

Análisis psicométrico de una escala de agotamiento emocional académico con la teoría de la respuesta al ítem

Auné, Sofía^{a,b}; Dominguez-Lara, Sergio^{c*}; Attorresi, Horacio^a

Artículo Original

Resumen	Abstract	Tabla de Contenido
<p>El objetivo fue analizar la Escala de Cansancio Emocional (ECE), que evalúa agotamiento emocional académico (AEA), con la Teoría de la Respuesta al Ítem (TRI). La muestra estuvo compuesta por 517 estudiantes peruanos de la carrera de psicología (76% mujeres). Ninguno de los ítems presentó funcionamiento diferencial del ítem según sexo. Con respecto a los parámetros de discriminación, estos tuvieron valores que representan la capacidad de las categorías de respuesta para distinguir entre participantes con diferentes niveles de AEA. Además, los ítems seis, siete y nueve presentaron niveles más elevados de discriminación. Por otro lado, la distancia entre los parámetros de umbral fue amplia, indicando que fue posible discriminar entre distintos niveles de AEA. El análisis con la función de información del test mostró una precisión aceptable en un rango amplio de niveles de AEA. Se puede concluir que la ECE obtuvo excelentes evidencias de calidad psicométrica mediante la TRI.</p>	<p>Psychometric analysis of an emotional exhaustion scale with the item response theory. The objective was to analyze the Emotional Exhaustion Scale (EES), which evaluates academic emotional exhaustion (AEE), using the Item Response Theory (IRT). The sample was comprised by 517 Peruvian psychology students (76% women). Furthermore, none of the items presented differential item functioning according to sex. Regarding the discrimination parameters, we found values that represent the ability of the response categories to distinguish between participants with different levels of AEE. In addition, items six, seven and nine showed the highest levels of discrimination. On the other hand, the distance between the threshold parameters was wide, indicating that it was possible to discriminate between different levels of AEE. The analysis with the test information function showed an acceptable precision in a wide range of AEE levels. It can be concluded that the EES obtained excellent evidence of psychometric quality using the IRT.</p>	<p>Introducción 41 Método 43 Participantes 43 Instrumento 43 Procedimiento 43 Análisis de datos 44 Resultados 44 Discusión 47 Referencias 49</p>
<p><i>Palabras clave:</i> cansancio emocional, agotamiento emocional académico, modelo de respuesta graduada, teoría de la respuesta al ítem.</p>	<p><i>Keywords:</i> emotional fatigue, academic emotional exhaustion, graded response model, item response theory.</p>	

Recibido el 21 de julio de 2020; Aceptado el 08 de octubre de 2020

Editaron este artículo: Tomas Caycho Rodriguez, Paula Abate, Melisa Díaz y Paula Aguirre

Introducción

Agotamiento emocional académico

El *burnout académico* (BA) hace referencia a un proceso insidioso caracterizado por el sentimiento de *agotamiento emocional* académico (AEA) a causa de las demandas del estudio, menguando el compromiso del estudiante generando una actitud cínica (*cinismo*) y disminuyendo sus creencias de autoeficacia (*ineficacia académica*; Schaufeli, Martínez, Pinto, Salanova, & Bakker, 2002). En ese sentido, su estudio surgió como una necesidad para entender

la forma en que los estudiantes universitarios experimentan ese agotamiento y distanciamiento de las actividades académicas como consecuencia de las exigencias propias del contexto educativo, y considerando que ha sido asociado con ansiedad, depresión y otros aspectos vinculados al malestar físico y emocional (Caballero, Breso, & González Gutiérrez, 2015; Caballero, Hederich, & Palacio, 2010; Marengo-Escuderos, Suárez-Colorado, & Palacio-Sañudo, 2017; Rosales Ricardo & Rosales Paneque,

^a Universidad de Buenos Aires, Instituto de Investigaciones, Buenos Aires, Argentina.

^b Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina

^c Universidad de San Martín de Porres, Lima, Perú

*Enviar correspondencia a: Dominguez-Lara, S. E-mail: sdominguezmpcs@gmail.com

2013), traería consigo consecuencias negativas a corto o largo plazo, a nivel personal, como el desarrollo de los trastornos antes mencionados, así como académicas, sea por la desaprobación de alguna asignatura o el abandono de la institución educativa.

