

Big Five Inventory abreviado con la teoría de respuesta al ítem en una muestra argentina

Big Five Inventory Abbreviated from the Item Response Theory

Recibido: mayo21/2022; **Concepto de evaluación:** febrero 21/2024; **Aceptado:** mayo 15/2024

Facundo Juan Pablo Abal¹

Universidad de Buenos Aires y Consejo Nacional de Investigaciones Científicas y Técnicas,
Buenos Aires, Argentina
ORCID: <http://orcid.org/0000-0001-7023-5380>

Juan Franco Sánchez González

Universidad de Buenos Aires, Buenos Aires, Argentina
ORCID: <https://orcid.org/0000-0003-1993-4981>

Horacio Félix Attorresi

Universidad de Buenos Aires, Buenos Aires, Argentina
ORCID: <http://orcid.org/0000-0002-3027-1069>

Resumen

Aunque se han elaborado numerosas versiones abreviadas del Big Five Inventory (BFI), las propiedades psicométricas obtenidas a nivel mundial presentan inconsistencias insoslayables. En este trabajo se propone reducir este inventario a partir de la teoría de respuesta al ítem como soporte metodológico, con una muestra de 987 adultos (55.3 % mujeres; $M_{edad} = 38$; $DE_{edad} = 13.9$) residentes en el área metropolitana de Buenos Aires, Argentina. La reducción de ítems se realizó de manera progresiva por medio del modelo de respuesta graduada —con los supuestos de unidimensionalidad de cada escala por separado, independencia local de los ítems, y ajuste al modelo—, con lo cual se alcanzó una versión de 20 ítems libres de funcionamiento diferencial según el género. Las correlaciones entre las escalas originales y las reducidas fueron superiores a .73. Adicionalmente, se replicó la estructura del modelo pentafactorial con un análisis factorial confirmatorio y se aportaron evidencias de validez basadas en la relación con tests que miden sintomatología y facetas de neuroticismo. Como resultado, los índices de consistencia interna globales de la versión abreviada mostraron valores aceptables, pero las funciones de información revelaron que las escalas disminuyen su precisión en los niveles altos de los rasgos. Se recomienda ensayar nuevos ítems para disminuir el error en estos niveles en futuros estudios.

Palabras clave

modelo de los cinco factores, modelo de los cinco grandes, teoría de respuesta al ítem, evaluación de la personalidad.

Abstract

Although numerous abbreviated versions of the Big Five Inventory (BFI) have been developed, the psychometric properties obtained worldwide present unavoidable inconsistencies. In this paper, it is proposed to reduce this inventory based on item response theory as methodological support, with a sample of 987 adults (55.3% women; $M_{age} = 38$; $SD_{age} = 13.9$) residing in the metropolitan area of Buenos Aires, Argentina. The reduction of items was carried out progressively through the graded response model—with the assumptions of unidimensionality of each scale separately, local independence of the items, and fit to the model—, which resulted in a 20-item version free of differential functioning according to gender. The correlations between the original and reduced scales were greater than .73. Additionally, the structure of the five-factor model was replicated with a confirmatory factor analysis and evidence of validity was provided based on the relationship with tests that measure symptomatology and facets of neuroticism. As a result, the global internal consistency indices of the abbreviated version showed acceptable values, but the information functions revealed that the scales decrease their precision at high levels of the traits. It is recommended to test new items to reduce the error at these levels in future studies.

Keywords

five factor model, big five model, item response theory, personality assessment.

Cómo citar [APA]:

Abal, F. J. P., Sánchez, J. F., & Attorresi, H. F. (2024). Big Five Inventory abreviado con la teoría de respuesta al ítem en una muestra argentina. *Acta Colombiana de Psicología*, 27(2), 57-74. <https://doi.org/10.14718/ACP.2024.27.2.4>

.....

1 Datos de contacto: Instituto de Investigaciones de la Facultad de Psicología de la Universidad de Buenos Aires, Lavalle 2353, Ciudad de Buenos Aires, C1052AAA. Correo electrónico: afjp79@gmail.com

Nota del autor: Proyecto UBACyT2023, código 20020220400196BA, de la Universidad de Buenos Aires, Argentina.

Conflicto de intereses: Los autores expresamos que no presentamos conflictos de interés al redactar el manuscrito.

Introducción

En el marco de las teorías de los rasgos, el modelo de los cinco grandes —*Big Five Model*, BFM— (Digman, 1990; Goldberg, 1993; McCrae & Costa, 2003) es actualmente reconocido como la taxonomía más apropiada para sistematizar los diferentes dominios de la personalidad normal.

Específicamente, el BFM propone la descripción de la personalidad por medio de cinco rasgos generales cuyas combinaciones darían lugar a la variabilidad individual manifiesta en cada ser humano. Más allá de algunas diferencias mínimas entre las distintas corrientes que integran este modelo, estos factores son conocidos como *extraversión* (vs. *introversión*), *agradabilidad* (vs. *antagonismo*), *responsabilidad* (vs. *falta de responsabilidad*), *neuroticismo* (vs. *estabilidad emocional*) y *apertura a la experiencia* (vs. *cerrado a la experiencia*).

Este modelo ha logrado convertirse en hegemónico para la psicología de la personalidad, en parte porque ha conseguido unificar los aportes de distintos teóricos —como Cattell, Guilford y Eysenck— (Colom, 2018; John, 2021), así como por su vinculación conceptual con la perspectiva dimensional propuesta en el *Manual diagnóstico y estadístico de los trastornos mentales* (quinta edición) —conocido como DSM-5— en relación con los trastornos de personalidad (Widiger et al., 2015).

Entre sus características, el estudio de la fundamentación biológica y bioquímica de las diferencias entre los factores (Allen & DeYoung, 2017), así como la replicación de la estructura penta factorial en diferentes culturas (McCrae, 2017), constituyen las bases empíricas que fundamentan el carácter universal del modelo.

Con el fin de operacionalizar el BFM, distintos investigadores han creado diversos instrumentos, pero el Big Five Inventory (John et al., 1991) es uno de los más elegidos cuando se requiere de una medición rápida (44 ítems) y confiable de los cinco dominios. De hecho, esta prueba cuenta con adaptaciones a múltiples idiomas (Alansari, 2016; Carciofo et al., 2016; Ubbiali et al., 2013; Rammstedt, 1997), lo que refleja la magnitud de su trascendencia en el plano internacional, y recientemente se encuentran adaptaciones de investigadores de América Latina (Dominguez-Lara et al., 2018; Reyes Zamorano et al., 2014; Salgado et al., 2016) que evidencian la vigencia del BFI en la región.

Ahora bien, en los últimos años se ha visto acrecentado el interés por mejorar las medidas derivadas

del BFI, sobre todo en dos líneas: (a) en consolidar la estructura jerárquica del modelo, al identificar las subdimensiones más relevantes de cada dominio y así alcanzar una mejor representatividad del contenido —objetivo con el cual se desarrolló el Big Five Inventory 2 (BFI-2; Soto & John, 2017)—; y (b) en conseguir formas aún más breves que el BFI para optimizar la medición de los dominios —línea clave para el presente estudio— (Minkov et al., 2019; Rammstedt et al., 2020; Soto & John, 2019), pues los instrumentos breves muestran potentes ventajas cuando las condiciones de evaluación no permiten o no requieren recolectar información exhaustiva sobre la personalidad; aspecto que se observa principalmente en estudios a gran escala en los que se busca reducir los tiempos de administración, o cuando la personalidad no es el objetivo principal de la investigación sino una de las variables de control.

Entre las versiones breves del BFI que han alcanzado más popularidad se encuentran el BFI-10 (Rammstedt, 2007; Rammstedt & John, 2007) y el BFI-15 (Gerlitz & Schupp, 2005), los cuales utilizan solo dos y tres ítems respectivamente para la medición de cada dominio.

No obstante, las adaptaciones en estudios internacionales del BFI-10 presentan resultados dispares, pues mientras que algunos autores publican propiedades psicométricas óptimas (Courtois et al., 2020; Guido et al., 2015; Rammstedt et al., 2013; Rammstedt et al., 2020), otros trabajos abren cuestionamientos importantes respecto a distintos aspectos, como la dificultad para replicar la estructura de cinco factores esperables a nivel teórico (Balgiu, 2018; Brown & Sotardi, 2019; Dominguez-Lara & Merino-Soto, 2018), posibles predicciones espurias con indicadores de salud (Chapman & Elliot, 2019; Weiss & Costa, 2014), e indicadores de confiabilidad deficitarios (Carciofo et al., 2016; Ludeke & Larsen, 2017; Sleep et al., 2021). Asimismo, a pesar de que cuenta con más ítems y más cantidad de opciones de respuesta que el BFI-10, la consistencia interna de las escalas no mejora para el BFI-15, pues las escalas de *agradabilidad*, *apertura a la experiencia* y *responsabilidad* reportaron valores de alfa de Cronbach iguales o inferiores a .60 en población alemana —para la que fue desarrollada esta versión— (Hahn et al., 2012).

