

Invarianza Longitudinal y Error Transitorio de una Medida de Burnout Académico en Universitarios

Sergio Dominguez-Lara¹

Universidad de San Martín de Porres, Lima, Perú

RESUMEN

El presente estudio evaluó la invarianza longitudinal (IL) y error transitorio (ET) de la Escala de Cansancio Emocional (ECE) en una muestra de 116 estudiantes de psicología (79 mujeres; $M_{edad}=23.88$) de una universidad privada en Lima Metropolitana. La IL fue evaluada mediante un análisis factorial confirmatorio, y el ET, utilizando indicadores basados en puntuaciones observadas (por ejemplo, alfa test-retest) y variables latentes (por ejemplo, omega jerárquico por subescala). Los resultados indican que, con la muestra del presente estudio, la medición realizada por la ECE se mantiene invariable y con cierta fiabilidad a lo largo del tiempo. Se llegó a la conclusión que la ECE sería de utilidad en proyectos pre y post test, así como en estudios empíricos, en los cuales podría proporcionar resultados estables. Algunas limitaciones y futuras perspectivas de investigación son discutidas.

Palabras clave: agotamiento emocional académico; invarianza longitudinal; error transitorio; fiabilidad; validez.

RESUMO – Invariância Longitudinal e Erro Transitório de uma Medida de Desgaste Acadêmico em Estudantes Universitários

O presente estudo avaliou a invariância longitudinal (IL) e o erro transitório (ET) da Escala de Fadiga Emocional (EFE) em uma amostra de 116 estudantes de psicologia (79 mulheres, $M_{idade}=23,88$) de uma universidade privada em Lima Metropolitana. A IL foi avaliada usando análise fatorial confirmatória, e ET, usando indicadores baseados em pontuações observadas (por exemplo, teste-reteste reativo) e variáveis latentes (por exemplo, omega hierárquico por subescala). Os resultados indicam que, pelo menos com a amostra do presente estudo, a medição realizada pela ECE permanece invariante, aceitável e confiável ao longo do tempo. Concluiu-se que a ECE seria útil em projetos pré e pós-teste, bem como em estudos empíricos, nos quais poderia fornecer resultados estáveis. Algumas limitações e futuras perspectivas de pesquisa são discutidas.

Palavras-chave: exaustão emocional acadêmica; invariância longitudinal; erro transitório; confiabilidade; validade.

ABSTRACT – Longitudinal Invariance and Transient Error from an Academic Burnout Measure in College Students

The present study evaluated longitudinal invariance (LI) and transient error (TE) in the Emotional Fatigue Scale (EFS) in a sample of 116 psychology students (79 women, $M_{age}=23.88$) from a private university in Lima metropolitan area. LI was assessed using confirmatory factor analysis, and TE using indicators based on observed scores (e.g., reactive test) and latent variables (e.g., omega hierarchical by subscale). Results indicate that, at least with the sample from the present study, the measurement performed by EFS remains invariant and acceptable over time. It was concluded that EFS would be useful in pre- and post-test projects, as well as in empirical studies, in which it could provide stable results. Some limitations and future research perspectives are discussed.

Keywords: academic emotional exhaustion; longitudinal invariance; transient error; reliability; validity.

El carácter competitivo de la educación superior hace que muchos estudiantes desarrollen sus labores de forma ininterrumpida, dejando de lado actividades que impliquen descanso y reparación psicológica para reiniciar una jornada académica, agotando emocionalmente a la persona y disminuyendo su grado de motivación. Debido a ello, surge el *burnout académico* (BA), definido como el sentimiento de agotamiento a causa de las demandas del estudio, generando una actitud cínica y poco comprometida con las labores académicas y con creencias de poca

capacidad para hacer frente a tales demandas (Schaufeli, Martínez, Marques, Salanova, & Bakker, 2002).

El agotamiento emocional académico (AEM) es la primera fase y núcleo del BA (Caballero, Hederich, & Palacio, 2010), y su presencia en niveles moderados o altos suele considerarse como indicio de burnout (Tuithof et al., 2017). Además, es frecuente en estudiantes de carreras relacionadas con ciencias de la salud (Atalayin, Balkis, Tezel, Onal, & Kayrak, 2015; Fares et al., 2016; Ríos-Risquez, García-Izquierdo, Sabuco-Tebar,

¹ Endereço para correspondência: Instituto de Investigación de Psicología – Universidad de San Martín de Porres, Avenida Tomás Marsano, 242, 5to piso, Lima 34 – Perú. E-mails: sdominguezmpcs@gmail.com; sdominguezl@usmp.pe

Carrillo-García, & Martínez-Roche, 2016), y destaca por su relación con constructos psicológicos que reflejan comportamientos poco adaptativos, como ansiedad, depresión, entre otros (ver Domínguez-Lara, Fernández-Arata, Manrique-Millones, Alarcón-Parco, & Díaz-Peñaloza, 2017),

La aparición y posterior cronicidad del BA es un riesgo latente en instituciones de educación superior debido a las consecuencias en el estudiante, por lo que el diagnóstico e intervención en esta primera etapa facilitarían la recuperación en la fase inicial (e.g., Bresó, Schaufeli, & Salanova, 2011). No obstante, independientemente del tipo de intervención que se realice, para analizar su eficacia es necesario establecer una línea base y evaluar los cambios después de un tiempo. En este sentido, se espera que los cambios sean *verdaderos*, es decir, que las diferencias entre las puntuaciones de los sujetos antes y después de la intervención sean producto de esta.

Si bien existen instrumentos que evalúan el AEM como parte de un modelo de BA más amplio (MBI-SS; Schaufeli et al., 2002), el uso y cambios que potencialmente se le pueda hacer a la estructura interna o fraseo de los ítems normalmente están limitados por derechos de autor de índole comercial, y su difusión tiende a ser limitada por la misma razón. Por ello, es necesario utilizar instrumentos que sean de libre acceso y de carácter breve, de tal forma que puedan integrarse fácilmente a baterías de evaluación extensas y sean utilizados en procesos de evaluación masivos (Domínguez-Lara & Merino-Soto, 2015), como la *Escala de Cansancio Emocional* (ECE; Ramos, Manga, & Moran, 2005).

