
ARTÍCULO DE INVESTIGACIÓN

EVALUACIÓN DE LA CONFIABILIDAD DEL CONSTRUCTO MEDIANTE EL COEFICIENTE H : BREVE REVISIÓN CONCEPTUAL Y APLICACIONES

EVALUATION OF THE CONSTRUCT RELIABILITY THROUGH THE INDEX H : BRIEF CONCEPTUAL REVIEW AND APPLICATIONS

SERGIO ALEXIS DOMINGUEZ-LARA*
UNIVERSIDAD DE SAN MARTÍN DE PORRES, LIMA - PERÚ

FECHA RECEPCIÓN: 15/1/2016 • FECHA ACEPTACIÓN: 25/4/2016

Para citar este artículo: Dominguez-Lara, S. (2016). Evaluación de la confiabilidad del constructo mediante el coeficiente H : breve revisión conceptual y aplicaciones. *Psychologia: Avances de la Disciplina*, 10(2), 87-94.

Resumen

El coeficiente H evalúa la confiabilidad del constructo desde el enfoque SEM, dado que puede ser inadecuado el uso de los estimadores de confiabilidad habituales si no se cumplen algunos supuestos básicos. Para la demostración de la técnica se usó la información de cuatro estudios empíricos y se compararon los coeficientes de confiabilidad obtenidos (α , ω , y H). Del mismo modo, se analizó su aplicación en modelamiento bifactor. Finalmente, se discuten las ventajas, limitaciones y las condiciones para la apropiada aplicación de la técnica en el contexto de la validación de instrumentos de evaluación psicológica.

Palabras clave: Coeficiente H , confiabilidad, psicometría, metodología.

Abstract

Index H evaluate the construct reliability from the SEM approach, since it may be inappropriate using usual reliability estimates if some basic assumptions are not met. To demonstrate the technique, was used information of four empirical studies and were compared the reliability coefficients obtained (α , ω and H). Similarly, application in bifactor modeling was analyzed. Finally, the advantages, limitations and conditions for the proper application of the technique in the context of the validation of psychological assessment instruments are discussed.

Keywords: Index H , reliability, psychometrics, methodology.

* Dirección de correspondencia: Sergio Dominguez Lara. Instituto de Investigación de Psicología, Universidad de San Martín de Porres, Av. Tomás Marsa-no 242 (5.o piso), Lima 34 - Perú; E-mail: sdominguezmpcs@gmail.com, sdominguezl@usmp.pe

Introducción

Desde la teoría clásica de los test, un coeficiente de confiabilidad es definido como la proporción de varianza verdadera que existe en la medición realizada (Lord & Novick, 1968). Los coeficientes de confiabilidad, entre ellos el α (Cronbach, 1951) que se usa de forma más frecuente, habitualmente trabajan con la suma directa de los puntajes (puntaje compuesto; X) de un conjunto de p medidas (X_p) (p .e., ítems). Es decir, $X = X_1 + X_2 + X_3 \dots X_p$.

Además, para el uso del coeficiente α , debe cumplirse que $X_i = \tau_i + \varepsilon_i$; donde τ y ε representan el puntaje verdadero y el error de medición, respectivamente; y para dos o más ítems sería: $\tau_1 = \tau_2 = \tau_3 = \dots = \tau_n$ (τ – equivalencia esencial) y $Cov(\varepsilon_n, \varepsilon_m) = 0$; $n \neq m$ (errores independientes). No obstante, a pesar de la popularidad del coeficiente α , existen autores que brindan argumentos en contra de su uso por diversos motivos, principalmente porque sus dos supuestos fundamentales, tau-equivalencia y ausencia de errores correlacionados, son difíciles de lograr en la práctica (Dunn, Baguley, & Brunsten, 2014; Green, & Yang, 2009a, Sijtsma, 2009), lo que lleva a resultados sesgados de la estimación de la confiabilidad de los puntajes. Si no se cumple la tau-equivalencia, se subestima el monto estimado de varianza verdadera, es decir, que el α obtenido sea considerado como el límite inferior de la confiabilidad (*lower bound*) (Eisinga, Grotenhuis, & Pelzer, 2012); y si existen errores correlacionados, se aprecia una sobreestimación del coeficiente (Dunn et al., 2014; Green & Yang 2009a, Yang & Green, 2010). Ante ello, es conveniente buscar alternativas más precisas desde lo empírico y lo conceptual, donde pueda analizarse la confiabilidad aun cuando esos supuestos no se cumplan (Dunn et al., 2014; Green & Yang, 2009b). Además, su uso no es recomendado para medidas ordinales como los ítems en escalamiento Likert (Dominguez, 2012; Elosua, & Zumbo, 2008), y aunque con opciones de respuesta mayores que cinco pueden considerarse como medidas continuas (Remthulla, Brosseau-Liard & Savalei, 2012), el exceso de asimetría y curtosis puede sesgar la estimación.

