo para cada factor por separado. Se usaron AFCs con idénticos procedimientos de estimación y criterios de interpretación mencionados para analizar la dimensionalidad de la prueba. Además, dado que los indicadores de ajuste no garantizan por sí mismos una solución factorial satisfactoria (Calderón Garrido, Navarro González, Lorenzo Seva, & Ferrando Piera, 2019) se complementaron con otros indicadores basados en múltiples fuentes para aportar evidencias de que cada una de las cuatro escalas es esencialmente unidimensional. Estos análisis se realizaron con el programa FACTOR 11.05 (Ferrando & Lorenzo-Seva, 2017). Se aplicó el análisis paralelo de Horn en su implementación óptima (Timmerman & Lorenzo Seva, 2011) para la extracción de cada factor y se evaluó la proximidad a la unidimensionalidad con los índices ECV (*Explained Common Variance*) y MIREAL (*Mean of Item Residual Absolute Loadings*) recomendados por Ferrando y Lorenzo-Seva (2018). Por último, para examinar la calidad y eficacia de las puntuaciones factoriales estimadas se obtuvieron los índices SR (*Sensitivity Ratio*) y EPTD (*Expected percentage of true differences*).

También se estudió el supuesto de independencia local con el estadístico χ^2_{LD} de Chen y Thissen (1997) y se adoptó el criterio recomendado por Reise y Rodriguez (2016) para su verificación ($\chi^2_{LD} < 10$). Adicionalmente, como puede resultar poco plausible asumir a priori que la asíntota inferior de la función logística (parámetro c) se ubica en 0, se comparó el ajuste relativo de los Modelos Logísticos de 2 y 3 parámetros a partir de la significación estadística de la diferencia

entre el índice -2 por valor del logaritmo de la verosimilitud ($-2 \log \text{likelihood}$, $-2LL$). Este estadístico se distribuye según χ^2 con grados de libertad igual a la diferencia de parámetros libres (De Ayala, 2009).

La estimación de los parámetros de los ítems (localización y discriminación) y de los sujetos se realizó por separado para cada factor con el método de Máxima Verosimilitud Marginal (MML). Posteriormente, se estudió ítem por ítem el ajuste al ML2p usando el índice $S\text{-}\chi^2$ de Orlando y Thissen (2003). Siguiendo a Toland (2013), se consideró un ajuste aceptable si el p asociado al $S\text{-}\chi^2$ es mayor a .01.

Finalmente, se empleó el test de Wald para analizar el posible Funcionamiento Diferencial de los Ítems (*Differential Item Functioning*, DIF) en función del género. Además, se buscaron evidencias de validez convergente y discriminante al estudiar la asociación entre los ϑ estimados con el ML2p y los puntajes de las escalas del BIN y SCL-90-R. Para complementar la información ofrecida por las tradicionales correlaciones bivariadas también se calcularon correlaciones canónicas entre el conjunto de las dimensiones del EPQ-RS y el conjunto de variables usadas como criterio utilizando el paquete CCA disponible para R (González, Déjean, Martin, & Baccini, 2008).

Para estimar la confiabilidad basada en la consistencia interna se aplicó el coeficiente Alfa de Cronbach. Desde la TRI se obtuvieron el coeficiente de fiabilidad marginal y las funciones de información y del error estándar de medida para los valores estimados de ϑ en cada faceta por separado.

Resultados

Estudio factorial confirmatorio

Se encontró un ajuste aceptable al modelo de cuatro factores de Eysenck (CFI = .95; TLI = .94; RMSEA = .048, 90 IC [.046, .050]). Todos los parámetros estimados fueron significativos al 5 % y los coeficientes de regresión estandarizados de cada uno de los ítems en sus respectivos factores fueron iguales o mayores a .40.

Aplicación ML2p

En la [Tabla 1](#) se resumen los resultados obtenidos en el estudio de la unidimensionalidad de cada factor por separado. En todos los casos las respectivas cargas factoriales para la solución unidimensional fueron estadísticamente significativas ($p < .05$). Los resultados reflejan un ajuste al modelo unidimensional aceptable en las cuatro variables que operacionaliza el instrumento (TLI y CFI > .90; RMSEA < .08). Los análisis paralelos de Horn sugieren la extracción de un único factor en cada escala con varianzas explicadas que oscilaron entre 48.9% y 63.8%. Los ECV ≥ 0.853 y MIREAL ≤ 0.238 reconocen la existencia de un único factor dominante y descartan la presencia de un factor residual relevante. Asimismo, los SR ≥ 2.06 y EPTD $\geq 90.5\%$ muestran

que las estimaciones de los puntajes factoriales tienen una calidad aceptable (Ferrando & Lorenzo-Seva, 2018).

Tabla 1. Supuestos del ML2p y estudios de confiabilidad.

	<i>Psicoticismo</i>	<i>Extraversión</i>	<i>Neuroticismo</i>	<i>Sinceridad</i>
Análisis de la unidimensionalidad				
CFI / TLI	.97 / .96	.94 / .92	.96 / .96	.97 / .96
RMSEA (90 IC)	.046 (.036 - .056)	.066 (.057 - .076)	.069 (.061 - .076)	.033 (.022 - .043)
Pesaje Mín / Máx	.45 / .76	.40 / .82	.48 / .85	.41 / .62
AP (var exp.%)	56.4%	54.2%	63.8%	48.9%
ECV	0.861	0.853	0.888	0.857
MIREAL	0.209	0.238	0.211	0.207
SR	2.81	3.06	4.06	2.06
EPTD	91.9%	92.7%	94.8%	90.5%
Independencia local				
χ^2_{LD} Mín / Máx	0.4 / 9.1	0.0 / 9.7	0.2 / 8.5	0.1 / 9.8
-2 log likelihood				
ML2p	11535.7	10863.9	13108.9	11984.3
ML3p	11521.5	10855.1	13096.3	11968.7
Consistencia interna				
Alfa de Cronbach	.72	.76	.86	.72
Fiabilidad Marginal	.75	.73	.82	.68

Nota: AP (var exp %) = Porcentaje de varianza explicada en el primer factor identificado en el Análisis Paralelo. ECV = Explained Common Variance. MIREAL = Mean of Item Residual Absolute Loadings. SR = Sensitivity Ratio. EPTD = Expected percentage of true differences.