El BA, como ya se mencionó, es un proceso gradual, y se conoce que la primera fase y núcleo de este proceso es el AEA (Maslach & Leiter, 2016; Nikodijević, Labrović, & Đoković, 2012; Tuithof et al., 2017), el cual se caracteriza por una sensación progresiva de pérdida de energía y de recursos emocionales para hacer frente a las demandas del contexto (García, Escorcía, & Pérez, 2017). Por ello, es necesaria su evaluación a nivel preventivo, dado que si el estudiante consolida el BA, cualquier intervención para revertir la situación tendrá algunos obstáculos.

Escala de Cansancio Emocional

Para la evaluación del AEA en habla hispana se elaboró la Escala de Cansancio Emocional (ECE; Ramos, Manga, & Morán, 2005), que es una medida unidimensional del constructo que consta de 10 ítems escalados en formato Likert de cinco opciones. Este instrumento ha recibido evidencia psicométrica favorable en estudiantes universitarios de Argentina (Fontana, 2011), México (González Ramírez & Landero Hernández, 2007), España (Herrera Torres, Mohamed Mohand, & Cepero Espinosa, 2016), Perú (Dominguez-Lara, Fernández-Arata, Manrique-Millones, Alarcón-Parco, & Díaz-Peñaloza, 2018) y Colombia (Dominguez-Lara et al., 2021), y el panorama es similar: estructura unidimensional, indicadores de confiabilidad elevados, grado aceptable de invarianza entre hombres y mujeres, estabilidad temporal y asociación coherente con constructos teóricamente relevantes.

Sin embargo, toda esa evidencia fue obtenida bajo los supuestos de la Teoría Clásica de los Test (TCT), lo que no representa una limitación en sí misma, pero si se considera que un supuesto de la TCT es que la relación entre los ítems y el constructo es lineal, sería necesario ampliar la perspectiva para tener un panorama sobre su funcionamiento psicométrico en vista de otros enfoques que podrían complementar o, inclusive, contradecir estos hallazgos. En esta línea de argumentación, es necesario conocer algunas limitaciones de la TCT (Embretson & Reise, 2000; Hambleton & Slater, 1997) a fin de lograr una

mejor comprensión del enfoque elegido: (a) los índices psicométricos obtenidos dependen de la muestra de participantes que se utilice; (b) las estimaciones calculadas tienen validez específicamente para esa muestra u otra similar; (c) si el instrumento presenta índices de dificultad y discriminación diferentes genera resultados distintos en la misma muestra y (d) los resultados de dos o más escalas, aunque midan el mismo constructo, no se pueden comparar directamente.

Aplicación de la Teoría de la Respuesta al Ítem (TRI)

La *Teoría de la Respuesta al Ítem* (TRI) es una familia de modelos no lineales para la medición psicológica y educativa (Birnbaum, 1968; Rasch, 1960), y se basa en el concepto de rasgo latente (θ ; Lazarsfeld, 1950). Los θ se estiman a partir de todos los ítems que configuran la escala y representan el nivel del constructo evaluado (e.g., el nivel de AEA), de tal forma que la puntuación de los sujetos en determinado instrumento de medición sirve para estimar θ (Muñiz Fernández, 1997). Existen algunos supuestos de los modelos más tradicionales de la TRI que deben considerarse: la unidimensionalidad, independencia local de los ítems, y la ausencia de funcionamiento diferencial de los ítems (DIF). El primero espera que todos los ítems representen al θ que está midiendo; el segundo, que no debe existir varianza irrelevante compartida entre los ítems, dado que, si existen pares de ítems localmente dependientes, puede existir otra dimensión (sustantiva o no); y el último, el DIF, se asocia con la evaluación de la invarianza de medición, es decir, que la ausencia de DIF implicaría que el θ se mide de forma distinta entre los grupos (e.g. Auné, Abal, & Attorresi, 2019). Actualmente, existen modelos que permiten romper con alguno de estos supuestos, por ejemplo, los modelos multidimensionales de la TRI (e.g. Auné, Abal, & Attorresi, 2022), aunque su implementación es mucho más compleja. Por último, la valoración de la confiabilidad en la TRI permite saber en qué niveles del θ las puntuaciones son más confiables, mientras que la confiabilidad entendida desde la TCT se refiere a la consistencia de las respuestas procurando la homogeneidad de los ítems que miden el constructo (Attorresi, Lozzia, Abal, Galibert, & Aguerri, 2009; Toland, 2014).

El presente estudio

El objetivo de este estudio fue evaluar las propiedades psicométricas de la ECE en una muestra de estudiantes peruanos utilizando la TRI. La justificación del trabajo es estrictamente metodológica, basándose en dos aspectos.