La ECE fue construida específicamente para evaluar el AEM en estudiantes universitarios de contextos hispanoparlantes, y presenta estudios de psicométricos en varios países de la región (e.g., Domínguez-Lara, 2013, 2014a; Fontana, 2011). Su brevedad e indicadores psicométricos robustos la hacen un instrumento que ofrece una medición confiable ($\alpha > .85$) y unidimensional del AEM. Con todo, hasta la fecha, en conocimiento del autor, no existen estudios que reporten su estabilidad temporal. Cabe precisar que se espera cierto grado de estabilidad debido a que se trata de una condición crónica más que transitoria (Maslach, Schaufeli, & Leiter, 2001). Además, está fuertemente asociado al *neuroticismo*, rasgo que posee estabilidad temporal documentada (Chmielewski & Watson, 2009).

Ante la carencia de estudios de estabilidad de las medidas de BA, fueron hallados reportes de estabilidad temporal considerando el burnout general. En cuanto a dichos estudios, la estabilidad para intervalos de un (1) año para el AEM fue elevada en docentes ($r = .77$; $n = 967$) y cuidadores ($r_{1,12} = .65$; $r_{2,13} = .73$; $r_{11,13} = .65$; $n = 218$) (Tàris, Le Blanc, Schaufeli, & Schereurs, 2005). En otro estudio (Balogun, Helgemoe, Pellegrini, & Hoerberlein, 1995), la relación entre la primera y segunda aplicación (intervalo: cuatro semanas siguientes) fue

elevada ($r = .82$; $n = 55$), pero su magnitud disminuye luego de un año ($r = .60$; $n = 55$) en la misma muestra; aunque en otro trabajo (Brouwers & Tomic, 2000) luego de cinco meses la relación sigue siendo elevada ($r = .84$; $n = 243$). Por otro lado, en un estudio de seguimiento de 10 años ($t1$: línea base; $t2$: luego de cinco años; $t3$: luego de 10 años) fueron obtenidas magnitudes similares al evaluar el AE en tres tiempos ($r_{1,12} = .66$; $r_{12,13} = .63$; $r_{11,13} = .59$; $n = 165$), apreciándose un ligero decremento mientras avanza el tiempo (Schaufeli, Maasen, Bakker, & Sixma, 2011). Un resultado similar ($r = .62$; $n = 713$) fue hallado para el AE en un estudio de estabilidad luego de ocho años de seguimiento (Toppinnen-Tanner, Kalimo, & Mutanen, 2002). Esto hace pensar que aunque en intervalos cortos la magnitud de r_p sea alta, posteriormente va disminuyendo. Esta aparente contradicción estaría relacionada con algunos aspectos conceptuales relativos a las fuentes de error al estimar la confiabilidad de las puntuaciones.

Desde la Teoría Clásica de los Tests (TCT), toda puntuación obtenida a partir de la aplicación de un test (X) presenta dos componentes: puntuación verdadera y error de medición; considerándose a este último como *error aleatorio* (EA). Esto podría resultar apropiado cuando se trata de una sola aplicación, pero cuando es evaluada la eficacia de una intervención deben realizarse al menos dos mediciones, y es necesario conocer si la estabilidad temporal de las medidas utilizadas es la adecuada para concluir apropiadamente sobre la utilidad de las intervenciones (de Beurs, Fokkema, de Groot, de Keijser, & Kerkhof, 2015), porque en el lapso entre una evaluación y otra pueden ocurrir cambios en la respuesta que deben ser evaluados (Fokkema, Smits, Kelderman, & Cuijpers, 2013).

La *consistencia* y *estabilidad* son evaluadas a partir de un coeficiente de correlación, pero difieren entre sí. La *consistencia* se define como “la correlación entre dos aplicaciones del mismo test cuando el lapso de tiempo es insuficiente para que la gente cambie por sí misma con respecto a lo que está siendo medido” (Cattell, Eber, & Tatsuoka, 1970, p. 30). Una medición consistente indica similitud entre las puntuaciones de dos (o más) ocasiones, y que no interviene el cambio verdadero. Por el contrario, la estabilidad informa sobre qué tan estables son las puntuaciones a través del tiempo, pero incorporando tanto el EA como el cambio verdadero (Chmielewski & Watson, 2009; Watson, 2004).

Dicho esto, es probable que la *consistencia* caracterice en mayor medida a constructos como inteligencia o personalidad (Chmielewski & Watson, 2009), donde no se esperaría un cambio sustancial entre mediciones, más aún cuando el intervalo de aplicación es corto; pero al ser el AEM teóricamente estable en el tiempo (Maslach et al., 2001), aunque con una carga afectiva evidente, su medición puede ser vulnerable a los cambios emocionales de la persona observados al momento de las

evaluaciones, sobre todo en el ámbito académico, donde uno de los principales estresores son los exámenes (Vizoso & Gundín, 2016), los que suelen exacerbar algunas reacciones, derivando en una fuente de error diferenciada: *el error transitorio* (ET).

El ET hace referencia a un tipo de error sistemático que refleja las fluctuaciones en los estados psicológicos (sentimientos, ánimo, atención, etc.) de los participantes en determinado momento (Green, 2003), lo que agrega varianza irrelevante a constructos estables considerados como *rasgos* (Chmielewski & Watson, 2009). En el caso específico de la evaluación del AEM, indica con cuánta varianza contribuye el estado psicológico de los estudiantes durante la evaluación.

A pesar de que en apariencia la evaluación del ET es fútil, sus consecuencias son importantes, debido a que al ignorarlo, los cambios observados podrían ser interpretados de forma errada (Fong et al., 2016), y más aún cuando se trata de un constructo basado en reacciones emocionales como el AEM, y sobre el que se tomarían decisiones respecto a la eficacia de un tratamiento. Por ejemplo, si no se aprecian diferencias estadísticamente significativas luego de la intervención podría concluirse que no fue eficaz; pero si no existe información sobre el grado de variabilidad debido a la diferencia sistemática de la personas en torno al constructo evaluado (varianza verdadera) o del monto de ET, podría concluirse de forma sesgada.

