

Validación transcultural y funcionamiento diferencial del Maslach Burnout Inventory – General Survey en docentes de tres países latinoamericanos

Cross-Cultural Validation and Differential Functioning of the MBI-GS in Teachers from Three Latin-American Countries

Validação transcultural e funcionamento diferencial do MBI-GS em professores de três países da América Latina

Arturo Juárez García*

Universidad Autónoma del Estado de Morelos

César Merino Soto

Universidad de San Martín de Porres

Manuel Fernández Arata

Universidad de San Martín de Porres

Cintha Anamia Flores Jiménez

Universidad Autónoma del Estado de Morelos

Miriam Caraballo

PDVSA Intevep Venezuela

Carmen Camacho Cristiá

Universidad Veracruzana

Doi: <https://doi.org/10.12804/revistas.urosario.edu.co/apl/a.6621>

Resumen

En América Latina, la investigación del síndrome de *burnout* (SB) es creciente; sin embargo, poco se han explorado las posibles diferencias culturales en sus mecanismos psicológicos y eficiencia de medición. En particular, el cuestionario MBI-GS ha sido reconocido como una medida

satisfactoria del SB para su uso en distintos sectores y grupos ocupacionales, y su validez ha sido probada en múltiples países e idiomas, empero en Latinoamérica se adolece de investigaciones al respecto. El objetivo de este estudio fue realizar una traducción y adaptación cultural del MBI-GS para la población latinoamericana, y

* Dirigir la correspondencia a Arturo Juárez García: Centro de Investigación Transdisciplinar en Psicología, Universidad Autónoma del Estado de Morelos, México. Correo electrónico: arturojuarezg@hotmail.com

Cómo citar este artículo: Juárez García, A., Merino Soto, C., Fernández Arata, M., Flores Jiménez, C. A., Caraballo, M., & Camacho Cristiá, C. (2020). Validación transcultural y funcionamiento diferencial del MBI-GS en docentes de tres países latinoamericanos. *Avances en Psicología Latinoamericana*, 38(1), 135-156. <https://doi.org/10.12804/revistas.urosario.edu.co/apl/a.6621>

analizar su estructura interna, su fiabilidad y su invarianza (DIF y MGFA) entre sexos y muestras de docentes de tres países latinoamericanos: México, Perú y Venezuela. Los resultados evidencian que la versión latina del MBI-GS propuesta presenta propiedades psicométricas satisfactorias, que incluyen la ausencia de funcionamiento diferencial entre sexos y países participantes. Se discuten aspectos semánticos y perspectivas futuras.

Palabras clave: burnout, MBI GS, docentes, Latinoamérica.

Abstract

In Latin-America, burnout syndrome (BS) is a growing issue. However, very little has been explored of its cultural differences in its psychological mechanisms and its measurement efficiency. In particular, the MBI-GS questionnaire has been widely recognized as a satisfactory measure of BS in different sectors and occupational groups, and its validity has been proved in multiple languages and countries, but in Latin-America, research remains scarce. The objective of this study was to carry out a Latin-American translation and cultural adaptation of the MBI-GS, to analyze its internal structure, its reliability, and its measurement invariance (DIF and MGFA) among genders and samples of teachers from three different Latin-American countries: México, Perú, and Venezuela. The results show that the proposed Latin version of the MBI-GS has satisfactory psychometric properties, including an equivalent item and scale functioning among genders and countries' sample participants. Semantic issues and future perspectives are discussed.

Keywords: Burnout, MBI GS, Teachers, Latin-América.

Resumo

Na América Latina a pesquisa da síndrome de *burnout* (SB) é crescente, no entanto, pouco se tem explorado sobre as possíveis diferenças culturais em seus mecanismos psicológicos e eficiência de medição. Em particular, o questionário MBI-GS tem sido reconhecido como uma medida satisfatória do SB para uso em diferentes setores e grupos ocupacionais, e sua validade tem sido provada em múltiplos países e línguas, mas na Latino-América

escasseiam as pesquisas ao respeito. O objetivo deste estudo foi realizar uma tradução e adaptação cultural do MBI-GS para população latino-americana, analisar sua estrutura interna, sua fiabilidade e sua invariância (DIF y MGFA) entre sexos e amostras de professores de três países latino-americanos: o México, o Peru e a Venezuela. Os resultados evidenciam que a versão do MBI-GS proposta apresenta propriedades psicométricas satisfatórias incluindo a ausência de funcionamento diferencial entre sexos e países participantes. Discutem-se aspectos semânticos e perspectivas futuras.

Palavras-chave: burnout, MBI GS, professores, Latino-América

El síndrome de *burnout* (SB) es quizás el fenómeno psicosocial y laboral más estudiado y difundido en el ámbito internacional en los últimos años. Maslach (2009) lo ha definido como un conjunto de experiencias interrelacionadas que incluyen una sintomatología tridimensional: agotamiento como respuesta al estrés (cansancio emocional y pérdida de energía), cambio de actitud negativo caracterizado por indiferencia hacia el trabajo y las personas, y sentimientos negativos sobre las propias competencias, capacidades, deseos y motivaciones para trabajar. De acuerdo con la misma autora, este modelo teórico tridimensional surgió desde la experiencia empírica directa con los trabajadores, aunque reconoce que se trata de una teoría aún en desarrollo (Juárez, 2014). Dicho modelo ha recibido críticas relacionadas con la discriminación conceptual de sus componentes respecto a otros fenómenos psicológicos, la débil justificación de su agrupación general en un solo conjunto de síntomas bajo el concepto de *burnout* y la pobre comprensión o manejo empírico de su uni o multidimensionalidad, entre otros (Shirom, 2009). A pesar de que este modelo conceptual tridimensional sigue siendo el más popular, al realizar una revisión a 35 años de investigación y práctica del SB, Schaufeli, Leiter y Maslach (2009) señalan que el contexto de emergencia económica

de los países hace una diferencia importante en el *burnout*, pero en particular, concluyen que el uso y significado de este difiere enormemente entre países y contextos, por lo que en sentido estricto no existe una concepción única o universal. Por ejemplo, señalan entre otros aspectos, diferencias en el uso del término para denominarlo, la perspectiva y el manejo de su naturaleza uni o multidimensional, la forma de diagnosticarlo o inclusive el hecho de que en algunos países se comprende como un fenómeno que requiere de un tratamiento clínico especializado, cuya gravedad implica un reconocimiento legal y posibilidades de compensación económica (e.g. Suecia y Holanda), mientras que en otros se concibe como una sintomatología más pasajera de impacto moderado o leve. Hasta ahora, la mayoría coincide en que el reto para comprender mejor este fenómeno consiste en incrementar la investigación en diferentes culturas.

Particularmente en América Latina existe un creciente número de artículos, tesis, ponencias y libros sobre el tema. Sin embargo, la cantidad y la calidad metodológica de estas publicaciones parece no ser suficiente para identificar con certeza la magnitud y prevalencia de este problema en la región. Por ejemplo, algunos análisis han encontrado limitaciones relacionadas con una masa de estudios de alcance solo descriptivo, falta de representatividad de las muestras, deficiente control de sesgos y serios problemas en los criterios para la determinación de prevalencias, entre otros (Juárez-García, 2015; Juárez-García, Idrovo-Velandia, Camacho-Ávila, & Placencia-Reyes, 2014). De igual forma, y a sabiendas de las particularidades económicas, políticas y contextuales de los países latinoamericanos en relación a los países altamente industrializados donde el constructo de *burnout* ha sido creado e investigado ampliamente, no se han estudiado las posibles diferencias culturales específicas en su manifestación o sus mecanismos psicológicos y, sobre todo, la validez de sus medidas.

En cuanto a la medición del SB, la escala denominada Maslach Burnout Inventory (MBI) en su

versión para servicios humanos —Human Services Survey -HSS—, es definitivamente la más frecuentemente utilizada en Latinoamérica, mientras que la más reciente versión MBI General Survey —encuesta general— (MBI-GS) (Schaufeli, Leiter, & Maslach, 2009), se ha utilizado escasamente. De acuerdo con sus autores, esta versión aplica a cualquier tipo de ocupación, no únicamente a las relacionadas con los servicios humanos, es más robusta, flexible y breve. Esto la convierte en la natural evolución de una mejor medida de *burnout* de las diferentes versiones disponibles con antelación, por lo que ahora es la más recomendable para cualquier ocupación (Juárez, 2014; Maslach, Jackson, & Leiter, 2010). A pesar de que el MBI-GS no fue creado propiamente en una perspectiva transcultural que asegure que sus expresiones psicoafectivas, actitudinales y comportamentales sean universales en todos los contextos, este ya ha sido ampliamente validado en múltiples continentes, países e idiomas. No obstante, en Latinoamérica existen muy pocos estudios al respecto y frecuentemente usan traducciones literales europeas (de España), que no parecen ser total y semánticamente equivalentes a las culturas latinoamericanas de acuerdo con un estudio donde se utilizaron entrevistas cognitivas (Sáenz, 2014).

Las publicaciones del MBI-GS en bases de datos internacionales indizadas con muestras latinas son prácticamente inexistentes, pero en una revisión en bases de datos latinoamericanas se encontraron algunos estudios contradictorios, unos concluyen un comportamiento virtuoso de la escala y otros señalan limitaciones psicométricas de la misma. La razón de esas inconsistencias no es clara y se observan áreas de oportunidad en estos estudios y sus conclusiones. Por ejemplo, Oramas, González y Vergara (2007) realizaron una adaptación lingüística para Cuba, que se utilizó también para analizar su validez psicométrica en Venezuela (Millán de Lange & D'Aubeterre, 2012) y Colombia (Guevara & Ocampo, 2014). Solo los dos primeros concluyen un satisfactorio desempeño

psicométrico del MBI-GS, pero en los tres estudios se presentó un comportamiento inconsistente, ya que la composición de los factores, ítems problemáticos, varianzas explicadas y coeficientes de confiabilidad no fueron totalmente similares entre estos estudios ni plenamente equivalentes a lo establecido teóricamente en la estructura original del MBI-GS. Dicha adaptación cubana, tiene cambios importantes en el sentido semántico de las palabras en algunos ítems y un orden diferente, pero en el artículo citado no se describe el detalle de la metodología de adaptación cultural realizada. En los estudios de Venezuela y Colombia se asumió la equivalencia de significados de ítems entre países y grupos ocupacionales sin realizar una adaptación independiente. Estas limitaciones pudieron influir en las diferencias de los resultados.