Así, desde el enfoque de ecuaciones estructurales (*structural equation modeling*, SEM), existen diversos

planteamientos independientes acerca de la confiabilidad compuesta para medidas congénicas. Por ejemplo, Raykov (1997) sugiere:

$$\frac{\left(\sum_{i=1}^p \lambda_i\right)^2 Var(T)}{\left(\sum_{i=1}^p \lambda_i\right)^2 Var(T) + \sum_{i=1}^p (Var(E_i))},$$

donde $Var(T)$ es la varianza de la variable latente, λ_i y $Var(E_i)$ son los coeficientes de configuración (cargas factoriales) no estandarizados, y las varianzas del error, respectivamente. Dicha expresión hace referencia a la proporción de la varianza explicada por el constructo en un puntaje compuesto, con respecto a la varianza total de ese compuesto. Más adelante, surgió otra medida de confiabilidad en el marco SEM (Fornell & Larcker, 1981) expresada como:

$$\frac{\left(\sum_{i=1}^p \lambda_i\right)^2}{\left(\sum_{i=1}^p \lambda_i\right)^2 + \sum_{i=1}^p (1 - \lambda_i^2)},$$

en la cual λ_i representa a los coeficientes de configuración estandarizados. En la ecuación del coeficiente ω (McDonald, 1999) figura $\sum_{i=1}^p (\theta_{ii})$ en lugar de $\sum_{i=1}^p (1 - \lambda_i^2)$ en el denominador, pero ambas hacen referencia a la varianza del error (Raykov, 2001a). Es una expresión equivalente a la anterior (Raykov, 1997) cuando todos los coeficientes de configuración son estandarizados, y por ende su interpretación es similar.

De este modo, estos métodos son útiles si se tiene en cuenta el grado de consistencia o interrelación de un conjunto de ítems que podrían ser usados para formar un compuesto, pero se concentran en *compuestos no ponderados* (Raykov, 2001a), es decir, estos procedimientos no consideran la variabilidad que puede existir entre los λ_i usados al momento de configurar el puntaje total de un compuesto, ya que se sabe que cada λ_i refleja la influencia del constructo en cuestión y está relacionado directamente con este. Por lo tanto, no proveen una evaluación de la confiabilidad del constructo *en sí misma*, lo que indicaría que los datos utilizados en estos métodos

no reflejan óptimamente la variable latente (Geldhof, Preacher, & Zyphur, 2014).

Además, esos métodos presentan algunas limitaciones. Primero, si un ítem es afectado negativamente por la variable latente (posee un λ_i con signo negativo), aunque este sea de magnitud aceptable y coherente desde un punto de vista psicométrico (p. e., se justifica que el ítem tenga puntaje invertido) impactará de forma negativa en la estimación de los coeficientes antes mencionados. Segundo, si existen ítems que no brinden mucha información acerca de la variable latente (coeficientes de configuración de magnitud baja), su permanencia impacta de forma significativa en la estimación de la confiabilidad (Hancock & Mueller, 2001).

De este modo, la mejor opción es la *combinación lineal ponderada con máxima confiabilidad* (*combinación lineal óptima*, CLO), es decir, la medida que se concentra en los indicadores y puede diferenciar en el más alto grado a los sujetos examinados en el rasgo latente y que se asocia con la menor varianza de error posible (Raykov & Hancock, 2005). Entonces, en el marco SEM, esta CLO se asocia de forma significativa con el puntaje del rasgo latente. Siguiendo esa lógica, para un compuesto la ponderación sería: $X = w_1X_1 + w_2X_2 + w_3X_3 \dots w_pX_p$ (para más detalles técnicos, ver Raykov, 2004). La expresión matemática es (Li, Rosenthal, & Rubin, 1996):

$$\frac{\sum_{i=1}^p (\lambda_i^2 / [1 - \lambda_i^2])}{1 + \sum_{i=1}^p (\lambda_i^2 / [1 - \lambda_i^2])}$$

donde λ_i^2 es la confiabilidad para cada X_i , $i = 1, 2, 3, \dots, p$ (Raykov, 2004). Dicha medida es la forma más general para casos de compuestos de medidas congénicas.