Existen dos medidas que evalúan esa variabilidad. En primer lugar, ha sido desarrollado el *alfa test-retest* (α_{T1T2} ; Green, 2003), que en esencia busca cuantificar la varianza verdadera en una medición hecha en dos tiempos controlando el ET. Adicionalmente, existe una variante del *Coficiente de Equivalencia y Estabilidad* (CEE) sin la restricción sobre su uso exclusivo en medidas paralelas (Schmidt, Le, & Ilies, 2003). El CEE informa de la magnitud de la confiabilidad de las puntuaciones controlando otra fuentes de error, y podría considerarse una mejor estimación de la confiabilidad (Schmidt et al., 2003). Pese a que no informan sobre la cantidad de varianza atribuida al ET, pueden dar información sobre el monto de confiabilidad que se deja de cuantificar de una ocasión a otra, lo que podría atribuirse hasta cierto punto al ET. Todos estos resultados son relevantes cuando se pretende analizar tanto la estabilidad de las puntuaciones (e.g., test-retest) y la evaluación de las fuentes de variabilidad.

Además de lo mencionado, si no se garantiza la equivalencia de la medición entre situaciones, cualquier comparación posterior podría estar sujeta a error (Byrne, 2008). De esta forma, es necesario evaluar una variante de la invarianza de medición para datos longitudinales (IL; Brown, 2015) de forma previa al análisis del ET.

El α_{T1T2} y el CEE poseen como insumos las puntuaciones observadas. Sin embargo, las aproximaciones mediante el modelamiento SEM pueden ser más robustas

debido a que trabajan con variables latentes. En este sentido, puede cuantificarse el ET proveniente de ambas ocasiones por separado por medio de un modelamiento bifactor (Canivez, 2016) teniendo como factor general (FG) a la puntuación verdadera, y como factores específicos al AEM de cada tiempo (en los cuales se estimarían los ET), en un sistema denominado *modelo de la puntuación verdadero test-retest* (Green, 2003).

De esta forma, el primer objetivo del presente estudio fue analizar la IL de la ECE en estudiantes universitarios. Luego de ello, determinar el grado de ET mediante las aproximaciones descritas anteriormente (α_{T1T2} , CEE y modelo bifactor), a fin de conocer cuánto podrían ser afectadas las puntuaciones por las diversas fuentes de error (e.g., ET) en diseños pre y post-intervención, así como en estudios empíricos (Chi & Yang, 2015).

Hasta donde se sabe, es la primera investigación que explora de forma simultánea tanto la IL como el ET en medidas relacionadas con el BA, y podría ser un aporte metodológico para la evaluación de constructos en diseños pre-postest.

Método

Participantes

Se contó con muestra intencional de 116 estudiantes universitarios (79 mujeres) de 3er a 7mo ciclo de la carrera de psicología de una universidad privada de Lima Metropolitana con edades comprendidas entre 17 y 56 años ($M=23.88$, $DE=6.45$), siendo esta similar respecto al género ($t_{[113]}=1.614$; $p>.05$; $d=.33$)

Instrumento

Escala de Cansancio Emocional (ECE; Fontana, 2011). Es una medida unidimensional de AEM que consta de 10 ítems que se puntúan de uno a cinco (desde *Raras veces* hasta *Siempre*), considerando los 12 últimos meses de vida estudiantil. Para el presente estudio se utilizó la versión adaptada a universitarios peruanos (Dominguez-Lara, 2013, 2014a).

Procedimiento

La participación voluntaria fue solicitada en horas de clases de forma presencial, explicando a los participantes que se les contactará por una segunda ocasión para realizar el cuestionario nuevamente dentro de cuatro semanas. Se les garantizó la confidencialidad de los resultados a fin de elevar la tasa de respuestas positivas. Los estudiantes que aceptaron participaron firmaron un consentimiento informado, y fueron evaluados inmediatamente (T1=Aplicación 1). Se les solicitó un código de participación (los cuatro últimos dígitos de su documento nacional de identidad) para identificar sus respuestas en la segunda aplicación (T2), en la cual fueron obtenidas las respuestas de 116 estudiantes (de los 227 que accedieron participar en T1). El proyecto fue

aprobado por el Instituto de Investigación de psicología de la institución en la cual se ejecutó el estudio, teniendo en consideración los aspectos éticos y procedimentales de la investigación. En cuanto a las consideraciones éticas en el trabajo, el autor declara que para esta investigación no se han realizado experimentos en seres humanos ni en animales.

Cabe precisar que el intervalo de tiempo elegido para T2 fue de cuatro semanas debido a que no se espera un cambio sustancial del AE en ese periodo temporal, y además otros estudios han utilizado intervalos similares (e.g., Balogun et al., 1995). Por último, la recolección de información se llevó a cabo en el marco de un trabajo orientado a la validación de otros instrumentos de medición.

Análisis de los datos

Inicialmente fueron calculados estadísticos descriptivos y de distribución, así como el índice estandarizados de asimetría (*Standardized Skew Index*, SSI; Malgady, 2007) esperando magnitudes moderadas ($SSI \leq .50$) para brindar un acercamiento a la normalidad de cada ítem. Finalmente, la normalidad multivariada fue estimada con el estadístico (G2) de Mardia esperando magnitudes < 70 (Rodríguez, & Ruiz, 2008).

Previo al análisis de la IL fue analizada la estructura interna de la ECE con los datos de la primera y segunda aplicación por separado mediante un análisis factorial confirmatorio ejecutado con el programa EQS 6.2 (Bentler & Wu, 2012), usando el método de máxima verosimilitud robusto con matriz de correlaciones policóricas (Dominguez-Lara, 2014b). El ajuste del modelo fue evaluado a partir de la evaluación conjunta de la prueba $SB-\chi^2$ (Satorra & Bentler, 1994), el RMSEA ($\leq .05$), CFI ($\geq .95$), y el SRMR ($\leq .08$). Este procedimiento es necesario porque si la configuración del instrumento (estructura interna) varía de un momento a otro, cualquier análisis de invarianza sería nulo. Además, debido a su carácter invariante respecto al género (Dominguez-Lara et al., 2017), no se realizó un análisis diferenciado.