Con respecto al supuesto de independencia local de los ítems es posible observar la aparición de valores máximos del estadístico χ^2_{LD} próximos al criterio planteado por Chen y Thissen (1997) aunque no ha sido superado por ningún par de ítems de los factores. Estos hallazgos son suficientes para aceptar el cumplimiento de este supuesto.

Las diferencias de los -2LL obtenidas en el ajuste del modelo más restringido (ML2p) y el modelo más general (ML3p) no registraron significación estadística para las dimensiones evaluadas en el EPQ-RS: Psicoticismo ($\chi^2 = 14.2$, $gl = 10$, $p = .29$), Extraversión ($\chi^2 = 8.85$, $gl = 10$, $p = .55$), Neuroticismo ($\chi^2 = 12.6$, $gl = 12$, $p = .40$), Sinceridad ($\chi^2 = 15.6$, $gl = 10$, $p = .11$). Estos resultados indican que la estimación adicional del parámetro c del ML3p no mejora sustantivamente el

ajuste de los datos al modelo y, en consecuencia, la parsimonia justifica la preferencia por la elección del ML2p.

En la [Tabla 2](#) se exhiben los resultados obtenidos en el proceso de calibración de los ítems de cada factor por separado. Se incluyen los parámetros de localización y discriminación del ML2p con sus respectivos errores de estimación. Los estadísticos $S\text{-}\chi^2$ que aparecen en la [Tabla 2](#) señalan que todos los ítems mostraron un ajuste satisfactorio al ML2p ($p > .01$). También es posible observar que no se detectaron ítems con DIF cuando se analizaron las diferencias de género. En efecto, todos los p -valores asociados al estadístico χ^2 del test de Wald fueron superiores a .01, permitiendo descartar la presencia de DIF en los ítems pertenecientes a los cuatro factores del EPQ-RS.

Escala Psicoticismo. Los parámetros a estimados para la escala P mostraron un valor promedio de 1.45 ($DE = 0.41$), lo que representa una alta capacidad discriminativa global. El ítem 1 - inverso - (*¿Se detiene a pensar las cosas antes de hacerlas?*) evidenció el valor más elevado de discriminación ($a = 2.03$) y el ítem 26 (*¿La mayoría de las cosas le son indiferentes?*) presentó el más bajo ($a = 0.85$). Aun siendo 0.85 el parámetro de pendiente más pequeño para los ítems de esta escala, su capacidad discriminativa resultó moderada con respecto a la clasificación establecida por Baker y Kim (2017). El ítem 26 también se destaca por requerir el menor nivel de rasgo para responder afirmativamente ($b = -0.66$). En contraposición, el ítem 16 (*¿Es mejor actuar como uno quiera antes que seguir las normas sociales?*) fue el que demandó un mayor nivel de Psicoticismo ($b = 1.43$) para contestar la opción clave. A nivel descriptivo se observa que los tres ítems con redacción inversa tendieron a presentar los valores de b más bajos.

Escala Extraversión. La capacidad discriminativa del conjunto de ítems de la escala de Extraversión también resultó elevada ($Media = 1.56$, $DE = 0.66$) oscilando entre $a = 0.70$ (ítem 40. *¿Le gusta el ruido y la agitación a su alrededor?*) y $a = 2.55$ (Ítem 11. *¿Tiende a mantenerse apartado/a en las situaciones sociales?*). El ítem 40 mostró, además, el parámetro b más elevado ($b = 1.63$) en tanto que el ítem 6 (*¿Es usted una persona animada, vital?*) requiere de un ϑ medio-bajo para estar de acuerdo con la frase del enunciado ($b = -1.27$). Los únicos dos ítems inversos de esta escala mostraron valores de discriminación comparativamente más elevados que el resto de los ítems.

Escala Neuroticismo. El valor promedio de los parámetros a obtenidos para la escala N fue de 2.01 ($DE = 0.73$) siendo 0.93 y 3.28 los respectivos valores mínimo (ítem 8: *¿Es usted una persona que se enoja con facilidad (irritable)?*) y máximo (ítem 36. *¿Es usted inestable en sus estados de ánimo?*). Los ítems se localizaron entre $\vartheta = -0.21$ (ítem 31. *A menudo ¿siente que la vida es muy monótona?*) y $\vartheta = 0.73$ (ítem 19. *¿Es usted una persona que sufre mucho (sufriente)?*) con un promedio que refleja la tendencia a ubicarse en los niveles que superan la media del rasgo ($Media = 0.14$, $DE = 0.28$).

Enero – abril de 2022

Escala Sinceridad. Los parámetros de discriminación de S adoptaron valores aceptables desde 0.62 (ítem 7. *Si usted asegura que hará una cosa ¿siempre mantiene su promesa, sin importarle las molestias que eso le pueda ocasionar?*) a 1.51 (ítem 13. *¿Ha tomado alguna vez alguna cosa (aunque no fuese más que un alfiler o un botón) que perteneciese a otra persona?*). El promedio de a para estos ítems resultó más bajo que para las demás escalas ($Media = 1.07$, $DE = 0.30$). Dos de los tres ítems inversos mostraron los índices de discriminación más bajos de la escala. Los ítems de S se localizaron mayoritariamente en los niveles medios del rasgo desde $\vartheta = -1.21$ (ítem 10. *¿Son todos sus hábitos buenos y deseables?*) a $\vartheta = 0.83$ (ítem 32. *¿Alguna vez se ha aprovechado de alguien, ha usado a alguien?*). La media de estos parámetros b se ubicó próxima a la media del rasgo ($Media = -0.08$, $DE = 0.93$).

Tabla 2. Parámetros de los ítems.

