

## Consistencia Interna del Eysenck Personality Questionnaire - Revised: Cuando Alfa de Cronbach No Es Suficiente

### Internal Consistency of Eysenck Personality Questionnaire - Revised: When Cronbach's Alpha Is Not Sufficient

César Merino Soto<sup>1</sup>

#### Resumen

Debido a la importancia de obtener una apropiada estimación de la consistencia interna e integrar procedimientos psicométricos útiles en la medición de la personalidad, el presente estudio tiene por objetivo evaluar la confiabilidad por consistencia del Cuestionario Revisado de Personalidad de Eysenck (EPQR). Los participantes fueron 115 adultos jóvenes de Lima, universitarios. Se aplicó el EPQR de 83 ítems, del cual se derivaron los puntajes para la versión completa, y las versiones breves de 48 y 24 ítems (EPQR-S y EPQR-A, respectivamente). Se calculó el coeficiente KR20 con modificación Horst, el error estándar de medición condicional directo y estandarizado. Los resultados indican que los puntajes E y N muestran adecuado nivel de confiabilidad, mientras que P y L tienden a mostrar bajos coeficientes. Las versiones breves muestran las estimaciones más bajas, especialmente EPQR-A. Se concluye que el EPQR puede requerir una reconstrucción respecto a la escala P y L.

**Palabras clave:** Eysenck, personalidad, evaluación, confiabilidad, alfa de Cronbach

#### Abstract

Due to the importance of obtaining an appropriate estimate of internal consistency, and the integration of useful psychometric procedures in the measurement of personality, the present study aims to evaluate the consistency reliability of the Eysenck Revised Personality Questionnaire (EPQR). The participants were 115 young adults from Lima, university students. The EPQR of 83 items was applied, from which the scores for the full version were derived, and the brief versions of 48 and 24 items (EPQR-S and EPQR-A, respectively). The coefficient KR20 was calculated with Horst modification, the standard direct and conditional standard error of measurement. The results indicate that the Extraversion (E) and Neuroticism (N) scores show an adequate level of reliability, while P and L tend to show low coefficients. The short versions show the lowest estimates, especially EPQR-A. It is concluded that the EPQR may require restructure in the P and L scales.

**Keywords:** Eysenck, personality, assessment, reliability, Cronbach's alpha

<sup>1</sup> Magister en Psicología Educativa. Docente investigador. Instituto de Investigación, Universidad Católica los Ángeles de Chimbote (Chimbote, Peru). Dirección postal: Calle vieja 332, Condominio Bugambilia II, interior 116, Lomas Tzompantle, CP62157, Cuernavaca, Morelos, México. Correo: sikayax@yahoo.com.ar

## Introducción

La presente investigación se orienta a evaluar un parámetro psicométrico del Cuestionario de Personalidad de Eysenck Revisado (EPQR, Eysenck, & Barret, 1985; Eysenck & Eysenck, 2001) en una muestra peruana, el mismo que ha sido uno de los criterios principales de su construcción: la consistencia interna. La versión revisada (EPQR) se origina de la modificación que hicieron Eysenck y Barret (1985) a la escala de Psicoticismo del EPQ (Eysenck & Eysenck, 1975). Esta modificación trató de superar sus debilidades psicométricas referidas a su pobre consistencia interna, el bajo rango de puntuaciones en los sujetos y, consecuentemente, una forma distribucional severamente asimétrica positiva, que podría modelarse muy bien con una distribución teórica de tipo Poisson (Eysenck & Barret, 1985), este tipo distribución teórica ayuda a describir eventos extremos o raros en una población, y en que la mayor densidad ocurre en las magnitudes cerca de cero, con una larga cola hacia la derecha.

La consistencia interna es uno de los criterios para construir ítems que aporten al contenido de la prueba, y que permitan modelar satisfactoriamente el puntaje de interés, pues se relaciona con las correlaciones entre los ítems de un constructo (Muñiz, 1998; Tang, Cui, & Babenko, 2014; Nunnally & Bernstein, 1995). El análisis de ítems, en este contexto, es un esencial componente para maximizar la consistencia interna de las escalas (Hofstee, 2003; Nunnally & Bernstein, 1995), que resultará en ítems que reflejen comportamientos relativamente homogéneos, consistentes, y que produzcan correlaciones inter-ítem elevados (Nunnally & Bernstein, 1995; Feldt & Brennan, 1989). Este aspecto de la confiabilidad también ha servido como criterio para delimitar el número de componentes sustanciales en un análisis de componentes principales (Hofstee, 2003).

Desde la Teoría Clásica de los Tests, se conoce que este estadístico es dependiente de aspectos como el número de ítems, la correlación entre ellos, y la dispersión de las respuestas en la muestra (Nunnally & Bernstein, 1995; Ponterotto & Ruckdeschel, 2007; Ponterotto & Charter, 2009). Esto último sugiere la dependencia de los coeficientes de confiabilidad de la muestra de

participantes. La variabilidad de la muestra en algún aspecto que afecte a sus características estadísticas no solo se observa en la dispersión de las respuestas, sino también en los puntajes promedio; esta similaridad de los puntajes promedio de los ítems es una de las condiciones básicas para aplicar apropiadamente el coeficiente alfa o su equivalente para ítems dicotómicos, KR20 (Feldt & Brennan, 1989; Meyer, 2010). Específicamente, esta condición se refiere al modelo *equivalente tau*, en que se asume una similitud en los puntajes promedio de los ítems y una moderada dispersión de las respuestas; y el incumplimiento de este supuesto produce que la magnitud de los coeficientes disminuyan (sesgo negativo) y que solo se obtenga el límite inferior de la confiabilidad (Feldt & Charter, 2003a, 2003b; Gilmer & Feldt, 1983; Meyer, 2010).

La importancia de obtener valores precisos de la confiabilidad no es solo un interés metodológico, pues también tiene impacto aplicado. Por ejemplo, se conoce que la confiabilidad de la variables sometidas a un análisis estadístico afecta el poder estadístico de las pruebas inferenciales usadas (Charter & Feldt, 2001; Ponterotto & Charter, 2009), pues el poder estadístico disminuye con el decremento de la confiabilidad de los puntajes. Por otro lado, si se obtienen confiabilidad bajas debido a la aplicación de un coeficiente no correspondiente con las características de los datos, entonces se inflará el error de Tipo II en los resultados de las pruebas estadísticas. También, la estimación del puntaje verdadero se afectará, creando intervalos de confianza espuriamente altos o diferencias grandes entre el puntaje hipotetizado y el puntaje obtenido (Charter & Feldt, 2001).