La IL fue realizada bajo un análisis de medias y estructuras de covarianzas, por lo que desde un principio fueron modelados los interceptos (τ). Este modelamiento hace inviable el uso de correlaciones policóricas, optando por las matrices de covarianzas (Dominguez-Lara et al., 2017). Esta decisión podría causar algunas dificultades, dado que para el análisis de instrumentos como la ECE (cuyos ítems son básicamente medidas ordinales), el uso de matrices de covarianzas tiende a infraestimar los índices de ajuste (e.g. Dominguez-Lara & Merino-Soto, 2016).

La IL fue realizada considerando las dos olas de datos (T1 y T2) de forma simultánea, pero el proceso para evaluar el grado de invarianza adquiere una forma

diferente de los procedimientos tradicionales realizados normalmente bajo el enfoque del análisis factorial de grupo múltiple (Byrne, 2008). En primer lugar, fueron modelados dos factores oblicuos (AET1 y AET2) con sus respectivos ítems, considerando adicionalmente las covarianzas entre residuales de ítems análogos (e.g., modelando la covarianza del residual del ítem 1 de T1, con el residual del ítem 1 de T2) (Figura 1). Las correlaciones entre residuales fueron modeladas porque las medidas repetidas agregan varianza irrelevante y mediante esa especificación se buscan controlar su efecto (Pitts, West, & Tein, 1996).

De forma anidada al modelo anterior, se evaluó la igualdad de cargas factoriales correspondientes a T1 y T2. Por ejemplo, se igualó λ_{11} (carga factorial del ítem 1 del T1) con λ_{12} (carga factorial del ítem 1 del T2), lo que evidencia que la relación de los ítems con la variable latente en T1 y T2 (Brown, 2015). Una vez corroborada la condición anterior, fueron igualados los interceptos de forma similar a las cargas factoriales. Al cumplir la igualdad de interceptos, es posible comprobar la igualdad estadística de las medidas latentes. En este contexto significaría que las puntuaciones verdaderas de T1 y T2 son similares. Por último, fue analizada la invarianza de los residuales entre ítems homólogos. Para concluir a favor de cada etapa de la IL fueron consideradas las variaciones en el CFI ($CFI_1 - CFI_2 < .01$; Cheung & Rensvold, 2002), el RMSEA ($RMSEA_2 - RMSEA_1 < .015$; Chen, 2007), y SRMR ($SRMR_2 - SRMR_1 < .030$; Chen, 2007).

En cuanto a la confiabilidad de las puntuaciones, a fin de establecer una línea base para su evaluación tanto en T1 como en T2 utilizando el coeficiente α , fue evaluado el modelo de medición tau-equivalente (Dunn, Baguley, & Brunson, 2014) en cada ola por separado. Para concluir sobre la equivalencia de las cargas factoriales de la ECE en T1 y T2 (modelo tau-equivalente) fueron consideradas las variaciones en CFI, RMSEA y SRMR expuestas anteriormente. Asimismo fueron implementados los coeficientes ω (McDonald, 1999) y H (Dominguez-Lara, 2016; Hancock & Mueller, 2001) para la evaluación de la confiabilidad del constructo.

Con relación a la evaluación de error transitorio, para la estimación del α_{T1T2} fue usado un módulo de cálculo basado en la propuesta de Green (2003), a partir de una matriz de covarianzas que considera de forma simultánea los ítems de T1 y T2. En cuanto al CEE (Schmidt et al., 2003), previamente fueron obtenidas mitades de cada ola de datos (T1 y T2), considerando los ítems pares e impares de la ECE. Fue calculado un coeficiente de equivalencia (CE) que consistió en el promedio de los coeficientes α de la ECE obtenidos en T1 y T2, y el ET es producto de la diferencia entre el CE y el CEE ($ET = CE - CEE$; Schmidt et al., 2003).

