

ARTÍCULO ORIGINAL

Índices estadísticos de modelos bifactor

Statistical indices from bifactor models

Sergio Alexis Dominguez-Lara ¹ y Anthony Rodriguez ²

¹ Universidad de San Martín de Porres, Perú

² University of California, Estados Unidos

* Dirección de correspondencia: Sergio Dominguez-Lara. Instituto de Investigación de Psicología, Universidad de San Martín de Porres, Av. Tomás Marsano 242 (5to piso), Lima 34 – Perú. Email: sdominguezmpcs@gmail.com, sdominguezl@usmp.pe

Recibido: 24 de febrero de 2017; Revisado: 25 de junio de 2017; Aceptado: 26 de junio de 2017; Publicado Online: 29 de junio de 2017.

CITARLO COMO:

Dominguez-Lara, S. & Rodriguez, A. (2017). Índices estadísticos de modelos bifactor. *Interacciones*, 3(2), 59-65. doi: 10.24016/2017.v3n2.51

RESUMEN

Muchos instrumentos se crean con el propósito principal de evaluar sujetos con relación a un solo rasgo. Sin embargo, los rasgos psicológicos son frecuentemente complejos y contienen manifestaciones de dominio específico. Como tal, muchos instrumentos brindan información que son consistentes tanto con las estructuras unidimensionales como las multidimensionales. Por desgracia, muchas veces, los investigadores aplicados hacen determinaciones sobre la estructura final basado únicamente en los índices de ajuste obtenidos a partir de modelos de ecuaciones estructurales. Dado que los índices de ajuste generalmente favorecen al modelo bifactor sobre los modelos de medición competidores, es imperativo que los investigadores hacen uso de la información disponible que los modelos bifactor tienen para ofrecer con el fin de calcular los índices informativos incluyendo coeficientes de confiabilidad omega, confiabilidad del constructo, varianza común explicada, y el porcentaje de correlaciones no contaminadas. Dichos índices proporcionan información acerca de la fuerza tanto de los factores generales como de los factores específicos con el fin de sacar conclusiones acerca de la dimensionalidad y la puntuación global de las escalas (y subescalas). En este documento, se describen estos índices y ofrecen un nuevo módulo que facilita su cálculo.

PALABRAS CLAVE

Análisis factorial confirmatorio;

Bifactor;

Omega;

Confiabilidad del constructo;

Varianza común explicada;

Porcentaje de correlaciones no contaminadas.



KEY WORDS

Confirmatory factorial analysis, bifactor;
 Omega;
 Construct reliability;
 Explained common variance;
 Percentage of uncontaminated correlations.

ABSTRACT

Many instruments are created with the primary purpose of scaling individuals on a single trait. However psychological traits are often complex and contain domain specific manifestations. As such, many instruments produce data that are consistent with both unidimensional and multidimensional structures. Unfortunately, oftentimes, applied researchers make determinations about the final structure based solely on fit indices obtained from structural equation models. Given that fit indices generally favor the bifactor model over competing measurement models it is imperative that researchers make use of the available information the bifactor has to offer in order to compute informative indices including omega reliability coefficients, construct reliability, explained common variance, and percentage of uncontaminated correlations. Said indices provide unique information about the strength of both the general and specific factors in order to draw conclusions about dimensionality and overall scoring of scales (and subscales). Herein, we describe these indices and offer a new module which easily facilitates their computation.

La construcción de instrumentos de evaluación psicológica bajo un enfoque analítico factorial normalmente busca la medición tanto de constructos amplios como de las dimensiones que los conforman, a fin de obtener información relevante para fines diagnósticos o de intervención. Ante ello, el supuesto de dimensionalidad teórica (p.e., que las dimensiones F_1 , F_2 y F_3 integran el constructo G) se traslada al plano empírico, y las puntuaciones de las tres dimensiones son adicionadas, configurando una puntuación total (puntuación F_1 + puntuación F_2 + puntuación F_3 = puntuación_G). En otras palabras, se asume que todos los ítems son influidos por la misma variable latente (p.e., Dominguez-Lara, 2014). Sin embargo, ello no es respaldado por información empírica que acredite la presencia de un factor de orden superior predominante, y la decisión se toma en base únicamente a

las correlaciones entre las dimensiones (Figura 1), lo que podría tener un claro sentido teórico, pero necesita ser refrendado con datos empíricos.