Con respecto al EPQR, la práctica usual es calcular los coeficientes de estabilidad test-retest y de consistencia interna, pero sin alguna referencia del cumplimiento de sus presupuestos, o de algún ajuste relevante para atenuar este problema. Los resultados psicométricos de la consistencia interna en la primera adaptación hispana del EPQ-R en población española (Aguilar, Tous, & Andrés, 1990), arrojaron coeficientes  $\alpha$  satisfactorios, en donde los puntajes N tuvieron mayores coeficientes (alrededor de .90), y los puntajes de E y P presentaron similares magnitudes (alrededor de .83). Aguilar et al. (1990) usaron coeficientes  $\alpha$

(Cronbach, 1951), que es un estadístico equivalente a KR20 (Kuder & Richardson, 1937) en ítems dicotómicos. La consistencia interna en el manual oficial de la versión hispana (Eysenck & Eysenck, 2001) fue similar entre las escalas P y L (alrededor de .74), y entre las escalas E y N (.84); mientras que para la forma corta (48 ítems), se mantuvo este patrón de magnitudes para E y N (alrededor de .81), pero bajos para P y L (alrededor de .68).

Otras investigaciones han usado solo algunas de las escalas del EPQ-R, reportando también aceptables niveles de consistencia interna para los fines de sus estudios. Por ejemplo, el uso independiente de las escalas N ( $\alpha=.79$ ) y E ( $\alpha=.84$ ) en sujetos de población general ( $n=254$ ) fue considerada adecuada (Hervás, 2008), mientras que otros reportan confiabilidades ligeramente bajas (alrededor de .74, Ferrer et al., 2010). En la adaptación catalana ( $n=574$ , Ortet, Roglà, & Ibáñez, 2001), la consistencia interna varió entre .78 (P) y .97 (N); y en un reciente estudio con la misma versión ( $n=559$ , Marqués, Ibáñez, Ruipérez, Moya, & Ortet, 2005), los coeficientes fueron ligeramente menores pero manteniéndose el mismo patrón: las escalas E y N tienden a ser más elevadas comparadas con P y L (Caruso, Witkiewitz, Belcourt, & Gottlieb, 2001)

El objetivo del presente artículo es obtener estimaciones de confiabilidad por consistencia interna y compararlas, usando varios métodos que traten de caracterizarlo de una manera más completa. Para ambos objetivos, la aplicación de la modificación Horst (1953) al KR puede probar ser una mejor estimación de la consistencia de los puntajes del EPQR debido que toma en cuenta el grado de dispersión de la respuesta promedio a los ítems (Merino & Charter, 2010), aspecto que no ha sido valorado apropiadamente en la investigación aplicada (Charter, 1995). La estimación apropiada de la consistencia interna usando este ajuste (KR20-Horst) no ha sido reportada en las publicaciones que han tratado las propiedades psicométricas de puntajes del EPQR, y por lo tanto es probable que las estimaciones publicadas puedan ser subestimadas debido a la variabilidad esperable en la dificultad de los ítems del EPQR.

De acuerdo a lo explicado en los párrafos anteriores, si los ítems de los puntajes del EPQR

no muestran similares valores promedio, entonces esto afectaría a la magnitud de los coeficientes de confiabilidad, subestimándolos. Han aparecido unos pocos reportes aplicando estimaciones apropiadas del coeficiente KR20-Horst en pruebas neuropsicológicas (por ejemplo, Charter, 2000, 2001; Charter & Webster, 1997; Lopez, Charter, & Newman, 2000), pero se desconoce de su uso en pruebas de personalidad.

Por otro lado, el error estándar de medición (EMM) es un componente que permite la transferencia entre la estimación teórica del error de medición y su aplicación práctica en la interpretación del puntaje, ya que su métrica ocurre en las unidades del puntaje observado. La estimación básica es un parámetro único derivado del coeficiente de confiabilidad elegido, y el parámetro de dispersión, usualmente, la desviación estándar (Nunnally & Bernstein, 1995). Pero desde hace años también se ha reconocido que EEM puede variar a lo largo del puntaje observado en un instrumento, situación que origina al error estándar de medición condicional (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999). Por lo tanto, otro aspecto en que el presente estudio aportará es la estimación de la *consistencia interna condicional*, obteniendo en cada nivel del puntaje de las escalas y derivado del error estándar de medición condicional. Este método propuesto por Raju, Price, Oshima y Nering (2007) es una variación del modelo de la teoría clásica de los test, y en que se asume que el error no es constante en los diferentes niveles del puntaje (Keats, 1957; Raju et al., 2007), y por lo tanto, la expresión de la confiabilidad puede seguir esta redefinición. La obtención del error estándar de medición condicional (EEMC) ha sido reconocida por los Estándares (American Educational Research Association, et al., 1999) como una información fundamental que debe acompañar la interpretación de los puntajes, pero que aún es una práctica poco frecuente en pruebas de personalidad. Excepto el reporte de Geemp y Cuesta (2007) para la adaptación del EPQR en Chile, hasta la fecha no hay más trabajos en Iberoamérica que involucren específicamente la estimación del error de medición en el EPQR usando métodos alternativos (coeficientes) y el

presente estudio obtuvo estas estimaciones relevantes a un reporte más informado de la consistencia interna del EPQR.

El EPQR, como medida breve y representó un cambio que ayudó a configurar los tres grandes constructos de los trabajos de Eysenck, y parece aun frecuentemente utilizado en la práctica profesional y formativa en pregrado, e investigación, en Latinoamérica (e.g., Altimó & Simões, 2016), y específicamente Perú. Dado que no hay estudios recientes sobre el uso de pruebas en Latinoamérica y Perú, la experiencia anecdótica podría corroborar el uso de EPQR, o incluso la aplicación de su modelo teórico operacionalizado en instrumentos anteriores al EPQR (véase, Fumero & Navarrate, 2016). Si un usuario pretendo aproximar la medición de estos tres constructos, el EPQR puede ser una opción razonable, más aún con estudios recientes de su adaptación (Altimó & Simões, 2016; Dominguez, 2014; Dominguez, Villegas, Yauri, Aravena, & Ramírez 2013; Gemmp & Cuesta, 2007), pero en los que no se puede evitar ser precisos y estimar el monto de esta precisión psicométrica.

## Método

### Participantes

Los participantes fueron 115 estudiantes de psicología (varones=32, 27.8%) de una universidad privada de Lima Metropolitana. Los participantes provienen, de un amplio rango de distritos de Lima, pero puede considerarse un grupo básicamente heterogénea y con una tendencia al nivel socioeconómico medio. De acuerdo a la *Clasificación Internacional Normalizada de la Educación (CINE-97, UNESCO, 1997)*, el nivel educativo de los padres o apoderados se clasifica mayormente en los niveles de educación básica obligatoria completa (Nivel ISCED 2) y niveles post secundarios sin estudios universitarios (Nivel ISCED 3 y 4), y en mucho menor frecuencia, estudios universitarios (Nivel ISCED 5). La mayor proporción de mujeres en la muestra está en la línea de la tendencia internacional de la participación femenina en la carrera de psicología (Michalski, Kohout, Wicherski, & Hart, 2011). Todos los alumnos eran del mismo ciclo de estudios (mitad

de carrera), y la edad promedio fue 21.18 años (d.e.=3.19, min=17, máx=33 años); no hubo diferencias de edad respecto al sexo ( $t[45.36] = 1.89, p > .05$ ).