Esta ausencia de resultados consistentes entre las adaptaciones de las versiones breves del BFI deja en evidencia, en principio, la necesidad de revisar los

criterios aplicados originalmente para reducir el instrumento, a lo cual se le suma el potencial impacto de las variaciones culturales en los contenidos que representan los ítems seleccionados. En este sentido, parece razonable proponer una nueva selección de ítems del BFI basada en criterios empíricos que se deriven de un análisis psicométrico pormenorizado.

Respecto al análisis psicométrico de las pruebas, el crecimiento que ha tenido la aplicación de la teoría de respuesta al ítem (TRI) en tests de comportamiento típico ha permitido importantes avances tanto para el desarrollo de nuevos instrumentos como para la revisión de la calidad psicométrica de pruebas validadas en el marco de la teoría clásica de los test (TCT) (Reise & Revicki, 2015; Thomas, 2019).

Una de las ventajas más relevantes de la TRI respecto a la perspectiva clásica es la posibilidad de estudiar el comportamiento de los ítems mediante el ajuste a un modelo

que explica la probabilidad de respuesta para cada una de las opciones en función del nivel de rasgo del sujeto (Muñiz, 2018). Esto permite realizar un análisis de ítems más riguroso y exhaustivo que el que se alcanza desde la teoría clásica y, por esta razón, es usado con frecuencia para decidir qué elementos conviene conservar cuando se busca generar versiones reducidas de los instrumentos (Chiesi et al., 2018; Chio et al., 2018; Colledani et al., 2018; Maples-Keller et al., 2019; Sekely et al., 2018; Stein et al., 2019).

En suma, el objetivo de este trabajo es analizar, a partir de la TRI, la calidad psicométrica de los ítems del BFI, y con ello alcanzar una versión reducida que optimice la medición de las dimensiones de la personalidad propuestas por el BFM. De esta manera, se busca brindar una nueva medida breve que contemple las características de la población local, basada en un modelo teórico reconocido a nivel internacional y con evidencias de validez y confiabilidad derivadas de una metodología de vanguardia.

Método

Tipo de estudio

Según los criterios clasificatorios de Ato et al. (2013), se llevó a cabo un estudio instrumental, pues el objetivo principal de la investigación fue analizar las propiedades psicométricas de un instrumento de medida.

Participantes

Se contó con la participación de 987 adultos de 18 a 81 años ($M = 38$, $DE = 13.9$), residentes en el área metropolitana de Buenos Aires, Argentina, seleccionados mediante un muestreo no probabilístico por accesibilidad. Específicamente, el 55.3 % indicó que se identificaba con el género femenino; el 79.5 % afirmó estar empleado en trabajos de medio tiempo o tiempo completo; la mayoría completó el nivel educativo secundario (52.2 %) y universitario (27.4 %); el 47.7 % registró que su estado civil era casado/a (o en unión civil); el 44 % manifestó tener al menos un hijo; y el 69.6 % afirmó tener un nivel socioeconómico medio.

Instrumentos

Big Five Inventory (John et al., 1991)

El BFI es un inventario compuesto por 44 ítems con formato de respuesta politómica de cinco opciones (desde

“completo desacuerdo” a “completo acuerdo”), en el que las dimensiones de *extraversión* y *neuroticismo* se miden a partir de ocho ítems, mientras que las de *agradabilidad* y *responsabilidad* cuentan con nueve ítems, y la de *apertura a la experiencia*, con 10 ítems.

Este instrumento cuenta con una versión en español desarrollada por Benet-Martínez y John (1998), pero se escogió la adaptación de Castro Solano (2002) por considerarla más ajustada a las características lingüísticas del medio local. Los ítems tienen un encabezado general (“Yo me veo a mí mismo/a como alguien...”) que se complementa con las sentencias específicas de cada elemento (p. ej., 15. “... que es ingenioso/a”).

Banco de ítems de neuroticismo (Abal et al., 2019)

Este instrumento consta de 54 ítems con formato de respuesta politómica (“en desacuerdo”, “ligeramente en desacuerdo”, “ligeramente de acuerdo” y “de acuerdo”) que operacionalizan las seis facetas propuestas por McCrae y Costa (2010): *ansiedad*, *hostilidad*, *depresión*, *autoconciencia*, *impulsividad* y *vulnerabilidad*. El instrumento cuenta con evidencias de validez basadas en la estructura interna, sus ítems están calibrados con el modelo de respuesta graduada (MRC) de la TRI, y los coeficientes de consistencia interna —alfa de Cronbach— en este estudio oscilaron

entre .70 (*hostilidad y vulnerabilidad*) y .83 (*depresión*), lo que indica valores aceptables de confiabilidad.

Inventario de síntomas SCL-90-R (Derogatis, 1994)

El SCL-90-R está compuesto por 90 ítems con formato de respuesta de cinco opciones (de “*nada*” a “*mucho*”), que examinan patrones de síntomas psicológicos, y que se agrupan para configurar nueve dimensiones clínicas (*somatización, obsesiones/compulsiones, sensibilidad interpersonal, depresión, ansiedad, hostilidad, ansiedad fóbica, ideación paranoide y psicoticismo*). Este instrumento cuenta con evidencias de validez y confiabilidad en el contexto local (Sánchez & Ledesma, 2009), y en el presente estudio el análisis de la consistencia interna de las nueve dimensiones sintomatológicas registró valores alfa de Cronbach de .77 (*ansiedad fóbica*) a .86 (*depresión*).

Procedimiento

Los participantes fueron contactados por psicólogos y alumnos avanzados de la carrera de Psicología de la Universidad de Buenos Aires, quienes colaboraron en las tareas de administración de manera supervisada. Los administradores fueron debidamente entrenados para garantizar que las tomas se realizaran de manera individual en entornos físicos acordes con las coordenadas deseables para un adecuado *setting* de evaluación.

Los evaluados contestaron un protocolo en formato de lápiz y papel que constaba, en primera instancia, de un consentimiento informado en el que se detallaban los objetivos de la investigación y las condiciones establecidas para la participación en el estudio, así como el carácter voluntario de su participación y la posibilidad de abandonar la evaluación en cualquier momento si así se deseaba, las garantías de anonimato y confidencialidad de las respuestas, y que no recibirían alguna devolución de resultados individuales ni retribución económica. Posteriormente, luego de firmar este consentimiento, los participantes podían acceder a la sección del protocolo que incluía los instrumentos. Al finalizar la aplicación, los evaluados recibieron información de contacto de los responsables de la investigación por si eventualmente surgían consultas. No se estableció un tiempo límite, pero se estima que todos los participantes respondieron el protocolo en un tiempo alrededor de los 20 a 25 minutos.

Aspectos éticos

El diseño de este estudio fue evaluado y aprobado por el Comité de Conductas Responsables de la Facultad de

Psicología de la Universidad de Buenos Aires. Se trata de un estudio de bajo riesgo, en virtud de que se realiza la recolección de la información mediante cuestionarios que no identifican a los sujetos, y no se efectúan intervenciones sobre variables psicológicas del evaluado. Se contemplaron pautas éticas locales del ejercicio profesional de la Psicología (Federación de Psicólogos de la República Argentina, 2013) y normativas internacionales para la adaptación y validación de tests usados en la práctica de la evaluación psicológica (International Test Commission, 2014).

Análisis de datos

Se aplicó el modelo de respuesta graduada (MRG) de Samejima, con base en la consideración de sus aspectos teóricos, metodológicos y empíricos (Penfield, 2014). Desde una perspectiva teórica, porque se ha demostrado ampliamente la utilidad de los parámetros del MRG para describir la respuesta a un ítem con respuesta tipo Likert (Reise & Revicki, 2015); a nivel metodológico, porque se alcanzó un tamaño muestral adecuado para responder a los requerimientos de estimación de los parámetros del modelo; y a nivel empírico, porque los datos recolectados en esta investigación permitieron examinar y corroborar empíricamente el ajuste del modelo.