Coefficiente H: medida de confiabilidad del constructo

Conforme a lo expresado anteriormente, el *coeficiente de máxima confiabilidad* puede considerarse como un índice de confiabilidad del constructo. Si cada λ_i refleja la influencia del constructo, y mientras mayor es su magnitud, el constructo se ve mejor representado. Entonces, ese coeficiente cuantifica el grado en que el constructo es «capturado» por la información encontrada

en sus indicadores mensurables (Hancock & Mueller, 2001; Greene & Brown, 2009; Raykov & Hancock, 2005). Dado que en un sistema de variables latentes se asume que un constructo es perfectamente confiable, la variabilidad de los valores obtenidos en distintas muestras se debe a la falta de precisión de los ítems que reflejan dicha variable latente (Hancock & Mueller, 2001). De esta forma, la modificación algebraica del coeficiente de máxima confiabilidad propuesta por Hancock, y Mueller (2001), el coeficiente H se expresa como:

$$H = \frac{1}{\left(1 + \left(\frac{1}{\sum_{i=1}^p (\lambda_i^2 / [1 - \lambda_i^2])}\right)\right)}$$

Donde $\lambda_i^2 / (1 - \lambda_i^2)$ es la proporción de la varianza de la i -ésima variable explicada por el constructo con respecto a la proporción no explicada (al elevar λ_i al cuadrado, queda al margen del signo original de λ_i), por lo cual el coeficiente H está en función de la confiabilidad del ítem (λ_i^2), recopilando esa información a través de las p variables que reflejan el constructo.

El coeficiente H funciona como una estimación de la confiabilidad del constructo, y se interpreta como el porcentaje de la variabilidad de la variable latente explicada por los indicadores (Mueller & Hancock, 2001), siendo deseable que sea $\geq .70$. En este sentido, supera la limitación de los métodos anteriores al no estar influenciada por el signo de los coeficientes de configuración (Geldhof et al., 2014), y además nunca será menor que la confiabilidad del mejor indicador ($H \geq \max(\lambda_i^2)$).

Finalmente, también puede aplicarse para medidas multidimensionales. De forma similar al $\alpha_{\text{estratificado}}$ (Cronbach, Schoenemann, & McKie, 1965), el coeficiente H puede calcularse para medidas que evalúen más de un constructo, e incluso parece obtener mejores resultados (Widhiarso, 2007). No obstante, ese incremento podría estar influido más por la cantidad de ítems.

Cálculo y aplicaciones en análisis psicométrico

El cálculo del coeficiente H se podría hacer manualmente, pero también puede hallarse en el output del programa EQS como *Maximal Weighted Internal*

Consistency Reliability. Cabe mencionar que si se trata de un instrumento multidimensional, el coeficiente H se considera con base en la totalidad de los ítems, por lo que se recomienda calcularlo para cada factor por separado. Asimismo, su cálculo no forma parte del output de resultados en programas como *Mplus* o *AMOS*. El autor del manuscrito ha creado un archivo en MS Excel para calcular el coeficiente H y está disponible al lector sin costo.

A continuación se brindarán tres ejemplos aplicativos para el coeficiente H . Con fines de comparación, adicionalmente fueron calculados el coeficiente α con intervalos de confianza (IC; Dominguez & Merino, 2015a) y el coeficiente ω .

El primer ejemplo se enfoca en una medida tau-equivalente. La Escala de Autoeficacia para Situaciones Académicas (EAPESA) evalúa las creencias que tiene el estudiante con respecto a sus capacidades para hacer frente a las demandas en el ámbito académico y cuenta con evidencias de validez en cuanto a su estructura interna y su relación con otros constructos (Dominguez, 2014a; Dominguez, Villegas, Yauri, Mattos, & Ramírez, 2012; Dominguez, Villegas, Cabezas, Aravena, & De la Cruz, 2013). Fueron evaluadas con la EAPESA 215 personas (79.5% mujeres). A través del programa EQS 6.2, fueron obtenidos índices de ajuste satisfactorios para el modelo unidimensional (congenérico): $SB-\chi^2(27) = 52.100$ ($p < .01$), $CFI = .969$, $RMSEA$ (IC 90 %) = .066 (.038, .092), y al modelar las restricciones para la tau-equivalencia el ajuste también fue bueno ($SB-\chi^2(35) = 65.840$ [$p < .01$], $CFI = .962$, $RMSEA$ [IC 90 %] = .064 [.040, .088]). Además, al ser las diferencias entre los CFI del modelo congenérico y tau-equivalente menores que .01 (Cheung & Rensvold, 2002), los ítems del EAPESA son tau-equivalentes. Para la EAPESA, el coeficiente α fue de .901 (IC95 % .872, .923), la confiabilidad compuesta (ω) obtuvo un índice de .902, y el coeficiente H de .906.