	Parámetros		Ajuste			Test de Wald	
	b (se)	a (se)	S- χ^2	gl	P	χ^2 Total (gl = 2)	p
Psicoticismo							
Ítem 01	0.11 (0.05)	2.03 (0.19)	5.17	8	.74	3.7	.16
Ítem 14	-0.66 (0.11)	0.85 (0.10)	12.43	8	.13	0.4	.80
Ítem 16	0.35 (0.06)	1.48 (0.14)	18.48	8	.02	0.8	.68
Ítem 22	0.72 (0.07)	1.39 (0.13)	6.32	8	.61	0.1	.93
Ítem 25	0.98 (0.07)	1.98 (0.19)	10.48	8	.23	0.5	.77
Ítem 26	1.43 (0.13)	1.09 (0.12)	10.09	8	.26	1.2	.54
Ítem 33	0.15 (0.06)	1.49 (0.14)	13.97	8	.08	0.5	.78
Ítem 35	0.66 (0.06)	1.66 (0.15)	16.55	8	.04	0.8	.66
Ítem 39	-0.08 (0.05)	1.66 (0.15)	19.01	8	.02	0.1	.95
Ítem 42	-0.12 (0.08)	0.89 (0.1)	14.32	8	.07	0.9	.63
Extraversión							
Ítem 03	-0.87 (0.11)	0.92 (0.10)	6.37	8	.61	1.1	.59
Ítem 06	-1.27 (0.09)	1.64 (0.17)	9.68	8	.29	4.2	.12
Ítem 11	-0.61 (0.05)	2.55 (0.25)	2.54	8	.96	2.7	.27
Ítem 15	-0.40 (0.09)	0.85 (0.10)	6.88	8	.55	0.8	.67
Ítem 21	-0.10 (0.06)	1.49 (0.14)	3.63	8	.89	2.2	.34
Ítem 24	-0.45 (0.05)	2.03 (0.19)	5.54	8	.70	0.9	.65
Ítem 28	-0.53 (0.05)	2.34 (0.23)	8.61	8	.38	0.7	.72
Ítem 34	-0.59 (0.08)	1.07 (0.11)	3.23	8	.92	0.1	.96
Ítem 40	1.63 (0.23)	0.70 (0.10)	4.29	8	.83	0.3	.85
Ítem 41	-1.05 (0.07)	2.02 (0.21)	3.21	8	.92	1	.61
Neuroticismo							
Ítem 02	-0.04 (0.05)	3.18 (0.30)	8.62	9	.47	2.2	.33
Ítem 04	0.05 (0.06)	2.02 (0.17)	7.40	10	.69	0	.99
Ítem 08	0.37 (0.09)	0.93 (0.09)	5.38	10	.87	0.6	.75
Ítem 12	-0.17 (0.05)	2.08 (0.17)	8.21	10	.61	2.1	.36
Ítem 17	0.10 (0.07)	1.40 (0.12)	6.86	10	.74	4.1	.18
Ítem 18	0.04 (0.07)	1.29 (0.11)	9.54	10	.48	0.7	.71
Ítem 19	0.73 (0.07)	2.51 (0.23)	16.40	9	.06	0.3	.88
Ítem 23	0.48 (0.07)	1.51 (0.13)	8.83	10	.55	0	.98
Ítem 29	-0.07 (0.05)	2.27 (0.19)	7.77	10	.65	0	.99
Ítem 31	-0.21 (0.06)	1.47 (0.13)	5.66	10	.84	1.1	.56
Ítem 36	0.26 (0.05)	3.28 (0.33)	11.16	10	.35	0.7	.72
Ítem 37	0.19 (0.06)	2.17 (0.19)	8.59	10	.57	0.5	.77

	Parámetros		Ajuste			Test de Wald	
	b (se)	a (se)	S- χ^2	gl	P	χ^2 Total (gl = 2)	p
Sinceridad							
Ítem 05	1.14 (0.13)	0.98 (0.12)	11.85	8	.16	2.5	.28
<u>Ítem 07</u>	1.07 (0.18)	0.62 (0.10)	6.13	8	.63	1.2	.56
Ítem 09	0.62 (0.08)	1.27 (0.14)	10.00	8	.26	1.9	.39
<u>Ítem 10</u>	-1.21 (0.13)	0.99 (0.12)	2.88	8	.94	1.8	.42
Ítem 13	-0.17 (0.06)	1.51 (0.16)	12.71	8	.12	4.1	.17
Ítem 20	-1.06 (0.13)	0.96 (0.12)	3.38	8	.91	0.2	.89
Ítem 27	-0.51 (0.10)	0.87 (0.11)	3.68	8	.89	2.6	.28
Ítem 30	-0.40 (0.06)	1.47 (0.15)	2.90	8	.94	1.2	.54
Ítem 32	0.83 (0.09)	1.29 (0.15)	7.21	7	.41	0.6	.75
<u>Ítem 38</u>	-1.10 (0.16)	0.74 (0.10)	3.45	8	.90	2.9	.23

Nota: Los ítems subrayados están redactados en sentido inverso. se = Error de estimación.

Evidencias basadas en la relación con otras variables

Al analizar las correlaciones canónicas entre el conjunto de dimensiones del modelo PEN y el conjunto de facetas de Neuroticismo del BIN se obtuvieron magnitudes considerables y significativas para la primera (correlación canónica = .81; λ de Wilks = 0.23; $F(24, 3929) = 86.9$; $p < .0001$) y segunda función canónica (correlación canónica = .53; λ de Wilks = 0.66; $F(15, 3111) = 33.7$; $p < .0001$). Como era de esperarse dado que se trata del mismo constructo, la ponderación estandarizada más relevante para la primera función canónica fue la dimensión Neuroticismo del EPQ-RS (.92) y, entre las variables de criterio, las contribuciones más elevadas se observaron para las facetas Depresión (.45), Ansiedad (.29) y Vulnerabilidad (.18). En cambio, en la segunda función canónica prevaleció la contribución de la dimensión Extraversión (-.96) del EPQ-RS y la faceta Autoconciencia (.92) del BIN.

En el estudio de la correlación canónica entre las dimensiones del modelo PEN y las dimensiones sintomatológicas de SCL-90-R también se encontraron asociaciones significativas para la primera (correlación canónica = .69; λ de Wilks = 0.41; $F(36, 713) = 5.31$; $p < .0001$) y segunda función (correlación canónica = .42; λ de Wilks = 0.77; $F(24, 554) = 2.1$; $p < .0001$). Con base en las ponderaciones estandarizadas, la contribución a la primera función es de la dimensión Neuroticismo (.97) del EPQ-RS, mientras que el orden de las variables criterio (SCL-90R) más relevantes fue: Depresión (.85), Ansiedad (.72), Obsesiones (.51) y Sensitividad interpersonal (.47). Para la segunda función los coeficientes más elevados fueron Extraversión (.94) del EPQ-RS y la dimensión Depresión (-.47) del SCL-90-R.