Previo a la aplicación de la TRI, se verificaron los supuestos requeridos por el MRG. Por tanto, se corroboró el supuesto de unidimensionalidad por separado para cada dimensión a partir de análisis factoriales confirmatorios (AFC); y se estimaron los parámetros con el método robusto de mínimos cuadrados ponderados (WLSMV) usando la matriz de correlaciones policóricas. Por otra parte, se analizó el ajuste con los criterios definidos por Byrne (2012): CFI y TLI \geq .90, y RMSEA \leq .08; y se estudió el supuesto de independencia local de los ítems con el estadístico χ^2 LD, en el que se aceptan los valores inferiores a 10 (Reise & Rodríguez, 2016).

También se estimaron los parámetros de los ítems del MRG con el método de máxima verosimilitud marginal. Por tanto, para cada ítem se estimó un parámetro de pendiente (a) y cuatro parámetros de umbral (b_1 , b_2 , b_3 y b_4) que permiten conocer la localización de cada una de las categorías de respuesta en el continuo del rasgo latente; y se analizó el ajuste del MRG a los datos, ítem por ítem, con el índice $S-\chi^2$, considerándose adecuados si obtenían un $p > .05$ (Kang & Chen, 2011).

Para la reducción de cada una de las dimensiones del BFI, se realizó una reducción progresiva, eliminando

los ítems que no cumplían con los requerimientos de calidad psicométrica de la TRI: supuestos de independencia local y unidimensionalidad, y, posteriormente, desajuste al MRG.

Luego de alcanzar la versión depurada, se realizaron estudios para aportar evidencias de validez y confiabilidad. Específicamente, se analizó el funcionamiento diferencial de los ítems (*Differential Item Functioning*, DIF) en función del género con el test de Wald modificado, se analizó el ajuste al modelo penta factorial mediante un AFC, y se estudió la asociación de los puntajes de cada escala original con la puntuación total obtenida luego de la depuración. Asimismo, se correlacionaron los θ estimados para cada sujeto con las variables que mide

el inventario SCL-90-R y el BIN. En lo que respecta a los estudios de confiabilidad, se analizó la consistencia interna con coeficientes globales derivados de la TCT (alfa de Cronbach, alfa ordinal y omega) y de la TRI (fiabilidad marginal). Aprovechando las ventajas que proporciona la TRI con medidas locales de precisión, se analizaron las funciones de información y de error estándar del test.

Finalmente, es importante mencionar que se utilizaron los programas Mplus (Muthén & Muthén, 2010) para el AFC, y el paquete *user friendly science* (Peters, 2014) del programa R para los coeficientes alfa de Cronbach, alfa ordinal y omega con sus respectivos intervalos de confianza al 95 %. El resto de los análisis se realizó con el IRTPRO 4.2 (Cai et al., 2011).

Resultados

Reducción de ítems

La depuración progresiva de los ítems de cada escala permitió reducir el inventario de 44 a 20 ítems. En términos generales, la principal causa que justificó la eliminación de los ítems fue la violación del supuesto de independencia local (15 ítems), resultado esperable en virtud de la presencia de pares de ítems con contenidos redundantes, ya sea por la similitud del indicador o por su negación en la redacción inversa. Solo un ítem perteneciente a la escala de *apertura* fue eliminado por presentar bajo pesaje (inferior a .30) en la aplicación del AFC. Al considerar la calibración de los ítems con el MRG, se descartaron seis ítems que evidenciaron problemas de ajuste según los valores del índice S_{χ^2} ($p < .001$). Por último, se decidió eliminar dos ítems con funcionamiento psicométrico adecuado de las escalas *agradabilidad* y *responsabilidad* con el objetivo de alcanzar la misma cantidad de elementos en la medición de todos los dominios. Por esta razón,

se escogieron reactivos con menores cargas factoriales. Estas decisiones mostraron resultados satisfactorios en los indicadores de la TRI; motivo por el cual no fue necesario eliminar otros ítems de las escalas.

Aplicación de la TRI a la versión reducida

En la Tabla 1 se resumen los resultados obtenidos para cada dimensión del BFI en la verificación de los supuestos de unidimensionalidad e independencia local. Los valores de CFI y TLI superiores a .90 y los límites superiores de los intervalos de confianza de los RMSEA por debajo de .08 demuestran que cada uno de los rasgos evaluados responde a un modelo unidimensional. Asimismo, todas las cargas factoriales fueron estadísticamente significativas ($p < .05$), con pesajes adecuados, y, finalmente, los valores de χ^2_{LD} registrados para cada dominio fueron suficientes como para aceptar que los ítems son localmente independientes ($\chi^2_{LD} < 10$).

Tabla 1. Ajuste de los modelos del análisis factorial confirmatorio, realizados para verificar los supuestos de unidimensionalidad y el análisis de la independencia local en la versión reducida de 20 ítems del Big Five Inventory

	Extraversión	Agradabilidad	Responsabilidad	Neuroticismo	Apertura
Análisis factorial confirmatorio					
CFI	.934	.917	.926	.964	.944
TLI	.919	.904	.914	.939	.933
RMSEA	.026	.029	.032	.015	.025
IC 90 %	[.051; .076]	[.054; .079]	[.056; .080]	[.040; .065]	[.049; .073]

	Extraversión	Agradabilidad	Responsabilidad	Neuroticismo	Apertura
Pesajes					
Mínimo	.59	.62	.56	.51	.40
Máximo	.72	.75	.69	.68	.66
Independencia local					
χ^2_{LD} mínimo	4.7	4.0	0.1	3.6	0.1
χ^2_{LD} máximo	9.6	9.0	7.0	9.1	8.6

Por otra parte, en la Tabla 2 se presentan los resultados de la calibración de los ítems de cada una de las dimensiones con el MRG. Como se puede corroborar, todos los estadísticos S_{χ^2} mostraron un *p-value* superior a .01, el cual refleja un ajuste satisfactorio del MRG a los datos. Además, el promedio de los parámetros *a* osciló entre 1.45 (*neuroticismo*) y 1.86 (*extraversión*); y cinco parámetros de umbral b_1 (dos pertenecientes a ítems de responsabilidad y tres a agradabilidad) se localizaron

por debajo del límite inferior esperado (-3). Ambos constructos también tendieron a presentar valores de b_4 en torno a $\theta = 0$. No obstante, los valores obtenidos en estas estimaciones parecen razonables considerando que los contenidos de estos ítems podrían ser sensibles a la deseabilidad social por el hecho de referirse a hábitos de trabajo valorados (autopercebirse como eficaz y digno de confianza) y comportamientos solidarios (cooperación, amabilidad y capacidad para perdonar).

Tabla 2. Calibración con el modelo de respuesta graduada (MRG) y estudio de funcionamiento diferencial de los 20 ítems de la versión reducida del Big Five Inventory

Ítem	Parámetros del MRG				Ajuste del MRG			Test de Wald		
	<i>a</i> (s.e.)	b_1 (s.e.)	b_2 (s.e.)	b_3 (s.e.)	b_4 (s.e.)	χ^2	<i>gl</i>	<i>p</i>	χ^2 (<i>gl</i> = 5)	<i>p</i>
Extraversión										
1	1.87 (0.16)	-2.27 (0.15)	-1.53 (0.10)	-0.61 (0.06)	0.25 (0.06)	51.8	42	.14	1.2	.87
21R	2.19 (0.19)	-1.20 (0.08)	-0.53 (0.05)	0.15 (0.05)	0.73 (0.06)	55.5	38	.03	1.7	.80
31R	1.31 (0.11)	-1.17 (0.10)	-0.06 (0.06)	0.71 (0.08)	1.59 (0.12)	57.9	40	.03	4.3	.37
36	2.06 (0.18)	-2.29 (0.15)	-1.53 (0.10)	-0.68 (0.06)	0.16 (0.05)	35.6	41	.71	3.0	.56
Agradabilidad										
7	2.14 (0.18)	-2.95 (0.20)	-2.08 (0.12)	-1.14 (0.07)	-0.04 (0.05)	52.5	36	.04	3.5	.49
17	1.46 (0.13)	-3.66 (0.30)	-2.90 (0.22)	-1.91 (0.14)	-0.50 (0.07)	51.3	34	.03	8.7	.07
32	2.11 (0.17)	-3.27 (0.24)	-2.11 (0.12)	-1.05 (0.07)	0.11 (0.05)	36.5	26	.08	3.3	.51
42	2.13 (0.18)	-3.07 (0.21)	-2.09 (0.12)	-1.23 (0.08)	0.02 (0.05)	55.1	36	.02	6.5	.16
Responsabilidad										
13	1.49 (0.15)	-3.44 (0.30)	-2.71 (0.22)	-1.87 (0.14)	-0.78 (0.07)	57.8	38	.02	5.8	.21
28	1.99 (0.18)	-2.28 (0.15)	-1.50 (0.10)	-0.65 (0.06)	0.30 (0.05)	52.4	33	.02	1.7	.80
33	1.54 (0.14)	-3.58 (0.30)	-2.68 (0.19)	-1.14 (0.09)	0.41 (0.07)	48.1	32	.03	2.5	.64
38	1.87 (0.16)	-2.67 (0.19)	-1.72 (0.11)	-0.88 (0.07)	0.21 (0.05)	47.4	34	.06	2.6	.63