Es necesario resaltar que cuando las correlaciones entre las dimensiones son elevadas, puede hipotetizarse que esa relación se debe a la presencia de un factor general (FG; Reise, 2012). Este FG explicaría una mayor proporción de variabilidad de los ítems en comparación a los factores específicos (FEs), es decir, las dimensiones que lo integran. Por el contrario, si la correlación entre FEs es débil, probablemente el FG no explique una mayor variabilidad de los ítems que la explicada por los FEs de forma individual.

La influencia de este FG se analiza en el marco de los modelos jerárquicos, siendo el modelamiento *bifactor* o *jerárqui-*

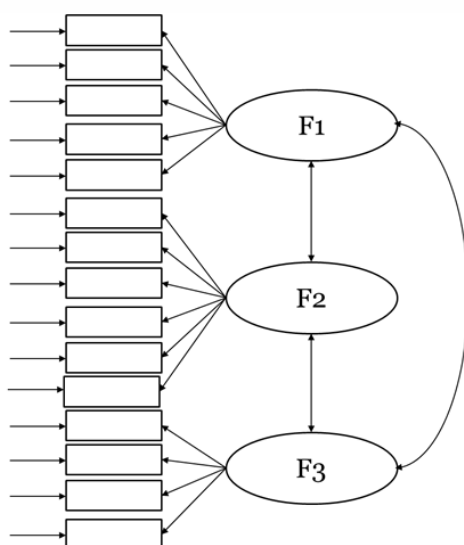


Figura 1. Modelo de tres factores oblicuos

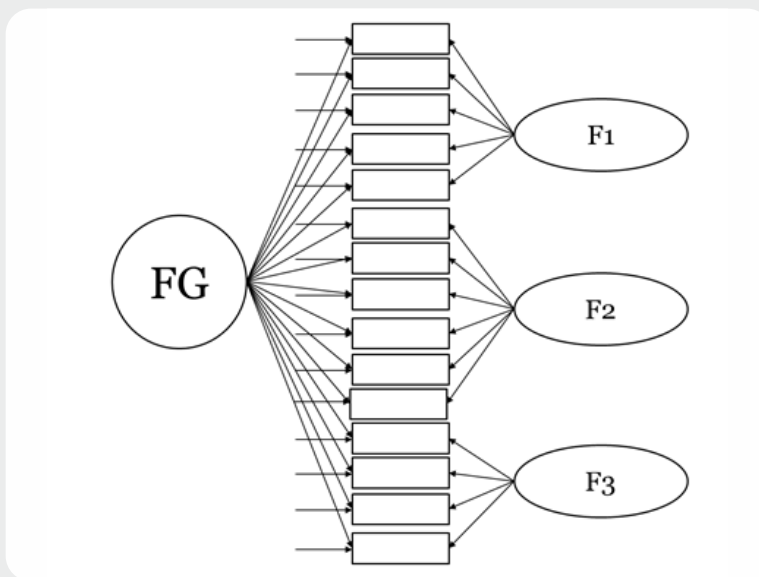


Figura 2. Modelo bifactor

co directo (Canivez, 2016; Holzinger & Harman, 1938, Holzinger & Swineford, 1937; Reise, 2012; Reise, Morizot, & Hays, 2007) el más adecuado para evaluar simultáneamente la influencia del FG y FEs sobre la variabilidad de cada ítem. De este modo, podría valorarse si es pertinente la consideración de un FG de forma exclusiva, o de diferentes FEs, al evaluar determinado instrumento.

Desde los modelos de ecuaciones estructurales (SEM), los modelos bifactor (Figura 2) se caracterizan por especificar en el modelamiento adicionalmente un FG que influye en todos los ítems de forma simultánea a los FEs (F₁, F₂, y F₃, en la Figura 2). Los FE son modelados ortogonalmente ya que se parte de la hipótesis de que la varianza compartida entre ellos se debe a la presencia de un FG (Reise, 2012). El supuesto general del modelo es que el FG explica una mayor cantidad de varianza de los ítems en comparación a la que explican los FEs. Sin embargo, en ocasiones su evaluación bajo SEM puede llevar a falsos positivos si son considerados únicamente los índices de ajuste.