### Instrumento

El estudio analizó la consistencia interna de las tres versiones principales del EPQR: la forma completa (versión española oficial, EPQR-S; Eysenck & Eysenck, 2001) y dos versiones breves: *Cuestionario de Personalidad de Eysenck Revisado, forma corta* (EPQR-S; Eysenck & Eysenck, 2001) y *Cuestionario de Personalidad de Eysenck Revisado, Abreviado* (EPQR-A; Francis, Brown & Philipchalk, 1992). Todas estas versiones se responden en ítems en formato dicotómico (Sí/No), distribuyéndose aleatoriamente en el cuadernillo de preguntas; también, evalúan las cuatro dimensiones del EPQR: Neuroticismo (N), Psicoticismo (P), Extraversión (E) y Sinceridad (L). El EPQR es la versión completa para el habla hispana, adaptado en España; consta de 83 ítems. La consistencia interna reportada en la estandarización española fue .81 (E), .86 (N), .72 (P) y .76 (L). Coeficientes parecidos se hallaron en España (Aguilar et al., 1990) y Chile (Gemmp & Cuesta (2007), lo cual puede indicar la potencial generalización de estas estimaciones en distintas culturas. En Perú, la versión completa fue parte de estudios de adaptación (Dominguez, 2014; Dominguez et al., 2013; Merino-Soto, 2016).

El EPQR-S fue desarrollado de la versión completa de 83 ítems, y presentado en el mismo manual; consta de 48 ítems (12 ítems en cada escala). Por otro lado, el EPQR-A tiene 24 ítems, 6 en cada subescala; y fue desarrollada desde el EPQR-S. Su adaptación española demostró buenas cualidades estructurales y de validez externa luego de algunas modificaciones a los ítems (Sandín, Valiente, Chorot, Olmedo, & Santed, 2002a; Sandín, Valiente, Olmedo, Chorot, & Santed, 2002b). La consistencia reportada en Sandín et al. (2002b) para E, N, P y L fue, respectivamente, .74, .78, .63, y .54.

### Procedimiento

La aplicación de la prueba se hizo durante el desarrollo normal del semestre académico y en el horario regular de clases. La administración de las

pruebas mantuvo un procedimiento uniforme (Merino, 2010), y de acuerdo con las recomendaciones generales y normativas para la aplicación de pruebas (International Test Commission Test, 2000). Esta consideración fue importante para minimizar en lo posible la varianza irrelevante al constructo proclive a ocurrir durante la administración de las pruebas (McCallin, 2006).

Previamente al análisis de la consistencia interna, se evaluaron univariadamente el sesgo potencial en los coeficientes, debido a la dispersión de las varianzas y respuestas promedio a los ítems; esto se hizo estimando la magnitud de las diferencias de las medias y desviación estándar de los ítems. De este modo, la magnitud de la diferencia en las varianzas de los ítems se hizo dividiendo la DE mayor sobre la DE menor; el resultado de esta operación se compararía con las recomendaciones de Feldt y Charter (2003a), y Helms, Henze, Sass y Mifsud (2006), que indican que una diferencia entre las desviaciones estándares mayor al 30% sugiere una dispersión sustancial de la varianza de los ítems. Respecto a la dispersión de la media de los ítems, se obtuvo la diferencia entre la media de mayor magnitud y la de menor magnitud, previa transformación de estas medias (básicamente, proporciones) a una variable de intervalo mediante la función

arcoseno: 
$$h = \left(2 \operatorname{sen} \sqrt{p_1}\right) - \left(2 \operatorname{sen} \sqrt{p_2}\right)$$
. Esta transformación permite estandarizar la diferencia de las medias de ítems dicotómicos (Cohen, 1992).

Aunque la obtención de las confiabilidades es indicada para el grupo total y la submuestras que la componen American Educational Research Association et al., 1999; Merino, 2011a, 2011b), los cálculos se harán únicamente para la muestra total, pues el tamaño muestral para los varones es pequeño y produciría mayor error muestral (Charter, 2003b). Se calculó el coeficiente de consistencia interna KR20 (Kuder & Richardson, 1937), sin corrección y con corrección (Horst, 1953; Merino & Charter, 2010); éste último toma en cuenta la dispersión de las medias de los ítems en cada escala. Sus intervalos de confianza (95%) se calcularon por el método de Fisher (1950), que logra una mejor cobertura en distribuciones no normales

(Romano, Kromrey, & Hibbard, 2010; Romano, Kromrey, Owens, & Scott, 2011).

Se obtuvo el error estándar de medición condicional (Conditional Standard Error of Measurement: CSEM) para cada puntaje, aplicando el modelo de error binomial de Lord (1955), con la modificación de Keats (1957). Este procedimiento también fue aplicado en el estudio de Gempp y Cuesta (2007), y asume que los ítems han sido extraídos aleatoriamente de un universo de ítems. La ecuación del CSEM es:

$$\sigma_E = \sqrt{\left(\frac{X(k-X)}{k-1}\right)\left(\frac{1-r_{xx}}{1-r_{21}}\right)}$$

La corrección Keats (1957) usa el coeficiente KR21 y otro método de consistencia interna, y para el presente análisis se usará el KR20-Horst. Finalmente, se aplicará una estimación de la confiabilidad condicional ( $\alpha_{CSEM}$ ), método propuesto por Raju et al. (2007) para expresar estandarizadamente la precisión en cada nivel del puntaje, usando como base de su cálculo el CSEM obtenido.

## Resultados

### Descripción univariada de los puntajes

En la Tabla 1, se observa que la locación del puntaje promedio en las versiones del EPQR indica que los participantes dan más respuestas afirmativas a la escala E, lo cual se confirma en su severa asimetría negativa. Por otro lado, las escalas P y N contienen más respuestas que niegan el atributo medido; comparativamente, la media de L es cercano al centro de su rango completo de puntaje. También se observa que la escala P muestra un fuerte efecto de techo, problema que es más severo en las versiones breves; por otro lado, E parece mostrar un leve efecto de piso.

Respecto a la variabilidad, las versiones breves tienen menos variabilidad que la versión completa (un resultado esperado), pero la variabilidad siempre es menor en las escalas P y L. Distribucionalmente, los puntajes son más densos hacia los extremos en E, P y N; y L es comparativamente más simétrico.