El segundo ejemplo tuvo como base la aplicación de la Escala de Cansancio Emocional (ECE), que evalúa el estado de agotamiento emocional del estudiante con respecto a sus actividades académicas, y presenta evidencias de confiabilidad y validez en estudios realizados con universitarios peruanos (Dominguez, 2013; 2014b). Se usó la muestra del ejemplo 1. Los índices de ajuste mostraron magnitudes aceptables para el modelo congenérico ($SB-\chi^2(35) = 77.680$ [$p < .01$],

$CFI = .976$, $RMSEA$ [IC 90 %] = .075 [.053, .098]), pero al modelar las restricciones para la tau-equivalencia, aunque el ajuste fue bueno ($SB-\chi^2(44) = 124.228$ [$p < .01$], $CFI = .954$, $RMSEA$ [IC 90 %] = .092 [.073, .111]), las diferencias en el CFI fueron mayores que .01 (Cheung & Rensvold, 2002), lo cual indica que los ítems del ECE no son tau-equivalentes. En este caso, el coeficiente α fue de .875 (IC95 % .839, .903), la confiabilidad compuesta (ω) obtuvo un índice de .878, y el coeficiente H , de .895.

Finalmente, para su uso en medidas ultrabreves de dos ítems, se tomaron los datos del análisis psicométrico del Cuestionario de Regulación Cognitiva de las Emociones (CERQ-18; Dominguez, & Merino, 2015b). El CERQ-18 evalúa nueve estrategias cognitivas de regulación emocional (dos ítems por subescala). Dentro del modelo, fueron consideradas para el presente ejemplo dos de las subescalas: *Culpar a otros*, y *Autoculparse*. Los λ_i de *Culpar a otros* fueron de .858 y .739; y de *Autoculparse*, .499 y .930. Se escogieron debido a que los λ_i al interior de estas difieren notoriamente entre sí. En el primer caso, el coeficiente α fue de .710 (IC95 % .647, .763), la confiabilidad compuesta (ω) obtuvo un índice de .780, y el coeficiente H de .800. Finalmente, en cuanto a la subescala *Autoculparse*, el coeficiente α fue de .604 (IC95 % .525, .673), la confiabilidad compuesta (ω) obtuvo un índice de .697, y el coeficiente H fue de .871.

Como puede apreciarse, a medida que los λ_i se van diferenciando en cuanto a su magnitud dentro de cada subescala, tanto el coeficiente α como el ω se ven afectados, mientras que el coeficiente H parece ser robusto ante esas discrepancias.

Coeficiente H y su aplicación en análisis bifactor confirmatorio

Para este ejemplo, se tomaron parte de los resultados preliminares de un estudio bifactor con el CERQ-18 (Dominguez, 2015). Debido a las elevadas correlaciones interfactoriales observadas en estudios previos en la versión completa (CERQ-36; Dominguez, & Medrano, 2016) y breve (CERQ-18; Dominguez & Merino, 2015b), se planificó un estudio que modele un factor general que sea capaz de explicar la variabilidad conjunta de esas subescalas (Reise, 2012).

De esta forma, se modelaron dos grandes factores generales: Estrategias desadaptativas (*Culpar a otros*, *Autoculparse*, *Rumiación*, y *Catastrofización*), y Estrategias

adaptativas (*Aceptación, Poner en perspectiva, Reinterpretación positiva, Focalización en los planes, y Focalización positiva*). La hipótesis general fue que los factores generales explicarían más varianza que los factores específicos. Si bien los resultados brindaron evidencias a favor de esta hipótesis mediante índices de ajuste aceptables, e índices adicionales favorables a ello como el ECV (*Explained Common Variance*; Reise, Scheines, Widaman, & Haviland, 2013), el PUC (*Percentage of uncontaminated correlations*; Reise et al., 2013) y el ω_h (Zinbarg, Yovel, Revelle, & McDonald, 2006), la aplicación del coeficiente H pudo ser un complemento importante.