Ítem	Parámetros del MRG				Ajuste del MRG				Test de Wald	
	<i>a</i> (s.e.)	<i>b</i> ₁ (s.e.)	<i>b</i> ₂ (s.e.)	<i>b</i> ₃ (s.e.)	<i>b</i> ₄ (s.e.)	χ^2	<i>gl</i>	<i>p</i>	χ^2 (<i>gl</i> = 5)	<i>p</i>
Neuroticismo										
4	1.31 (0.13)	0.02 (0.07)	0.74 (0.08)	1.67 (0.14)	2.82 (0.23)	43.0	40	.34	4.2	.38
14	1.33 (0.12)	-1.57 (0.13)	-0.76 (0.08)	0.30 (0.07)	1.61 (0.13)	50.2	43	.21	2.7	.61
19	1.01 (0.10)	-2.93 (0.27)	-1.70 (0.16)	-0.49 (0.09)	0.75 (0.10)	56.9	44	.09	2.1	.72
29	2.14 (0.24)	-1.43 (0.13)	-0.75 (0.07)	0.11 (0.05)	1.02 (0.08)	46.1	39	.21	2.3	.68
Apertura										
5	2.37 (0.21)	-2.29 (0.14)	-1.54 (0.09)	-0.38 (0.05)	0.71 (0.06)	37.9	34	.30	3.1	.55
15	2.24 (0.20)	-2.77 (0.18)	-1.85 (0.11)	-0.57 (0.06)	0.64 (0.06)	43.8	34	.12	2.6	.63
20	1.63 (0.13)	-2.95 (0.21)	-1.99 (0.13)	-0.83 (0.07)	0.23 (0.06)	41.8	37	.27	1.4	.84
44	0.54 (0.08)	-2.34 (0.33)	-1.01 (0.18)	0.83 (0.17)	2.68 (0.38)	62.7	40	.02	1.4	.85

Nota. R = ítems redactados en sentido inverso.

Estudios de validez de la versión reducida

En la Tabla 2 también se incluyen los resultados obtenidos en el test de Wald al analizar el funcionamiento diferencial de los ítems según el género del evaluado. Como se puede observar, dado que todos los *p-values* son superiores a .05, es posible descartar la presencia de DIF en los ítems de la versión reducida. Esto quiere decir que no se registran diferencias estadísticamente significativas al comparar los parámetros de los ítems estimados para varones y mujeres, por lo que se asume que el instrumento arroja mediciones invariantes entre estos grupos.

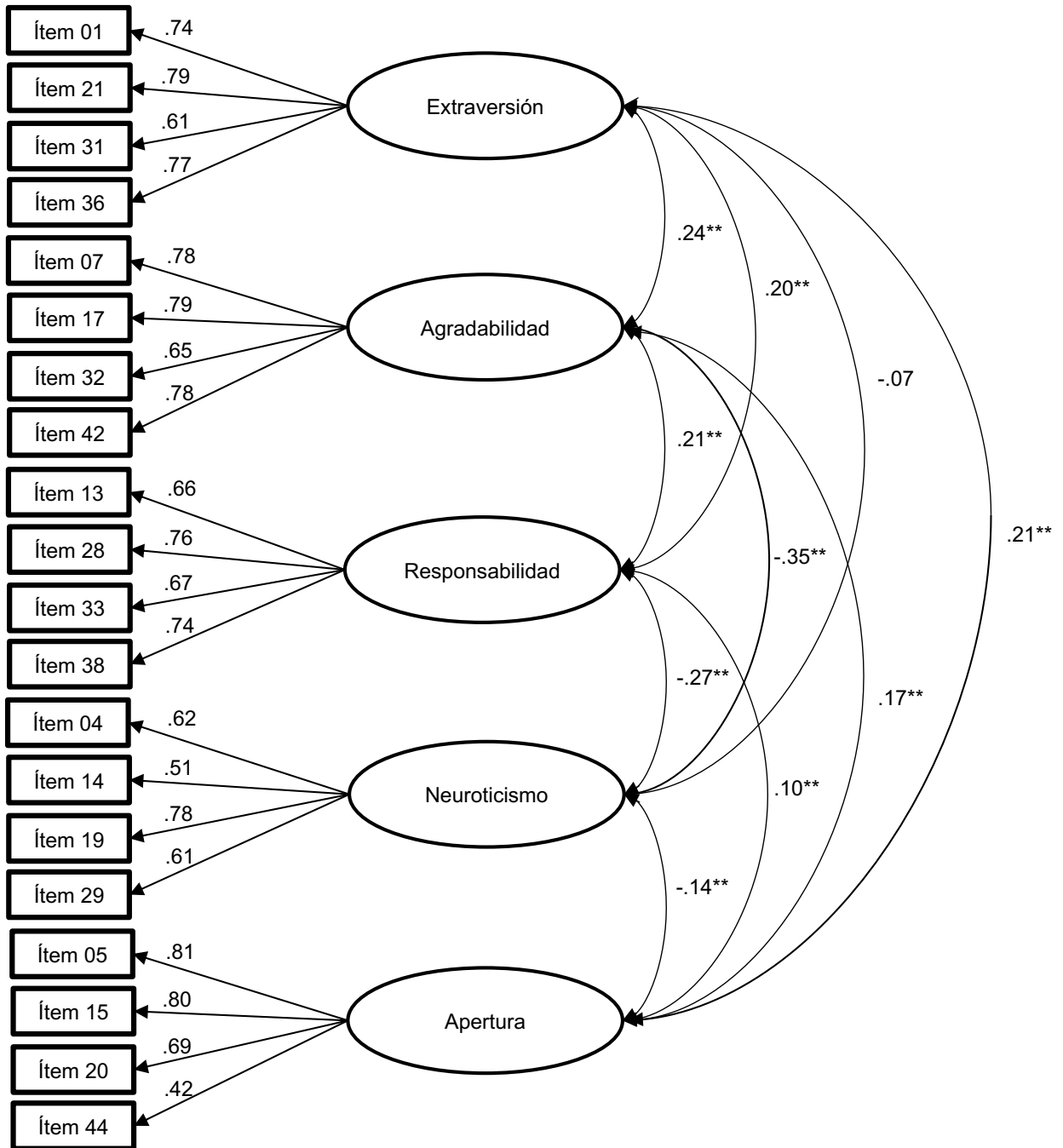
Ahora bien, para analizar el impacto de la variación de contenido ocasionado por la reducción de ítems, se correlacionaron, para cada dimensión, los puntajes totales de la prueba en su versión original y en la versión abreviada en este estudio. Con esto, se encontró que todas las correlaciones resultaron iguales o superiores a .90, que también se redujeron al controlar la naturaleza espuria de estas asociaciones con la corrección de Levy (1967): *responsabilidad* ($r = .79$), *agradabilidad* ($r = .78$), *extraversión* ($r = .75$), *neuroticismo* ($r = .73$) y *apertura* ($r = .73$).

Por otra parte, el estudio factorial confirmatorio con los ítems de la versión reducida mostró un ajuste aceptable al modelo penta factorial (χ^2 (*gl* = 160) = 456.1; $p < .001$; CFI = .93; TLI = .92; RMSEA = .048, 90 IC [.024; .072]).

Como era esperable, el χ^2 resultó significativo, aunque el tamaño muestral es lo suficientemente grande como para aceptar que el rechazo del ajuste se debe a la sensibilidad del estadístico. Además, las saturaciones factoriales fueron mayores a .40 y significativas para un valor $p \leq .05$ (Figura 1); y las correlaciones interfactores presentaron intensidades entre bajas y moderadas. El rasgo de *neuroticismo* registró las asociaciones más elevadas, junto con *agradabilidad* ($r = -.27$) y *responsabilidad* ($r = -.35$).