Por ejemplo, al evaluar un modelo bifactor y obtener índices de ajuste favorables (p.e., CFI > .95, RMSEA < .05), puede concluirse que es razonable suponer que el FG es la mejor explicación. En este caso, lo que se evalúa es el ajuste del modelo, pero no el grado de influencia del FG, en comparación a los FEs, sobre los ítems. Además, existe evidencia de que los índices de ajuste tradicionales (p.e. CFI, RMSEA, etc.) suelen favorecer a los modelos bifactor (Gignac, 2016; Morgan, Hodge, Wells, & Watkins, 2015).

Ese panorama se observa en estudios sobre engagement

académico (Stefansson, Gestsdottir, Geldhof, Skulason, & Lerner, 2016), burnout académico (Morgan, de Bruin, & Bruin, 2014; Tsubakita, & Shimazaki, 2016), etc.; donde los índices de ajuste del bifactor son favorables, incluso por encima de los otros modelos competidores, pero las magnitudes de las cargas factoriales de los FEs no difieren significativamente de las cargas factoriales del FG en el factor correspondiente (p.e., Stefansson et al., 2016), por lo que la consideración de un FG, y por ende una puntuación total, es cuestionable.

En este sentido, es necesario contar con índices estadísticos que permitan valorar la robustez del FG, así como evaluar si la contribución de los FEs que está por encima del FG es relevante, ya que frecuentemente el FG es lo suficientemente bueno, y no sería necesaria la evaluación de FE.

Existen diversos índices para valorar la fortaleza del FG. Por ejemplo, es utilizado el omega jerárquico (ω_H ; Zinbarg, Yovel, Revelle, & McDonald, 2006) para informar el monto de varianza total que puede ser atribuida al FG. Su cálculo se realiza con la siguiente fórmula:

$$\omega_H = \frac{\sum \lambda_{iFG}^2}{(\sum \lambda_{iFG}^2) + (\sum \lambda_{iFE_1}^2) + (\sum \lambda_{iFE_2}^2) + \dots + (\sum \lambda_{iFE_n}^2) + \sum \theta_i^2}$$

Por su parte, cada FE posee su omega jerárquico (ω_{hs}) que refleja la varianza confiable de las puntuaciones por encima del FG (Rodríguez, Reise, & Haviland, 2016a). Su cálculo obedece a la siguiente expresión:

$$\omega_{hs} = \frac{\sum \lambda_{iFE_n}^2}{(\sum \lambda_{iFG})^2 + (\sum \lambda_{iFE_n})^2 + \sum \theta_i^2}$$

Para las dos fórmulas anteriores, λ_{iFG} se refiere a la carga factorial de cada ítem en el FG, λ_{iFE} a las cargas factoriales en cada FE, y θ^2 es la varianza del error de cada ítem.

Como puede apreciarse, para el caso del ω_{hr} , mientras mayores sean las cargas factoriales en el FG, en contraste con las cargas en los FE, su magnitud será mayor. Del mismo modo, para ω_{hs} , si las cargas factoriales de los ítems que pertenecen al FE son mayores en el FG, la magnitud de ω_{hs} irá disminuyendo.

En cuando al ω_h se esperan magnitudes $\geq .70$ para concluir, al menos parcialmente, a favor de la unidimensionalidad (Reise, Scheines, Widaman, & Haviland., 2013); y para el caso del ω_{hs} valores $\geq .30$ podrían considerarse como significativas (Smits, Timmerman, Barelds, & Meijer, 2015).