Tabla 1. Estadísticos descriptivos para las versiones del EPQR

	Estadísticos descriptivos					
	M	DE	Min	Máx	As	Cu
EPQR-83						
E (19)	14.278	3.655	4	19	-1.031	.556
P (23)	5.130	2.795	0	12	.320	-.613
N (23)	8.634	5.416	1	23	.652	-.287
L (18)	8.373	3.285	0	17	-.006	.184
EPQR-48						
E (12)	8.939	2.623	1	12	-1.235	1.141
P (12)	2.904	1.982	0	8	.575	-.469
N (12)	3.556	3.162	0	12	.996	.186
L (12)	5.460	2.514	0	12	.033	-.395
EPQR-24						
E (6)	4.713	1.400	0	6	-1.326	1.425
P (6)	1.634	1.141	0	5	.255	-.552
N (6)	2.165	1.669	0	6	.769	-.222
L (6)	2.826	1.718	0	6	-.031	-.901

Nota. E: Extroversión. P: Psicoticismo. N: Neuroticismo. L: Mentira. As y Cu: coeficiente de asimetría y curtosis de Fisher. El error estándar del coeficiente de asimetría (As.) y curtosis es .226 y .447, respectivamente.

Tabla 2. Sumario de dificultad de los ítems en cada forma del EPQR

Media	EPQR-83			EPQR-48			EPQR-24		
	M <sub>max</sub>	M <sub>min</sub>	.h	M <sub>max</sub>	M <sub>min</sub>	.h	M <sub>max</sub>	M <sub>min</sub>	.h
E	.91	.35	1.469	.913	.356	1.263	.9130	.6696	.626
P	.74	.01	1.871	.530	.078	1.064	.5304	.0435	1.211
N	.83	.16	1.469	.478	.156	.714	.5913	.2087	.806
L	.84	.10	1.675	.756	.173	1.249	.6087	.3826	.456
DE	DE <sub>max</sub>	DE <sub>min</sub>	F.	DE <sub>max</sub>	DE <sub>min</sub>	F.	DE <sub>max</sub>	DE <sub>min</sub>	F.
E	.488	.283	1.724	.481	.283	1.699	.472	.283	1.669
P	.502	.093	5.397	.501	.269	1.859	.501	.204	2.449
N	.502	.365	1.375	.501	.364	1.374	.501	.408	1.229
L	.502	.093	5.397	.502	.380	1.319	.502	.488	1.028

Nota. M<sub>max</sub> y M<sub>min</sub>: media máxima y mínima. E: Extroversión. P: Psicoticismo. N: Neuroticismo. L: Mentira. DE<sub>max</sub> y DE<sub>min</sub>: desviación estándar máxima y mínima. F: razón DE<sub>max</sub> y DE<sub>min</sub>. h: magnitud del efecto entre medias de proporciones (ver texto).

### Exploración de los ítems

En la Tabla 2 se reportan la M y DE máximo y mínimo de los ítems en cada versión del EPQR. Tomando en cuenta las sugerencias de Cohen (1992) para valorar las diferencias estandarizadas (.20, .50 y .80, para pequeño, moderado y grande, respectivamente), hay grandes discrepancias de las medias de los ítems en todas las versiones del EPQR. Con respecto a la desviación estándar, P y L tienden a presentar ítems con mucha restricción de la dispersión, coherente con la marcada asimetría de la distribución de sus puntajes. En resumen, se observan gran heterogeneidad en los ítems respecto a la respuesta promedio y variabilidad en las escalas del EPQR, sugiriendo que la aplicación rutinaria del KR20 no sería apropiada.

### Confiabilidad de los puntajes

En la correlación inter-ítem promedio (Tabla 3), P fue generalmente baja y más cerca de cero, mientras que en los otros puntajes fueron iguales o mayores a .11. Los puntajes E y N han mostrado valores  $r_{ii}$  generalmente altos comparado con L y P, por lo tanto es claro que entre sus ítems hay mayor covariación comparado con los ítems de L y P. Los coeficientes KR20 (Tabla 3) siempre fueron más bajos en comparación con el KR20-H. En cada versión, la magnitud de la consistencia interna de los puntajes ha mantenido un patrón inalterable, pues el orden de la magnitud de la confiabilidad fue (de mayor a menor magnitud): N, E, L y P. Considerando la variación por el muestreo (ver los intervalos de confianza), la variación de la magnitud de estos coeficientes de E y N está sobre .70 en el EPQR-83 y EPQR-48, pero no en EPQR-24. También se puede observar

Tabla 3. Estimaciones de consistencia interna para las versiones del EPQR (83, 48 y 24 ítems)

	$R_{ii}$	$KR_{21}$	$KR_{20}$	$KR_{20-H}$	I.C. 95%	EEM
EPQR-83						
E (19)	.177	.775	.795	.816	[.763, .861]	1.568
P (23)	.068	.512	.619	.647	[.546, .733]	1.661
N (23)	.225	.853	.869	.885	[.852, .913]	1.837
L (18)	.110	.619	.692	.723	[.643, .791]	1.729
EPQR-48						
E (12)	.222	.729	.763	.803	[.745, .852]	1.164
P (12)	.097	.480	.551	.583	[.460, .687]	1.280
N (12)	.288	.818	.827	.849	[.804, .886]	1.229
L (12)	.123	.577	.631	.664	[.565, .748]	1.457
EPQR-24						
E (6)	.224	.581	.608	.659	[.559, .744]	.818
P (6)	.081	.104	.336	.402	[.227, .551]	.882
N (6)	.243	.604	.652	.708	[.622, .781]	.902
L (6)	.201	.592	.602	.643	[.538, .732]	1.026

Nota. E: Extroversión. P: Psicoticismo. N: Neuroticismo. L: Mentira.  $R_{ii}$ : correlación inter-ítem.  $KR_{21}$ : coeficiente de confiabilidad  $KR_{21}$ .  $KR_{20}$ : coeficiente de confiabilidad  $KR_{20}$ .  $KR_{20-H}$ : coeficiente de confiabilidad  $KR_{20}$  con modificación Horst. EEM: error estándar de medición.