Finalmente, en la Tabla 3 se presentan los resultados de las correlaciones de los cinco dominios con las variables del BIN y el SCL-90-R. Aquí, destaca la asociación entre los puntajes totales de *neuroticismo* del BIN y BFI reducido, aun cuando existe una gran diferencia en la extensión de ambos instrumentos (54 ítems vs. 4 ítems). A su vez, al medir el *neuroticismo* con el BIN, también se registraron correlaciones similares a las reportadas con *agradabilidad* ($r = -.37$) y *responsabilidad* ($r = -.40$). Además, si se repasa en las facetas de *neuroticismo* del BIN, es posible identificar hallazgos acordes a lo esperable a nivel teórico, como las correlaciones entre *agradabilidad* y *hostilidad*, entre *extraversión* y *autoconciencia*, y entre *responsabilidad* y *vulnerabilidad*. Finalmente, *neuroticismo* y *responsabilidad* mostraron las correlaciones más intensas con respecto a las dimensiones sintomáticas medidas con el SCL-90-R.

Figura 1. Análisis factorial confirmatorio realizado para aportar evidencias de validez basadas en la estructura interna de la versión reducida del Big Five Inventory: saturaciones factoriales estandarizadas y correlaciones interfactoriales



Nota. ** $p < .01$.

Tabla 3. Asociación entre las dimensiones que mide la versión reducida del Big Five Inventory y variables del Banco de ítems de neuroticismo (BIN) y el Inventario de síntomas SCL-90-R para aportar evidencias de validez basadas en la relación con otras variables

	Extraversión	Agradabilidad	Responsabilidad	Neuroticismo	Apertura
BIN	-.222**	-.372**	-.395**	.729**	-.135**
Ansiedad	-.150**	-.264**	-.269**	.723**	-.036
Hostilidad	-.074	-.521**	-.185**	.607**	-.144**
Depresión	-.230**	-.289**	-.368**	.647**	-.095
Autoconciencia	-.478**	-.247**	-.330**	.440**	-.127*
Impulsividad	.070	-.317**	-.296**	.572**	-.089
Vulnerabilidad	-.174**	-.188**	-.441**	.527**	-.168**
SCL-90-R					
Somatización	-.068*	-.061	-.147**	.420**	-.006
Obsesiones/ Compulsiones	-.142**	-.036	-.264**	.474**	-.040
Depresión	-.123**	-.135**	-.234**	.437**	-.008
Sensitividad interpersonal	-.148**	.012	-.253**	.532**	-.025
Ansiedad	-.038	-.091**	-.155**	.529**	-.017
Hostilidad	.005	-.160**	-.098**	.414**	.000
Ansiedad fóbica	-.182**	-.107**	-.169**	.361**	-.046
Ideación paranoide	-.042	-.070*	-.106**	.427**	.109**
Psicoticismo	-.075*	-.071*	-.219**	.504**	.021

Nota. * $p < .01$; ** $p < .001$.

Estudios de confiabilidad de la versión reducida

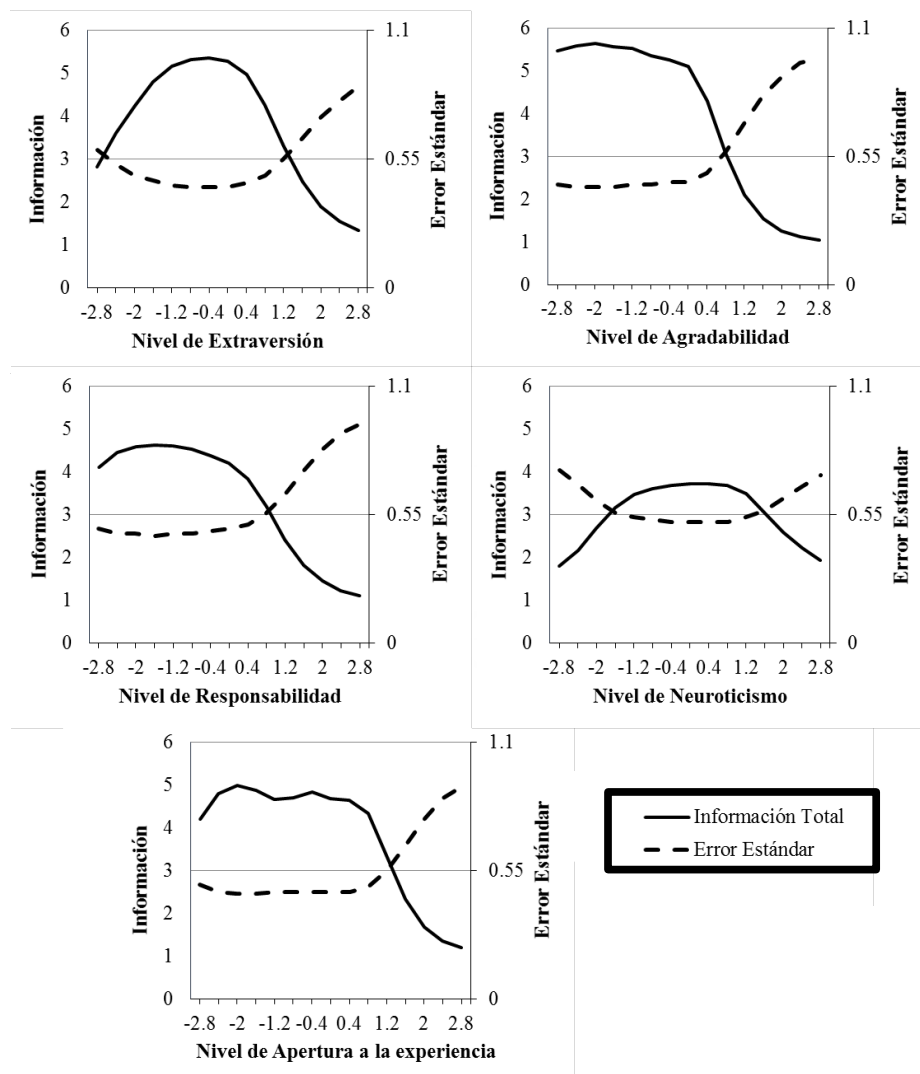
Por último, y como era previsible, dada la corta extensión de las escalas, los coeficientes de consistencia interna alfa de Cronbach mostraron valores al límite de ser considerados aceptables, como sucede con *neuroticismo* (.66) y *apertura* (.63). No obstante, se registran mejoras si se emplean coeficientes que contemplan la naturaleza ordinal del formato de respuesta de los ítems (Tabla 4); con este ajuste, los alfas y omegas ordinales resultaron mayores a .70 para todas las dimensiones. En el marco TRI, los indicadores globales de confiabilidad (fiabilidad marginal) también fueron aceptables.

Finalmente, para profundizar en el estudio de la precisión de las mediciones, se representaron las funciones de los errores estándares de estimación de cada uno de los cinco factores medidos por el BFI reducido (Figura 2). En el gráfico, se destacan los rangos de los respectivos θ en los que las funciones presentan errores superiores a .55 (equivalente a una confiabilidad clásica de .70). Aunque con mayor error, la función de *neuroticismo* es la única relativamente simétrica y se encuentra centrada en los niveles medios del rasgo. En el resto de las curvas, los errores tendieron a mantenerse constantes para los niveles medio y bajo. Ninguna de las escalas mostró precisión aceptable para la medición de los valores extremos del polo superior.

Tabla 4. Coeficientes obtenidos en el estudio de confiabilidad de la versión reducida del Big Five Inventory

	Extraversión	Agradabilidad	Responsabilidad	Neuroticismo	Apertura
Alfa	.74 [.71; .77]	.74 [.72; .77]	.70 [.67; .74]	.66 [.62; .69]	.63 [.60; .67]
Alfa ordinal	.80 [.78; .82]	.81 [.79; .83]	.78 [.75; .80]	.71 [.68; .74]	.72 [.69; .75]
Omega ordinal	.80 [.78; .83]	.81 [.79; .83]	.78 [.75; .80]	.71 [.68; .74]	.74 [.71; .77]
Fiabilidad marginal	.77	.72	.71	.71	.75

Figura 2. Funciones de información y error estándar para el estudio de la precisión de la medición según los diferentes niveles de cada uno de los factores medidos en la versión reducida del Big Five Inventory



Discusión

Los tests abreviados de personalidad ofrecen una solución tentadora ante una situación de evaluación que requiera medir un constructo por demás complejo, apelando a una economía en los tiempos de administración. Sin embargo, el beneficio que se registra en la práctica se consigue necesariamente a expensas de resignar calidad psicométrica, tratando de alcanzar un equilibrio entre la exhaustividad en la cobertura del contenido y un grado aceptable de consistencia interna a partir de ítems heterogéneos (Balasaro et al., 2013; Credé et al., 2012; Milojev et al., 2013; Morizot, 2014; Sibley, 2012; Ziegler et al., 2014). En estas circunstancias es que cobra relevancia la explicitación de los criterios tomados en la reducción de los ítems y se impone la necesidad de hallar evidencia empírica rigurosa que legitime la calidad de las medidas abreviadas.