Por su parte, Pando, Aranda y López (2015) utilizaron una traducción española para identificar la validez psicométrica del MBI-GS en ocho países latinoamericanos con resultados que no coinciden con la estructura tridimensional original (obtuvieron solo 2 de 3 factores), con una pobre consistencia interna en la mayoría de los países ($\alpha.60$). Sin embargo, las limitaciones del estudio no permiten una conclusión certera, por ejemplo, no se realizó una traducción con algún estándar de adaptación cultural del instrumento, se empleó un solo análisis factorial en una muestra total que fue demasiado heterogénea en diversos aspectos sociodemográficos tales como el sexo, la ocupación o el país de origen. Además, se asumió indebidamente la inexistencia de diferencias semánticas y de funcionamiento diferencial de los ítems entre países y tampoco se realizó un análisis factorial confirmatorio. Asimismo, este estudio, como muchos otros en la región, usó solamente el coeficiente α (Cronbach, 1951) como estimador de consistencia interna sin analizar el supuesto de *tau-equivalencia* que difícilmente puede cumplirse (Dunn, Baguley, & Brunsten, 2014) y llevó a cabo un análisis factorial exploratorio bajo el procedimiento que incluyó la regla Kaiser, componentes principales y rotación Varimax, que

han sido fuertemente criticados por sus déficits (Dominguez & Merino, 2016; Lloret-Segura et al., 2014). Finalmente, el estudio dominicano de Tomás, de los Santos, Alonso-Andrés y Fernández (2016) tuvo resultados a favor de la validez de la escala con base en una muestra homogénea (solo docentes) y con una metodología estadística más robusta (e.g. análisis factoriales confirmatorios). Sin embargo, ello sigue confirmando la inconsistencia de resultados entre estudios del MBI-GS en la región.

En general, la mayoría de los estudios latinos se caracterizan por análisis estadísticos exploratorios, muestras ocupacional y demográficamente demasiado heterogéneas y no se han llevado a cabo análisis de funcionamiento diferencial de ítems ni se han realizado procedimientos de adaptación rigurosos del MBI-GS que aseguren su equivalencia cultural. En este sentido, y de acuerdo con metodologías sugeridas para la adaptación lingüística y transcultural de instrumentos, además del procedimiento de traducción-re-traducción (*back-translation*), es necesario incluir un paso cualitativo que requiere de un comité experto revisor para asegurar la equivalencia cultural de la escala (Beaton, Bombardier, Guillermin, & Ferraz, 2000; Gjersing, Caplehorn, & Clausen, 2010). Esto coincide con las sugerencias de Hwang, Scherer y Ainina (2003) con respecto a que una metodología eficiente de adaptación y de equivalencia en la medición del MBI-GS, exige un proceso combinado de jueceo acompañado con procedimientos empírico-psicométricos. De esta forma, complementariamente se ha considerado realizar análisis psicométricos basados en la teoría de respuesta al ítem como el de funcionamiento diferencial del ítem (DIF) y análisis de invarianza factorial confirmatoria multi-grupo (MGCFAs) entre muestras de distintos grupos culturales (Teresi, 2001). Dichos análisis permiten analizar la equivalencia semántica, métrica y estructural de instrumentos entre distintos grupos (e.g. sexo, ocupación, país, cultura) lo que contribuye con mayor contundencia a la evidencia de validez. Aunque típicamente se utiliza uno u otro procedimiento (DIF vs. MGCFAs),

su uso combinado permite una perspectiva más integral, complementaria y confirmatoria de la invarianza de un instrumento. Para este estudio, el uso de ambos procedimientos se vislumbró como una aproximación más robusta para confirmar qué ítems, estructuras, cargas factoriales y otras propiedades psicométricas del MBI-GS son invariantes entre grupos, para cumplir así con un requisito metodológico fundamental que evita el sesgo en la comparación entre grupos y problemas éticos debido a conclusiones erróneas.

En cuanto a la diferencia de grupos, se ha señalado que las condiciones sociales, psicológicas y biológicas alrededor del sexo pueden contribuir a la experiencia diferenciada del *burnout* (Purvanova & Muros, 2010), por lo que es necesario confirmar la invarianza del MBI-GS entre sexos más allá de las diferencias de medias que tradicionalmente se reportan también en población latina. Desde el punto de vista cultural, dicha equivalencia es importante analizarla en términos de la comparación entre países latinos, ya que, a pesar de tener idiomas y costumbres similares, sus culturas y usos del lenguaje no son idénticos. Hasta ahora, en Latinoamérica son inexistentes estudios de análisis de funcionamiento diferencial de ítems o del comportamiento estructural del MBI-GS entre grupos diferenciados por sexo, ocupación o país, que pueden tener un rol relevante en una escala destinada a medir una respuesta de estrés crónico.

El funcionamiento psicométrico del MBI-GS en países desarrollados ha sido frecuentemente consistente no solo en su estructura factorial (Bakker, Demerouti, & Schaufeli, 2002; Poghosyan, Aiken, & Sloane, 2009; Schutte Toppinen, Kalimo, & Schaufeli, 2000), sino que también se ha demostrado que no presenta funcionamiento diferencial del ítem y es invariante en su estructura factorial en diferentes grupos o categorías tales como el rol profesional, la ocupación, el sexo, la edad, la antigüedad, el país de residencia —europeos— o el simple paso del tiempo —4 meses— (Bria, Spânu, Băban, & Dumitrașcu, 2014; Foster, 2015; Langballe, Falkum, Innstrand

& Aasland, 2006; Mäkikangas, Hättinen, Kinnunen, & Pekkonen, 2011; Xanthopoulou, Bakker, Kantas, & Demerouti, 2012).

Debido a las limitaciones metodológicas señaladas en los estudios latinos que han utilizado el MBI-GS, se desconoce si las inconsistencias reportadas en el funcionamiento psicométrico de la escala se deben a diferencias o problemas en las traducciones utilizadas, adaptaciones lingüísticas deficientes o incluso a verdaderas cuestiones culturales que restan universalidad a su validez o la teoría trifactorial del SB que la sustenta. Asimismo, no debiera asumirse homogeneidad semántica y cultural entre países ibero o latinoamericanos, pues la diversidad de costumbres y uso del lenguaje son muy variados y la comprensión de palabras o ítems en cuestionarios puede ser disímil, lo que puede afectar la validez psicométrica de cualquier instrumento. Dado lo anterior, el objetivo de este estudio es realizar una traducción y adaptación cultural del MBI-GS para población latinoamericana, analizar su estructura interna, su fiabilidad y su invarianza entre sexos y culturas latinas. Esto último mediante el análisis multi-grupo de invarianza de la estructura factorial (MGCF) y el funcionamiento diferencial de ítems (DIF) en tres muestras de países latinoamericanos: México, Perú y Venezuela.

Método y materiales

Participantes

Participaron de manera voluntaria 806 maestros de nivel básico, medio y superior, de distintos países latinoamericanos (México $n=324$, Perú $n=346$, Venezuela $n=135$), quienes fueron seleccionados con base en un muestreo no aleatorio, guiado por participantes y efectuado mediante gestiones entre diversos centros educativos públicos, carteles públicos e invitación verbal. Todos los participantes firmaron un consentimiento informado y se siguieron plenamente los principios éticos estipulados

en la declaración de Helsinki para investigación con seres humanos (Asociación Médica Mundial, 2017). A sabiendas de que el MBI-GS puede ser utilizado en cualquier ocupación, se eligió una muestra ocupacional de docentes porque la homogeneidad en la actividad laboral realizada permite analizar el funcionamiento diferencial entre culturas sin la influencia de la ocupación. El estudio del SB en maestros es importante porque pese a su desempeño en labores y tareas propias del ámbito pedagógico y su mayor vulnerabilidad a los síntomas del *burnout* (Fernet, Guay, Senécal, & Austin, 2012), las condiciones laborales y el ambiente de trabajo pueden variar de acuerdo a las políticas educativas de los países y del tipo de gestión de la escuela (pública o privada), lo que podría incrementar los niveles de estrés y derivar en dichos síntomas, independientemente del contexto que lo antecede. De esta forma,

la tasa de respuesta de los maestros participantes voluntarios que originalmente fueron invitados superó el 80% en todos los casos. Se procuró respetar al menos una relación ítem/participante de 5:1 como han sugerido algunos autores (de Winter, Dodou & Wieringa, 2009). El perfil sociodemográfico fue similar a la población de profesores en cada región y sin diferencias entre países respecto al sexo ($\chi^2 = 3.95$, $gl = 2$, $p > .10$, $V_{Cramer} = .07$), y leve diferencia en el estado civil ($\chi^2 = 76.49$, $gl = 8$, $p < .01$, $V_{Cramer} = .21$) y la participación en trabajos adicionales ($\chi^2 = 46.32$, $gl = 2$, $p < .01$, $V_{Cramer} = .24$). Diferencias moderadamente grandes fueron halladas en el tipo de contratación (tiempo parcial vs. tiempo completo), ($\chi^2 = 279.12$, $gl = 2$, $p < .01$, $V_{Cramer} = .591$). Fuertes discrepancias distribucionales se detectaron en el nivel de enseñanza, ($\chi^2 = 649.92$, $gl = 6$, $p < .01$, $V_{Cramer} = .63$).