En la [Tabla 3](#) se exhiben las correlaciones bivariadas obtenidas al estudiar las asociaciones de los factores del modelo PEN con las escalas de SCL-90-R y las facetas de Neuroticismo del BIN. La intensidad y dirección de las correlaciones son acordes a lo esperable desde una perspectiva teórica para aportar evidencias de validez convergente y discriminante. La asociación más fuerte de la escala P fue con la faceta Impulsividad (BIN) y no se registraron correlaciones significativas con las dimensiones de SCL-90-R. Finalmente, la escala S presentó correlaciones

relativamente bajas con todos los constructos usados como criterio externo. Se destacan algunas asociaciones con Impulsividad (BIN) y Hostilidad (BIN y SCL-90-R), que podrían vincularse con contenidos de los ítems de S en los que debe reconocerse la transgresión de normas sociales.

Tabla 3. Relación del EPQ-RS con otras variables.

	<i>Psicoticismo</i>	<i>Extraversión</i>	<i>Neuroticismo</i>	<i>Sinceridad</i>
Banco de ítems de Neuroticismo				
Ansiedad	.14**	-.19**	.70**	.16**
Hostilidad	.19**	-.19**	.55**	.25**
Depresión	.19**	-.38**	.73**	.16**
Autoconciencia	.11**	-.52**	.52**	.09**
Impulsividad	.31**	.08*	.40**	.30**
Vulnerabilidad	.18**	-.27**	.66**	.21**
SCL-90-R				
Somatización	.06*	-.18**	.44**	-.11**
Obsesiones/Compulsiones	.09**	-.19**	.56**	.09**
Sensitividad Interpersonal	.01	-.19**	.53**	.05
Depresión	.04	-.28**	.64**	-.09**
Ansiedad	.10**	-.22**	.56**	-.02
Hostilidad	.05	-.05	.40**	.22**
Ansiedad Fóbica	.07*	-.23**	.34**	-.16*
Ideación Paranoide	.08*	-.10**	.45**	.05
Psicoticismo	.03	-.21**	.48**	.07*
Índice Severidad Global	.07*	-.25**	.64**	.00
Total Síntomas Positivos	.08*	-.23**	.65**	.02
Índice Malestar Positivo	.08*	-.21**	.67**	.02

Estudios de confiabilidad

Los coeficientes de confiabilidad globales obtenidos desde la teoría clásica (Alfa de Cronbach) y la TRI (Fiabilidad Marginal) fueron aceptables para todas las escalas (Tabla 1). Aun así, resulta evidente que Sinceridad presentó coeficientes más bajos que el resto de los factores.

En los gráficos de la Figura 1 se pueden comparar las funciones de información y error estándar de estimación correspondientes a cada uno de los factores evaluados con el EPQ-RS. La escala N sobresale por aportar la mayor información predominantemente entre los valores medios del rasgo que mide. Las escalas P y E brindan información considerablemente más baja para la medición de sus respectivos constructos. La máxima precisión de P se alcanza para evaluar a sujetos con valores medio-bajos de Psicoticismo. En cambio, la escala E resulta más precisa cuando los evaluados se localizan en niveles medio-altos de Extraversión. Por último, la escala S mostró una función de información relativamente plana en todo el recorrido del rasgo latente.

Como se aprecia en la función del error estándar de S, su curva presenta valores de estimación superiores a .50 (equivalente a una confiabilidad clásica mínima de 0.75) para todos los niveles de Sinceridad.

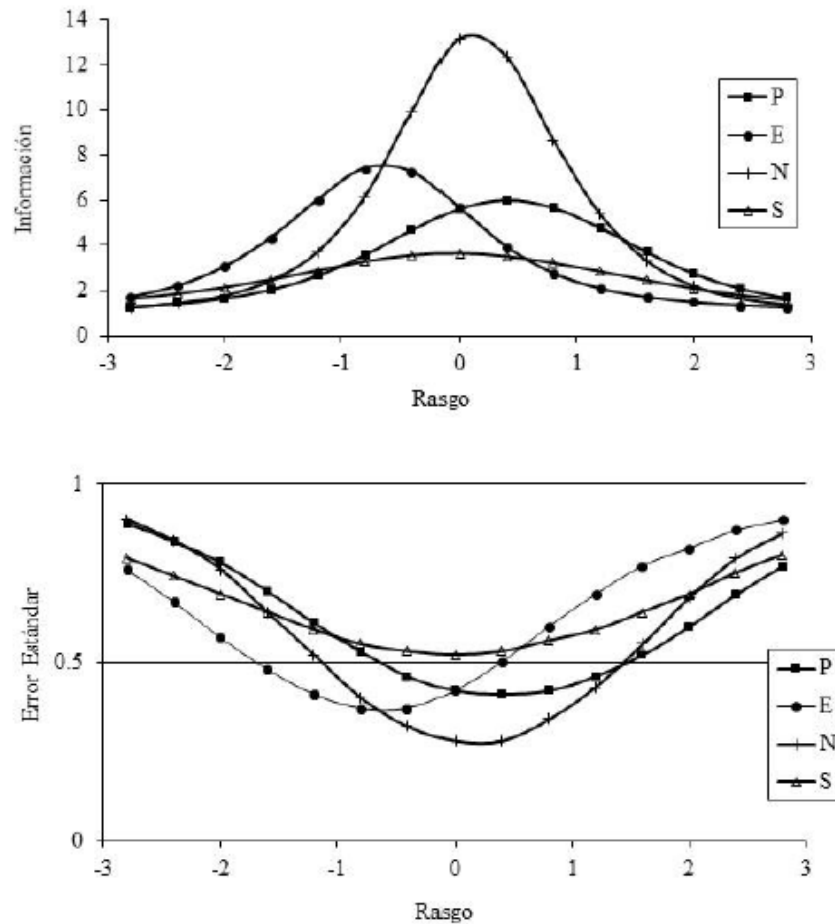


Figura 1. Funciones de Información y Error Estándar.