El análisis psicométrico realizado en este estudio permitió reducir el BFI de 44 ítems a 20 ítems. Si bien existen escalas más breves, como el BFI-10 o el BFI-15, se decidió que eliminar una mayor cantidad de elementos podría afectar a la consistencia interna de la prueba. En efecto, dos escalas (*neuroticismo* y *apertura*) registraron coeficientes de consistencia interna al límite de lo aceptable según el criterio de DeVellis (2016). No obstante, dada la escasa cantidad de ítems que componen cada una de las escalas, parece razonable que se alcancen coeficientes con estos valores. A partir de esto, se puede afirmar que una adecuada representatividad de los contenidos de factores que tienen una gran complejidad conceptual exige la inclusión de indicadores con un menor grado de asociación, lo cual repercute en la consistencia interna de la prueba.

También se decidió eliminar dos ítems con funcionamiento psicométrico adecuado para que todos los factores tuvieran la misma cantidad de elementos. Aunque este criterio no resulta estrictamente necesario, dado que no es preciso que los dominios se midan con el mismo número de ítems, se priorizó un criterio práctico que lleva a reducir los tiempos de administración, sin poner en riesgo las propiedades psicométricas de las escalas.

Respecto a las asociaciones entre las escalas originales y las versiones abreviadas, se encontró que aportan una importante evidencia de validez concurrente y permiten concluir que la reducción de contenido no incidió sustantivamente en la medición de los dominios. Además, el estudio del DIF en función del género mostró

que la versión reducida arrojó mediciones invariantes para varones y mujeres, lo cual aporta mayor evidencia de su validez.

Ahora bien, los criterios usados en esta investigación para seleccionar los ítems de la versión abreviada se asientan principalmente en las herramientas analíticas que brinda la TRI, siendo la violación del supuesto de independencia local la razón que propició la mayor cantidad de elementos eliminados (62.5 %). Al respecto, se debe tener en cuenta que el BFI fue desarrollado desde la teoría clásica, y en esta perspectiva es habitual el uso y abuso de ítems redundantes (Lloret Segura et al., 2014), empleados con frecuencia para comprobar la coherencia de las respuestas de los evaluados, aumentar artificialmente la consistencia interna de la escala o conseguir una estructura factorial clara y bien definida (Díaz-Morales, 2006; Ferrando & Anguiano-Carrasco, 2010). En cambio, desde la TRI, la inclusión de estos ítems puede afectar de manera significativa a la estimación de los parámetros, dado que la respuesta a un elemento no es estadísticamente independiente de la respuesta a otro elemento (Attorresi et al., 2009).

Un hallazgo que conviene discutir es la escasa cantidad de ítems inversos conservados en esta versión reducida (solo dos ítems pertenecientes a la dimensión *extraversión*, de un total de 15 ítems de la prueba original). Para Rammstedt y John (2007), la utilidad de incluir ítems redactados en ambos sentidos del rasgo busca mejorar la cobertura del contenido y permite una mejor representación de cada uno de los polos; no obstante, el análisis con la TRI reveló que los ítems pueden localizarse en los extremos inferior o superior del constructo, con independencia del sentido de la redacción. En efecto, los ítems inversos (21 y 31) de *extraversión* mostraron capacidad discriminativa en niveles del rasgo más altos que los ítems directos (1 y 36).

A esto se suma que la bibliografía especializada más tradicional recomienda combinar ítems directos e inversos para controlar el sesgo en el estilo de respuesta aquiescente, aunque en los últimos años esta estrategia recibió cuestionamientos importantes (Suárez-Álvarez et al., 2018). Siguiendo esta línea, en el proceso de reducción del instrumento se priorizó la selección de ítems con mejores propiedades psicométricas, sin considerar el sentido de su redacción. En su mayoría, los ítems inversos del BFI presentaron algún par que expresaba una

idea similar, pero redactada en sentido positivo (p. ej., 27R. “... que puedo ser frío/a y distante” y 32. “...que es muy considerado/a y amable con los demás”). Consecuentemente, también registraron valores elevados de χ^2_{LD} que justificaban su eliminación. Desde la teoría clásica, otros autores también refieren dificultades psicométricas en ítems inversos, o deciden *a priori* no incluirlos para evitar introducir varianza irrelevante en la medición de los constructos (Dominguez-Lara & Merino-Soto, 2018; Gouveia et al., 2021).

Por otra parte, tanto las correlaciones interfactores como las asociaciones halladas con las facetas de *neuroticismo* del BIN responden a un patrón de relaciones coherentes a las expectativas teóricas (McCrae & Costa, 2010), de manera que, por ejemplo, niveles elevados de *neuroticismo* tenderán a mantener vínculos interpersonales conflictivos (baja *agradabilidad*) y a autoperibirse como poco eficientes (baja *responsabilidad*), basándose en su pobre autoestima. Asimismo, las asociaciones con las dimensiones sintomatológicas del SCL-90-R son consistentes con los hallazgos reportados en la literatura, siendo *neuroticismo* y *responsabilidad* los dominios mencionados con más frecuencia como factores de riesgo (Castellanos-Ryan et al., 2016; Jeronimus et al., 2016; Menéndez et al., 2019) y factores protectores (Bogg & Roberts, 2004; Sutin et al., 2018), respectivamente, para el desarrollo y mantenimiento de diversas psicopatologías.

En cuanto al análisis de las Funciones de Información de los Tests (FIT), se encontró una debilidad de la prueba en su conformación actual que permanece oculta para

los coeficientes de confiabilidad globales de la teoría clásica. Idealmente, cuando un instrumento está destinado para la población general, como el BFI, se espera que los ítems se ubiquen a lo largo del espectro de cada dimensión y la FIT resulte relativamente simétrica en torno a $\theta = 0$. No obstante, en términos generales, los ítems conservados tendieron a ubicarse por debajo de las medias de los respectivos rasgos, en tanto que, como se observa en las FIT, en cuatro dimensiones la medición de los niveles más elevados tuvo una menor precisión —aunque debe exceptuarse el factor *neuroticismo*, cuya FIT es menos elevada que la del resto de las dimensiones, pero más simétrica—. En consecuencia, aunque los coeficientes de fiabilidad clásicos fueron aceptables, en futuros estudios se recomienda ensayar nuevos elementos que puedan reemplazar alguno de los ítems actuales, con el objetivo de alcanzar una mayor precisión para medir niveles elevados de los rasgos.

Finalmente, entre las limitaciones propias de la metodología TRI aplicada, se encuentra que el estadístico $S\text{-}\chi^2$, empleado en el análisis del ajuste del MRC, puede verse afectado por el tamaño de la muestra y la extensión de la prueba (Sueiro & Abad, 2009). Además, la presente investigación se asienta en la aplicación de modelos de la TRI unidimensionales para el análisis de un constructo multidimensional (Makransky et al., 2013; Reckase, 2009); pero el estudio del instrumento a través de un modelo multidimensional de la TRI podría introducir mejoras en la estimación de los parámetros y alcanzar interpretaciones más exhaustivas que reconozcan la complejidad conceptual de un fenómeno como la personalidad.