Tabla 1
Distribución demográfica de los participantes

	Venezuela		Perú		México		Total	
	n	%	n	%	n	%	n	%
Sexo								
Masculino	50	37.3	123	35.5	139	42.9	312	38.8
Femenino	84	62.7	223	64.5	185	57.1	492	61.2
Estado civil								
Soltero(a)	49	36.8	92	27.0	193	59.8	334	41.9
Casado (a)	71	53.4	204	59.8	106	32.8	381	47.8
Unión Libre	7	5.3	19	5.6	9	2.8	35	4.4
Divorciado(a)	5	3.8	19	5.6	12	3.7	36	4.5
Viudo (a)	1	.8	7	2.1	3	.9	11	1.4
Tipo de contratación								
Completo	67	52.8	288	83.0	60	18.5	415	52.0
Parcial	60	47.2	59	17.0	264	81.5	383	48.0
Otros trabajos actuales								
Sí	73	56.2	84	24.3	90	28.5	247	31.2
No	57	43.8	261	75.7	226	71.5	544	68.8
Nivel de enseñanza								
Preescolar	0	0.0	0	0.0	74	22.8	74	9.2
Primaria	0	0.0	146	42.6	69	21.3	215	26.8
Secundaria	135	100.0	197	57.4	16	4.9	348	43.4
Medio superior	0	0.0	0	0.0	0	0.0	0	0.0
Superior	0	0.0	0	0.0	165	50.9	165	20.6

Instrumento y metodología de adaptación transcultural

Se adaptó y tradujo al español el MBI-GS© original en inglés bajo la autorización de sus autores y la editorial Mind Garden (Maslach, Jackson, & Leiter, 2010). Dicha versión consta de 16 ítems organizados en tres dimensiones de la siguiente manera: agotamiento emocional (AE) —cinco ítems—, indiferencia (IN) —cinco ítems— y eficacia profesional (EP) —seis ítems—. Estos se valoran en una escala tipo Likert que refleja la frecuencia con la que experimentan la situación que se describe en el ítem y tiene siete grados que oscilan desde 0 (Nunca) hasta 6 (Todos los días).

El proceso de traducción se realizó de acuerdo con las recomendaciones internacionales para la adaptación transcultural de autoinformes (Beaton et al., 2000; Gjersing et al., 2010): primero, y de acuerdo al procedimiento de traducción-re-traducción (*back-translation*), se realizó una traducción del inglés al español y luego la re-traducción del español al inglés por separado, por dos especialistas (con lengua materna de acuerdo a la etapa de traducción); se revisaron discrepancias con ayuda de un tercer especialista, hasta ajustarla a una sola versión en español final. En un siguiente paso, dicha versión se canalizó a un comité revisor de expertos de 6 diferentes países latinoamericanos para su revisión y ajuste final, quienes fueron elegidos principalmente por su experiencia en la evaluación de *burnout* en trabajadores de diferentes contextos. El comité analizó diferencias y similitudes en la comprensión semántica de cada ítem hasta acordar una versión con palabras entendibles en los diferentes contextos y países de cada juez (Chile, Colombia, México, Costa Rica, Perú y Venezuela). Cabe mencionar que de esta versión se realizaron algunos estudios piloto que ya han sido publicados en muestras mexicanas y peruanas (Fernández-Arata, Juárez-García, & Merino-Soto, 2015; Flores, Merino, Camacho, Juárez, & Placencia, 2015) aunque el presente estudio integra por primera vez la eviden-

cia de todo el proceso de traducción, adaptación y validez psicométrica e invarianza de la escala en las muestras de tres países latinos.

Estrategia de análisis psicométricos

En primer lugar, se evaluaron preliminarmente la calidad de los datos respecto a la aceptabilidad, patrones de sesgo de respuesta y el piso y techo de los ítems. La aceptabilidad fue evaluada cuantitativamente mediante la tasa de respuesta y valores perdidos (Menard, Hinds, Jacobs, Cranston, Wang, DeWalt, & Gross, 2014). Los patrones de sesgo de respuesta se inspeccionaron con la identificación de la respuesta control, debido a que el MBI-GS contiene constructos que sistemáticamente demandan respuestas correlacionadas negativamente, y por lo tanto la elección de la misma opción de respuesta en ítems de constructos linealmente negativos es inesperada (e.g. los ítems de eficacia profesional e indiferencia). Finalmente, el piso y techo se examinaron para describir el potencial de los ítems para diferenciar conductas en determinadas regiones de la distribución de puntajes.

Para examinar el posible funcionamiento diferencial en ítems y estructuras de las escalas se implementaron dos metodologías que examinan la equivalencia de los parámetros de los ítems: el DIF y el MGCF. En la implementación del primero, el marco analítico fue exploratorio, orientado por la situación inédita e intercultural del posible DIF en la población hispana (Walker, 2011). Además, se consideró el puntaje total de cada subescala como un indicador próximo y suficiente del constructo latente, adicional a que este puntaje es rutinariamente aceptado para su interpretación en la práctica profesional y de investigación. Se analizó tanto el DIF uniforme como el no uniforme para verificar si existía algún tipo de diferencia entre grupos con respecto a la probabilidad de respuestas en cada ítem a lo largo de todos los niveles de los constructos medidos. Dado que la implementación del DIF requiere primero identificar la unidimensionalidad

latente de los ítems, lo cual ayuda a disminuir la inflación del error Tipo I (Li, Brooks, & Johanson, 2012), se verificó la estructura interna del MBI-GS en cada país antes de pasar al examen del DIF entre países. Cabe aclarar que, aunque la escala no es unidimensional, el enfoque de análisis del supuesto se refirió a la unidimensionalidad de los ítems respecto al constructo de cada subescala. De esta forma, la dimensionalidad fue probada mediante el análisis factorial confirmatorio, aplicando la prueba χ^2 con corrección Satorra y Bentler (1994; SB- χ^2) para atenuar el efecto de la no normalidad de los ítems; la significancia práctica del ajuste se hizo por medio de índices de ajuste, como el CFI ($\geq .95$) y el RMSEA ($\leq .05$).

El análisis estructural en cada país también requirió internamente alguna verificación de invarianza dentro de cada país evaluado, por lo que se comprobó la equivalencia de los parámetros de los ítems de acuerdo al sexo, como se señaló previamente, por la importancia de esta variable en el síndrome (Purvanova & Muros, 2010). El problema que se tuvo, fue que el tamaño muestral dentro cada país se caracterizó por ser pequeño; esto fue más evidente en el grupo venezolano (n° mujeres = 83, 63.4%; n° hombres = 48, 36.6%). Para hacer frente a ello, se implementó una metodología no paramétrica basada en DIF con tablas de contingencia, la cual está indicada para situaciones de tamaño muestral no significativo (Lai, Teresi, & Gershon, 2005). Se estableció la hipótesis nula de no DIF mediante la prueba $M - \chi^2$ (Mantel, 1963) con $gl = 1$, y la estimación estandarizada de *cummulative common log odds ratio* de Liu-Agresti (Z_{LA-LOR}) (Liu & Agresti, 1996). En ambos estadísticos se eligió el criterio conservador de $p < .01$ ($M - \chi^2 > 6.63$) y $Z_{LA-LOR} > |2.50|$, en los que se trató de controlar el error Tipo I en esta fase de la investigación. Cuando se identificó un ítem con DIF, se continuó con la valoración de su significancia práctica por medio del estimador LA-LOR (Penfield, 2005); se identificaron tres niveles: trivial ($Z_{LA-LOR} < 0.43$), moderado ($Z_{LA-LOR} < 0.63$) y grande ($Z_{LA-LOR} \geq 0.64$).

Para evaluar la invariancia de medición entre las muestras totales de cada país, se implementó un procedimiento de regresión logística para la detección de DIF (Swaminathan, & Rogers, 1990), aplicado a ítems ordinales (Zumbo, 1999), en el cual se asumió que la variable dependiente (ítem) es subyacente continua. Para su implementación, se ajustaron los datos a tres modelos conteniendo como variables independientes el atributo medido (o puntaje, θ), la membresía del sujeto al grupo (G) y el término de interacción entre θ y G ; la variable dependiente es la respuesta al ítem, Z , estandarizada en *logits* y condicional a las variaciones del atributo. El modelo 1 (MRL₁) corresponde a $Z = \beta_o + \beta_1\theta + \beta_2G + \beta_3\theta*G$, cuyo término de interacción ($\theta*G$) representa la presencia de DIF no uniforme; el siguiente modelo (MRL₂), $Z = \beta_o + \beta_1\theta + \beta_2G$, introduce el efecto principal de dependencia al grupo (G) como fuente para evaluar el DIF uniforme. Finalmente, se construyó el modelo de ítem sin DIF (MRL₃), $Z = \beta_o + \beta_1\theta$, que representa la variación de las respuestas al ítem dependiente del atributo medido, es decir, sin DIF. La estrategia de detección de DIF uniforme y no uniforme fue de naturaleza comparativa entre modelos anidados, correspondiente a una secuencia de pasos; el enfoque elegido en el presente estudio fue el de Crane y colaboradores (Crane, Gibbons, Jolley, & Van Belle, 2006; Crane, Gibbons, Narasimhalu, Lai, & Cella, 2007), debido a que es eficiente e integrativo respecto a la significancia práctica y estadística, y está dentro del marco de una secuencia razonable de pruebas que incluyen el término de interacción como el primer componente que debe verificar. En esta estrategia, primero se evaluó el DIF no-uniforme comparando el $-2 \log likelihood$ de los modelos MRL₁ y MRL₂, contra la distribución χ^2 ($gl = 1$) en un nivel $\alpha = 0.05$. Sin embargo, se aplicó la corrección Bonferroni para ajustar este valor α de acuerdo al número de pruebas estadísticas aplicadas (k); eso es, el número de ítems examinados en cada subescala (Crane et al., 2007; Crane, Van Belle, & Larson, 2004). De ese modo, para las subescalas AE, EP e

IN, α se ajustó respectivamente en $.05/5 = .01$, $.05/6 = .008$, y $.05/5 = 0.01$; resultados debajo de estos niveles indicarían el rechazo de la hipótesis nula de no diferencia entre MRL_3 y MRL_2 , lo cual indica que el modelo con el término de interacción ($\theta * G$) mejora la predicción de la respuesta condicional al ítem. Segundo, se verificó la presencia de DIF uniforme; para ello se computó la diferencia entre los coeficientes beta de los modelos MRL_1 (β_θ) y MRL_2 (β_G) (Crane et al., 2007). Un cambio del $\geq 10\%$ está asociado a la significancia estadística en el nivel $\alpha = 0.20$, que generalmente confluye en la identificación de variables moderadoras (Crane et al., 2004). Se hizo un análisis de sensibilidad, cambiando el criterio de detección de DIF, de 10 a 1 o 5 % (Crane et al., 2007). Se aplicó un procedimiento iterativo de purificación del criterio del atributo (French, & Maller, 2007), mediante el cual el impacto del DIF individual no podría valorarse en tal nivel, sino en la puntuación de la escala; de este modo, se evaluó el impacto en el nivel de los puntajes, usándolos en las subescalas con y sin los ítems detectados con DIF (Teresi, Ramírez, Lai, & Silver, 2008), para obtener correlaciones lineales de Pearson (r) y de concordancia (ρ_c), y estimar el error estándar de medición; estas estimaciones sirvieron para compararlos (Teresi, Ramírez, Jones, Choi, & Crane, 2012) y ver su equivalencia.