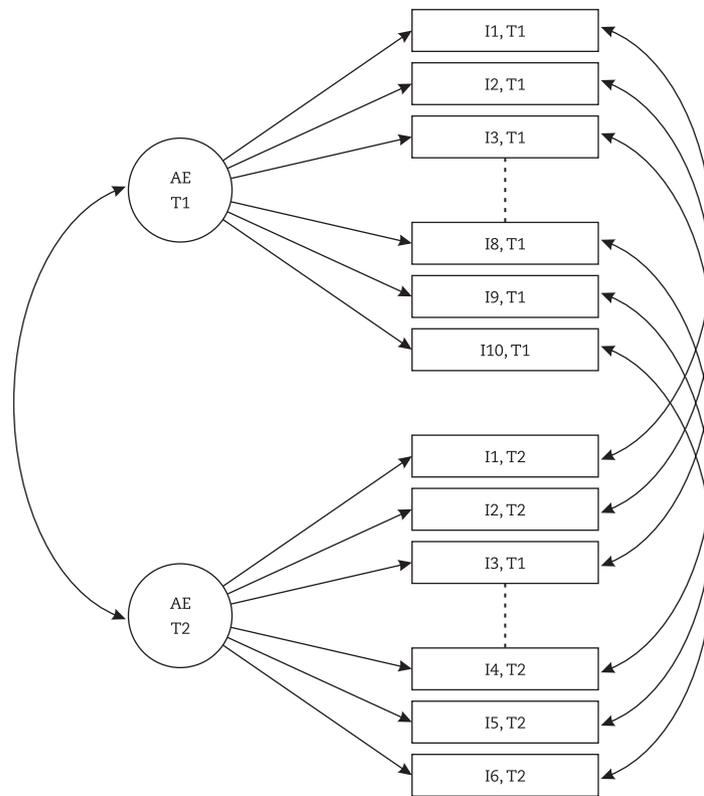


Figura 1. Modelo de invarianza longitudinal de la ECE

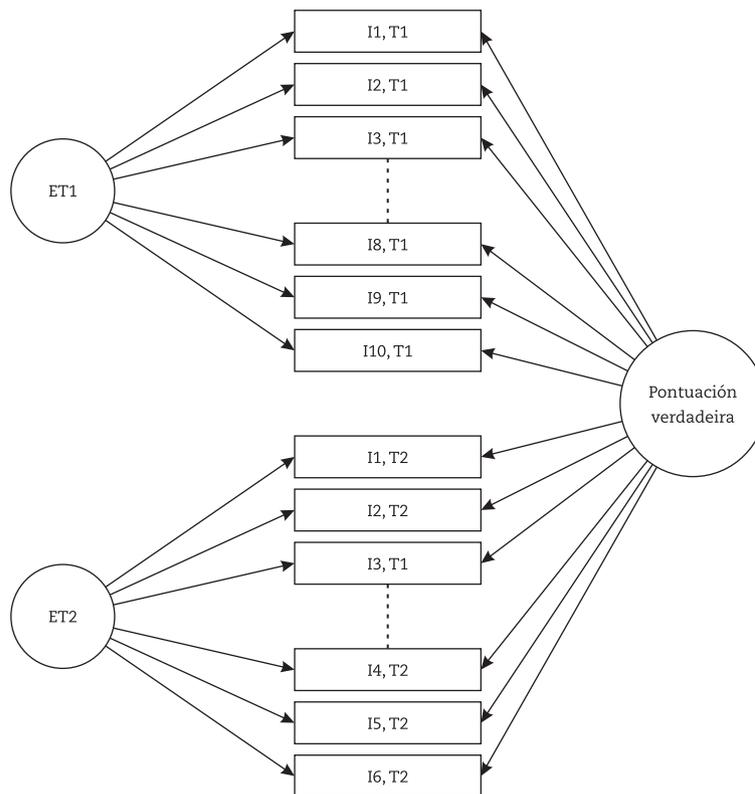


Figura 2. Modelo de puntuación verdadera de la ECE. Nota: por motivos de simplicidad del gráfico se omitieron las correlaciones entre residuales

El modelo de la puntuación verdadera test-retest (Green, 2003) fue evaluado bajo la forma de un modelo bifactor (Rodríguez, Reise, & Haviland, 2016), incluyendo la correlación entre los residuales de los ítems equivalentes entre tiempos (Figura 2), siguiendo los mismos criterios del procedimiento inicial previo a la IL (máxima verosimilitud, matrices policóricas, e índices de ajuste). Adicionalmente fueron empleados otros estadísticos de interés dado que el objetivo es cuantificar la varianza debido a la *puntuación verdadera*, representada en el modelo por el FG. En este sentido, fue utilizado el ω_i (Zinbarg, Yovel, Revelle, & McDonald, 2006), que cuantifica el monto de varianza total atribuida al FG; el ω_{hs} que refleja la varianza de los factores específicos controlando la influencia del FG (Rodríguez et al.,

2016); y el ECV (*Explained Common Variance*; Sijtsma, 2009) que representa el monto de varianza común que se debe al FG. Los cálculos se realizaron utilizando un módulo especializado (Dominguez-Lara, & Rodríguez, 2017).

Resultados

Con respecto al análisis descriptivo realizado, los ítems tienen hacia la respuesta central, excepto en el ítem tres (*Me siento bajo de ánimo, como triste, sin motivo aparente*), que mostró la menor puntuación en ambas aplicaciones. Del mismo modo, el alejamiento de la normalidad univariada ($SSI < .25$) y multivariado ($G2 < 70$) no fue significativo (Tabla 1).

Tabla 1
Estadísticos Descriptivos de la ECE en T1 y T2

	Primera aplicación (T1)					Segunda aplicación (T2)				
	M	DE	g_1	SSI	g_2	M	DE	g_1	SSI	g_2
Ítem 1	2.672	1.053	-.126	.057	-.676	2.647	1.09	.047	.020	-.605
Ítem 2	2.138	.968	.479	.255	-.467	2.250	1.102	.518	.213	-.451
Ítem 3	1.845	.938	.767	.436	-.505	1.776	.876	.851	.555	-.196
Ítem 4	2.388	1.094	.312	.131	-.745	2.526	1.034	.002	.001	-.949
Ítem 5	2.086	1.191	.837	.295	-.323	2.103	1.182	.760	.272	-.506
Ítem 6	2.491	1.075	.193	.084	-.585	2.612	1.086	.080	.034	-.720
Ítem 7	2.112	1.020	.673	.324	-.199	2.388	1.109	.580	.236	-.264
Ítem 8	2.69	1.107	.095	.039	-.785	2.922	1.12	-.109	.044	-.552
Ítem 9	2.552	1.175	.25	.091	-.615	2.534	1.233	.302	.099	-.884
Ítem 10	2.371	1.115	.561	.226	-.283	2.397	1.094	.289	.121	-.608
G2	8.587					16.386				

Nota. M=Media; DE=Desviación estándar; g_1 =Asimetría; SSI=Índice estandarizado de asimetría; g_2 =Curtosis; G2=Coefficiente de Mardia

En relación a la IL, los índices de ajuste hallados confirman la estructura unidimensional de la ECE en T1 y T2 por separado (Tabla 2). En cuanto a la IL, los índices de ajuste del primer modelo (M1: Dos factores oblicuos y errores correlacionados) servirán como línea base para las siguientes restricciones. La correlación entre factores fue de magnitud elevada ($\phi_{T1T2}=.729$). Las variaciones en cuanto al CFI, RMSEA y SRMR apoyan la IL en sus diferentes grados (Tabla 2). Adicionalmente, no fueron halladas diferencias significativas entre las puntuaciones latentes de T1 y T2 (M4).

En cuanto a la evaluación de error transitorio, la ECE cumple con el supuesto de tau-equivalencia tanto en T1 como en T2 (Tabla 2), por lo que el uso del coeficiente α está justificado. El α_{T1T2} fue .657, lo que implica que poco menos del 70% de la variabilidad entre

puntuaciones, es verdadera. El CEE alcanzó una magnitud de .660, y el CE fue .887 ($\alpha_{T1}=.873$, $\alpha_{T2}=.900$), siendo el ET de .227. La confiabilidad del constructo para T1 ($\omega=.878$, $H=.894$) y T2 ($\omega=.902$, $H=.921$) fue elevada.