Por su parte, el coeficiente H_h (Hancock, 2001; Hancock & Mueller, 2001) que es un método estadístico para evaluar cuán bien una variable latente es representada por un conjunto de ítems (Dominguez-Lara, 2016a). El coeficiente H no está limitado por la cantidad de grupos, y puede ser calculado también para el FG. En este caso, evalúa la confiabilidad de cada FE controlando el efecto del FG ($> .70$; Raykov, & Hancock, 2005). El cálculo del coeficiente H_{hr} , sea para el FG o para cada FE, se realiza con la siguiente expresión:

$$H_h = \frac{1}{1 + \frac{1}{\sum_{i=1}^{I_F} \left[\frac{\lambda_i^2}{1 - \lambda_i^2} \right]}}$$

Donde I_f representa al número de ítems en cada factor, tanto del FG como de los FEs; y el λ^2 es la carga factorial al cuadrado (varianza) de cada ítem.

También es usual reportar el *ECV* (Explained Common Variance; Sijtsma, 2009; Ten Berge & Socan, 2004) que se interpreta como el monto de varianza común que se debe al FG. Un *ECV* mayor que .60 es indicador de que hay poca varianza común entre FE más allá que la del FG (Reise et al., 2013), y a nivel de ítem, el *ECV-I* (Stucky, Thissen, & Edelen, 2013) indica qué porcentaje de la varianza verdadera de cada ítem es explicada por el FG, esperando valores $\geq .80$ para concluir sobre una influencia significativa del FG (Stucky, & Edelen, 2015). Como puede apreciarse, se excluye del análisis a la varianza del error, por lo que su análisis debe

considerarse de forma conjunta con el ω_{hr} . Las expresiones matemáticas son:

$$ECV = \frac{\sum_{i=1}^{I_{FG}} \lambda_{iFG}^2}{\sum_{i=1}^{I_{FG}} \lambda_{iFG}^2 + \sum_{i=1}^{I_{FE_1}} \lambda_{iFE_1}^2 + \sum_{i=1}^{I_{FE_2}} \lambda_{iFE_2}^2 + \dots + \sum_{i=1}^{I_{FE_n}} \lambda_{iFE_n}^2}$$

$$ECV - I = \frac{\lambda_{iFG}^3}{\lambda_{iFE_n}^2 + \lambda_{iFG}^2}$$

Por su parte, el *PUC* (Percentage of Uncontaminated Correlations; Reise et al., 2013) brinda información sobre el porcentaje de correlaciones no contaminadas por la multidimensionalidad (Rodriguez, Reise, & Haviland, 2016a), calculado con:

$$PUC = \frac{\frac{I_{FG}(I_{FG}-1)}{2} - \left(\frac{I_{FE_1}(I_{FE_1}-1)}{2} + \frac{I_{FE_2}(I_{FE_2}-1)}{2} + \dots + \frac{I_{FE_n}(I_{FE_n}-1)}{2} \right)}{\frac{I_{FG}(I_{FG}-1)}{2}}$$

En las tres expresiones anteriores, I_{FG} reflejan la cantidad ítems en el FG, I_{FE} la cantidad de ítems en cada FE, y λ^2 , la carga factorial al cuadrado (sea en FG o FE, según corresponda).

El *PUC* ayuda a interpretar prudentemente el *ECV*. Por ejemplo, si el *ECV* es $> .70$ y el *PUC* $> .70$, se podría concluir a favor de la unidimensional (Rodriguez et al., 2016a); inclusive si el *PUC* es alto ($> .80$), incluso con magnitudes de *ECV* $> .60$ y $\omega_{hr} \geq .70$ informan favorablemente sobre la unidimensionalidad (Reise et al., 2013).

Esto es importante debido a que generalmente solo se observa el reporte del ω_{hr} , pero incluso si cuenta con una magnitud moderada o elevada, por sí mismo no es evidencia suficiente para concluir sobre la fortaleza del FG, ya que debe existir adicionalmente un *ECV* elevado.

Adicionalmente, es recomendable reportar la carga factorial promedio ($\lambda_{promedio}$) tanto de los FEs como del FG con el objetivo de dar un índice general sobre la influencia de cada factor sobre los ítems.