Como estrategia adicional para confirmar la invarianza del MBI-GS entre los tres países, se aplicó complementariamente MGFA (Byrne, 2008; Byrne, Shavelson, & Muthén, 1989), con una estrategia secuencial desde un modelo sin restricciones (invarianza configuracional) hacia otros consecutivamente más restringidos (invarianza métrica o de las cargas factoriales, e invarianza escalar o de los interceptos, respectivamente), y la diferencia mínima propuesta por Cheung y Rensvold (2002) entre CFI ($\Delta_{CFI} \leq .01$) y RMSEA ($\Delta_{RMSEA} \leq .01$) para aceptar los modelos de invarianza analizados.

Finalmente, para estimar la consistencia interna, además del coeficiente α , se calculó el coeficiente de α ordinal (α_o), el porcentaje de atenuación y el

coeficiente ω como indicadores adicionales y procedimientos que se han sugerido recientemente como alternativas más robustas para determinar la consistencia interna (Dunn, Baguley, & Brunsten, 2014).

Resultados

Análisis preliminar

Los ítems del MBI-GS mostraron excelente aceptabilidad; respectivamente, el 97.4 y el 100 % de los participantes respondieron a todos los ítems. El 2.6 % de sujetos no respondieron entre 1 a 3 ítems del MBI-GS, y estos parecieron distribuirse aleatoriamente. Dado esta baja tasa de valores perdidos y tomando en cuenta el patrón de respuestas de los ítems dentro de la dimensión del ítem imputado, estos fueron reemplazados por la respuesta modal. Por otro lado, la respuesta aquiescente se detectó en tres participantes (de Venezuela) que respondieron a todos los ítems con la opción 0 y un sujeto (de Perú) con la opción 6; estos cuatro sujetos fueron removidos de la base de datos antes de los siguientes análisis.

Análisis descriptivos de los ítems

Las respuestas a los ítems en todos los países alcanzaron el rango completo de respuesta, desde el mínimo (0) hasta máximo valor posible (6). Al observar las respuestas de los sujetos de los países muestreados, en agotamiento emocional (AE) e indiferencia (IN) tienden a concentrarse en los niveles de respuesta 1 (Esporádicamente: Pocas veces al año o menos) y 3 (Regularmente: algunas veces al mes), mientras que en eficacia profesional (EP), alrededor de las opciones 4 (Frecuentemente: una vez por semana) y 5 (Muy frecuentemente: varias veces por semana). También existe aparente similitud en la dispersión de los ítems. La consistencia y acuerdo del patrón de medias de los ítems en el instrumento fueron altamente similares entre los

países ($\rho_{c_{\text{Mex-Pe}}} = .97, p < .01; \rho_{c_{\text{Mex-Ven}}} = .96, p < .01; \rho_{c_{\text{Pe-Ven}}} = .98, p < .01$), lo que indica la inexistencia de diferencias relevantes en la sintomatología de *burnout* entre las muestras de los tres países. Como un punto a destacar, vale la pena señalar que el ítem 13 (...*not to be bothered*) exhibió un comportamiento diferente en todos sus estadísticos univariados del resto de los ítems de su escala (Indiferencia), en las tres muestras estudiadas (tabla 2).

Análisis de invariancia o funcionamiento diferencial del ítem dentro de los países (por sexo)

En la tabla 3 se presentan los resultados del análisis DIF dentro de cada país por sexo. Con los dos métodos aplicados, ninguno de los ítems ex-

hibió DIF uniforme; sin embargo, con la regresión logística ($p = \acute{o} < 0.05$), mostraron DIF no uniforme dos ítems en México (ítem 1 y 4), uno en Venezuela (ítem 2) y otro en Perú (ítem 11). Aplicando el α con ajuste Bonferroni el ítem 4 en México e ítem 2 en Venezuela mantuvieron su estatus de DIF no uniforme. Estos resultados contrastaron con el análisis, usando tablas de contingencia, en que ninguno de los ítems fue detectado con alguna forma de DIF aplicando los criterios elegidos.

Antes de concluir el status de DIF de los ítems 4 (en México) y 2 (en Venezuela), se aplicó el procedimiento de purificación y su impacto en el puntaje de AE. En el grupo de México, se removió el ítem 4 (AE_4) y nuevamente se aplicó RLO y el ajuste Bonferroni; el ítem aun fue detectado como DIF no uniforme ($p_{\text{ALL}} = .003$). En la evaluación de

Tabla 2
Análisis descriptivos de los ítems del invarianza factorial confirmatoria multi-grupo

	México (n = 324)						Perú (n = 346)						Venezuela (n = 132)					
	M	DE	As	Cu	Piso	Techo	M	DE	As	Cu	Piso	Techo	M	DE	As	Cu	Piso	Techo
MBI																		
Agotamiento																		
1	2.42	1.83	0.47	-0.85	66.7	33.3	1.95	1.42	0.89	0.38	82.2	17.8	1.73	1.62	0.79	-0.29	92.1	7.9
2	2.91	1.84	0.31	-1.10	32.3	67.7	2.26	1.49	0.63	-0.09	73.3	26.7	2.30	1.87	0.49	-0.91	72.7	27.3
3	2.10	1.80	0.62	-0.67	82.2	17.8	1.51	1.40	0.90	0.41	95.0	5.0	1.63	1.58	1.02	0.52	88.4	11.6
4	2.05	1.87	0.60	-0.74	82.3	17.7	1.77	1.53	0.82	0.00	91.6	8.4	1.58	1.64	1.02	0.26	91.7	8.3
5	1.63	1.76	0.89	-0.24	91.2	8.8	.85	1.26	1.78	3.03	98.5	1.5	1.05	1.52	1.74	2.48	94.4	5.6
Eficacia profesional																		
6	4.45	1.66	-1.01	0.20	9.3	90.7	4.30	1.71	-0.85	-0.32	9.3	90.7	4.70	1.54	-1.42	1.65	8.9	91.1
7	4.72	1.57	-1.34	1.16	6.8	93.2	4.64	1.55	-1.24	0.77	5.1	94.9	4.86	1.64	-1.68	2.07	8.5	91.5
10	4.72	1.50	-1.36	1.51	7.1	92.9	4.93	1.40	-1.79	2.92	3.9	96.1	5.36	1.20	-2.65	7.62	2.3	97.7
11	5.21	1.15	-2.19	5.95	2.8	97.2	5.19	1.12	-2.21	6.16	2.3	97.7	5.39	1.20	-2.69	7.79	2.2	97.8
12	4.92	1.29	-1.55	2.69	3.6	96.4	4.78	1.25	-1.35	2.07	3.5	96.5	5.33	1.14	-2.19	5.18	1.2	98.8
16	4.80	1.57	-1.48	1.49	6.4	93.6	4.98	1.30	-1.88	3.95	5.1	94.9	5.21	1.38	-1.99	3.32	1.2	98.8
Indiferencia																		
8	1.35	1.79	1.27	0.52	90.8	9.2	.86	1.23	1.83	3.37	98.9	1.1	.73	1.25	2.01	3.80	98.8	1.2
9	1.38	1.69	1.16	0.23	95.3	4.7	.88	1.18	1.69	2.90	99.4	.6	.86	1.25	1.89	3.72	98.6	1.4
14	1.28	1.73	1.20	0.28	94.5	5.5	1.09	1.56	1.46	1.25	96.4	3.6	.64	1.24	2.40	5.53	98.9	1.1
13	3.64	2.07	-.46	-1.10	12.3	25.6	3.90	2.00	-.60	-1.38	7.8	28.6	3.76	2.29	-.48	-1.38	12.9	37.1
15	1.07	1.57	1.48	1.28	96.8	3.2	1.21	1.64	1.44	1.19	94.5	5.5	.83	1.46	1.91	2.91	97.7	2.3

Nota. M: media. DE: desviación estándar. As: asimetría. Cu: curtosis. Piso y Techo: corresponden a las categorías Nunca y Diariamente, de manera respectiva.

su impacto, se correlacionaron el puntaje AE sin el ítem en cuestión, con los puntajes completos de AE, EP e IN. Primero se halló muy alta convergencia entre AE y AE₄ ($r = .98, p < .01, \rho_c = .93, p < .01$); con EP e IN, las correlaciones r fueron esencialmente iguales (respectivamente, AE y AE₄, para EP e IN): $-.25$ y $-.24, .59$ y $.57$. Esta similaridad ocurrió también con el error estándar de medición en AE y AE₄, respectivamente: 2.54 ($\alpha_{Cronbach} = .89, DE = 7.67$) y 2.59 ($\alpha_{Cronbach} = .86, DE = 6.12$); así, con respecto al ítem AE₄, se puede concluir un efecto insustancial del DIF. Por otro lado, respecto a los participantes de Venezuela, luego de remover el ítem 2 en el proceso de purificación, y aplicando el

ajuste Bonferroni, el ítem dejó de ser identificado con DIF ($p_{ALL} = .003$). En conjunto, los resultados expuestos sugieren que se puede aceptar la equivalencia de los ítems del MBI-GS entre maestros varones y mujeres dentro de cada país.