Tabla 4. Error estándar de medición y confiabilidad condicionales para el EPQR (83 ítems)

	E		P		N		L	
	CSEM	< csem	CSEM	< csem	CSEM	< csem	CSEM	< csem
0	-	-	-	-	-	-	-	-
1	.905	.939	.851	.933	.885	.973	.853	.933
2	1.243	.884	1.175	.872	1.223	.949	1.171	.873
3	1.477	.837	1.405	.817	1.462	.927	1.388	.821
4	1.652	.796	1.581	.768	1.645	.908	1.548	.778
5	1.784	.762	1.720	.726	1.790	.891	1.668	.742
6	1.883	.735	1.831	.689	1.906	.876	1.756	.714
7	1.954	.714	1.919	.659	1.997	.864	1.816	.694
8	2.000	.700	1.986	.634	2.067	.854	1.851	.683
9	2.023	.694	2.036	.616	2.119	.847	1.862	.679
10	2.023	.694	2.068	.604	2.152	.842	1.851	.683
11	2.000	.700	2.083	.598	2.168	.840	1.816	.694
12	1.954	.714	2.083	.598	2.168	.840	1.756	.714
13	1.883	.735	2.068	.604	2.152	.842	1.668	.742
14	1.784	.762	2.036	.616	2.119	.847	1.548	.778
15	1.652	.796	1.986	.634	2.067	.854	1.388	.821
16	1.477	.837	1.919	.659	1.997	.864	1.171	.873
17	1.243	.884	1.831	.689	1.906	.876	.853	.933
18	.905	.939	1.720	.726	1.790	.891	-	-
19	-	-	1.581	.768	1.645	.908	-	-
20	-	-	1.405	.817	1.462	.927	-	-
21	-	-	1.175	.872	1.223	.949	-	-
22	-	-	.851	.933	.885	.973	-	-
23	-	-	-	-	-	-	-	-

Nota. CSEM: error estándar de medición condicional (Conditional Standard Error Measurement); < csem: confiabilidad condicional. E: Extroversión. P: Psicoticismo. N: Neuroticismo. L: Mentira.

que la consistencia interna de cada subescala del EPQR-24 comparado con el EPQR-83 y EPQR-48 fue siempre más baja. Dado que el coeficiente  $KR_{20}$  es afectado por el número de ítems, el número de ítems en las subescalas del EPQR24 no parece ser suficiente para obtener puntajes satisfactoriamente confiables. Las estimaciones usando  $KR_{21}$  son consistentemente bajas en comparación con  $KR_{20}$  y  $KR_{20-H}$ , como es

predicible; sin embargo, para E y N, las diferencias entre ambos coeficientes tienden a ser pequeñas comparadas con P y L. Las discrepancias entre las estimaciones  $KR_{20}$  y  $KR_{20-H}$  claramente indican que la usual aplicación de  $KR_{20}$  produce una infraestimación de la consistencia interna.

Tabla 5. Error estándar de medición y confiabilidad condicionales para el EPQR-48 y EPQR-24

	E		P		N		L	
	CSEM	< $\alpha_{csem}$	CSEM	< $\alpha_{csem}$	CSEM	< $\alpha_{csem}$	CSEM	< $\alpha_{csem}$
EPQR-48								
0	-	-	-	-	-	-	-	-
1	.853	.894	.895	.796	.911	.917	.892	.874
2	1.150	.808	1.207	.629	1.228	.849	1.202	.771
3	1.337	.740	1.402	.499	1.427	.796	1.397	.691
4	1.455	.692	1.527	.407	1.553	.759	1.521	.634
5	1.522	.663	1.597	.351	1.624	.736	1.590	.600
6	1.543	.654	1.619	.332	1.647	.729	1.613	.588
7	1.522	.663	1.597	.351	1.624	.736	1.590	.600
8	1.455	.692	1.527	.407	1.553	.759	1.521	.634
9	1.337	.740	1.402	.499	1.427	.796	1.397	.691
10	1.150	.808	1.207	.629	1.228	.849	1.202	.771
11	.853	.894	.895	.796	.911	.917	.892	.874
12	-	-	-	-	-	-	-	-
EPQR-24								
0	-	-	-	-	-	-	-	-
1	.902	.585	.817	.487	.859	.735	.936	.703
2	1.141	.336	1.033	.180	1.086	.577	1.184	.525
3	1.210	.252	1.096	.077	1.152	.524	1.255	.466
4	1.141	.336	1.033	.180	1.086	.577	1.184	.525
5	.902	.585	.817	.487	.859	.735	.936	.703
6	-	-	-	-	-	-	-	-

Nota. CSEM: error estándar de medición condicional (Conditional Standard Error Measurement);  $\alpha_{csem}$ : confiabilidad condicional. E: Extroversión. P: Psicoticismo. N: Neuroticismo. L: Mentira.

### Confiabilidad condicional

La distribución de CSEM y  $\alpha_{csem}$  muestra una forma de “U”, característica recurrente en este tipo de cálculo (Meyer, 2010). Se observa que el error de medición tiende a elevarse en los extremos de los puntajes, mientras que desciende en alrededor de la media. P y L muestran los niveles de confiabilidad más bajos comparados (menos de .70) con las otras subescalas. La baja magnitud teórica de la confiabilidad fue aún más seria en las versiones breves de P y L, en que llegan a niveles no recomendables para la práctica profesional y de investigación (debajo y alrededor de .60).

### Discusión

Los resultados de la presente investigación revelan que los puntajes del EPQR y sus versiones cortas, tienen cuestionable confiabilidad por consistencia interna para los puntajes de P y L; en cambio las escalas E y N presentan mejor consistencia interna, pero comparativamente algo bajos respecto a la literatura internacional. Un examen formal de esta discrepancia es necesario con una muestra de mayor tamaño, para reducir el error de muestreo. Los resultados son peores para

la versión más abreviada (EPQR24; Francis et al., 1992), lo que supone una limitación para su uso en aplicaciones clínicas, e inclusive para investigación dado el mayor monto de error de medición alrededor de sus puntajes. También se halló un consistente patrón de coeficientes bajos para P, un problema corroborado en los estudios realizados en diferentes culturas (por ejemplo, Caruso et al., 2001; Tiwari, Singh, & Singh, 2012).

Hay varios motivos para que la consistencia interna muestre valores debajo de los recomendados, y uno de ellos son las propiedades distribucionales de los ítems. En ese sentido, la tendencia de respuesta hallada (o dificultad del ítem) para estos ítems problemáticos tendieron a no satisfacer los criterios que se usaron para la selección de ítems en la adaptación española del EPQR (Eysenck & Eysenck, 2001), efectuada también en muestra universitaria. De este modo, en la versión larga, las escalas E, N y L tuvieron ítems con valores  $p$  mayor a .80 y menor a .20. En la forma corta EPQR-48, E tuvo ítems altamente frecuentes; por el contrario, este patrón fue opuesto para P, N y L. Finalmente, en la versión EPQR-24, el  $p$  máximo y mínimo fue de menor magnitud relativo a las otras formas (EPQR-83 y



EPQR-48), pero de similar patrón a la forma EQPR-48. En este punto, y con respecto a la dificultad de los ítems (p), los puntajes parecen violar la presunción de homogeneidad de la respuesta media a los ítems, y por lo tanto no cumplen con el modelo equivalente tau sobre el cual de fundamenta el coeficiente alfa (Cronbach, 1951).