Discusión

En este estudio se han revisado y ampliado las propiedades psicométricas de la adaptación argentina del EPQ-RS (versión abreviada) de Squilace et al. (2013). Desde una perspectiva instrumental, las evidencias de validez y confiabilidad obtenidas ofrecen garantías de calidad suficientes para la aplicación de este instrumento en el contexto local.

A nivel teórico, la revisión de la estructura interna del EPQ-RS ha permitido replicar el modelo con las tres dimensiones básicas de la personalidad desarrolladas por Eysenck y la escala de Sinceridad del cuestionario. Este hallazgo está en línea con lo registrado para la prueba original

y posteriores adaptaciones a otras culturas (Almiro et al., 2016; Alexopoulos & Kalaitzidis, 2004; Aluja et al., 2003; Eysenck et al., 1985). No fueron registrados aquí los problemas psicométricos señalados recurrentemente en la escala P al ser analizada desde la perspectiva clásica (e.g. Colledani, Anselmi et al., 2018; Forrest, Lewis, & Shevlin, 2000; Francis et al., 2006). Tanto los pesos factoriales como la confiabilidad medida a través de los coeficientes de consistencia interna adoptaron valores comparables al resto de las dimensiones.

El ajuste del ML2p demostró que todos los ítems son indicadores útiles para predecir los niveles del rasgo en las dimensiones que operacionalizan. La capacidad discriminativa de los ítems resultó entre moderada y alta, aunque, comparativamente, L registró valores más bajos de α . En cuanto a los ítems inversos, no se observó una tendencia clara en los valores estimados de los parámetros que los distinguiera de los ítems redactados en sentido directo. No obstante, los ítems inversos de cada dimensión son escasos como para alcanzar interpretaciones concluyentes. El desbalance entre ítems directos e inversos se corresponde con el proceso de depuración efectuado en estudios previos (Squillace et al., 2013) en los que se ha priorizado la selección de ítems con mejores propiedades psicométricas sin tomar recaudos sobre el control del estilo de respuesta aquiescente.

En consonancia con los hallazgos de otros autores (Escorial & Navas, 2007), los ítems del EPQ-RS no registraron DIF en función género. Esto permite concluir que cualquier diferencia observada en las mediciones de mujeres y varones es real y no ha sido generada artificialmente por el instrumento. Es decir, las mediciones son invariantes en función del género.

Uno de los resultados más relevantes que se desprenden de la modelización con la TRI remite al análisis de la localización de los parámetros b en los continuos de cada una de las dimensiones. En una situación ideal es esperable que los parámetros b oscilen entre -2 y 2 para garantizar que la escala presente capacidad discriminativa en un rango considerable de niveles del rasgo (Baker & Kim, 2017). Sin embargo, estos parámetros no se distribuyeron de manera pareja en el recorrido de los respectivos continuos. La situación más extrema se observó en la escala E, cuyos ítems se concentraron mayoritariamente en los niveles medio-bajos del rasgo. En las escalas P y N, en cambio, los ítems tendieron a localizarse por encima de la media del rasgo.

La concentración de los parámetros b en rangos acotados del rasgo tiene como consecuencia un problema que permanece invisibilizado en el análisis global de la confiabilidad que se realiza desde la teoría clásica. Como se demuestra a partir de la FIT, la precisión de las escalas varía en el continuo de los constructos. Concretamente, la escala E aporta más información para la medición de valores cercanos al polo Introversión, en tanto que su error aumenta hacia los niveles próximos al extremo Extraversión. La localización de los b en un extremo del rasgo es habitual en la medición de variables psicopatológicas en donde sólo es clínicamente relevante un polo mientras que el otro extremo del rasgo es poco informativo (Ferrando &

Navarro-González, 2018; Reise & Rodriguez, 2016). En este sentido, cabría discutir si se debe a una falencia de la medición o si efectivamente las variables del modelo PEN no son bipolares y pueden ser consideradas cuasi-rasgos en los que sólo es interpretable un extremo (Reise & Waller, 2009).

Las posibles soluciones para medir con precisión un rango más amplio del rasgo no son simples y requieren sacrificar algún aspecto valorado de la versión actual. En principio, se podrían incorporar ítems que evalúen los niveles del rasgo no cubiertos, aunque resulta complejo anticipar la localización que tendrá un ítem en el rasgo (Abal et al., 2010; Reise & Rodriguez, 2016) y, además, aumentaría la extensión de la escala. En cambio, si se decidiera mantener fija la cantidad de ítems reemplazando algunos de los elementos por otros que discriminen en niveles del rasgo que no son cubiertos con la versión actual se podría llevar a FITs aplanadas y poco informativas.

Otra mejora potencial se alcanzaría al modificar el formato de respuesta de los ítems con un aumento en la cantidad de opciones. Estudios empíricos y de simulación sostienen que una escala tipo Likert con aproximadamente cuatro categorías podría optimizar la calidad psicométrica de un instrumento optimizando los indicadores de confiabilidad y validez (e.g. Abal, Auné, Attorresi, 2018; Muñiz, García-Cueto, & Lozano, 2005; Lozano et al., 2008). En efecto, la modelización de ítems dicotómicos revela que cada elemento brinda información en torno a un único nivel del rasgo latente (que coincide con el valor del parámetro b). Por el contrario, los modelos de la TRI para ítems politómicos, como los Modelos Respuesta Graduada (Samejima, 2016) o Crédito Parcial Generalizado (Muraki & Muraki, 2016), ofrecen una mayor cantidad de parámetros de localización por ítem que se distribuyen en un rango más amplio del espectro del rasgo. De esta manera, se aprovecha la información contenida en cada una de las categorías de respuesta de los ítems sin tener que aumentar la cantidad de elementos.