Estructura factorial de la invarianza factorial confirmatoria multi-grupo en cada país

Antes del análisis estructural resultó necesario considerar tres antecedentes en relación al ítem 13. La primera es que se ha documentado como el ítem más problemático en traducciones al español (Gil Monte, 2002; Moreno-Jiménez, Rodríguez-Carva-

Tabla 3

Resultados del DIF (regresión logística ordinal y tablas de contingencia) dentro cada país por sexo.

	DIF Regression logística ordinal						DIF Tabla de contingencia					
	México (n = 324)		Perú (n = 346)		Venezuela (n = 132)		México (n = 324)		Perú (n = 346)		Venezuela (n = 132)	
	no-Unif. P _{ALL}	Unif. ΔEst.	no-Unif. P _{LL}	Unif. ΔEst.	no-Unif. P _{LL}	Unif. ΔEst.	M-χ ²	LA LOR	M-χ ²	LA LOR	M-χ ²	LA LOR
Agotamiento Emocional												
MBI1	.0203*	-.007	.817	-.000	.073	-.002	3.175	-.439	0.046	-.051	0.051	-.100
MBI2	.4737	-.001	.663	.0023	.009**	-.000	0.176	.101	2.450	-.367	0.611	.335
MBI3	.2805	.010	.816	.0056	.397	-.000	0.179	.104	0.974	.251	0.002	-.019
MBI4	.002**	.027	.472	.0001	.879	-.004	3.969	.479	0.232	.114	0.443	-.281
MBI6	.7599	-.002	.696	.0020	.912	-.003	0.653	-.210	0.180	.109	0.001	.013
Eficacia Profesional												
MBI5	.155	.001	.405	-.004	.682	-.001	0.843	-.220	1.916	-.333	0.103	.160
MBI7	.090	-.016	.534	.019	.233	.006	4.648	.570	1.438	.304	0.463	.300
MBI10	.66	.000	.880	.004	.709	-.00	0.521	-.190	0.267	.141	0.001	.021
MBI11	.954	.021	.026*	.000	.626	.046	1.801	-.366	0.011	.029	3.544	-1.221
MBI12	.632	-.004	.403	.008	.428	.001	0.092	.078	0.446	.158	0.418	-.301
MBI16	.435	-.009	.201	.000	.416	.000	0.013	.031	0.684	-.222	0.655	.470
Indiferencia												
MBI8	.534	-.002	.9670	.000	.229	.003	0.272	.143	0.358	-.167	0.038	-.117
MBI9	.708	.005	.5615	.002	.334	-.002	0.454	-.176	1.098	-.273	3.174	-1.111
MBI14	.599	-.001	.333	.000	.671	-.000	1.106	.308	0.961	-.275	0.00	-.006
MBI13	.878	-.033	.497	-.000	.820	-.028	.528	-.185	0.234	.114	1.081	.508
MBI15	.206	-.002	.839	.000	.651	.000	0.003	.018	2.425	0.423	0.102	0.208

Nota. M-χ²: prueba de Mantel (gl = 1). LA LOR: Liu-Agresti Log odd ratio. *DIF detectado en $p < 0.05$. **DIF detectado con ajuste Bonferroni. No-Unif: DIF no uniforme. Unif: DIF uniforme.

jal, & Escobar, 2001; Salanova & Schaufeli, 2000), la segunda, que este comportamiento también se confirmó en los estudios preliminares de la adaptación propuesta (Fernández et al., 2015; Flores et al., 2015) y la tercera, que su tendencia en los estadísticos descriptivos de este estudio se caracterizó por ser bastante diferente al resto (tabla 2). Por lo anterior, primero se verificó su validez en el factor indiferencia antes de incluirlo en el modelamiento final. El ajuste estadístico del modelo MBI-GS, con el ítem 13 incluido, fue en México, $SB-\chi^2$ (gl: 101) = 161.317 ($p < .01$), CFI = .991, RMSEA = .043 (IC 90% = .030, .055), SRMR = .055 y CAIC = -523.225; en Perú, $SB-\chi^2$ (gl: 101) = 264.437 ($p < .01$), CFI = .967, RMSEA = .069 (IC 90% = .059, .079), SRMR = .072 y CAIC = -426.467, y Venezuela, $SB-\chi^2$ (gl: 101) = 58.564 ($p > .10$), CFI = 1.000, RMSEA = .000 (IC 90% = .000, .000), SRMR = .087 y CAIC = -534.831. Comparativamente, este ajuste es solo ligeramente inferior a las estimaciones sin el ítem 13 (tabla 4). No obstante, su baja carga factorial con su constructo (IN) en México ($\lambda = .227$), Perú ($\lambda = .007$) y Venezuela ($\lambda = -.006$), y la baja confiabilidad (α) del puntaje IN con el ítem 13 incluido (respectivamente, .770, .646 y .610), sugieren que es el más débil psicométricamente. El porcentaje de atenuación comparándolo con los coeficientes α finales (tabla 6) fue de 13.4%, 11.5% y 18.6%. Dadas estas evidencias interculturales sobre la falta de validez del ítem 13, este se retiró del análisis factorial dentro de cada país (tabla 4), aunque se decidió probar su funcionamiento diferencial entre países (tabla 5).

Los resultados del análisis estructural en cada país muestran que el ajuste en cada una de las muestras es satisfactorio. En general, este es elevado, y es excepcionalmente óptimo en Venezuela. En este grupo, el modelo muestra ser un inusual resultado de ajuste completo de acuerdo con el CFI y el RMSEA. Aunque el SRMR no estuvo cerca al valor cero, su magnitud no es discrepante con las conclusiones de los índices anteriores. Las cargas factoriales son elevadas ($\lambda > .50$) y estadísticamente significativas

($Z > 6.0$) en cada factor y homogéneas en AE, pero moderadamente inestables en EP e IN. Los datos de México obtuvieron el segundo mejor resultado, pues todos los índices de ajuste fueron satisfactorios; las cargas factoriales fueron estadísticamente significativas ($Z > 12.0$) y elevadas ($\lambda > .50$); la moderada heterogeneidad también se halló en las escalas EP e IN. Finalmente, el grupo de Perú logró también un ajuste satisfactorio, pero fue ligeramente menor comparado con los grupos anteriores. Se observó discrepancia entre el resultado del CFI y RMSEA (este último, $> .05$), lo que condujo a explorar los índices de modificación para alguna mala especificación trivial o significativa. Se detectó potencial error correlacionado entre los ítems 14 y 15, para evaluar si esta mala especificación tiene significancia práctica, se aplicó el método de Saris, Satorra y van der Veld (2009) con un mínimo valor de carga significativa no-trivial de .40 y potencia estadística de .80, se halló que esta nueva especificación tiene alta potencia estadística (.92), es significativa en términos prácticos y estadísticos (IC 95%: .43, .66), y es 2.73 más grande que el criterio (.40). Dado este nuevo parámetro únicamente en el grupo peruano, se añadió este error correlacionado. Los índices de ajuste presentados en la tabla 4 son los resultados finales de este análisis.

Análisis de invarianza entre países

Funcionamiento diferencial de ítems. Al examinar la equivalencia entre países mediante los procedimientos DIF (RLO y de tabla de contingencia), esencialmente se halló que la mayoría de los ítems no presentaron ninguna forma de DIF (tabla 5); todos los resultados estuvieron muy por debajo de los criterios de detección. Aunque algunos ítems (2, 14, 15 y 8) en algunas comparaciones entre países fueron inicialmente detectados con DIF no-uniforme en el nivel .05, el ajuste Bonferroni aplicado a estas comparaciones quitó esta identificación. Solo el caso de los ítems 14 y 15 en la comparación México-Perú, permanecieron con DIF luego del ajuste.

Tabla 4
Parámetros de la estructura factorial del MBI-GS en cada país

	México (n = 324)				Perú (n = 346)				Venezuela (n = 132)			
	AGO	EP	IN	h^2	AGO	EP	IN	h^2	AGO	EP	IN	h^2
MBI1	.76			.58	.71			.50	.79			.63
MBI2	.80			.69	.67			.45	.72			.52
MBI3	.78			.60	.76			.58	.75			.57
MBI4	.86			.70	.76			.58	.83			.68
MBI6	.82			.62	.77			.59	.75			.56
MBI5		.60		.26		0.55		.31		.52		.27
MBI7		.60		.26		0.59		.35		.67		.45
MBI10		.79		.50		0.73		.54		.81		.66
MBI11		.83		.60		0.83		.69		.87		.76
MBI12		.77		.50		0.79		.62		.86		.75
MBI16		.73		.38		0.75		.56		.83		.69
MBI8			.83	.64			0.83	.67			.88	.77
MBI9			.85	.69			0.91	.80			.82	.68
MBI14			.76	.43			0.56	.37			.76	.57
MBI15			.81	.53			0.57	.39			.79	.63
Correlaciones												
AGO												
EP		-.42				-.37				-.51		
IN		.73	-.65			.73	-.62			.75	-.80	
Índices de ajuste												
SB- χ^2 (gl: 87)				137.72*				243.83*				83.68
CFI				.993				.978				1.00
RMSEA				.042				.061				0.00
(IC 90%)				(.02, .05)				(.05, .07)				(.00, .04)
SRMR				.048				.063				.062
CAIC				-452.19				-392.27				-428.11

Nota. AGO: agotamiento emocional. EP: eficacia profesional. IN: indiferencia. * $p < 0.01$.