La consistencia interna de P y L puede sugerir también problemas de unidimensionalidad, pero está relacionado a ella (Delboy et al., 2011; Meyer, 2010). La baja consistencia interna de P sugiere ítems estadística y, posiblemente, también conceptualmente heterogéneos. Esto llevaría a suponer que en la muestra de estudio, una sola dimensión podría no ser suficiente para explicar la varianza de sus ítems; sin embargo, se requiere la confirmación de esta conjetura en una futura investigación. También debe considerarse que, como varios los ítems de P son controversiales en su contenido, los sujetos pueden generar respuestas aquiescentes, específicamente hacia el rechazo de estos ítems (Hanley, 1965).

Los resultados obtenidos sugieren también algunos aspectos del uso de la adaptación española del EPQR en otro contexto. La consistencia interna hallada no ha replicado satisfactoriamente lo reportado en estudios españoles y latinoamericanos (por ejemplo, Gempp & Cuesta, 2007). Aunque la presente muestra de estudio no es representativa de la población peruana, se puede valorar los resultados como una aproximación aceptable para sugerir el inicio de un proceso de re-adaptación del EPQR. Esta adaptación debería procurar una mayor variabilidad de las respuestas en el rango completo del puntaje en P y L, pues su impacto sobre confiabilidad deberá ser positivo (Ponterotto & Ruckdeschel, 2007).

Por otro lado, el estudio efectuó ajustes razonables que mejoraron las estimaciones de confiabilidad; y que podrían ser de valor metodológico en futuras aplicaciones del instrumento. Por ejemplo, si la variabilidad de las medias de los ítems no es un problema, entonces KR20 y KR20-Horst deberían arrojar el mismo resultado, por lo tanto el usuario aplicaría rutinariamente la modificación Horst al KR20. La modificación de Horst al KR20 asume que el modelo subyacente a los ítems es aún equivalente

tau, pero que ocurren variaciones en la respuesta media necesitan ser ajustadas para estimar apropiadamente la confiabilidad (Merino & Charter, 2010). Sin embargo, al detectarse las diferencias en la varianza de los ítems, una mejor estimación de la consistencia interna debe considerar otros modelos, como el modelo congénico (Feldt & Brennan, 1989; Meyer, 2010). Un ejemplo, de este coeficiente es el propuesto por Gilmer y Feldt (1983).

Se debe observar también que las distribuciones de los puntajes del EPQR en nuestros participantes generalmente han variado desde el nivel de moderada a severa asimetría, y con variados grados de leptocurtosis (frecuente en P) y platicurtosis (a menudo en E). Varios de estos puntajes claramente no pueden identificarse como distribuciones normales o aún cercanas a ellas. Parece que la forma como se distribuyen los puntajes del EPQR y sus versiones puede recordar las limitaciones del EPQ, hacia la cual las críticas y sus propios autores (Eysenck et al., 1985) reconocieron como debilidades psicométricas. Estos aspectos fueron la distribución asimétrica, baja confiabilidad y la restricción del rango de los puntajes (Eysenck et al., 1985), aspectos que limitan la utilidad de un instrumento psicométrico (Feldt & Brennan, 1989; Helms et al., 2006; Nunnally & Bernstein, 1995).

Una limitación del presente estudio es que el aspecto evaluado en este artículo (confiabilidad por consistencia interna) solo es un tipo de cuantificación del error de medición, una faceta de otros métodos de confiabilidad que deben ser estimados para una mejor comprensión e interpretación de los puntajes obtenidos en el EPQR. Otra limitación es que no se verificó formalmente si el modelo de medición equivalente tau es apropiado para calcular el KR20. Sin embargo, se adoptaron criterios prácticos para obtener un respaldo cuantitativo en la identificar la heterogeneidad de las varianzas y medias de los ítems, los mismos que sirvieron como una aproximación razonable del modelo de medición subyacente de los ítems (Feldt & Charter, 2003; Helm et al., 2006). Por otro lado, es probable que la muestra de estudiantes (estudios de pregrado en psicología) del presente estudio influyó para restringir la varianza de los ítems y de los puntajes, pues este grupo puede considerarse

homogéneo. Pero esta impresión es más intuitiva que fáctica, pues algunos hallazgos han revelado que las muestras de estudiantes producen coeficientes más elevados comparados con otras muestras (Caruso et al., 2001). Involucrada con la anterior limitación, la representación de la muestra no está asegurada, y el alcance de las conclusiones debe considerar esto.

Un futuro estudio en una muestra de mayor tamaño (por ejemplo, mayor o igual a 300; Charter, 2003a, 2003b) ayudará una mejor precisión de los coeficientes de confiabilidad y tener conclusiones más robustas. Finalmente, el posible efecto del método no fue analizado, y se desconoce si su impacto puede alterar la varianza de error estimada, y el monto de esta posible influencia. Este problema del efecto del método parece ser habitual en escalas que contienen ítems fraseados positiva y negativamente (Tomas, Meléndez, Oliver, Navarro, & Zaragoza, 2010), y se requerirá modelarlo con métodos avanzados.

Dada la importancia de evaluar los atributos psicológicos y de generar instrumentos con la mejor precisión posible, se requiere que la versión EPQR posiblemente sea re-adaptado hacia un lenguaje hispano que tome en cuenta las diferencias culturales como un plausible paso para mejorar o reducir el error de medición.