Las evidencias de validez basadas en la relación con otras variables vincularon de manera coherente las dimensiones del Modelo PEN con los criterios externos considerados. En términos generales, las asociaciones del EPQ-RS con el BIN son más elevadas porque ambas medidas fueron concebidas para la medición de rasgos de la personalidad normal. En cambio, las correlaciones con SCL-90-R son en su mayoría más bajas presumiblemente porque este inventario recoge información sobre dimensiones sintomáticas que se relacionan conceptualmente a rasgos desadaptativos. Las correlaciones moderadas de N con sintomatología psicológica e índices de riesgo del SCL-90-R corroboran su utilidad en el ámbito clínico y de la salud (Almiro et al., 2016; Jylhä & Isometsä, 2006). Además, se encuentran en línea con planteos actuales que reconocen el valor de N como un factor de riesgo amplio e inespecífico para el desarrollo y mantenimiento de múltiples psicopatologías (e.g. Castellanos-Ryan et al., 2016).

En este trabajo se efectuó un análisis de ítems con la TRI aplicando modelos unidimensionales por separado a cada una de las cuatro escalas del cuestionario. Este procedimiento se considera aceptable en la medida en que el AFC corroboró una estructura con factores incorrelados (Toland, Sulis, Giambona, Porcu, & Campbell, 2017) y ha sido usado tradicionalmente para estos fines (e.g. Colledani, Anselmi et al., 2018; Ebesutani et al., 2012; Nieto et al., 2017). Sin embargo, también se han reportado evidencias de que la aplicación de modelos multidimensionales de la TRI podría proporcionar estimaciones más precisas de los parámetros de los ítems y de los niveles del rasgo (Makransky, Mortensen, & Glas, 2013; Wiberg, 2012). Entendiendo que este aspecto podría considerarse una limitación del presente estudio, acciones futuras se encaminarán en profundizar el análisis de la calidad psicométrica de los ítems a partir de modelos multidimensionales.

Agradecimientos

Proyectos de la Universidad de Buenos Aires (UBACyT2018 20020170100200BA y UBACyT2020 20020190200156BA) y Agencia Nacional de Promoción Científica y Tecnológica (PICT-2017-3226).

Referencias

- Abal, F.J.P., Auné, S.E., & Attorresi, H.F. (2015). Relación entre la Deseabilidad Social y los parámetros del Modelo Logístico de tres parámetros. *Psicología, Conocimiento y Sociedad*, 5(1), 5-24.
- Abal, F. J. P., Auné, S. E., & Attorresi, H. F. (2018). Variación de la escala Likert en el test de Utilidad de la Matemática. *Interacciones*, 4(3), 177-189. <https://doi.org/10.24016/2018.v4n3.134>
- Abal, F. J. P., Auné, S. E., & Attorresi, H. F. (2019). Construcción de un banco de ítems de Facetas de Neuroticismo para el desarrollo de un test adaptativo. *Psicodebate*, 19(1), 31-50. <https://doi.org/10.18682/pd.v1i1.854>
- Abal, F. J. P., Lozzia, G. S., Aguerri, M. E., Galibert, M. S., & 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. <https://doi.org/10.35670/1667-4545.v19.n1.23877>
- Alexopoulos, D. S., & Kalaitzidis, I. (2004). Psychometric properties of Eysenck Personality Questionnaire-Revised (EPQ-R) Short Scale in Greece. *Personality and Individual Differences*, 37(6), 1205-1220. <https://doi.org/10.1016/j.paid.2003.12.005>
- 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. <https://doi.org/10.1016/j.paid.2015.08.050>

- Aluja, A., García, O., & García, F. A. (2003). Psychometric analysis of the Revised Eysenck Personality Questionnaire Short Scale. *Personality and Individual Differences, 35*(2), 449-460. [https://doi.org/10.1016/S0191-8869\(02\)00206-4](https://doi.org/10.1016/S0191-8869(02)00206-4)
- Baker, F. B., & Kim, S. H. (2017). *The Basics of Item Response Theory Using R*. Cham, Suiza: Springer International Publishing. <https://doi.org/10.1007/978-3-319-54205-8>
- Bauermeister, S., & Gallacher, J. (2020). A psychometric evaluation of the 12-item EPQ-R neuroticism scale in 502,591 UK Biobank participants using item response theory (IRT). *BioRxiv 741249*, <https://doi.org/10.1101/741249>
- Bech, P. (2016). Neuroticism (Eysenck's Theory). En V. Zeigler-Hill & T. K. Shackelford (Eds.), *Encyclopedia of personality and individual differences* (pp. 1-4). Cham: Springer International Publishing AG. https://doi.org/10.1007/978-3-319-28099-8_1094-1
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. En F. M. Lord y M. R. Novick (Eds.). *Statistical Theories of Mental Test Scores*. Reading, MA: Addison Wesley.
- Bowden, S. C., Saklofske, D. H., van de Vijver, F. J. R., Sudarshan, N. J., & Eysenck, D. B. G. (2016). Cross-cultural measurement invariance of the Eysenck Personality Questionnaire across 33 countries. *Personality and Individual Differences, 103*, 53-60. <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: Routledge.
- Cai, L., Thissen, D., & du Toit, S.H.C. (2011). IRTPRO 4.2 for Windows [Computer software]. Skokie, IL: Scientific Software International, Inc.
- Calderón Garrido, C., Navarro González, D., Lorenzo Seva, U., & Ferrando Piera, P. J. (2019). Multidimensional or essentially unidimensional? A multi-faceted factor-analytic approach for assessing the dimensionality of tests and items. *Psicothema, 31*(4), 450-457. <https://doi.org/10.7334/psicothema2019.153>
- Castellanos-Ryan, N., Brière, F.N., O'Leary-Barrett, M., Banaschewski, T., Bokde, A., & Bromberg, U. (2016). The structure of psychopathology in adolescence and its common personality and cognitive correlates. *Journal of Abnormal Psychology, 125*(8), 1039-1052. <http://dx.doi.org/10.1037/abn0000193>
- Casullo, M. (2004). Síntomas psicopatológicos en adultos urbanos. *Psicología y Ciencia Social, 6*(1), 49-57.
- Chen, W., & Thissen, D. (1997). Local dependence indices for item pairs using item response theory. *Journal of Educational and Behavioral Statistics, 22*, 265-289. <https://doi.org/10.2307/1165285>
- 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*, 1834. <https://doi.org/10.3389/fpsyg.2018.01834>