En cuanto al modelo de la puntuación verdadera test-retest, el ajuste inicial del modelo fue adecuado ($SB-\chi^2_{[140]}=161.206$, $p=.106$; CFI=.995; RMSEA [IC 90%]=.036 [.000, .059]; SRMR=.063). Asimismo, el 64% de la varianza en los ítems se debe a la puntuación verdadera ($\omega_i=.800$), mientras que solo un 7.78% se debe al ET de T1 ($\omega_{hs}=.279$), y un 4% al ET de T2 ($\omega_{hs}=.200$), explicando de forma conjunta poco menos del 12% de variabilidad. Asimismo, con un ECV de .701 se concluye que una cantidad importante de la varianza común se debe a la puntuación verdadera.

Tabla 2
Índices de Ajuste de la ECE: Modelos de Medición e Invarianza Longitudinal

Modelos de medición ^a	SB- χ^2 (gl, p)	CFI	RMSEA (IC 90%)	SRMR	Δ CFI	Δ RMSEA	Δ SRMR
Unidimensional (T1)	38.646 (35, .308)	.996	.030 (.000, .075)	.055			
Tau-equivalencia (T1)	52.752 (44, .172)	.991	.042 (.000, .078)	.081	-.005	.012	.026
Unidimensional (T2)	59.486 (35, .006)	.984	.078 (.042, .111)	.061			
Tau-equivalencia (T2)	76.932 (44, .002)	.978	.081 (.049, .110)	.094	-.006	.003	.033
Invarianza longitudinal ^b	SB- χ^2 (gl, p)	CFI	RMSEA ^c (IC 90%)	SRMR	Δ CFI	Δ RMSEA	Δ SRMR
M1: Dos factores oblicuos y errores correlacionados	249.644 (157, <.001)	.909	.072 (.054, .088)	.082			
M2: M1 + Igualdad de cargas factoriales	262.631 (167, <.001)	.906	.071 (.054, .086)	.093	-.003	-.001	.011
M3: M2 + Igualdad de interceptos	285.918 (177, <.001)	.896	.073 (.057, .088)	.095	-.010	.002	.002
M4: M3 + Igualdad de medias latentes	288.429 (178, <.001)	.896	.073 (.057, .088)	.094	.000	.000	-.001
M5: M3 + Igualdad de residuales	297.182 (187, <.001)	.898	.072 (.056, .086)	.095	.002d	-.001d	.001d

Nota. T1=Primera aplicación; T2=Segunda aplicación; ^a=Realizado en base a la matriz de correlaciones policóricas; ^b=Realizado en base a la matriz de covarianzas; ^c=basada en la matriz de covarianza y las medias; ^d=comparación respecto a M3

Discusión

Los objetivos del presente estudio se enfocaron en analizar la IL de la ECE en estudiantes universitarios, así como en cuantificar el ET tanto en función de las puntuaciones observadas como de variables latentes (en cada etapa) ya que no es una práctica frecuente en los estudios empíricos ligados al BA y esa evidencia es condición necesaria para arribar a conclusiones menos sesgadas al momento de evaluar la eficacia de las intervenciones realizadas.

Con relación al análisis de la IL, se evidencia que la medición es equivalente en las dos ocasiones (T1 y T2), llegando incluso a considerarse paralelas. A pesar de que no existen investigaciones previas en el ámbito académico que evalúen la IL de medidas de BA, un estudio longitudinal llevado a cabo en población de trabajadores considera como un paso necesario este procedimiento a fin de concluir de sobre la estabilidad de las puntuaciones y las conclusiones derivadas de estas (Fong et al., 2015). Por otro lado, la fiabilidad de las puntuaciones observadas y variables latentes es adecuada (>.80).

Con respecto a la evaluación del ET, los dos métodos enfocados en puntuaciones observadas muestran resultados similares, indicando que poco más del 20% de

la varianza de las puntuaciones corresponderían a cambios en el estado anímico de los participantes, y cerca del 70% correspondería a la varianza verdadera. Aunque según algunos autores el valor mínimo aceptable para el α_{T1T2} es .70 (e.g., Axford & Hobbs, 2011), la magnitud hallada en el presente estudio puede ser considerada adecuada a decir del tamaño muestral y la cantidad de ítems de la escala (.65 < α < .70; Ponterotto & Charter, 2009). Asimismo, cuando se explora el ET a nivel de variables latentes (modelamiento bifactor), su variabilidad disminuye casi a la mitad (12%) respecto al hallazgo previo. Si bien el porcentaje de varianza verdadera disminuye de forma mínima, los resultados son equivalentes al método anteriormente descrito. En conclusión, debido a que los resultados basados en variables latentes tienden a ser más robustos que los obtenidos con puntuaciones, el método basado en modelamiento bifactor diferencia en mayor grado el EA del ET, concluyendo de forma más específica sobre este último.

En este sentido, la ausencia de diferencia entre las puntuaciones latentes de T1 y T2 con relación al AEM experimentado por los estudiantes (M4), y el porcentaje similar de varianza vinculada al ET en cada etapa informa que las fuentes de variabilidad intrasujeto son similares en ambas ocasiones, quizás se deba a la relación

significativa entre el AEM y las diferentes manifestaciones de afecto negativo (Bianchi, Schonfeld, & Laurent, 2015; Caballero, Gutiérrez, & Palacio, 2015), las cuales podrían aparecer en la forma de *estado* como algo circunstancial (Geiser & Lockhart, 2012).

Entonces, de acuerdo con lo descrito, la medición realizada por la ECE se mantiene invariante y aceptablemente confiable a través del tiempo. De este modo, y de forma provisional, la ECE sería de utilidad en diseños pre y post test, así como en estudios empíricos, en los cuales podría brindar resultados estables.