Análisis factorial confirmatorio de múltiple grupo. Con el procedimiento MGCF, en el primer nivel de invarianza (configuracional), el resultado fue aceptable, SB- χ^2 (gl: 261) = 490.99, CFI = .933, RMSEA = .058 (IC90% = .050, .065). En el siguiente nivel de invarianza (métrica), el resultado también fue aceptable, SB- χ^2 (gl: 286) = 505.07, CFI = .936, RMSEA = .054 (IC90% = .046, .061) y su diferencia con el nivel anterior fue pequeña ($\Delta_{CFI} = -.009$; $\Delta_{RMSEA} = .002$). En el último nivel de invarianza (escalar), los resultados permitieron aceptar la res-

tricción de igualdad de interceptos, SB- χ^2 (gl: 303) = 561.22, CFI = .941, RMSEA = .057 (IC 90% = .049, .064); $\Delta_{CFI} = .005$; $\Delta_{RMSEA} = .003$). En conjunto, los resultados sugieren que la invarianza configuracional, métrica y escalar del MGI-GS se mantiene entre las tres muestras de países latinoamericanos. Asimismo, los ítems detectados con posible funcionamiento diferencial en el análisis anterior (14 y 15 entre Perú y México), no se replicaron en este análisis, y alguna diferencia psicométrica en ellos no mostró impacto sobre los índices de ajuste.

Tabla 5
DIF entre países

	Perú (n = 346)		Venezuela (n = 142)	
	no-Unif. P(Δ_{LL})	Unif Δ Est.	no-Unif. P(Δ_{LL})	Unif Δ Est.
México (n = 324)				
AGO				
MBI1	.654	.009	.140	-.005
MBI2	.693	-.003	.021*	-.004
MBI3	.163	.000	.639	.009
MBI4	.698	.04	.984	.009
MBI6	.421	-.012	.803	-.004
EP				
MBI5	.686	.003	.323	.014
MBI7	.984	.001	.078	.018
MBI10	.323	.012	.271	-.004
MBI11	.644	.000	.230	-.000
MBI12	.787	.002	.270	-.007
MBI16	.193	.003	.62	-.009
IN				
MBI8	.006*	-.003	.551	-.003
MBI9	.218	-.001	.918	.002
MBI14	.006**	.013	.609	-.006
MBI15	.000**	.066	.021*	.022
MBI13	.372	.037	.707	.070
Perú				
AGO				
MBI1	-	-	.146	.001
MBI2	-	-	.01*	-.000
MBI3	-	-	.439	.005
MBI4	-	-	.80	.006
MBI6	-	-	.368	-.001
EP				
MBI5	-	-	.243	-.000
MBI7	-	-	.055	.005
MBI10	-	-	.708	-.011
MBI11	-	-	.372	-.004
MBI12	-	-	.195	-.005
MBI16	-	-	.187	-.006
IN				
MBI8	-	-	.011*	.000
MBI9	-	-	.454	.015
MBI14	-	-	.212	-.001
MBI15	-	-	.858	-.004
MBI13	-	-	.853	.010

Nota. AGO: agotamiento emocional. EP: eficacia profesional. IN: indiferencia. *: DIF detectado en el nivel $p < .05$. **: DIF detectado luego del ajuste Bonferroni.

Confiabilidad por consistencia interna

Los coeficientes calculados fueron esencialmente aceptables y se observaron patrones relevantes teóricamente; por ejemplo, los coeficientes α fueron comparativamente bajos, mientras que los coeficientes ω fueron elevados y más estables entre sí. El alfa ordinal (α_o) produjo atenuación diferentemente en cada país en el puntaje IN, mientras que fue similar en AGO y EP. En cada país, la diferencia entre las estimaciones de α y ω —diferencia significativa: $\Delta \geq |.06|$ — (Gignac, Bates & Jang, 2007) fueron consistentemente pequeñas en el puntaje de AGO ($\Delta \leq .01$) y grandes en el de EP ($\Delta \geq .07$); en el puntaje IN, las diferencias notorias fueron en Perú y Venezuela ($\Delta \geq .10$).

Discusión

El objetivo de este estudio fue realizar una traducción y adaptación cultural del MBI-GS para población latinoamericana, analizar su estructura interna, su fiabilidad y su invarianza entre sexos en muestras de docentes de tres países latinoamericanos: México, Perú y Venezuela. En cuanto a la adaptación obtenida, la mayoría de los ítems fueron similares a las versiones españolas (e.g. Gil Monte, 2002; Moreno-Jiménez et al, 2001; Salanova, & Schaufeli, 2000); sin embargo, el 25% de los ítems presentaron diferencias semánticas y de traducción trascendentales, en particular en los

ítems 2 (*used up*), 6 (*burned out*), 11 (*exhilarated*) y 14 (*become more cynical*), los que parecen tener significados diferentes en poblaciones latinas. Por ejemplo, mientras que en España la traducción de *burned out* (ítem 6) puede ser comprendida y traducida de forma más literal por los españoles (quemado = *burned*), en Latinoamérica tiene un significado ligado principalmente a una quemadura por el fuego, y en lugares como México, puede hacer alusión a alguien con “mala fama”. Por ello parece necesario recurrir a conceptos como “consumido” (Juárez, Toledo, Camacho, & Arias, 2009) o “acabado” = *worn out* (en el presente estudio), los que cobran más sentido idiosincrático en el contexto de los síntomas del SB. Esto se confirma con el desempeño positivo obtenido en nuestra adaptación de este ítem al compararlo con lo encontrado en otros estudios latinos donde se utilizó la palabra “quemado” o “hastiado”, que tuvo problemas psicométricos en la mayoría de los casos. Similarmente, los demás ítems adaptados que mostraron distinciones semánticas a las versiones españolas (2, 11 y 14) no fueron problemáticos psicométricamente en este estudio, lo que sí sucedió en otros estudios latinos. Asimismo, la palabra *cynicism* ha tenido una traducción española literal como “cinismo”, aun cuando esta última tiene un significado en el diccionario de la real académica española (2018) como “desvergüenza” (*effrontery*) o “insolencia” (*impudence*). El sentido original en el MBI-GS tiene una connotación más hacia el escepticismo, la incredulidad, duda o desconfianza hacia otros o al

Tabla 6
Estimaciones de confiabilidad de consistencia interna

	México			Perú			Venezuela		
	AGO	EP	IN ^a	AGO	EP	IN ^a	AGO	EP	IN ^a
α	.89	.79	.89	.85	.78	.73	.87	.77	.75
α_o	.96	.95	.97	.94	.94	.93	.95	.95	.96
$\%_{at}$	8.3	20.9	9.1	11.4	21.3	28.6	9.85	23.1	27.0
ω	.90	.86	.88	.85	.86	.83	.87	.89	.88

Nota. α : coeficiente alfa. α_o : coeficiente alfa ordinal. $\%_{at}$: porcentaje de atenuación. ω : coeficiente omega. ^aEstimaciones realizadas sin el ítem 13. AGO: agotamiento emocional. EP: eficacia profesional. IN: indiferencia.

trabajo mismo, sin llegar a la insolencia o falta de respeto. Por ello, hemos considerado que la traducción latina de *cynicism* debiera ser “indiferencia”.

En cuanto a la estructura trifactorial del MBI-GS y de consistencia interna, en general se obtuvieron resultados satisfactorios: las saturaciones factoriales estuvieron arriba de .50 con la configuración tridimensional esperada, y los valores α , α ordinal y ω estuvieron por arriba de .70 en la muestra de los tres países lo que indica la replicabilidad del modelo de *burnout* medido por nuestra propuesta de traducción y adaptación del MBI-GS en las muestras de los tres países participantes. Implícitamente se confirma la necesidad de estos procedimientos metodológicos de adaptación semántica en América Latina para el empleo de escalas o test desarrollados en otros contextos.

Con respecto a la equivalencia psicométrica del contenido del MBI-GS dentro y entre los tres países, en general el presente estudio arrojó resultados favorables. Esto fue verificado por dos metodologías, los análisis DIF y MGCFa. En el MGCFa, el ajuste de los modelos con restricciones en la igualdad de cargas factoriales e interceptos fueron aceptables, pero no muy altos (e.g., CFA > .95, RMSEA > .05). Estas variaciones irrelevantes al constructo pueden provenir de múltiples fuentes independientes o en interacción (Byrne, Shavelson, & Muthén, 1989), e iniciarse potencialmente en el proceso de recolección de datos de cada país; esto puede ser particularmente más influyente en uno de los grupos cuyo tamaño muestral fue menor a 200. Sin embargo, nuestros resultados podrían ser razonablemente satisfactorios respecto a la invarianza de los parámetros analizados, más aún cuando otras características ocurrieron en los datos. Por ejemplo, la moderada similaridad de las cargas factoriales en los datos tiende a ser un protector contra el error Tipo I (Li, Brooks, & Johanson, 2012), especialmente entre cargas entre .60 y .90. En segundo lugar, los índices de ajuste alcanzaron un modesto ajuste —cerca, pero debajo de .95— (Bentler, & Wu, 2012; Tabachnick, & Fidell, 2007),

pero pueden representar resultados robustos dado su aceptable desempeño en muestras pequeñas —por ejemplo, el CFI— (Hu, & Bentler, 1999).

Con respecto a los resultados de la metodología DIF, su impacto sobre los puntajes del instrumento no fue grande, y no es una sorpresa que el impacto del DIF individual sea trivial. Efectivamente, el tamaño del impacto ha sido puesto en relevancia en otros estudios en que el número de ítems con DIF fue mayor a lo hallado en el presente estudio, y sus consecuencias sobre otros parámetros importantes para la interpretación del constructo, como las diferencias entre grupos o correlaciones, no tuvo importancia práctica, aun en instrumentos cognitivos y no cognitivos (Meiring, Van de Vijver, Rothmann, & Barrick, 2005).

Una discusión aparte merece el ítem 13, ya que, aunque pocas veces ha mostrado problemas en estudios anglosajones, en todas las traducciones españolas (y por ende en los estudios latinos que las utilizaron) ha sido el más deficiente desde el punto de vista psicométrico (Gil Monte, 2002; Moreno et al., 2001; Salanova, & Schaufeli, 2000). Ello también sucedió en nuestra adaptación latina al presentar el rendimiento psicométrico más endeble en términos de sus cargas factoriales e impacto negativo en la fiabilidad de la escala. Uno de los hallazgos más interesantes de nuestro estudio es que los análisis de DIF e invarianza no presentaron anomalías en este ítem, confirmando así el riesgo de que en todos los contextos (e.g. entre sexos o entre países) se está midiendo “un algo disperso” que tampoco es lo mismo que se pretende medir en su acepción original, problema que no parece relacionado con su traducción, por lo que se sugiere valorar cuidadosamente su inclusión en futuros estudios y profundizar en su pertinencia teórica.

