

Consistencia Interna de los Cinco Marcadores Reducidos de Hauck y Colaboradores: Un Re-Análisis

César Merino-Soto¹  & Conrad Stanisław Zygmunt^{2,*} 

¹Universidad de San Martín de Porres, Lima, Peru

²Helderberg College of Higher Education & Stellenbosch University, Cape Town, África do Sul

RESUMEN – La presente carta tiene por finalidad exponer información complementaria sobre las estimaciones de consistencia interna ($\hat{I}\pm$; Cronbach, 1951) obtenidas en el estudio de Hauck, de Lara, Pereira y Ruschel (2012), para la validación de los cinco marcadores del Big Five en Brasil (Hutz et al., 1998). Esta nueva información puede tener un efecto crítico sobre las conclusiones psicométricas de los autores, pues el error de medición debería tomarse en cuenta para evaluar la equivalencia entre los resultados. Los autores declaran que los niveles obtenidos de los coeficientes de confiabilidad variaron entre magnitudes razonables hasta satisfactorias, sin embargo, no pusieron algún énfasis en que la variada magnitud de estos coeficientes sugiere que no pueden considerarse homogéneas y que diferentes montos de error de medición están presentes en los puntajes del instrumento. El presente aporte involucra dos tipos de precisión: de los puntajes de los cinco marcadores reducidos examinados, y la precisión del coeficiente α .

PALABRAS-CLAVE: Confiabilidad, Cinco Grandes, error de medición, coeficiente alfa

Internal Consistency of the Five Reduced Markers of Hauck and Collaborators: A Re-Analysis

ABSTRACT – The purpose of this letter is to present commentary on the internal consistency estimates (α ; Cronbach, 1951) obtained in the study by Hauck, de Lara, Pereira and Ruschel (2012) towards the validation of the five Big Five Personality factors in Brazil (Hutz et al., 1998). This new information may have a critical impact on the psychometric conclusions drawn by the authors, since measurement error should always be taken into account when evaluating the equivalence between test forms. The authors state that the levels of internal consistency obtained from the reliability coefficients varied from reasonable to satisfactory in magnitude; however, they did not put any emphasis on the fact that the variance in magnitude of these coefficients suggests that they cannot be considered homogeneous, nor that different amounts of measurement error are present in subscale scores. The present critique involves two types of measurement precision: of the scores of the five reduced scales examined, and the precision of the α coefficient.

KEYWORDS: Reliability, Big Five, measurement error, alpha coefficient

Consistência Interna dos Cinco Marcadores Reduzidos de Hauck e Colaboradores: Uma Re-Análise

RESUMO – O objetivo desta carta é apresentar informações complementares sobre as estimativas de consistência interna (α ; Cronbach, 1951) obtidas no estudo de Hauck, de Lara, Pereira e Ruschel (2012), para a validação dos cinco marcadores Big Five em Brasil (Hutz et al., 1998). Esta nova informação pode ter um efeito crítico nas conclusões psicométricas dos autores, uma vez que o erro de medição deve ser levado em consideração para avaliar a equivalência entre os resultados.

* E-mail: sikayax@yahoo.com.ar

■ Submetido: 04/06/2017; Revisado: 07/04/2020; Aceito: 30/06/2020.

Os autores afirmam que os níveis obtidos a partir dos coeficientes de confiabilidade variaram de magnitudes razoáveis a satisfatórias, porém não enfatizam o fato de que a magnitude variada desses coeficientes sugere que eles não podem ser considerados homogêneos e que diferentes valores de erro de medição, estão presentes nas pontuações do instrumento. Essa contribuição envolve dois tipos de precisão: das pontuações dos cinco marcadores reduzidos examinados e a precisão do coeficiente α .

PALAVRAS-CHAVES: Confiabilidade, cinco grandes, erro de medição, coeficiente alfa

ESTIMADO EDITOR

La presente carta tiene por finalidad exponer información complementaria sobre las estimaciones de consistencia interna (α ; Cronbach, 1951) obtenidas en el estudio de Hauck et al. (2012), para la validación de los cinco marcadores del Big Five en Brasil (Hutz et al., 1998). Esta nueva información puede tener un efecto crítico sobre las conclusiones psicométricas de los autores, pues el error de medición debería tomarse en cuenta para evaluar la equivalencia entre los resultados. Los autores declaran que los niveles obtenidos de los coeficientes de confiabilidad variaron entre magnitudes razonables hasta satisfactorias, sin embargo, no pusieron algún énfasis en que la variada magnitud de estos coeficientes sugiere que no pueden considerarse homogéneas y que diferentes montos de error de medición están presentes en los puntajes del instrumento. El presente aporte involucra dos tipos de precisión: de los puntajes de los cinco marcadores reducidos examinados, y la precisión del coeficiente α .