El estudio presentó algunas limitaciones. En primer lugar, si bien el tamaño muestral puede considerarse pequeño, un re-análisis realizado con información de estudios antecedentes (magnitud de correlaciones, y tamaño muestral), muestran una baja correlación entre ambas ($r=.138$), por lo que no impactaría de forma significativa. En segundo lugar,

pudo afectar los resultados mostrados que poco más del 50% de los estudiantes evaluados en la primera ocasión participen en la segunda etapa, sobre todo las predicciones de los resultados obtenidos (Wolke et al., 2009). Por último, la participación exclusiva de estudiantes de psicología podría implicar un sesgo de *deseabilidad social*, al no reportar algunas características que podrían juzgarse como indeseables.

Los hallazgos presentados brindan la posibilidad de otros estudios. Por ejemplo, además de ampliar la muestra y contar con la participación de estudiantes de otras carreras, podrían considerarse las mediciones en diversos periodos de tiempo (Taris et al., 2005; Schaufeli et al., 2011), porque de esa forma la estabilidad de la medición realizada podría tener mayores evidencias. Otro aspecto a destacar es la exploración de las fuentes de ET, ya que al reflejar las fluctuaciones en los estados psicológicos, es posible evaluarlas.

Referencias

- Atalayin, C., Balkis, M., Tezel, H., Onal, B., & Kayrak, G. (2015). The prevalence and consequences of burnout on a group of preclinical dental students. *European Journal of Dentistry*, *9*(3), 356-363. doi: 10.4103/1305-7456.163227.
- Axford, N., & Hobbs, T. (2011). Getting the measure of child health and development outcomes (1): A method for use in children's services settings. *Child Indicators Research*, *4*(1), 59-80. doi: 10.1007/s12187-010-9074-2
- Balogun, J. A., Helgemoe, S., Pellegrini, E., & Hoerberlein, T. (1995). Test-retest reliability of a psychometric instrument designed to measure physical therapy students' burnout. *Perceptual and Motor Skills*, *81*(2), 667-672. doi: 10.2466/pms.1995.81.2.667
- Bianchi, R., Schonfeld, I.S., & Laurent, E. (2015). Burnout-depression overlap: A review. *Clinical Psychology Review*, *36*, 28-41. doi: 10.1016/j.cpr.2015.01.004
- Bentler, P. M., & Wu, E. J. C. (2012). *EQS 6.2 for windows* [Statistical Program]. Encino, CA: Multivariate Software, Inc.
- Bresó, E., Schaufeli, W. B., & Salanova, M. (2011). Can a self-efficacy-based intervention decrease burnout, increase engagement, and enhance performance? A quasi-experimental study. *Higher Education*, *61*(4), 339-355. doi: 10.1007/s10734-010-9334-6
- Brouwers, A., & Tomic, W. (2000). A longitudinal study of teacher burnout and perceived self-efficacy in classroom management. *Teaching and Teacher Education*, *16*(2), 239-253. doi: 10.1016/S0742-051X(99)00057-8
- Brown, T. (2015). *Confirmatory Factor Analysis for Applied Research* (2nd Ed.). New York: The Guilford Press.
- Byrne, B. (2008). Testing for multigroup equivalence of a measuring instrument: A walk through the process. *Psicothema*, *20*(4), 872-882.
- Caballero, C., Gutiérrez, O., & Palacio, J. (2015). Relación del burnout y el engagement con depresión, ansiedad y rendimiento académico en estudiantes universitarios. *Salud Uninorte*, *31*(1), 59-69. doi: 10.14482/sun.31.1.5085
- Caballero, C., Hederich, C., & Palacio, J. (2010). El burnout académico: delimitación del síndrome y factores asociados a su aparición. *Revista Latinoamericana de Psicología*, *42*(1), 131-146.
- 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). Gottingen, Germany: Hogrefe.
- Cattell, R. B., Eber, H. W., & Tatsuoka, M. M. (1970). *Handbook for the Sixteen Personality Factor Questionnaire (16PF)*. Champaign, IL: Institute for Personality and Ability Testing.
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, *14*(3), 464-504. doi: 10.1080/10705510701301834
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*(2), 233-255. doi: 10.1207/S15328007SEM0902_5
- Chi, S. C., & Yang, M. Y. (2015). How does negative mood affect turnover intention? The interactive effect of self-monitoring and conflict perception. *European Journal of Work and Organizational Psychology*, *24*(1), 31-43. doi: 10.1080/1359432X.2013.856299
- Chmielewski, M., & Watson, D. (2009). What is being assessed and why it matters: The impact of transient error on trait research. *Journal of Personality and Social Psychology*, *97*(1), 186-202. doi: 10.1037/a0015618.
- de Beurs, D. P., Fokkema, M., de Groot, M. H., de Keijser, J., & Kerkhof, A. J. F. M. (2015). Longitudinal measurement invariance of the Beck Scale for Suicide Ideation. *Psychiatry Research*, *225*(3), 368-373. doi: 10.1016/j.psychres.2014.11.075
- Dominguez-Lara, 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. doi: 10.19083/ridu.7.186
- Dominguez-Lara, S. (2014a). 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-Lara, S. (2014b). ¿Matrices Policóricas/Tetracóricas o Matrices Pearson? Un estudio metodológico. *Revista Argentina de Ciencias del Comportamiento*, 6(1), 39-48.
- Dominguez-Lara, S. (2016). 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., & Merino-Soto, C. (2015). 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-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
- Dominguez-Lara, S., & Rodriguez, A. (2017). Índices estadísticos de modelos bifactor. *Interacciones*, 3(2), 59-65. doi: 10.24016/2017.v3n2.51
- Dominguez-Lara, S., Fernández-Arata, M., Manrique-Millones, D., Alarcón-Parco, D., & Díaz-Peñaloza, M. (2017). Datos normativos de una escala de agotamiento emocional académico en estudiantes universitarios de psicología de Lima (Perú). *Educación Médica*. Avance online. doi: 10.1016/j.edumed.2017.09.002
- Dunn, T. J., Baguley, T., & Brunsden, 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. doi: 10.1111/bjop.12046