Referencias

- 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. *Revista psicodebate: psicología, cultura y sociedad*, 19(1), 31-50. <https://doi.org/10.18682/pd.v19i1.854>
- Alansari, B. (2016). The Big Five Inventory (BFI): Reliability and validity of its Arabic translation in non clinical sample. *European Psychiatry*, 33(S1), S209-S210. <https://doi.org/10.1016/j.eurpsy.2016.01.500>
- Allen, T. A., & DeYoung, C. G. (2017). Personality neuroscience and the five factor model. En T. A. Widiger (Ed.), *The Oxford Handbook of the Five Factor Model* (pp. 319-352). Oxford Library of Psychology. <https://doi.org/10.1093/oxfordhb/9780199352487.013.26>
- Ato, M., López, J., & Benavente, A. (2013). Un sistema de clasificación de los diseños de investigación en psicología. *Anales de Psicología*, 29(3), 1038-1059. <https://doi.org/10.6018/analesps.29.3.178511>
- Attorresi, H. F., Lozzia, G. S., Abal, F. J., Galibert, M. S., & Aguerri, M. E. (2009). Teoría de Respuesta al Ítem. Conceptos básicos y aplicaciones para la medición de constructos psicológicos. *Revista Argentina de Clínica Psicológica*, 18(2), 179-188. https://www.cienciaried.com.ar/ra/usr/35/825/racp_xviii_2_pp179_188.pdf
- 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. <https://doi.org/10.1080/00223891.2012.700466>
- Balgiu, B. A. (2018). The psychometric properties of the Big Five inventory-10 (BFI-10) including correlations with subjective and psychological well-being. *Global Journal of Psychology Research: New Trends and Issues*, 8(2), 61-69. <https://doi.org/10.18844/gjpr.v8i2.3434>
- Benet-Martínez, V., & John, O. P. (1998). Los Cinco Grandes across cultures and ethnic groups: Multitrait multi-method analyses of the Big Five in Spanish and English. *Journal of Personality and Social Psychology*, 75(3), 729-750. <https://doi.org/10.1037/0022-3514.75.3.729>
- Bogg, T., & Roberts, B. W. (2004). Conscientiousness and health-related behaviors: a meta-analysis of the leading behavioral contributors to mortality. *Psychological Bulletin*, 130(6), 887-919. <https://doi.org/10.1037/0033-2909.130.6.887>
- Brown, G. T. L., & Sotardi, V. A. (2019, 7-9 de marzo). *Evaluation of the BFI-10 personality inventory with New Zealand University students: Failing psychometric tests* [Póster, 7-9 de marzo, 2019]. International Congress of Psychological Science (ICPS), Paris, Francia. <https://www.psychologicalscience.org/conventions/archive/2019-icps>
- Byrne, B. M. (2012). *A primer of LISREL: Basic applications and programming for confirmatory factor analytic models*. Springer Science & Business Media. <https://doi.org/10.1007/978-1-4613-8885-2>
- Cai, L., Thissen, D., & du Toit, S. (2017). *IRTPRO 3 for Windows* [Computer Software]. Skokie, IL: Scientific Software International, Inc.
- Carciofo, R., Yang, J., Song, N., Du, F., & Zhang, K. (2016). Psychometric Evaluation of Chinese-Language 44-Item and 10-Item Big Five Personality Inventories, Including Correlations with Chronotype, Mindfulness and Mind Wandering. *Plos One*, 11(2), e0149963. <https://doi.org/10.1371/journal.pone.0149963>
- Castellanos-Ryan, N., Brière, F. N., O'Leary-Barrett, M., Banaschewski, T., Bokde, A., Bromberg, U., Büchel, C., Flor, H., Frouin, V., Gallinat, J., Garavan, H., Martinot, J.-L., Nees, F., Paus, T., Pausova, Z., Rietschel, M., Smolka, M. N., Robbins, T. W., Whelan, R., ... The IMAGEN Consortium. (2016). The structure of psychopathology in adolescence and its common personality and cognitive correlates. *Journal of Abnormal Psychology*, 125(8), 1039-1052. <http://doi.org/10.1037/abn0000193>

- Castro Solano, A. (2002). *Técnicas de evaluación psicológica en los ámbitos militares*. Paidós.
- Chapman, B. P., & Elliot, A. J. (2019). Brief report: How short is too short? An ultra-brief measure of the big-five personality domains implicates “agreeableness” as a risk for all-cause mortality. *Journal of health psychology*, 24(11), 1568-1573. <https://doi.org/10.1177/1359105317720819>
- Chiesi, F., Morsanyi, K., Donati, M. A., & Primi, C. (2018). Applying Item Response Theory to Develop a Shortened Version of the Need for Cognition Scale. *Advances in Cognitive Psychology*, 14(3), 75-86. <https://doi.org/10.5709/acp-0240-z>
- Chio, F. H., Lai, M. H., & Mak, W. W. (2018). Development of the Nonattachment Scale-Short Form (NAS-SF) using item response theory. *Mindfulness*, 9(4), 1299-1308. <https://doi.org/10.1007/s12671-017-0874-z>
- 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>
- Colom, R. (2018). *Manual de psicología diferencial*. Pirámide.
- Courtois, R., Petot, J. M., Plaisant, O., Allibe, B., Lignier, B., Réveillère, C., Lecocq, G., & John, O. (2020). Validation of the French version of the 10-item Big Five Inventory. *L'encephale*, 46(6), 455-462. <https://doi.org/10.1016/j.encep.2020.02.006>
- 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. <https://doi.org/10.1037/a0027403>
- Derogatis, L. R. (1994). *SCL-90-R: Administration, scoring and procedures manual*. National Computer Systems.
- DeVellis, R. F. (2016). *Scale Development: Theory and Applications* (4.ª ed.). SAGE.
- Díaz-Morales, J. F. (2006). Estructura factorial y fiabilidad del Inventario de Perspectiva Temporal de Zimbardo. *Psicothema*, 18(3), 565-571. <https://reunido.uniovi.es/index.php/PST/article/view/8473>
- Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology*, 41(1), 417-440. <https://doi.org/10.1146/annurev.ps.41.020190.002221>
- Dominguez-Lara, S., & Merino-Soto, C. (2018). Dos versiones breves del Big Five Inventory en universitarios peruanos: BFI-15p y BFI-10p. *Liberabit*, 24(1), 81-96. <https://doi.org/10.24265/liberabit.2018.v24n1.06>
- Dominguez-Lara, S., Merino-Soto, C., Zamudio, B., & Guevara-Cordero, C. (2018). Big Five Inventory en Universitarios Peruanos: Resultados Preliminares de su Validación. *Psykhé (Santiago)*, 27(2), 1-12. <https://doi.org/10.7764/psykhe.27.2.1052>
- Federación de Psicólogos de la República Argentina [FePRA]. (2013). *Código de Ética de la Federación de Psicólogos de la República Argentina*. FePRA. <https://fepra.org.ar/wp-content/uploads/2021/06/Codigo-de-E%CC%81ti-ca-de-la-FePRA.pdf>
- Ferrando, P. J., & Anguiano-Carrasco, C. (2010). El análisis factorial como técnica de investigación en psicología. *Papeles del psicólogo*, 31(1), 18-33. <https://www.papelesdelpsicologo.es/resumen?pii=1793>
- Gerlitz, J. Y., & Schupp, J. (2005). Zur Erhebung der Big-Five-basierten persönlichkeitsmerkmale im SOEP DIW Research Notes, 4, 1-36. <https://www.diw.de/documents/publicationen/73/43490/rn4.pdf>
- Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist*, 48(1), 26-34. <https://doi.org/10.1037/0003-066X.48.1.26>
- Gouveia, V., Vasconcelos de Oliveira, I. C., Moura Grangeiro, A. S. D., Pereira Monteiro, R., & Lins de Holanda Coelho, G. (2021). The bright side of the human personality: evidence of a measure of prosocial traits. *Journal of Happiness Studies*, 22(3), 1459-1480. <https://doi.org/10.1007/s10902-020-00280-2>

- Guido, G., Peluso, A. M., Capestro, M., & Miglietta, M. (2015). An Italian version of the 10-item Big Five Inventory: An application to hedonic and utilitarian shopping values. *Personality and Individual Differences*, 76, 135-140. <https://doi.org/10.1016/j.paid.2014.11.053>
- Hahn, E., Gottschling, J., & Spinath, F. M. (2012). Short measurements of personality-Validity and reliability of the GSOEP Big Five Inventory (BFI-S). *Journal of Research in Personality*, 46(3), 355-359. <https://doi.org/10.1016/j.jrp.2012.03.008>
- International Test Commission [ITC]. (2014). *ITC Statement On the Use of Tests and Other Assessment Instruments for Research Purposes*. ITC. <http://www.intestcom.org/page/22>
- 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. <https://doi.org/10.1017/S0033291716001653>
- John, O. P. (2021). History, Measurement, and Conceptual Elaboration of the Big-Five Trait Taxonomy: The Paradigm Matures. En O. P. John & R. W. Robins (Eds.), *Handbook of personality: Theory and research* (4.^a ed., pp. 35-82). The Guilford Press.
- John, O. P., Donahue, E. M., & Kentle, R. L. (1991). *Big Five Inventory (BFI)* [Database record]. University of California, APA PsycTests. <https://doi.org/10.1037/t07550-000>
- Kang, T., & Chen, T. T. (2011). Performance of the generalized S-X2 item fit index for the graded response model. *Asia Pacific Education Review*, 12(1), 89-96. <https://doi.org/10.1007/s12564-010-9082-4>
- Levy, P. (1967). The correction for spurious correlation in the evaluation of short-form tests. *Journal of Clinical Psychology*, 23(1), 84-86. [https://doi.org/10.1002/1097-4679\(196701\)23:1<84::AID-JCLP2270230123>3.0.CO;2-2](https://doi.org/10.1002/1097-4679(196701)23:1<84::AID-JCLP2270230123>3.0.CO;2-2)
- 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. <https://doi.org/10.6018/analesps.30.3.199361>
- Ludeke, S. G., & Larsen, E. G. (2017). Problems with the Big Five assessment in the world values survey. *Personality and Individual Differences*, 112, 103-105. <https://doi.org/10.1016/j.paid.2017.02.042>
- 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(1), 3-13. <https://doi.org/10.1177/1073191112437756>
- Maples-Keller, J. L., Williamson, R. L., Sleep, C. E., Carter, N. T., Campbell, W. K., & Miller, J. D. (2019). Using item response theory to develop a 60-item representation of the NEO PI-R using the International Personality Item Pool: Development of the IPIP-NEO-60. *Journal of Personality Assessment*, 101(1), 4-15. <https://doi.org/10.1080/00223891.2017.1381968>
- McCrae, R. R. (2017). The Five-Factor Model across cultures. En A. T. Church (Ed.), *The Praeger Handbook of Personality Across Cultures: Trait Psychology Across Cultures* (pp. 47-71). Praeger/ABC-CLIO.
- McCrae, R. R., & Costa, P. T. (2003). *Personality in Adulthood, Second Edition: A Five-Factor Theory Perspective*. Guilford Press.
- McCrae, R. R., & Costa, P. T. (2010). *NEO Inventories professional manual*. Psychological Assessment Resources.
- Menéndez, J. A., Attorresi, H. F., & Abal, F. J. (2019). Aportes de la medición del Neuroticismo en los ámbitos clínico y de la salud. *Anuario de Investigaciones*, 26, 385-394. https://www.psi.uba.ar/investigaciones/revistas/anuario/trabajos_completos/26/menendez.pdf
- 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. <https://doi.org/10.1016/j.jrp.2013.09.004>