Las estimaciones de confiabilidad, en general, mostraron que los puntajes alcanzan aceptables niveles de consistencia interna para usarlos en investigación científica aplicada, así como en la descripción de grupos. Para la identificación de síntomas de *burnout*, en un contexto individual, usualmente

se tiende a requerir confiabilidades mayores (por ejemplo, $\geq .90$), y aunque los puntajes del MBI-GS no son exactamente iguales a este criterio, nuestras estimaciones están próximas y pueden no diferenciarse estadísticamente. Por lo tanto, podrían ser utilizadas para identificar los síntomas de *burnout* si se combina con criterios de selección más altos, para reducir el error Tipo II (identificar falsamente a sujetos sin *burnout*) debido al error de medición. En el presente estudio, el coeficiente ω es estable entre las subescalas y los países y representa una estimación más apropiada (Dunn et al., 2014).

Con todo lo anterior, podemos concluir que nuestra versión traducida y adaptada para población latinoamericana del MBI-GS presenta propiedades psicométricas satisfactorias e invarianza de medición en las 3 muestras de profesores latinos participantes, además de su invarianza entre sexos, como ha ocurrido en el ámbito internacional (Poghosyan et al., 2009), por lo que su uso parece eficaz para la medición del SB en estos contextos.

Limitaciones y perspectivas futuras

En el presente estudio, no se incluyó un criterio de validación convergente, lo que podría confirmar, en mayor medida, la evidencia empírica de la validez del MBI-GS. Sin embargo, la contribución psicométrica de este estudio representa la construcción de cimientos para estudios futuros con este objeto, pues puede ser no fructífero probar la validez convergente de un instrumento con estructura o fiabilidad inestable. Debe reconocerse como limitación adicional un aspecto probablemente problemático relacionado a la aplicación de la metodología DIF en pequeñas muestras y también en instrumentos con pequeño número de elementos; por ejemplo, entre 100 a 200 por grupo comparado y escalas con menos de 10 ítems respectivamente. Empero, algunos estudios han sugerido que estos no son necesariamente problemáticos, especialmente si la muestra no es exageradamente pequeña (Lai et al.,

2005); asimismo, algunos procedimientos convergen bien y con resultados esperados con medidas de pocos ítems (Scott et al., 2009). En todo caso, la posibilidad de inflación del error Tipo I al incluir un grupo de tamaño muestral pequeño (Venezuela) puede ser riesgoso para obtener espuriamente ítems con DIF (Gómez-Benito, Hidalgo, & Zumbo, 2013).

Ineludiblemente el tamaño de muestra y su heterogeneidad (en particular las diferencias en el nivel de enseñanza), así como la falta de representatividad poblacional del presente trabajo, pueden ser limitantes que deberán enfrentarse en posteriores estudios. Finalmente, es necesario señalar la importancia de investigaciones futuras que confirmen la validez semántica de la adaptación del MBI-GS aquí propuesta, con entrevistas cognitivas y otras estrategias cualitativas, así como la exploración de posibles mecanismos psicológicos culturalmente diferenciados alrededor del SB para confirmar su estructura tridimensional y la consolidación de su entidad como algo que se discrimina eficazmente de otras alteraciones de la salud mental.

Referencias

- Asociación Médica Mundial (2017). *Declaración de Helsinki de la AM M-Principios éticos para las investigaciones médicas en seres humanos*. Recuperado de www.wma.net/es/polices-post/declaracion-de-helsinki-de-la-amm-principios-eticos-para-las-investigaciones-medic-as-en-seres-humanos
- Bakker, A. B., Demerouti, E., & Schaufeli, W. B. (2002). Validation of the Maslach burnout inventory-general survey: An internet study. *Anxiety, Stress & Coping*, 15(3), 245-260. <https://doi.org/10.1080/1061580021000020716>
- Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine*, 25(24), 3186-3191. <https://doi.org/10.1097/00007632-200012150-00014>

- Bentler, P. M., & Wu, E. J. C. (2012). *EQS 6.2 for windows* [Statistical Program]. Encino, CA: Multivariate Software, Inc.
- Bria, M., Spânu, F., Băban, A., & Dumitrașcu, D. L. (2014). Maslach burnout inventory—general survey: Factorial validity and invariance among Romanian healthcare professionals. *Burnout Research, 1*(3), 103-111. <https://doi.org/10.1016/j.burn.2014.09.001>
- Byrne, B. M. (2008). Testing for multigroup equivalence of a measuring instrument: A walk through the process. *Psicothema, 20*(4), 872-882. <https://doi.org/10.1037/e508732013-001>
- Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement equivalence. *Psychological Bulletin, 105*, 456-466. <https://doi.org/10.1037//0033-2909.105.3.456>
- Cheung, G.W., & Rensvold, R.B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*, 233-255. https://doi.org/10.1207/s15328007sem0902_5
- Crane, P. K., Gibbons, L. E., Jolley, L., & Van Belle, G. (2006). Differential item functioning analysis with ordinal logistic regression techniques. *Medical Care, 44*(11-3), S115-S213.
- Crane, P. K., Gibbons, L. E., Narasimhalu, K., Lai, J. S., & Cella, D. (2007). Rapid detection of differential item functioning in assessments of health-related quality of life: The functional assessment of cancer therapy. *Quality of Life Research, 16*, 101-114. <https://doi.org/10.1007/s11136-006-0035-7>
- Crane, P. K., Van Belle, G., & Larson, E. B. (2004). Test bias in a cognitive test: Differential item functioning in the CASI. *Statistics in Medicine, 23*(2), 241-256. <https://doi.org/10.1002/sim.1713>
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297-334. <https://doi.org/10.1007/bf02310555>
- Dominguez, S., & Merino, C. (2016). Sobre o uso do Little Jiffy na validação dos testes: comentários a Ávila e colaboradores. *Jornal Brasileiro de Psiquiatria, 65*(2), 196-197. <https://doi.org/10.1590/0047-2085000000123>
- Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology, 105*(3), 399-412. <https://doi.org/10.1111/bjop.12046>
- Fernández-Arata, M., Juárez-García, A., & Merino-Soto, C. (2015). Análisis estructural e invarianza de medición del MBI-GS en trabajadores peruanos. *Liberabit, 21*(1), 9-20. <http://www.scielo.org.pe/pdf/liber/v21n1/a02v21n1.pdf>
- Fernet, C., Guay, F., Senécal, C., & Austin, S. (2012). Predicting intraindividual changes in teacher burnout: The role of perceived school environment and motivational factors. *Teacher and teacher education, 28*(4), 514-525. <https://doi.org/10.1016/j.tate.2011.11.013>
- Flores, C., Merino, C., Camacho, A., Juárez, A., & Placencia, O. (2015). Síndrome de *burnout* en instructores comunitarios: propiedades psicométricas del Maslach Burnout Inventory General Survey (MBI-GS). En A. Juárez-García (Coord.), *Investigaciones psicométricas de escalas psicosociales en trabajadores mexicanos* (pp. 279-308). Cuernavaca: UAEM / Plaza y Valdés Editores.
- Foster, G. C. (2015). *Measurement invariance of burnout inventories across sex* (Tesis doctoral, Bowling Green State University, Ohio).
- French, B. F., & Maller, S. J. (2007). Iterative purification and effect size use with logistic regression for differential item functioning detection. *Educational and Psychological Measurement, 67*, 373-393. <https://doi.org/10.1177/0013164406294781>
- Gil-Monte, P. R. (2002). Validez factorial de la adaptación al español del Maslach Burnout Inventory-General Survey. *Salud pública de México, 44*(1), 33-40. <https://doi.org/10.1590/s0036-36342002000100005>

- Gjersing, L., Caplehorn, J. R., & Clausen, T. (2010). Cross-cultural adaptation of research instruments: Language, setting, time and statistical considerations. *BMC Medical Research Methodology*, *10*(1), 13. <https://doi.org/10.1186/1471-2288-10-13>
- Gignac, G. E., Bates, T. C., & Jang, K. (2007). Implications relevant to CFA model misfit, reliability, and the Five Factor Model as measured by the NEO-FFI. *Personality and Individual Differences*, *43*, 1051-1062. <https://doi.org/10.1016/j.paid.2007.02.024>
- Gómez-Benito, J., Hidalgo, M. D., & Zumbo, B. D. (2013). Effectiveness of combining statistical tests and effect sizes when using logistic discriminant function regression to detect differential item functioning for polytomous items. *Educational and Psychological Measurement*, *73*(5), 875-897. <https://doi.org/10.1177/0013164413492419>
- Guevara, B. L., & Ocampo, A. N. (2014). Propiedades psicométricas de confiabilidad y validez del Maslach Burnout Inventory-General Survey. *Revista Interamericana de Psicología Ocupacional*, *33*(2), 128-142. <https://doi.org/10.21772/ripo.v33n2a04>
- Hu, L. T., & Bentler, P. M. (1999). Cut-off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1-55. <https://doi.org/10.1080/10705519909540118>
- Hwang, C. E., Scherer, R. F., & Ainina, M. F. (2003). Utilizing the Maslach Burnout Inventory in cross-cultural research. *International Journal of Management*, *20*(1), 3-10. <https://www.questia.com/library/journal/1P3-342900761/utilizing-the-maslach-burnout-inventory-in-cross-cultural>
- Juárez, G. A. (2014). Entrevista con Christina Maslach: reflexiones sobre el síndrome de *burnout*. *Librerabit*, *20*(2), 199-208. Recuperado de http://www.scielo.org.pe/scielo.php?script=sci_arttext&pid=S1729-48272014000200001&lng=es&tlng=es
- Juárez-García, A. (2015). *Síndrome de Burnout: Un análisis del estado del arte en 5 países latinoamericanos*. Memorias del XXV Congreso Interamericano de Psicología de la Sociedad Interamericana de Psicología, Lima, Perú.
- Juárez-García A, Toledo B, Camacho A, & Arias F. (2009). Validez factorial del MBI versión General Survey en población mexicana. *Revista Enseñanza e Investigación en Psicología*. Número especial.
- Juárez-García, A., Idrovo-Velandia, Á. J., Camacho-Ávila, A., & Placencia-Reyes, O. (2014). Síndrome de *burnout* en población mexicana: una revisión sistemática. *Salud Mental*, *37*(2), 159-176. <https://doi.org/10.17711/sm.0185-3325.2014.020>
- Lai, J. S., Teresi, J., & Gershon, R. (2005). Procedures for the analysis of differential item functioning (DIF) for small sample sizes. *Evaluation & Health Professions*, *28*(3), 283-294. <https://doi.org/10.1177/0163278705278276>
- Langballe, E. M., Falkum, E., Innstrand, S. T., & Aasland, O. G. (2006). The factorial validity of the Maslach Burnout Inventory-General Survey in representative samples of eight different occupational groups. *Journal of Career Assessment*, *14*(3), 370-384. <https://doi.org/10.1177/1069072706286497>
- Li, Y., Brooks, G. P., & Johanson, G. A. (2012). Item discrimination and Type I Error in the detection of differential item functioning. *Educational and Psychological Measurement*, *72*(5) 847-861. <https://doi.org/10.1177/0013164411432333>
- Liu, I., & Agresti, A. (1996). Mantel-Haenszel type inference for cumulative odds ratios with a stratified ordinal response. *Biometrics*, *52*(4), 1223-1234. <https://doi.org/10.2307/2532838>
- Lloret-Segura, S., Ferreres-Traver, A., Hernández-Baeza, A., & Tomás-Marco, I. (2014). El análisis factorial exploratorio de los ítems: una guía práctica, revisada y actualizada. *Anales de Psicología*. *30*(3), 1151-1169. <https://doi.org/10.6018/analesps.30.3.199361>