- Fares, J., Saadeddin, Z., Al Tabosh, H., Aridi, H., El Mouhayyar, C., Koleilat, M. K., Chaaya, M., & El Asmar, K. (2016). Extracurricular activities associated with stress and burnout in preclinical medical students. *Journal of Epidemiology and Glob Health*, 6(3), 177-185. doi: 10.1016/j.jegh.2015.10.003
- Fokkema, M., Smits, N., Kelderman, H., & Cuijpers, P. (2013). Response shifts in mental health interventions: An illustration of longitudinal measurement invariance. *Psychological Assessment*, 25(2), 520-531. doi: 10.1037/a0031669
- Fong, T.C.T, Ho, R.T.H., Au-Yeung, F.S.W., Sing, C.Y., Law, K.Y., Lee, L.F., & Ng, S.M. (2016). The relationships of change in work climate with changes in burnout and depression: a 2-year longitudinal study of Chinese mental health care workers. *Psychology, Health & Medicine*, 21(4), 401-412. doi: 10.1080/13548506.2015.1080849.
- Fontana, S. (2011). Estudio preliminar de las propiedades psicométricas de la escala de desgaste emocional para estudiantes universitarios. *Revista Argentina de Ciencias del Comportamiento*, 3(2), 44-48.
- Geiser, C., & Lockhart, G. (2012). A comparison of four approaches to account for method effects in latent state-trait analyses. *Psychological Methods*, 17(2), 255-283. doi: 10.1037/a0026977
- Green, S. B. (2003). A coefficient alpha for test-retest data. *Psychological Methods*, 8(1), 88-101. doi: 10.1037/1082-989X.8.1.88
- 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.
- Malgady, R. (2007). How skew are psychological data? A standardized index of effect size. *The Journal of General Psychology*, 134(3), 355-359. doi: 10.3200/GENP.134.3.355-360
- Maslach, C., Schaufeli, W.B., & Leiter, M.P. (2001). Job burnout. *Annual Review of Psychology*, 52(1), 397-422. doi: 10.1146/annurev.psych.52.1.397
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, N.J.: L. Erlbaum Associates.
- Pitts, S. C., West, S. G., & Tein, J. Y. (1996). Longitudinal measurement models in evaluation research: Examining stability and change. *Evaluation and Program Planning*, 19(4), 333-350. doi: 10.1016/S0149-7189(96)00027-4
- 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(3), 878-886. doi: 10.2466/PMS.108.3.878-886
- Ramos, F., Manga, D., & Moran, C. (2005). Escala de Cansancio Emocional (ECE) para estudiantes universitarios: Propiedades psicométricas y asociación. *Interpsiquis*. Recuperado de <http://psiqui.com/2-8938>
- Ríos-Risquez, M. I., García-Izquierdo, M., Sabuco-Tebar, E. A., Carrillo-García, C., & Martínez-Roche, M. E. (2016). An exploratory study of the relationship between resilience, academic burnout and psychological health in nursing students. *Contemporary Nurse*, 52(4), 430-439. doi: 10.1080/10376178.2016.1213648
- Rodriguez, M., & Ruiz, M. (2008). Atenuación de la asimetría y de la curtosis de las puntuaciones observadas mediante transformaciones de variables: Incidencia sobre la estructura factorial. *Psicológica*, 29(2), 205-227.
- Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating bifactor models: calculating and interpreting statistical indices. *Psychological Methods*, 21(2), 137-150. doi: 10.1037/met0000045
- Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399-419). Thousand Oaks, CA: Sage.
- Schaufeli, W. B., Maasen, G. H., Bakker, A. B., & Sixma, H. J. (2011). Stability and change in burnout: a 10-year follow-up study among primary care physicians. *Journal of Occupational and Organizational Psychology*, 84(2), 248-267. doi: 10.1111/j.2044-8325.2010.02013.x
- Schaufeli, W. B., Martínez, I. M., Marques, A., Salanova, M., & Bakker, A. B. (2002). Burnout and engagement in university students. A cross-national study. *Journal of Cross-Cultural Psychology*, 33(5), 464-481. doi: 10.1177/0022022102033005003
- Schmidt, F. L., Le, H., & Ilies, R. (2003). Beyond alpha: An empirical examination of the effects of different sources of measurement error on reliability estimates for measures of individual-differences constructs. *Psychological Methods*, 8(2), 206-224. doi: 10.1037/1082-989X.8.2.206
- Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, 74(1), 107-120. doi: 10.1007/s11336-008-9101-0
- Taris, T. W., Le Blanc, P., Schaufeli, W. B. & Schreurs, P. J. G. (2005). Are there relationships between the dimensions of the Maslach Burnout Inventory? A review and two longitudinal tests. *Work & Stress*, 19(3), 256-262. doi: 10.1080/02678370500270453
- Toppin-Tanner, S., Kalimo, R., & Mutanen, P. (2002). The process of burnout in white-collar and blue-collar jobs. Eight year prospective study of exhaustion. *Journal of Organizational Behavior*, 23(5), 255-270. doi: 10.1002/job.155
- Tuithof, M., ten Have, M., Beekman, A., van Dorsselaer, S., Kleinjan, M., Schaufeli, W., & de Graaf, R. (2017). The interplay between emotional exhaustion, common mental disorders, functioning and health care use in the working population. *Journal of Psychosomatic Research*, 100, 8-14. doi: 10.1016/j.jpsychores.2017.06.018.

- Vizoso, C. M., & Gundín, O. A. (2016). Estresores académicos percibidos por estudiantes universitarios y su relación con el burnout y el rendimiento académicos. *Anuario de Psicología*, *46*(2), 90-97. doi: 10.1016/j.anpsic.2016.07.006
- Watson, D. (2004). Stability versus change, dependability versus error: Issues in the assessment of personality over time. *Journal of Research in Personality*, *38*(4), 319-350. doi: 10.1016/j.jrp.2004.03.001
- Wolke, D., Waylen, A., Samara, M., Steer, C., Goodman, R., Ford, T., Lamberts, K. (2009). Selective drop-out in longitudinal studies and non-biased prediction of behaviour disorders. *The British Journal of Psychiatry*, *195*(3), 249-256. doi: 10.1192/bjp.bp.108.053751
- 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

recebido em setembro de 2017
aceito em maio de 2018

Sobre o autor

Sergio Dominguez-Lara es Psicólogo y Doctor en Psicología. Docente del Instituto de Investigación de Psicología (USMP, Perú).