En este contexto, si uno de los factores específicos obtiene un coeficiente $H \geq .70$, puede considerarse como un factor de grupo robusto, además del factor general (Rodríguez, Reise, & Haviland, 2015). De este modo, los resultados detallados en la tabla 1 indican que luego del modelamiento bifactor, los coeficientes de configuración de los factores específicos y general difieren en magnitudes, trayendo como secuencia $H < .70$ para todos los factores específicos, brindando de ese modo evidencia adicional a favor del factor general.

Tabla 1. Parámetros de los ítems del cerq-18: análisis bifactor

	F ₁	F ₂	F ₃	F ₄	F _G
F ₁					
Ítem 9	.540				.525
Ítem 29	.586				.448
F ₂					
Ítem 17		.747			.199
Ítem 33		.452			.475
F ₃					
Ítem 15			.389		.508
Ítem 27			.310		.631
F ₄					
Ítem 10				-.243	.721
Ítem 35				.181	.741
ω_h	.409	.502	.170	.001	.721
H _{bifactor}	.483	.603	.222	.088	.810

Nota: F₁: Culpar a otros; F₂: Autoculparse; F₃: Rumiación; F₄: Catastrofización; F_G: Factor general

Comentarios finales

En vista de que el uso adecuado del coeficiente α se ve restringido por sus supuestos, y si se utiliza al margen de su cumplimiento la estimación puede estar sesgada. Por ello se han desarrollado alternativas para la evaluación de la confiabilidad desde el marco SEM, aunque esos métodos pasan por alto la heterogeneidad de los coeficientes de configuración en una estructura factorial. Para superar esta limitación, se planteó el coeficiente H y su aplicación como medida de confiabilidad del constructo, como cálculo complementario a los realizados con las técnicas habituales.

Por otro lado, aunque los ejemplos presentados no son representativos de todas las condiciones posibles, pueden ser útiles para ejemplificar el comportamiento del coeficiente H . Dicho indicador parece ser robusto cuando los coeficientes de configuración difieren entre sí (incumplimiento de la tau-equivalencia), ya que si bien los ítems evalúan el mismo constructo, este puede afectarlos en diferente intensidad.

No obstante, existen algunas observaciones vinculadas con su cálculo e interpretación. En primer lugar, de forma similar al coeficiente α , parece que su magnitud está influida por la cantidad de ítems que existen, ya que a más ítems, la sumatoria $H = \frac{1}{\alpha + \frac{1}{\sum_{i=1}^p \frac{1}{(1 - \lambda_i^2)}}}$, que forma parte la expresión propuesta (Hancock & Mueller, 2001) es mayor, por lo que el coeficiente H aumentaría de forma espuria. A modo de ejemplo breve, con $\lambda_i = .40$ para cada ítem, cuando se cuenta con 10 ítems la magnitud de H es .656; con 20 ítems, $H = .792$; y con 30 ítems, $H = .851$. No obstante, son necesarios estudios de simulación que reflejen diferentes condiciones experimentales para reafirmar este argumento, y desarrollos posteriores que permitan lograr un ajuste apropiado de su magnitud en función del número de ítems a fin de no sobredimensionar su valor.

En segundo lugar, al prescindir del signo del coeficiente de configuración para su cálculo, si el factor analizado cuenta con ítems en escala invertida (fraseo en dirección opuesta al constructo), se debe realizar un análisis previo del efecto del método (EM) asociado a esos ítems (Lance, Dawson, Birkelbach, & Hoffman, 2010; Spector, 2006), toda vez que no sea posible hacer un fraseo

que evite esa dirección. Una vez que se determina que el EM es irrelevante, puede calcularse el coeficiente H .

Finalmente, no existen reportes de la alteración de su magnitud ante la presencia de errores correlacionados como ocurre con el coeficiente α (Pascual-Ferrá & Beatty, 2015) y ω (Raykov, 2001b). En ese caso, cabe la posibilidad de que en futuros estudios pueda agregarse un factor corrector a su expresión matemática a fin de lograr una estimación más precisa de la confiabilidad del constructo.

Para concluir, queda claro que el coeficiente H es una medida complementaria que puede ser de ayuda en los procesos analíticos orientados al reporte de las propiedades psicométricas de los instrumentos de evaluación, y aunque algunos desarrollos metodológicos quedan pendientes, es una alternativa interesante dentro del marco analítico de los modelos de ecuaciones estructurales.