No obstante, a pesar de su uso y utilidad en el ámbito de validación de instrumentos (ver Rodriguez, Reise, & Haviland, 2016b), no se hallan integrados en paquetes estadísticos (sea comerciales o de libre acceso), para que el usuario pueda obtenerlos fácilmente de forma conjunta. Existen

programas informáticos ejecutables para procedimientos específicos (P.e., ω_{μ} y ω_{hs} ; Watkins, 2013), o extensiones compatibles con el entorno R (Rodríguez et al., 2016a), pero eso implica una configuración particular de la PC, o conocimiento de lenguaje de programación, lo que podría desalentar al investigador aplicado.

PROPUESTA DE MÓDULO

Bajo ese argumento fue construido un módulo en MS Excel (*IndicesBifactor.xls*), que considera el cálculo de todos los índices mencionados anteriormente. La ventaja es que solo requiere ingresar las cargas factoriales estandarizadas de cada uno de los FEs y del FG.

Por ejemplo, en la columna S1 deben colocarse las cargas factoriales estandarizadas del primer FE, y al lado las cargas factoriales de esos ítems en el FG, y repetir el mismo procedimiento para todos los FE existentes (hasta un máximo de cinco FE). Una vez realizado aquello, los cálculos son automáticos. El módulo está disponible al lector sin costo en la sección de material complementario de la revista o escribiendo un correo electrónico al primer autor.

Para ejemplificar la funcionalidad del módulo, fueron usados los datos de la tabla 2 de Rodríguez et al. (2016a), donde evalúan una escala de ansiedad para niños, siendo la Ansiedad el FG, y los FEs las dimensiones *Síntomas Físicos*, *Evitación del Daño*, *Ansiedad Social*, y *Ansiedad de Separación/Pánico*. Todos los índices obtenidos son similares a los obtenidos en el citado estudio, por lo que los cálculos fueron programados eficientemente.

APLICACIÓN E INTERPRETACIÓN

El uso del módulo *IndicesBifactor.xls* puede orientarse tanto a la investigación instrumental, como a los re-análisis o análisis complementarios a investigaciones que usaron modelamiento bifactor con el objetivo de extraer información relevante, pero que por diferentes razones no fue reportada.

Por ejemplo, luego de realizar un análisis factorial confirmatorio a un instrumento que evalúa engagement, Stefansson y colaboradores (2016) concluyeron que “...the school engagement is characterized by both a single and multiple dimension (...). Our results suggest that a bifactor model is the best way to represent a comprehensive measure of school engagement...” (p. 479), aunque no se reportaron los índices estadísticos que den soporte empírico a tales afirmaciones. Los indicadores de forma conjunta permiten concluir que el FG es más fuerte que los FEs (Figura 4), de forma similar a los datos presentados por Rodríguez et al. (2016a) (Figura 3).

En ambos casos se aprecia una clara influencia del FG. En el caso de Stefansson et al. el 64% de la varianza en los ítems se debe al FG ($\omega_{\mu} = .806$). Asimismo, sólo 7% se debe a FE₁, 12% a FE₂, y un 8% a FE₃, explicando de forma conjunta un 27% de la varianza total. También el *PUC* (.714), *ECV* (.631) y *H₁* (.911) evidencian un FG fuerte; pero resulta claro que, a nivel general, el aporte de los FEs a los ítems no es significativo, por lo que la interpretación de los tres FEs de forma independiente sería forzada. Sin embargo, el ω_{hs} del segundo FE presenta una magnitud no despreciable (> .40), e inclusive *ECV-I* de magnitudes bajas (Figura 4).