- Minkov, M., van de Vijver, F. J., & Schachner, M. (2019). A test of a new short Big-Five tool in large probabilistic samples from 19 countries. *Personality and Individual Differences*, 151, 109519. <https://doi.org/10.1016/j.paid.2019.109519>
- Morizot, J. (2014). Construct validity of adolescents' self-reported big five personality traits: Importance of conceptual breadth and initial validation of a short measure. *Assessment*, 21(5), 580-606. <https://doi.org/10.1177/1073191114524015>
- Muñiz, J. (2018). *Introducción a la Psicometría: teoría clásica y TRI*. Pirámide.
- Muthén, L. K., & Muthén, B. O. (2010). *Mplus user's guide* (Versión 6) [Computer software]. Muthén & Muthén.
- Penfield, R. D. (2014). An NCME instructional module on polytomous item response theory models. *Educational Measurement: Issues and Practice*, 33(1), 36-48. <https://doi.org/10.1111/emip.12023>
- Peters, G. J. Y. (2014). The alpha and the omega of scale reliability and validity: Why and how to abandon Cronbach's alpha and the route towards more comprehensive assessment of scale quality. *European Health Psychologist*, 16(2), 56-69. <https://doi.org/10.31234/osf.io/h47fv>
- Rammstedt, B. (2007). The 10-item big five inventory. *European Journal of Psychological Assessment*, 23(3), 193-201. <https://doi.org/10.1027/1015-5759.23.3.193>
- Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, 41(1), 203-212. <https://doi.org/10.1016/j.jrp.2006.02.001>
- Rammstedt, B., Kemper, C. J., Klein, M. C., Beierlein, C., & Kovaleva, A. (2013). A short scale for assessing the big five dimensions of personality: 10 item big five inventory (BFI-10). *Methods, Data, Analyses*, 7(2), 233-249. <https://doi.org/10.12758/mda.2013.013>
- Rammstedt, B., Lechner, C. M., & Danner, D. (2020). Short Forms Do Not Fall Short. A Comparison of Three (Extra-) Short Forms of the Big Five. *European Journal of Psychological Assessment*, 37(1), 23-32. <https://doi.org/10.1027/1015-5759/a000574>
- Rammstedt, O. (1997). Das Durkheim-Simmelsche Projekteiner „rein wissenschaftlichen Soziologie“ im Schatten der Dreyfus-Affäre. *Zeitschrift für Soziologie*, 26(6), 444-457. <https://doi.org/10.1515/zfsoz-1997-0605>
- Reckase, M. D. (2009). *Multidimensional item response theory*. Springer
- Reise, S. P., & Revicki, D. A. (2015). *Handbook of Item Response Theory Modeling Applications to Typical Performance Assessment*. Routledge.
- 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>
- Reyes Zamorano, E. R., Carrillo, C. Á., Silva, A. P., Sandoval, A. M., & Pastrana, I. M. R. (2014). Psychometric properties of the big five inventory in a Mexican sample. *Salud mental*, 37(6), 491-497. <https://doi.org/10.17711/SM.0185-3325.2014.059>
- Salgado, E., Vargas-Trujillo, E., Schmutzler, J., & Wills-Herrera, E. (2016). Uso del Inventario de los Cinco Grandes en una muestra colombiana. *Avances en Psicología Latinoamericana*, 34(2), 365-382. <https://doi.org/10.12804/apl34.2.2016.10>
- 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. https://www.cienciarred.com.ar/ra/usr/35/856/racp_xviii_3_pp265_274.pdf

- Sekely, A., Taylor, G. J., & Bagby, R. M. (2018). Developing a short version of the Toronto Structured Interview for Alexithymia using item response theory. *Psychiatry Research*, 266, 218-227. <https://doi.org/10.1016/j.psychres.2018.03.002>
- 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), 20-30. <https://www.psychology.org.nz/journal-archive/Sibley-IPIP61.pdf>
- Sleep, C. E., Lynam, D. R., & Miller, J. D. (2021). A comparison of the validity of very brief measures of the Big Five/Five-Factor Model of personality. *Assessment*, 28(3), 739-758. <https://doi.org/10.1177/1073191120939160>
- Soto, C. J., & John, O. P. (2017). 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. <https://doi.org/10.1037/pspp0000096>
- Soto, C. J., & John, O. P. (2019). Optimizing the length, width, and balance of a personality scale: How do internal characteristics affect external validity? *Psychological Assessment*, 31(4), 444-459. <https://doi.org/10.1037/pas0000586>
- Stein, I., Asher, M., Erez, S., Shechner, T., Marom, S., Hermesh, H., & Aderka, I. M. (2019). Developing a brief version of the social thoughts and beliefs scale (STABS) using item response theory. *Cognitive Therapy and Research*, 43(4), 792-801. <https://doi.org/10.1007/s10608-019-10000-7>
- Suárez-Álvarez, J., Pedrosa, I., Lozano, L. M., García-Cueto, E., Cuesta Izquierdo, M., & Muñoz-Fernández, J. (2018). Using reversed items in Likert scales: A questionable practice. *Psicothema*, 30(2), 149-158. <https://doi.org/10.7334/psicothema2018.33>
- Sueiro, M. J., & Abad, F. J. (2009). Bondad de ajuste en ítems politómicos: tasas de error de tipo I y potencia de tres índices de ajuste. *Psicothema*, 21(4), 639-645. <https://www.psicothema.com/pdf/3684.pdf>
- Sutin, A. R., Stephan, Y., & Terracciano, A. (2018). Facets of conscientiousness and objective markers of health status. *Psychology & Health*, 33(9), 1100-1115. <https://doi.org/10.1080/08870446.2018.1464165>
- Thomas, M. L. (2019). Advances in applications of item response theory to clinical assessment. *Psychological Assessment*, 31(12), 1442-1455. <https://doi.org/10.1037/pas0000597>
- Ubbiali, A., Chiorri, C., Hampton, P., & Donati, D. (2013). Italian Big Five Inventory. Psychometric properties of the Italian adaptation of the Big Five Inventory (BFI). *Applied Psychology Bulletin*, 59(266), 37-48. <https://psycnet.apa.org/record/2013-22456-005>
- Weiss, A., & Costa Jr, P. T. (2014). Re: "Personality and all-cause mortality: individual-participant meta-analysis of 3,947 deaths in 76,150 adults". *American Journal of Epidemiology*, 179(6), 791-792. <https://doi.org/10.1093/aje/kwt300>
- Widiger, T. A., Gore, W. L., Crego, C., Rojas, S. L., & Oltmanns, J. R. (2015). Five Factor Model and Personality Disorder. En T. A. Widiger (Ed.), *The Oxford Handbook of the Five Factor Model* (pp. 449-478). Oxford Academic. <https://doi.org/10.1093/oxfordhb/9780199352487.013.4>
- 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. <https://doi.org/10.1027/1614-0001/a000148>