Referencias

- Cheung, G., & Rensvold, R. (2002). Evaluating Goodness-Of-Fit Indexes for Testing Measurement Invariance. *Structural Equation Modeling*, 9, 233-255.
- Cronbach, L. J. (1951). Coeficient Alpha and The Internal Structure of A Test. *Psychometrika*, 16, 297-334.
- Cronbach, L. J., Schoenemann, P., & McKie, D. (1965). Alpha Coefficient por Stratified-Parallel Tests. *Educational and Psychological Measurement*, 25, 291-312.
- Dominguez, S. (2012). Propuesta para el cálculo del alfa ordinal y theta de armor. *Revista de Investigación en Psicología*, 15(1), 213-217.
- Dominguez, S. (2013). Análisis psicométrico de la escala de cansancio emocional en estudiantes de una universidad privada. *Revista Digital de Investigación en Docencia Universitaria*, 7(1), 45-55.
- Dominguez, S. (2014a). Autoeficacia para situaciones académicas en estudiantes universitarios peruanos: un enfoque de ecuaciones estructurales. *Revista de Psicología-Universidad Católica San Pablo*, 4(4), 45-53.
- Dominguez, S. (2014b). Escala de cansancio emocional: estructura factorial y validez de los ítems en estudiantes de una universidad privada. *Avances en Psicología*, 21(1), 89-97.
- Dominguez, S. (2015, noviembre). *¿Qué hay más allá de los factores específicos?: modelos jerárquicos en el análisis de instrumentos de evaluación*. Ponencia presentada en la Semana Académica de la Facultad de Psicología de la UNMSM, Facultad de Psicología, Universidad Nacional Mayor de San Marcos, del 17 al 20 de noviembre, Lima, Perú.
- Dominguez, S., & Medrano, L. (2016). Propiedades psicométricas del Cognitive Emotional Regulation Questionnaire (CERQ) en estudiantes universitarios de Lima. *Psychologia: Avances en la Disciplina*, 10(1), 53-67.
- Dominguez, S. & Merino C. (2015a) ¿Por qué es importante reportar los intervalos de confianza del coeficiente alfa de Cronbach? *Revista Latinoamericana de Ciencias Sociales, Niñez y Juventud*, 13(2), 1326-1328.
- Dominguez, S., & Merino, C. (2015b). Una versión breve del Cognitive Emotional Regulation Questionnaire: análisis estructural del CERQ-18 en estudiantes universitarios limeños. *Revista Peruana de Psicología y Trabajo Social*, 4(1), 25-36.
- Dominguez, S., Villegas, G., Yauri, C., Mattos E. & Ramírez F. (2012). Propiedades psicométricas de una escala de autoeficacia para situaciones académicas en estudiantes universitarios peruanos. *Revista de Psicología-Universidad Católica San Pablo*, 2(1), 27-39.
- Dominguez, S., Villegas, G., Cabezas, M. Aravena, S. & De la Cruz, M. (2013). Autoeficacia académica y ansiedad ante los exámenes en estudiantes de psicología de una universidad privada. *Revista de Psicología-Universidad Católica San Pablo*, 3(3), 13-23.
- Dunn, T. J., Baguley, T., & Brunsdén, V. (2014). From Alpha to Omega: A Practical Solution to the Pervasive Problema of Internal Consistency Estimation. *British Journal of Psychology*, 105(3), 399-412.
- Eisinga, R., Te Grotenhuis, M., & Pelzer, B. (2012). The Reliability of A Two-Item Scale: Pearson, Cronbach, or Spearman-Brown. *International Journal of Public Health*, 58, 637-642.
- Elosua, P., & Zumbo, B. (2008). Coeficientes de fiabilidad para escalas de respuesta categórica ordenada. *Psicothema*, 20(4), 896-901.