S1	FG	S2	FG	S3	FG	S4	FG	S5	FG	ECV	0.795	S1: ECV-I	S2: ECV-I	S3: ECV-I	S4: ECV-I	S5: ECV-I
-0.13	0.38	0.59	0.24	0.75	0.55	0.24	0.6			PUC	0.765	0.895	0.142	0.35	0.862	
0.1	0.82	0.43	0.56	0.47	0.83	0.36	0.84			ω_{μ}	0.934	0.985	0.629	0.757	0.845	
0.38	0.82	0.78	0.06	0.06	0.97	0.26	0.66			ω_{h1}	0.076	0.823	0.006	0.996	0.866	
0.35	0.77	0.71	0.51	0.31	0.9	0.09	0.73			ω_{h2}	0.427	0.829	0.34	0.894	0.985	
0.28	0.74	0.45	0.46	0.51	0.81	0.1	0.7			ω_{h3}	0.166	0.875	0.511	0.716	0.98	
0.02	0.98	0.13	0.68	0.36	0.9	0.28	0.76			ω_{h4}	0.078	1	0.965	0.862	0.88	
0.28	0.77	0.43	0.59	0.27	0.6	0.16	0.74			ω_{h5}	0.078	0.883	0.653	0.832	0.955	
0.32	0.79	0.17	0.71	0.28	0.78	0.11	0.86			H_{11}	0.991	0.859	0.946	0.886	0.984	
0.39	0.66	0.5	0.6	0.17	0.65	0.41	0.76			H_{12}	0.515	0.741	0.59	0.936	0.775	
0.19	0.73									H_{13}	0.807	0.937				
0.2	0.82									H_{14}	0.704	0.792				
0.11	0.98									H_{15}	0.385	0.988				
										H_{16}						
										$\lambda_{\text{promedio G}}$	0.7					
										$\lambda_{\text{promedio S1}}$	0.226					
										$\lambda_{\text{promedio S2}}$	0.466					
										$\lambda_{\text{promedio S3}}$	0.353					
										$\lambda_{\text{promedio S4}}$	0.223					
										$\lambda_{\text{promedio S5}}$						

Figura 3. Vista general del módulo *IndicesBifactor*

COMENTARIOS FINALES

El uso cada más vez extendido de instrumentos multidimensionales y la disposición de métodos orientados a evaluar esa multidimensionalidad, obliga al investigador aplicado a brindar todas las evidencias posibles para garantizar un uso apropiado. De este modo, el modelamiento bifactor es una alternativa viable para dicho fin. Adicionalmente, puede ser poco recomendable sustentar todo el peso de la decisión en los índices de ajuste dado que existen otros elementos, como los índices de modificación, que brindaría información adicional y evitarían conclusiones sesgadas (Dominguez-Lara, 2016b). Por tal motivo, es necesario el cálculo y reporte de índices adicionales que permitan ampliar el panorama del investigador, y el módulo presentado (*IndicesBifactor.xls*), puede ser una alternativa útil en el marco de un modelamiento con estas características.

Si bien en el presente manuscrito el uso del modelamiento bifactor fue ejemplificado en el caso de un FG robusto en el contexto de instrumentos multidimensionales, también puede usarse cuando la correlación entre FEs sea tan elevada, que no sea recomendable una interpretación independiente (p.e., Dominguez-Lara, 2016c).

Por ello, si el investigador decide utilizar un instrumento multidimensional para utilizar posteriormente una puntuación general, es recomendable obtener evidencias empíricas que sustenten su proceder, por lo que el uso de modelos bifactor debe ser más frecuente.

Con todo, es importante mencionar que cualquier procedimiento estadístico debe ser ejecutado a la luz de una teoría

que provea significado a los resultados. La razón es sencilla: aunque el FG resulte más robusto que los FEs, si carece de una interpretación teórica coherente por sí mismo, no tiene sentido considerarlo y sería mejor interpretar cada FE por separado (Bonifay, Lane, & Reise, 2017; Dominguez-Lara, & Merino-Soto, 2016; Moscoso, Merino-Soto, Dominguez-Lara, Chau, & Claux, 2016).

CONFLICTOS DE INTERÉS

Los autores expresan que no hay conflictos de interés al redactar el manuscrito.

REFERENCIAS

Bonifay, W., Lane, S.P., & Reise, S.P. (2017). Three concerns with applying a bifactor model as a structure of psychopathology. *Clinical Psychological Science*, 5(1), 184–186. doi: 10.1177/2167702616657069

Canivez, G. L. (2016). Bifactor modeling in construct validation of multifaceted tests: Implications for multidimensionality and test interpretation. In K. Schweizer & C. DiStefano (Eds.), *Principles and methods of test construction: Standards and recent advancements* (pp. 247–271). Göttingen, Germany: Hogrefe.