## Referencias

- Aguilar, A., Tous, J. M., & Andrés-Pueyo, A. (1990). Adaptación y estudio psicométrico del EQP-R. *Anuario de Psicología*, *46*, 101-118.
- Almiro, P. A., & Simões, M. R. (2016). Estudo das Propriedades Psicométricas da Versão Experimental Portuguesa do Questionário de Personalidade de Eysenck – Forma Revista (EPQ-R). *Revista Iberoamericana de Diagnóstico y Evaluación – e Avaliação Psicológica*, *41*(1), 159-173.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Barrett, P., & Eysenck, S. B. G. (1984). The assessment of personality factors across twenty-five countries. *Personality and Individual Differences*, *5*, 615-632. [https://doi.org/10.1016/0191-8869\(84\)90110-7](https://doi.org/10.1016/0191-8869(84)90110-7)
- Brennan, R. L. (2001). An essay on the history and future of reliability from the perspective of replications. *Journal of Educational Measurement*, *38*(4), 295-317. <https://doi.org/10.1111/j.1745-3984.2001.tb01129.x>
- Caruso, J. C., Witkiewitz, K., Belcourt, A., & Gottlieb, J. D. (2001). Reliability of scores from the Eysenck Personality Questionnaire: A reliability generalization study. *Educational and Psychological Measurement*, *61*, 675-689. <https://doi.org/10.1177/00131640121971437>
- Charter, R. A. (1995). The under-representation of Horst's modification of the KR-20 reliability coefficient. *Perceptual and Motor Skills*, *81*, 770. <https://doi.org/10.2466/pms.1995.81.3.770>
- Charter, R. A. (2000). An alternate short form of the Speech-Sounds Perception Test. *Perceptual and Motor Skills*, *90*(2/3), 1184-1186. <https://doi.org/10.2466/pms.90.3.1184-1186>
- Charter, R. A. (2001). Speech-sounds perception test: Long- and shortforms reliability adjusted for item difficulty. *Perceptual and Motor Skills*, *92*(1), 31-34. <https://doi.org/10.2466/pms.2001.92.1.31>
- Charter, R. A. (2003a). A breakdown of reliability coefficients by test type and reliability method, and clinical implications of low reliability. *The Journal of General Psychology*, *30*(3), 290-304. <https://doi.org/10.1080/00221300309601160>
- Charter, R. A. (2003b). Study samples are too small to produce sufficiently precise reliability coefficients. *The Journal of General Psychology*, *130*(2), 117-129.
- Charter, R. A. (2007). A practical use for the KR-21 reliability coefficient. *Psychological Reports*, *101*, 673-674. <https://doi.org/10.1080/00221300309601280>
- Charter, R. A., & Feldt, L. S. (2001). Meaning of reliability in terms of correct and incorrect clinical decisions: The art of decision making is still alive. *Journal of Clinical and Experimental Neuropsychology*, *23*, 530-537. <https://doi.org/10.1076/jcen.23.4.530.1227>

- Charter, R. A., & Webster, J. S. (1997). Psychometric structure of the Seashore Rhythm Test. *The Clinical Neuropsychology*, *11*(2), 167-173. <https://doi.org/10.1080/13854049708407046>
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155-159. <https://doi.org/10.1037//0033-2909.112.1.155>
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *6*, 297-334. <https://doi.org/10.1007/bf02310555>
- Delboy, A., Black, C., & Merino, C. (2011). Consistencia interna del Inventario de Resolución de Problemas Sociales-Revisado. *Revista Científica*, *8*(2), 134-143.
- Dominguez, S., Villegas, G., Yauri, C., Aravena, S. & Ramírez F. (2013). Análisis psicométrico preliminar de la forma corta del EPQ-R en estudiantes universitarios. *Avances en Psicología*, *21*(1), 73-82.
- Dominguez, S. (2014). Análisis psicométrico preliminar del EPQ-R en estudiantes universitarios peruanos. *Revista de Psicología de Arequipa*, *4*(1), 35-46.
- Eysenck, H. J., & Eysenck, S. B. G. (1975). *Manual of the Eysenck Personality Questionnaire*. London: Hodder & Stoughton.
- Eysenck, H. J., & Eysenck, S. B. G. (2001). *EPQ-R, Cuestionario Revisado de Personalidad de Eysenck (2da ed.)*. Madrid: TEA.
- Feldt, L. S., & Brennan, R. L. (1989). Reliability. In R. L. Linn (Ed.). *Educational measurement*, (3rd ed., pp. 105-146). New York: American Council on Education Measurement.
- Feldt, L. S., & Charter, R. A. (2003a). Estimating the reliability of a test split into two parts of equal or unequal length. *Psychological Methods*, *8*, 102-109. <https://doi.org/10.1037/1082-989x.8.1.102>
- Feldt, L. S., & Charter, R. A. (2003b). Estimation of internal consistency reliability when test parts vary in effective length. *Measurement and Evaluation in Counseling and Development*, *36*, 23-27.
- Ferrer, M., Carbonell, X., Josep, J., Cebrià, J., Virgili, C., & Castellana, M. (2010). Distinguishing male juvenile offenders through personality traits, coping strategies, feelings of guilt and level of anger. *The Spanish Journal of Psychology*, *13*(2), 751-764. <https://doi.org/10.1017/s1138741600002419>
- Fisher, R. A. (1950). *Statistical methods for research workers*. Edinburgh, UK: Oliver & Boyd.
- 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*, 443-449. <https://doi.org/10.1037/t08986-000>
- Fumero, A., & Navarrete, G. (2016). Personalidad y bienestar psicológico: aplicación de un modelo de redes neuronales. *Revista Iberoamericana de Diagnóstico y Evaluación – e Avaliação Psicológica*, *41*(1), 28-38.
- Gemmp, R., & Cuesta, S. (2007). Errores estándar de medida condicionales para las normas metropolitanas de la adaptación chilena del EPQ-R: Aplicación de un modelo binomial a un test de personalidad. *Terapia Psicológica*, *25*(1), 51-62. <https://doi.org/10.4067/s0718-48082007000100004>
- Gilmer, J. S., & Feldt, L. S. (1983). Reliability estimation for a test with parts of unknown length. *Psychometrika*, *48*, 99-111. <https://doi.org/10.1007/bf02314679>
- Hanley, C. (1965). Personality item difficulty and acquiescence. *Journal of Applied Psychology*, *49*(3), 205-208. <https://doi.org/10.1037/h0022107>
- Helms, J. E., Henze, K. T., Sass, T. L., & Mifsud, V. A. (2006). Treating Cronbach's alpha reliability coefficients as data in counseling research. *The Counseling Psychologist*, *34*(5), 630-660. <https://doi.org/10.1177/0011000006288308>
- Hervás, G. (2008). Adaptación al castellano de un instrumento para evaluar el estilo rumiativo: La escala de respuestas rumiativas. *Revista de psicopatología y Psicología Clínica*, *13*(2), 111-121. <https://doi.org/10.5944/rppc.vol.13.num.2.2008.4054>
- Hofstee, W. K. B. (2003). Structures of personality traits. In I. B. Weiner (Series Ed.) & T. Millon & M. Lerner (Vol. Eds.), *Handbook of Psychology: Vol. 5. Personality and Social Psychology* (pp. 231-254).