- Fornell, C., & Larcker, D.F. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research*, 18(1), 39-50.
- Greene, J. A., & Brown, S. C., (2009). The Wisdom Development Scale: Further Validity Investigations. *International Journal of Aging and Human Development*, 68(4), 289- 320.
- Green, S. B., & Yang, Y (2009a). Commentary on Coefficient Alpha: A Cautionary Tale. *Psychometrika*, 74(1), 121-135.
- Green, S. B., & Yang, Y (2009b). Reliability of Summed Items Scores Using Structural Equation Modeling: and Alternative to Coefficient Alpha. *Psychometrika*, 74(1), 155- 167.
- Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability Estimation in A Multilevel Confirmatory Analysis Framework. *Psychological Methods*, 19(1), 72-91.
- Hancock, G. R., & Mueller, R. O. (2001). Rethinking Construct Reliability Within Latent Variable Systems. En R. Cudeck, S. H. C. du Toit & D. Sörbom (Eds.), *Structural Equation Modeling: Past and Present. A Festschrift in Honor of Karl G. Jöreskog* (pp. 195-261). Chicago: Scientific Software International.
- Lance, C., Dawson, B., Birkelbach, D., & Hoffman, B. (2010). Method Effects, Measurement Error, and Substantive Conclusions. *Organizational Research Methods*, 13(3), 435-455.
- Li, H., Rosenthal, R., & Rubin, D. (1996). Reliability of Measurements in Psychology: From Spearman-Brown to Maximal Reliability. *Psychological Methods*, 5, 98-107.
- Lord, F. M., & Novick, R. (1968). *Statistical Theories of Mental Tests Scores*. Reading: Addison-Wesley.
- McDonald, R. P (1999). *Test Theory: A United Treatment*. Mahwah, NJ: Lawrence Erlbaum.
- Mueller, R. O., & Hancock, G. R., (2001). Factor Analysis and Latent Structure. En N. J. Smelser, & P. B. Baltes (Eds.), *International Encyclopedia of Social and Behavioral Sciences* (pp. 5239-5244). Oxford: Pergamon.
- Pascual-Ferrá, P., & Beatty, M. J. (2015). Correcting Internal Consistency Estimates Inflated by Correlated Item Errors. *Communication Research Reports*, 32(4), 347-352.
- Raykov, T. (1997a). Estimation of Composite Reliability for Congeneric Measures. *Applied Psychological Measurement*, 22, 173-184.
- Raykov, T. (2001a). Estimation of Congeneric Scale Reliability Via Covariance Structure Analysis with Nonlinear Constraints. *British Journal of Mathematical and Statistical Psychology*, 54, 315-323.
- Raykov, T. (2001b). Bias in Coefficient Alpha for Fixed Congeneric Measures with Correlated Errors. *Applied Psychological Measurement*, 25(1), 69-76.
- Raykov, T. (2004). Estimation of Maximal Reliability: A Note on A Covariance Structure Modelling Approach. *British Journal of Mathematical and Statistical Psychology*, 57, 21-27.
- Raykov, T., & Hancock, G. R. (2005). Examining Change in Maximal Reliability for Multiple-Component Measuring Instruments. *British Journal of Mathematical and Statistical Psychology*, 58, 65-82.
- Reise, S. P. (2012). The Rediscovery of Bifactor Measurement Models. *Multivariate Behavioral Research*, 47, 667-696.
- Reise, S.P. Scheines, R., Widaman, K.F., & Haviland, M.G. (2013). Multidimensionality and Structural Coefficient Bias in Structural Equation Modeling: A Bifactor Perspective. *Educational and Psychological Measurement*, 73(1), 5-26.
- Remthulla, M., Brosseau-Liard, P.É., & Savalei, V. (2012). When Can Categorical Variables Be Treated As Continuous? A Comparison of Robust Continuous and Categorical SEM Estimation Methods under Suboptimal Conditions. *Psychological Methods*, 17(3), 354-373.
- Rodriguez, A., Reise, S. P., & Haviland, M. G. (2015). Evaluating Bifactor Models: Calculating and Interpreting Statistical Indices. *Psychological Methods*. Avance Online.
- Sijtsma, K. (2009). On the Use, Misuse, and The Very Limited Usefulness On Cronbach 'S Alpha. *Psychometrika*, 74(1), 107-120.
- Spector, P. (2006). Method Variance in Organizational Research. Truth or Urban Legend? *Organizational Research Methods*, 9(2), 221-232.

- Yang, Y., & Green, S.B. (2010). A Note on Structural Equation Modeling Estimates of Reliability. *Structural Equation Modeling*, 17(1): 66-81.
- Widhiarso, W. (2007). *Estimating Reliability for Multidimensional Measure*. Unpublished Research Summary. Faculty of Psychology. Gadjah Mada University.
- 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 Ω_h . *Applied Psychological Measurement*, 30(2), 121-144.