- Colledani, D., Anselmi, P., & Robusto, E. (2019). Development of a new abbreviated form of the Eysenck Personality Questionnaire-Revised with multidimensional item response theory. *Personality and Individual Differences, 149*, 108–117. <https://doi.org/10.1016/j.paid.2019.05.044>
- 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. <https://doi.org/10.1016/j.paid.2017.08.037>
- De Ayala, R. (2009). *The Theory and Practice of Item Response Theory*. New York: Guilford Press.
- Derogatis, L. (1994). *SCL-90-R. Symptom Checklist-90-R. Administration, Scoring and Procedures Manual*. Minneapolis: National Computer System.
- DeVellis, R. F. (2016). *Scale development: Theory and applications*. Newbury Park: Sage publications.
- Dominguez, S., Villegas, G., Yauri, S., Aravena, S., & Ramírez, F. (2013). Análisis psicométrico preliminar de la forma corta del EPQ-R en una muestra de estudiantes universitarios de Lima Metropolitana. *Avances en Psicología, 21*(1), 73-82. <https://doi.org/10.33539/avpsicol.2013.v21n1.308>
- Ebesutani, C., Regan, J., Smith, A., Reise, S., Higa-McMillan, C., & Chorpita, B. F. (2012). The 10-Item Positive and Negative Affect Schedule for Children, Child and Parent Shortened Versions: Application of Item Response Theory for More Efficient Assessment. *Journal of Psychopathology and Behavioral Assessment, 34*(2), 191–203. <https://doi.org/10.1007/s10862-011-9273-2>
- Escorial, S., & Navas, M. J. (2007). Analysis of the gender variable in the Eysenck Personality Questionnaire–revised scales using differential item functioning techniques. *Educational and Psychological Measurement, 67*(6), 990-1001. <https://doi.org/10.1177/0013164406299108>
- Eysenck, H. J., & Eysenck, M. W. (1985). *Personality and Individual Differences: A natural science approach*. New York: Plenum.
- Eysenck, H. J. (1947). *Dimensions of Personality*. Londres: Routledge y Kegan Paul.
- Eysenck, H. J. (1952). *The scientific study of personality*. Londres: Routledge y Kegan Paul.
- Eysenck, H. J. (1990). Genetic and environmental contributions to individual differences: the three major dimensions of personality. *Journal of Personality, 58*(1), 245-261. <https://doi.org/10.1111/j.1467-6494.1990.tb00915.x>
- Eysenck, H. J., & Eysenck, S. B. G. (1975). *Manual of the Eysenck Personality Questionnaire*. Londres: Hodder and Stoughton.
- Eysenck, H. J., Wilson, G. D., & Jackson, C. J. (1996). *Manual of the Eysenck Personality Profiler (Short)*. Guilford: 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. <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. [http://dx.doi.org/10.1016/0191-8869\(85\)90026-1](http://dx.doi.org/10.1016/0191-8869(85)90026-1)
- Ferrando, P. J. (2003). The accuracy of the E, N and P trait estimates: An empirical study using the EPQ-R. *Personality and Individual Differences*, 34(4), 665–679. [https://doi.org/10.1016/S0191-8869\(02\)00053-3](https://doi.org/10.1016/S0191-8869(02)00053-3)
- Ferrando, P. J., & Lorenzo-Seva, U. (2017). Program FACTOR at 10: Origins, development and future directions. *Psicothema*, 29, 236-240. <https://doi.org/10.7334/psicothema2016.304>
- Ferrando, P. J., & Lorenzo-Seva, U. (2018). Assessing the Quality and Appropriateness of Factor Solutions and Factor Score Estimates in Exploratory Item Factor Analysis. *Educational and Psychological Measurement*, 78, 762-780. <https://doi.org/10.1177/0013164417719308>
- Ferrando, P. J., & Navarro-González, D. (2018). Assessing the quality and usefulness of factor-analytic applications to personality measures: A study with the statistical anxiety scale. *Personality and Individual Differences*, 123, 81–86. <https://doi.org/10.1016/j.paid.2017.11.014>.
- Forrest, S., Lewis, C.A., & Shevlin, M. (2000). Examining the factor structure and differential functioning of the Eysenck Personality Questionnaire Revised-Abbreviated. *Personality and Individual Differences*, 29(3), 579-588. [https://doi.org/10.1016/S0191-8869\(99\)00220-2](https://doi.org/10.1016/S0191-8869(99)00220-2)
- 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 USA and Australia. *Personality and Individual Differences*, 13(4), 443–449. [https://doi.org/10.1016/0191-8869\(92\)90073-X](https://doi.org/10.1016/0191-8869(92)90073-X)
- Francis, L. J., Lewis, C. A., & Ziebertz, H. (2006). The short-form revised Eysenck personality Questionnaire (EPQ-S): A German edition. *Social Behavior and Personality: An International Journal*, 34(2), 197–204. <https://doi.org/10.2224/sbp.2006.34.2.197>
- 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, y D. H. Saklofske (Eds.), *The Sage handbook of personality theory and assessment: Vol. 2 — Personality measurement and testing* (pp. 199–218). Los Angeles, CA: Sage. <https://doi.org/10.4135/9781849200479.n10>
- González, I., Déjean, S., Martin, P., & Baccini, A. (2008). CCA: An R Package to Extend Canonical Correlation Analysis. *Journal of Statistical Software*, 23(12), 1 - 14. <https://doi.org/10.18637/jss.v023.i12>
- Jylhä, P., & Isometsä, E. (2006). *The relationship of neuroticism and extraversion to symptoms of anxiety and depression in the general population*. *Depression and Anxiety*, 23(5), 281–289. <https://doi.org/10.1002/da.20167>
- Katz, Y. J., & Francis, L. J. (1991). The dual nature of the EPQ Lie scale? A study among university students in Israel. *Social Behavior and Personality: An International Journal*, 19(4), 217–222. <https://doi.org/10.2224/sbp.1991.19.4.217>