- Hoboken, NJ: Wiley.  
<https://doi.org/10.1002/0471264385.wei0510>
- Hogan, T. P., Benjamin, A., & Brezinski, K. L. (2000). Reliability methods: A note on the frequency of use of various types. *Educational and Psychological Measurement*, *60*, 523-531.  
<https://doi.org/10.1177/00131640021970691>
- Horst, P. (1953). Correcting the Kuder-Richardson Reliability for dispersion of item difficulties. *Psychological Bulletin*, *50*, 371-374. <https://doi.org/10.1037/h0062012>
- International Test Commission (ITC) (2000). *Guidelines on Test Use: Spanish Version*. ITC: Author.
- Keats, J. A. (1957). Estimation of error variances in test scores. *Psychometrika*, *22*, 29-41.  
<https://doi.org/10.1007/bf02289207>
- Kline, P. (1993). *Personality: the psychometric view*, Londres, Routledge.
- Kuder, G., & Richardson, M. (1937). The theory of estimation of test reliability. *Psychometrika*, *2*, 151-160.  
<https://doi.org/10.1007/bf02288391>
- Lopez, M. N., Charter, R. A., & Newman, R. J. (2000). Psychometric properties of the Halstead Category Test. *Clinical Neuropsychologist*, *14*(2), 157-161.  
[https://doi.org/10.1076/1385-4046\(200005\)14:2;1-z;ft157](https://doi.org/10.1076/1385-4046(200005)14:2;1-z;ft157)
- Lord, F. M. (1955). Estimating test reliability. *Educational and Psychological Measurement*, *15*, 325-336. <https://doi.org/10.1177/001316445501500401>
- Marqués, M. J., Ibáñez, M. I., Ruipérez, M. A., Moya, J., & Ortet, G. (2005). The Self-Regulation Inventory (SRI): Psychometric properties of a health related coping measure. *Personality and Individual Differences*, *39*, 1043-1054. <https://doi.org/10.1016/j.paid.2005.02.030>
- Merino, C., & Charter, R. (2010). Modificación Horst al coeficiente KR-20 por dispersión de la dificultad de los ítems. *Interamerican Journal of Psychology*, *44*(2), 274-278.
- Merino, C. (2010). *Lista de chequeo para la administración grupal de cuestionarios*. Documento no publicado. Universidad Científica del Sur.
- Merino, C. (2011a). Consistencia interna de una escala de evaluación de las estrategias de afrontamiento del estrés. *Revista Científica*, *4*(1), 260-271.
- Merino, C. (2011b, Julio). *Ayudando al coeficiente alfa de Cronbach: Nuevos ingredientes para el análisis de la confiabilidad*. Ponencia presentada en el XV Congreso Nacional de Psicología, 21-23 de julio, Lima.
- Merino-Soto, C. (2016). Percepción de la claridad de los ítems: Comparación del juicio de estudiantes y jueces-expertos. *Revista Latinoamericana de Ciencias Sociales, Niñez y Juventud*, *14*(2), 1469-1477.
- Meyer, P. (2010). *Understanding measurement: Reliability*. New York: Oxford University. <https://doi.org/10.1093/acprof:oso/9780195380361.001.0001>
- Michalski, D., Kohout, J., Wicherski, M., & Hart, B. (2011). 2009 Doctorate Employment Survey. Washington, DC: APA Center for Workforce Studies. Disponible en: <http://www.apa.org/workforce/publications/09-doc-empl/report.pdf>.
- Muñiz, J. (1998). *Teoría clásica de los tests*. Madrid: Pirámide.
- Nunnally, J. C., & Bernstein, I. J. (1995). *Teoría Psicométrica* (3ra ed.). México, D. F: McGraw-Hill.
- Ortet, G., Roglà, R. M., & Ibáñez, M. I. (2001). Adaptació catalana del qüestionari revisat de personalitat. In H. J. Eysenck & S. B. G. Eysenck (Eds.), *EPQ-R: Cuestionario Revisado de Personalidad de Eysenck (2da ed.)*. Madrid: TEA.
- Ponterotto, J. G., & Ruckdeschel, D. E. (2007). An overview of coefficient alpha and a reliability matrix for estimating adequacy of internal consistency coefficients with psychological research measures. *Perceptual and Motor Skills*, *105*, 997-1014.  
<https://doi.org/10.2466/pms.105.3.997-1014>
- Ponterotto, J. G., & Charter, R. A. (2009). Statistical extensions of Ponterotto and Ruckdeschel's (2007) reliability matrix for estimating the adequacy of internal consistency coefficients. *Perceptual and Motor Skills*, *108*, 878-886.  
<https://doi.org/10.2466/pms.108.3.878-886>
- Raju, N. S., Price, L. R., Oshima, T. C., & Nering, M. L. (2007). Standardized conditional SEM:

- A case for conditional reliability. *Applied Psychological Measurement*, 31, 169-180.  
<https://doi.org/10.1177/0146621606291569>
- Romano, J. L., Kromrey, J. D., & Hibbard, S. T. (2010). A Monte Carlo study of eight confidence interval methods for coefficient alpha. *Educational and Psychological Measurement*, 70(3), 376-393.  
<https://doi.org/10.1177/0013164409355690>
- Romano, J. L., Kromrey, J. D., Owens, C. M., & Scott, H. M. (2011). Confidence interval methods for coefficient alpha on the basis of discrete, ordinal response items: Which one, if any, is the best? *The Journal of Experimental Education*, 79, 382-403.  
<https://doi.org/10.1080/00220973.2010.510859>
- Sandin, B., Valiente, R. M., Chorot, P., Olmedo, M., & Santed, M. A. (2002a). Versión española del cuestionario EPQR-Abreviado (EPQR-A) (I): Análisis exploratorio de la estructura factorial. *Revista de Psicopatología y Psicología Clínica*, 7, 195-205.  
<https://doi.org/10.5944/rppc.vol.7.num.3.2002.3933>
- Sandin, B., Valiente, R. M., Olmedo, M., Chorot, P., & Santed, M. A. (2002b). Versión española del cuestionario EPQR-abreviado (EPQR-A) (II): Replicación factorial, fiabilidad y validez. *Revista de Psicopatología y Psicología Clínica*, 7(3), 207-216.  
<https://doi.org/10.5944/rppc.vol.7.num.3.2002.3934>
- Tang, W., Cui, Y., & Babenko, O. (2014). Internal consistency: Do we really know what it is and how to assess it? *Journal of Psychology and Behavioral Science*, 2(2), 205-220.
- Thorndike, R. L. (1989). *Psicometría Aplicada*. México, DF: Limusa.
- Tomas, J. M., Meléndez, J. C., Oliver, A., Navarro, E., & Zaragoza, G. (2010). Efectos de método en las escalas de Ryff: Un estudio en población de personas mayores. *Psicológica*, 31, 383-400.
- Tiwari, T., Singh, A. L., & Singh, I. L. (2012). The short-form revised Eysenck Personality Questionnaire. A Hindi edition (EPQRS-H). *Industrial Psychiatry Journal*, 18, 27-31.  
<https://doi.org/10.4103/0972-6748.57854>
- UNESCO (2007). *Clasificación Internacional Normalizada de la Educación, CINE-97*. UNESCO-UIS. En línea:  
<http://www.uis.unesco.org/Library/Documents/isced97-es.pdf>