- Mäkikangas, A., Hättinen, M., Kinnunen, U., & Pekkonen, M. (2011). Longitudinal factorial invariance of the Maslach Burnout Inventory-General Survey among employees with job-related psychological health problems. *Stress and Health, 27*(4), 347-352. <https://doi.org/10.1002/smi.1381>
- Mantel N. (1963) Chi-square tests with one degree of freedom: Extensions of the Mantel-Haenszel procedure. *Journal of American Statistic Association, 58*, 690-700. <https://doi.org/10.1080/01621459.1963.10500879>
- Maslach, C. (2009). Comprendiendo el *burnout*. *Ciencia & Trabajo, 11*(32), 37-43.
- Maslach, C., Jackson, S. E., & Leiter, M. P. (2010). *Maslach Burnout Inventory: Manual and non-reproducible instrument and scoring guides*. California: Mind Garden.
- Meiring, D., Van De Vijver, A. J. R., Rothmann, S., & Barrick, M. R. (2005). Construct, item, and method bias of cognitive and personality tests in South Africa. *SA Journal of Industrial Psychology, 31*(1), a182. <https://doi.org/10.4102/sajip.v31i1.182>
- Menard, J. C., Hinds, P. S., Jacobs, S. S., Cranstont, K., Wang, J., DeWalt, D. A., & Gross, H. E. (2014). Feasibility and acceptability of the PROMIS measures in children and adolescents in active cancer treatment and survivorship. *Cancer Nursing, 37*(1), 66-74. <https://doi.org/10.1097/NCC.0b013e3182a0e23d>
- Millán de Lange, A. C., & D'Aubeterre, L. M. (2012). Propiedades psicométricas del Maslach Burnout Inventory-GS en una muestra multiocupacional venezolana. *Revista de Psicología (PUCP), 30*(1), 103-128. Doi: http://www.scielo.org.pe/scielo.php?pid=S0254-92472012000100005&script=sci_abstract
- Moreno-Jiménez, B., Rodríguez-Carvajal, R., & Escobar R. E. (2001). La evaluación del burnout profesional. Factorialización del MBI-GS. Un análisis preliminar. *Ansiedad y Estrés, 7*(1), 69-78. Recuperado de <http://www.integra.org.com/wp-content/docs/La%20Evaluacion%20del%20Burnout%20Profesional.pdf>
- Oramas, A., González, A., & Vergara, A. (2007). El desgaste profesional y factorialización del MBI-GS. *Revista Cubana de Salud y Trabajo, 8*(1), 37-45. Recuperado de <http://www.sld.cu/galerias/pdf/sitios/insat/rst06107.pdf>
- Pando, M., Aranda, C., & López, M. (2015). Validez factorial del Maslach Burnout Inventory-General Survey en ocho países latinoamericanos. *Ciencia & Trabajo, 17*(52), 28-31. <https://doi.org/10.4067/s0718-24492015000100006>
- Penfield, R. D. (2005). DIFAS: Differential item functioning analysis system. Computer program exchange. *Applied Psychological Measurement, 29*(2), 150-151. <https://doi.org/10.1177/0146621603260686>
- Poghosyan, L., Aiken, L. H., & Sloane, D. M. (2009). Factor structure of the Maslach Burnout Inventory: An analysis of data from large scale cross-sectional surveys of nurses from eight countries. *International Journal of Nursing Studies, 46*(7), 894-902. <https://doi.org/10.1016/j.ijnurstu.2009.03.004>
- Purvanova, R. K., & Muros, J. P. (2010). Gender differences in burnout: A meta-analysis. *Journal of Vocational Behavior, 77*, 168-185. <https://doi.org/10.1016/j.jvb.2004.04.006>
- Sáenz, V. K. (2014). Adaptación y estandarización del Maslach Burnout Inventory versión General Survey (MBI-GS) a una muestra de la población trabajadora costarricense. (Tesis de pregrado, Universidad de Costa Rica).
- Salanova, M., & Schaufeli, W. B. (2000). Exposure to information technologies and its relation to burnout. *Behavior & Information Technology, 19*, 385-392. <https://doi.org/10.1080/014492900750000081>
- Saris, W. E., Satorra, A., & van der Veld, W. M. (2009). Testing structural equation models or detection of misspecifications? *Structural Equation Modeling, 16*, 561-582. <https://doi.org/10.1080/10705510903203433>

- 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). California: Sage.
- Schaufeli, W. B., Leiter, M. P., & Maslach, C. (2009). Burnout: 35 years of research and practice. *Career Development International*, 14(3), 204-220. <https://doi.org/10.1108/13620430910966406>
- Schutte, N., Toppinen, S., Kalimo, R., & Schaufeli, W. (2000). The factorial validity of the Maslach Burnout Inventory-General Survey (MBI-GS) across occupational groups and nations. *Journal of Occupational and Organizational psychology*, 73(1), 53-66. <https://doi.org/10.1348/096317900166877>
- Scott, N. W., Fayers, P. M., Aaronson, N. K., Bottomley, A., de Graeff, A., Groenvold, M., ... Sprangers, M. A. (2009). A simulation study provided sample size guidance for differential item functioning (DIF) studies using short scales. *Journal of Clinical Epidemiology*, 62(3), 288-295. <https://doi.org/10.1016/j.jclinepi.2008.06.003>
- Shirom, A. (2009). Acerca de la validez del constructo, predictores y consecuencias del burnout en el lugar de trabajo. *Ciencia & Trabajo*, 32(11), 44-54. Recuperado de http://www.imbiomed.com.mx/1/1/articulos.php?method=showDetail&id_articulo=73875&id_seccion=4332&id Ejemplanar=7372&id_revista=275
- Swaminathan, H., & Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement*, 27(4), 361-370. <https://doi.org/10.1111/j.1745-3984.1990.tb00754.x>
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using Multivariate Statistics*. New York: Allyn and Bacon.
- Teresi, J. A. (2001). Statistical methods for examination of differential item functioning (DIF) with applications to cross-cultural measurement of functional, physical and mental health. *Journal of Mental Health and Aging*, 7, 31-40. Recuperado de <http://psycnet.apa.org/record/2001-06354-003>
- Teresi, J. A., Ramirez, M., Lai, J.-S., & Silver, S. (2008). Occurrences and sources of differential item functioning (DIF) in patient-reported outcome measures: Description of DIF methods, and review of measures of depression, quality of life and general health. *Psychology Science Quarterly*, 50(4), 538. Recuperado de http://www.psychologie-aktuell.com/fileadmin/download/PsychologyScience/4-2008/05_Teresi.pdf
- Teresi, J.A., Ramirez, M., Jones, R. N., Choi, S., & Crane, P. K. (2012). Modifying measures based on differential item functioning (DIF) impact analyses. *Journal of Aging and Health*, 24(6), 1044-1076. <https://doi.org/10.1177/0898264312457750>
- Tomás, J. M., de los Santos, S., Alonso-Andres, A., & Fernández, I. (2016). Validation of the Maslach Burnout Inventory-General Survey on a representative sample of Dominican teachers: Normative data. *The Spanish Journal of Psychology*, 19. <https://doi.org/10.1017/sjp.2016.91>
- Walker, C. M. (2011). What's the DIF? Why differential item functioning analyses are an important part of instrument development and validation. *Journal of Psychoeducational Assessment*, 29(40), 364-376. <https://doi.org/10.1177/0734282911406666>
- De Winter, J. D., Dodou, D., & Wieringa, P. A. (2009). Exploratory factor analysis with small sample sizes. *Multivariate Behavioral Research*, 44(2), 147-181. <https://doi.org/10.1080/00273170902794206>
- Xanthopoulou, D., Bakker, A. B., Kantas, A., & Demerouti, E. (2012). Measuring burnout and work engagement: Factor structure, invariance, and latent mean differences across Greece and the Netherlands. *International Journal of Business Science and Applied Management*, 7(2), 40-52. Recuperado de <http://www.business-and-management.org/paper.php?id=86>

Zumbo, B. A. (1999). *Handbook on the theory and methods of differential item functioning (DIF): Logistic regression modeling as a unitary framework for binary and Likert-type (Ordinal)*

item scores. Ottawa: Directorate of Human Resources Research and Evaluation, Department of National Defense.

Recibido: abril 25, 2018
Aceptado: mayo 17, 2019