- Knežević, G., Lazarević, L. B., Purić, D., Bosnjak, M., Teovanović, P., Petrović, B., & Opačić, G. (2019). Does Eysenck's personality model capture psychosis-proneness? A systematic review and meta-analysis. *Personality and Individual Differences, 143*, 155-164. <https://doi.org/10.1016/j.paid>.
- Lewis, C. A. & Musharraf, S. (2014). The short form Eysenck Personality Questionnaire-Revised (EPQR-S) and the revised abbreviated Eysenck Personality Questionnaire (EPQR-A): Urdu translations. *Journal of the Pakistan Medical Association 64*(2), 225-226.
- Lewis, C. A., Francis, L. J., Shevlin, M., & 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*, 179-185. <https://doi.org/10.1027//1015-5759.18.2.179>
- Lloret-Segura, S., Ferreres-Traver, A., Hernández-Baeza, A., & Tomás-Marco, I. (2014). El análisis factorial exploratorio de los ítems: una guía práctica, revisada y actualizada. *Anales de Psicología, 30*(3), 1151-1169. <http://dx.doi.org/10.6018/analesps.30.3.199361>
- Lozano, L.M., García-Cueto, E. & Muñiz, J. (2008). Effect of the Number of Response Categories on the Reliability and Validity of Rating Scales. *Methodology, 4* (2), 73–79. <https://doi.org/10.1027/1614-2241.4.2.73>
- Makransky, G., Mortensen, E. L., & Glas, C. A. (2013). Improving personality facet scores with multidimensional computer adaptive testing: An illustration with the NEO PI-R. *Assessment, 20*, 3-13. <https://doi.org/10.1177/1073191112437756>
- Massey, A. (1980). The Eysenck Personality Inventory Lie Scale: Lack of insight or ...? *The Irish Journal of Psychology, 4*(3), 172-174.
- McCrae, R. R. & Costa P. T. (2010). *NEO Inventories professional manual*. Odessa, FL: Psychological Assessment Resources.
- Min, H., Zickar, M. & Yankov, G. (2018). Understanding item parameters in personality scales: An explanatory item response modeling approach, *Personality and Individual Differences, 128*, 1-6. <https://doi.org/10.1016/j.paid.2018.02.012>
- 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. <https://doi.org/10.1016/j.paid.2016.04.009>
- Muñiz, J. (2018). *Introducción a la psicometría: teoría clásica y TRI*. Madrid: Pirámides.
- Muñiz, J., García-Cueto, E. & Lozano, L. M. (2005). Item format and the psychometric properties of the Eysenck Personality Questionnaire. *Personality and Individual Differences, 38* (1), 61-69. <https://doi.org/10.1016/j.paid.2004.03.021>
- Muraki, E. & Muraki, M. (2016). Generalized Partial Credit Model. En W. J. van der Linden (Ed.), *Handbook of item response theory, Volume One: Models* (pp. 127-135). Boca Raton, FL: CRC Press
- Muthén, L. & Muthén, B. (2010). *Mplus User's Guide, 6th Edn*. Los Angeles, CA: Muthén & Muthén.

- Nieto, M. D., Abad, F. J., Hernández-Camacho, A., Garrido, L. E., Barrada, J. R., Aguado, D. & Olea, J. (2017). Calibrating a new item pool to adaptively assess the Big Five. *Psicothema*, 29(3), 390-395. <https://doi.org/10.7334/psicothema2016.391>
- Orlando, M., & Thissen, D. (2003). Further Investigation of the Performance of $S-\chi^2$: An Item Fit Index for Use With Dichotomous Item Response Theory Models. *Applied Psychological Measurement*, 27(4), 289- 298. <https://doi.org/10.1177/0146621603027004004>
- Reise, S. & Waller, N. (2009). Item response theory and clinical measurement. *Annual Review of Clinical Psychology*, 5, 27-48. <https://doi.org/10.1146/annurev.clinpsy.032408.153553>
- Reise, S. P. & Revicki, D. A. (2015). *Handbook of Item Response Theory Modeling Applications to Typical Performance Assessment*. Nueva York: Routledge.
- Reise, S. P. & Waller, N. G. (1990). Fitting the two-parameter model to personality data. *Applied Psychological Measurement*, 14(1), 45-58. <https://doi.org/10.1177/014662169001400105>
- Reise, S. P., & Rodriguez, A. (2016). Item response theory and the measurement of psychiatric constructs: some empirical and conceptual issues and challenges. *Psychological Medicine*, 46(10), 2025–2039. <https://doi.org/10.1017/S0033291716000520>
- Reise, S. P., & Waller, N. G. (2009). Item response theory and clinical measurement. *Annual review of clinical psychology*, 5, 27-48. <https://doi.org/10.1146/annurev.clinpsy.032408.153553>
- Revelle, W. (2016). Hans Eysenck: Personality theorist. *Personality and Individual Differences*, 103, 32–39. <https://doi.org/10.1016/j.paid.2016.04.007>
- Samejima, F. (2016). Graded response models. En W. J. van der Linden (Ed.), *Handbook of item response theory, Volume One: Models* (pp. 95-108). Boca Raton, FL: CRC Press
- Sanchez, 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.
- Squillace, M., Picón Janeiro, J., & Schmidt, V. (2013). Adaptación local del Cuestionario Revisado de Personalidad de Eysenck (Versión abreviada). *Evaluar*, 13, 19-37. <https://doi.org/10.35670/1667-4545.v13.n1.6794>
- Timmerman, M. E., & Lorenzo-Seva, U. (2011). Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychological methods*, 16(2), 209. <https://doi.org/10.1037/a0023353>
- 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. <https://doi.org/10.4103/0972-6748.57854>
- Toland, M. (2013). Practical guide to conducting an item response theory analysis. *The Journal of Early Adolescence*, 34(1), 120-151. <https://doi.org/10.1177/0272431613511332>
- Toland, M. D., Sulis, I., Giambona, F., Porcu, M., & Campbell, J. M. (2017). Introduction to bifactor polytomous item response theory analysis. *Journal of School Psychology*, 60, 41–63. <https://doi.org/10.1016/j.jsp.2016.11.001>

Wiberg, M. (2012). Can a multidimensional test be evaluated with unidimensional item response theory? *Educational Research and Evaluation*, 18(4), 307–320. <https://doi.org/10.1080/13803611.2012.670416>