Dominguez-Lara, S. (2014). Análisis Psicométrico de la *Escala de Bienestar Psicológico para Adultos* en estudiantes universitarios de Lima: un enfoque de ecuaciones estructurales. *Psychologia: Avances en la disciplina*, 8(1), 23-31. doi: 10.21500/19002386.1211

Dominguez-Lara, S. (2016a). Evaluación de la confiabilidad del constructo mediante el Coeficiente H: breve revisión conceptual y aplicaciones. *Psychologia: Avances en la disciplina*, 10(2), 87 – 94. doi: 10.21500/19002386.2134

Dominguez-Lara, S. (2016b). Evaluación de modelos estructurales, más allá de los índices de ajuste. *Enfermería Intensiva*, 27, 84-85. doi: 10.1016/j.enfi.2016.03.003

Dominguez-Lara, S. (2016c). Inventario de la Ansiedad ante Exámenes-

S1	FG	S2	FG	S3	FG	S4	FG	S5	FG	ECV	0.631	S1: ECV-I	S2: ECV-I	S3: ECV-I	S4: ECV-I	S5: ECV-I
0.82	0.4	0.45	0.64	0.21	0.76					PUC	0.714	0.192	0.669	0.929		
0.46	0.67	0.67	0.64	0.39	0.74					ω_{11}	0.806	0.68	0.477	0.783		
0.3	0.64	0.56	0.62	0.51	0.6					ω_{12}	0.253	0.82	0.551	0.581		
0.11	0.44	0.41	0.47	0.47	0.64					ω_{13}	0.348	0.941	0.568	0.65		
0.2	0.74	0.29	0.56	0.67	0.6					ω_{14}	0.284	0.932	0.789	0.445		
										ω_{15}						
										ω_{21}						
										ω_{22}						
										ω_{23}						
										ω_{24}						
										ω_{25}						
										ω_{31}						
										ω_{32}						
										ω_{33}						
										ω_{34}						
										ω_{35}						
										ω_{41}						
										ω_{42}						
										ω_{43}						
										ω_{44}						
										ω_{45}						
										ω_{51}						
										ω_{52}						
										ω_{53}						
										ω_{54}						
										ω_{55}						
										ω_{61}	0.911					
										ω_{62}	0.712					
										ω_{63}	0.645					
										ω_{64}	0.626					
										ω_{65}						
										$\lambda_{\text{ajustado}}$	0.611					
										$\lambda_{\text{ajustado S1}}$	0.378					
										$\lambda_{\text{ajustado S2}}$	0.476					
										$\lambda_{\text{ajustado S3}}$	0.45					
										$\lambda_{\text{ajustado S4}}$						
										$\lambda_{\text{ajustado S5}}$						

Figura 4. Análisis complementario a Stefansson et al. (2016)