La diferente precisión de los puntajes puede probarse empíricamente aplicando un método estadístico de comparación de coeficientes α para muestras dependientes (Feldt et al., 1987), efectuado mediante el programa *AlphaTest* (Lautenschlager & Meade, 2008). Afortunadamente, toda la información necesaria para este procedimiento aparece en el trabajo de Hauck et al., (2012). Hallamos un resultado igual a $\chi^2(\text{gl: } 4) = 101.0737, p < 0.0001$. Este resultado indica que, en la potencial población de referencia, los coeficientes no pueden considerarse iguales, y cada puntaje contiene diferentes montos de precisión. Este diferente monto de error de medición puede sugerir, entre otras cosas, la presencia de sesgos en el instrumento que desafían la interpretación de la equivalencia de los puntajes entre grupos (Merino, & Lautenschlager, 2003; Reynolds, 2000) y la precisión de las correlaciones entre puntajes de los cinco marcadores.

Esta afirmación se respalda inequívocamente por las estimaciones de este coeficiente en cada subescala. Sin embargo, desde un punto de vista estadístico, estos coeficientes están sujetos a error de muestreo, y solo representan la estimación en la muestra del estudio, dejando sin resolver el grado de precisión del propio coeficiente α . La información de la precisión envuelve a la población de

referencia, en el que se pueden estimar los valores probables del coeficiente en tal población. Basado en un método asintótico usando la distribución teórica F (Feldt, 1965), se puede estimar el intervalo de confianza (95%) para los coeficientes α del estudio de Hauck et al. (2012). Se obtuvo lo siguiente para cada una de las subescalas: Extraversión = .808, .849; Socialización = .763, 0.814; Conciencia = .763, .814; Neuroticismo = .651, .725; y Apertura = .561, .654. Este resultado indica que la escala Apertura puede ser más baja que .60, lo que involucraría mayor sesgo en la interpretación de este puntaje. Aunque el instrumento no está referido para utilizarse como diagnóstico para describir eficazmente la variabilidad individual, puede pensarse que coeficientes más bajos que .60 (incluso debajo de .70) no garantizan precisión en la interpretación de los puntajes. Por lo tanto, es difícil que un investigador se encuentre satisfecho con coeficientes bajos (i.e., en Neuroticismo y Apertura) en un contexto donde el resto de sus puntajes muestran coeficientes aceptables (i.e., Extraversión, Socialización y Conciencia).

Foxcroft y Roodt (2013) sugieren que instrumentos estandarizados de personalidad deben tener coeficientes de α entre .80 y .85; y al mismo tiempo, Hauck et al. (2012) tienen razón que esto no debe ser a expensas de validez de contenido (Widaman et al., 2011). Pero primero necesitamos una buena definición y estimación de los coeficientes de confiabilidad. Aunque el coeficiente α es muy popular, no es ideal porque es influenciado por saturación total de los factores presentes en el instrumento, es decir, un factor general que influye en cada factor específico (o subescala) y la variabilidad de las cargas factoriales dentro de cada factor específico (o subescala) (Zinbarg et al., 2005). El coeficiente α tiende a inclinarse a sobreestimar la confiabilidad cuando los indicadores son multidimensionales, y subestimar cuando no cumple con el modelo tau-equivalente (no tienen cargas factoriales iguales). En vista de que las subescalas en Hauck et al. (2012) probablemente contienen ítems que pueden ser subsumidos por más que un factor y no tienen tau-equivalencia, un método más preciso sería para estimar la saturación del general factor presente en cada subescala (e.g., coeficiente jerárquico, ω_H o coeficiente beta, β). Ambos están incluidos en la función *omega* del *psych* modulo en

el programa estadístico gratuito *R* (Revelle, 2021; Revelle & Zinbarg, 2008). Adicionalmente, se pueden generar intervalos de confianza para estos coeficientes con la función *ci.reliability* o *ci.reliability.bs* en el módulo MBESS (Kelley, 2020), también en *R*. Estos métodos han demostrado superioridad respecto al coeficiente α en una serie de estudios (e.g., Zinbarg et al., 2006).

Como nota adicional, exponemos que un examen más riguroso del error de medición debería tomarse en cuenta al evaluar la viabilidad de la versión breve desarrollada comparada con la versión completa. Ya que las formas breves generalmente tienen confiabilidades más bajas, es predecible que el error de medición atenúe las correlaciones entre la versión breve y otras variables. Por esta razón,

consideramos necesario la implementación de métodos formales para evaluar las diferencias entre coeficiente α , estimar la precisión de este coeficiente mediante intervalos de confianza (Dominguez & Merino, 2015), usar coeficientes más aplicables, y evaluar la divergencia entre correlaciones atenuadas o desatenuadas por error de medición (Widaman et al., 2011). Finalmente, cada vez que se utiliza una versión breve para generar puntajes que son intercambiables con los de la escala original, los métodos de equiparación de prueba pueden utilizarse en la determinación de resultados de las pruebas de la versión breve (Holland & Dorans, 2006; Holland et al., 2007; Kolen, 2006; Kolen & Brennan, 2004). Un ejemplo de estos métodos pueden verse en Han et al. (2015).

REFERENCIAS

- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334. doi: 10.1007/bf02310555.
- Dominguez-Lara, S., & Merino-Soto, C. (2015). ¿Por qué es importante reportar los intervalos de confianza del coeficiente alfa de Cronbach? *Revista Latinoamericana de Ciencias Sociales, Niñez y Juventud*, *13*(2), 1326-1327.
- Feldt, L. S. (1965). The approximate sampling distribution of Kuder-Richardson reliability coefficient twenty. *Psychometrika*, *30*(3), 357–370. doi: 10.1007/bf02289499
- Feldt, L. S., Woodruff, D. J., & Salih, F. A. (1987). Statistical Inference for Coefficient Alpha. *Applied Psychological Measurement*, *11*(1), 93–103. doi:10.1177/014662168701100107
- Foxcroft, C., & Roodt, G. (2013). *Introduction to psychological assessment in the South African context* (4th ed.). Oxford University Press.
- Han, J., Bao, L., Chen, L., Cai, T., Pi, Y., Zhou, S., Tu, Y., & Koenig, K. (2015). Dividing the Force Concept Inventory into two equivalent half-length tests. *Physical Review Special Topics – Physics Education Research*, *11*(1), 1-9. doi: 10.1103/PhysRevSTPER.11.010112
- Hauck, N., de Lara, W., Pereira, M., & Ruschel, D. (2012). Evidências de validade de marcadores reduzidos para a avaliação da personalidade no modelo dos cinco grandes fatores. *Psicologia: Teoria e Pesquisa*, *28*(4), 417-423. doi:10.1590/S0102-37722012000400007
- Holland, P. W., & Dorans, N. J. (2006). Linking and equating. In R. L. Brennan (Ed.), *Educational measurement* (4th ed.). ACE/Praeger.
- Holland, P. W., Dorans, N. J., & Petersen, N. S. (2007). Equating test scores. In C. R. Rao, & S. Sinharay (Eds.), *Handbook of Statistics* (Vol. 26: Psychometrics; pp. 169-204). Elsevier. doi: 10.1016/S0169-7161(06)26006-1
- Hutz, C. S., Nunes, C. H. S. S., Silveira, A. D., Serra, J., Anton, M., & Wiczorek, L. S. (1998). The development of the big five markers for personality assessment in Brazil. *Psicologia: Reflexão e Crítica*, *11*, 395-411. doi: 10.1590/S0102-79721998000200015
- Kelley, K. (2020). MBESS: The MBESS R package. (R package version 4.8.0). Retrieved August 04, 2020, from <http://CRAN.R-project.org/package=MBESS>.
- Kolen, M. J. (2006). Scaling and norming. In R. L. Brennan (Ed.), *Educational measurement* (4th ed.). Westport, CT: ACE/Praeger.
- Kolen, M. J., & Brennan, R. L. (2004). *Test equating, linking and scaling: Methods and practices* (2nd ed.). Springer-Verlag. doi: 10.1007/978-1-4757-4310-4
- Kristof, W. (1963). The statistical theory of stepped-up reliability coefficients when a test has been divided into several equivalent parts. *Psychometrika*, *28*, 221–238. doi: 10.1007/BF02289571
- Lautenschlager, G. J., & Meade, A. W. (2008). AlphaTest: A Windows program for tests of hypotheses about coefficient alpha. *Applied Psychological Measurement*, *32*, 502-503. doi: 10.1177/0146621607312307
- Merino, C., & Lautenschlager, G. (2003). Comparación estadística de la confiabilidad alfa de Cronbach: aplicaciones en la medición educacional. *Revista de Psicología de la Universidad de Chile*, *12*(2), 129-139.
- Revelle, W. (2021). *psych: Procedures for personality and psychological research*. (R package versión 2.1.3). Retrieved from <http://CRAN.R-project.org/package=psych>
- Revelle, W., & Zinbarg, R. E. (2008). Coefficients alpha, beta, omega and the glb: Comments on Sijtsma. *Psychometrika*, *74*(1), 145-154. doi: 10.1007/s11336-008-9102-z
- Reynolds, C. R. (2000). Methods for detecting and evaluating cultural bias in neuropsychological tests. In S. Fletcher-Janzen, L. S. Strickland & C. R. Reynolds (Eds.), *Handbook of Cross-Cultural Neuropsychology* (pp. 249-285). Kluwer Academic / Plenum Publishers. doi: 10.1007/978-1-4615-4219-3_15
- Widaman, K. F., Little, T. D., Preacher, K. J., & Sawalani, G. M. (2011). On creating and using short forms of scales in secondary research. In K. H. Trzesniewski, M. B. Donnellan, & R. E. Lucas (Eds.), *Secondary data analysis: An introduction for psychologists* (pp. 39-61). Washington, DC: APA. doi: 10.1037/12350-003
- Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α , Revelle's β , and McDonald's ω_H : Their relations with each other and two alternative conceptions of reliability. *Psychometrika*, *70*(1), 123-133. doi: 10.1007/s11336-003-0974-7
- Zinbarg, R. E., Yovel, I., Revelle, W., & McDonald, R. P. (2006). Estimating generalizability to a latent variable common to all of a scale's indicators: A comparison of estimators for omega_h. *Applied Psychological Measurement*, *30*(2), 121-144. doi: 10.1177/0146621605278814