- Estado: análisis preliminar de validez y confiabilidad en estudiantes de psicología. *Liberabit*, 22(2), 219–228.
- Dominguez-Lara, S., & Merino-Soto, C. (2016). Análisis Estructural de la Escala de Afrontamiento ante la Ansiedad e Incertidumbre Pre-examen (COPEAU) en universitarios peruanos. *Revista Digital de Investigación en Docencia Universitaria*, 10(2), 32 – 47. doi: 10.19083/ridu.10.474
- Gignac, G.E. (2016). The higher-order model imposes a proportionality constraint: That is why the bifactor model tends to fit better. *Intelligence*, 55, 57–68. doi: 10.1016/j.intell.2016.01.006
- Hancock, G. R. (2001). Effect size, power, and sample size determination for structured means modeling and MIMIC approaches to between-groups hypothesis testing of means on a single latent construct. *Psychometrika*, 66, 373–388. doi: 10.1007/BF02294440
- Hancock, G. R., & Mueller, R. O. (2001). Rethinking construct reliability within latent variable systems. In R. Cudeck, S. du Toit, & D. Sörbom (Eds.), *Structural equation modeling: Present and future—A Festschrift in honor of Karl Jöreskog* (pp. 195–216). Lincolnwood, IL: Scientific Software International
- Holzinger, K. J., & Harman, H. H. (1938). Comparison of two factorial analyses. *Psychometrika*, 3(1), 45-60. doi: 10.1007/BF02287919
- Holzinger, K. J., & Swineford, F. (1937). The bi-factor method. *Psychometrika*, 2(1), 41-54. doi: 10.1007/BF02287965
- Morgan, B., de Bruin, G.P., & de Bruin, K. (2014). Operationalizing burnout in the Maslach Burnout Inventory – Student Survey: personal efficacy versus personal inefficacy. *South African Journal of Psychology*, 44(2), 216–227. doi: 10.1177/0081246314528834
- Morgan, G.B., Hodge, K.J., Wells, K.E., & Watkins, M.W. (2015). Are fit indices biased in favor of bi-factor models in cognitive ability research?: A comparison of fit in correlated factors, higher-order, and bi-factor models via Monte Carlo simulations. *Journal of Intelligence*, 3, 2–20. doi: 10.3390/jintelligence3010002
- Moscoso, M., Merino-Soto, C., Dominguez-Lara, S., Chau, C., & Claux, M. (2016). Análisis factorial confirmatorio del Inventario Multicultural de la Expresión de la Ira y Hostilidad. *Liberabit*, 22(2), 137–152.
- Raykov, T., & Hancock, G. R. (2005). Examining change in maximal reliability for multiple-component measuring instruments. *British Journal of Mathematical and Statistical Psychology*, 58(1), 65–82. doi: 10.1348/000711005X38753
- Reise, S.P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, 47(5), 667-696. doi: 10.1080/00273171.2012.715555
- Reise, S.P., Morizot, J., & Hays, R.D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research*, 16(1), 19–31. doi: 10.1007/s11136-007-9183-7
- 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. doi: 10.1177/0013164412449831
- Rodriguez, A., Reise, S.P., & Haviland, M.G. (2016a). Evaluating bifactor models: calculating and interpreting statistical indices. *Psychological Methods*, 21(2), 137–150. doi: 10.1037/met0000045
- Rodriguez, A., Reise, S.P., & Haviland, M.G. (2016b). Applying Bifactor Statistical Indices in the Evaluation of Psychological Measures. *Journal of Personality Assessment*, 98(3):223-237. doi: 10.1080/00223891.2015.1089249
- Stefansson, K.K., Gestsdottir, S., Geldhof, G.J., Skulason, S., & Lerner, R.M. (2016). A Bifactor Model of School Engagement. Assessing General and Specific Aspects of Behavioral, Emotional and Cognitive Engagement among Adolescents. *International Journal of Behavioral Development*, 40(5), 471–480. doi: 10.1177/0165025415604056
- Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, 74, 107 – 120. doi: 10.1007/s11336-008-9101-0
- Smits, I.A.M., Timmerman, M.E., Barelds, D.P.H., & Meijer, R.R. (2015). The Dutch symptom checklist-90-revised: is the use of the subscales justified? *European Journal of Psychological Assessment*, 31(4), 263-271. doi: 10.1027/1015-5759/a000233
- Stucky, B. D. & Edelen, M. O. (2015). Using hierarchical IRT models to create unidimensional measures from multidimensional data. In S. P. Reise & D. A. Revicki (Eds.), *Handbook of item response theory modeling: Applications to typical performance assessment*, (pp. 183-206). New York: Routledge.
- Stucky, B.D., Thissen, D., & Edelen, M.O. (2013). Using logistic approximations of marginal trace lines to develop short assessments. *Applied Psychological Measurement*, 37(1), 41 – 57. doi: 10.1177/0146621612462759
- Ten Berge, J. M., & Sočan, G. (2004). The greatest lower bound to the reliability of a test and the hypothesis of unidimensionality. *Psychometrika*, 69(4), 613-625. doi: 10.1007/BF02289858
- Tsubakita, T., & Shimazaki, K. (2016). Constructing the Japanese version of the Maslach Burnout Inventory – Student Survey: confirmatory factor analysis. *Japan Journal of Nursing Science*, 13, 183 – 188. doi: 10.1111/jjns.12082
- Watkins, M. W. (2013). Omega [Computer software]. Phoenix, AZ: Ed & Psych Associates.
- 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. doi: 10.1177/0146621605278814