El primero está referido a la potencial invariabilidad de las propiedades psicométricas, ya que si bien la ECE cuenta con evidencias favorables de confiabilidad y validez desde la TCT, quedaba como tarea pendiente evaluar su desempeño bajo otros enfoques en psicometría como la TRI (Dominguez-Lara et al., 2021). De ese modo, este análisis permitiría conocer algunos indicadores relevantes, como su capacidad discriminativa en distintos niveles del θ , así como su confiabilidad.

Esto es importante porque el supuesto de la relación lineal entre ítems y constructo de la TCT en ocasiones no aplica a algunos constructos psicológicos dado que no siempre existe una relación de homogeneidad entre todos los ítems con relación al rasgo que se está evaluando. Por ejemplo, dos ítems respondidos con la misma opción de respuesta (e.g., *siempre*) pueden tener un contenido que se diferencie en cuanto a la representatividad o intensidad del constructo. Esto puede apreciarse, por ejemplo, en escalas como el *Patient Health Questionnaire – 9* (Kroenke, Spitzer, & Williams, 2001) donde los primeros ítems representan los síntomas relevantes de depresión (1. *Poco interés o placer en las cosas*; 2. *Sentirse desanimado(a) deprimido(a) o sin esperanza*, otros más leves (e.g., 7. *Dificultad para concentrarse en las cosas, tales como leer el periódico o ver la televisión*), y otro más grave (9. *Pensamientos de que sería mejor estar muerto[a] o que quisiera lastimarse de alguna forma*); pero todos se califican en la misma escala (de *cero a cuatro*). De este modo, es probable que presente dentro de la ECE ocurra algún tipo de variación considerando que los ítems no son completamente homogéneos.

Asimismo, la invarianza de medición es un procedimiento recomendado en la actualidad para realizar comparaciones justas entre grupos (Pendergast, von der Embse, Kilgus, & Eklund, 2017) pero aún no se ha consolidado como una exigencia a nivel de estudios psicométricos bajo la TCT que garantice su uso en cada caso. Respecto a este punto, un estudio previo con el ECE evidenció que la medida del AEA es invariante según el sexo (Dominguez-Lara et al., 2018), por

lo que sería esperable que en este estudio cumpla el supuesto del DIF.

Por último, desde la TCT fueron hallados indicadores aceptables de confiabilidad, pero en la medida en que la concepción del procedimiento es distinta en TRI, esta podría verse afectada cuando existen ítems que evalúan diferente intensidad del rasgo que se pretende medir.

El segundo aspecto a considerar es el práctico, ya que al tener la posibilidad de conocer qué ítems poseen mayor capacidad discriminativa, podría aumentarse el valor predictivo de la ECE y así detectar estudiantes que estén en verdadero riesgo de desarrollar BA.

Método

El estudio es de tipo instrumental (Ato, López, & Benavente, 2013), el cual busca analizar las propiedades psicométricas, confiabilidad y validez, de la ECE mediante la TRI.

Participantes

La muestra intencional se conformó por 517 estudiantes peruanos (76.21% mujeres) de psicología del 3er al 8vo ciclo entre los 17 y 53 años ($M_{\text{edad}} = 20.76$ años; $DE_{\text{edad}} = 3.143$), predominantemente solteros (95.16%) y sin trabajo actual (77.76%).

Instrumento

Escala de Cansancio Emocional (ECE; Ramos et al., 2005). Es una medida unidimensional del AEA compuesta por 10 ítems en respuesta categórica ordenada de cinco opciones que va desde *raras veces* [1] hasta *siempre* [5]. A mayor puntuación, mayor presencia de AEA. Fue utilizada la versión peruana (Dominguez-Lara et al., 2018).

Procedimiento

Esta investigación forma parte de un proyecto mayor aprobado por la universidad del segundo autor, en el cual fueron incluidas otras escalas, además de la ECE, y fue conducida según principios de la declaración de Helsinki (Asociación Médica Mundial, 1964) y del código de ética del Colegio de Psicólogos del Perú (Colegio de Psicólogos del Perú, 2017). De forma previa a la administración de los instrumentos en horario de clase se solicitó el permiso respectivo a las instancias pertinentes, lo que posteriormente fue coordinado con los docentes de cada asignatura. Entonces, después de firmar el

consentimiento informado, los estudiantes completaron los cuestionarios de manera voluntaria y anónima.

Análisis de datos

De forma previa al análisis principal, se analizó si los datos cumplían con los supuestos de los modelos de la TRI, la unidimensionalidad, independencia local de los ítems, y la ausencia de funcionamiento diferencial de los ítems (DIF). La evaluación del cumplimiento del primer supuesto fue realizada mediante un Análisis Factorial Exploratorio (AFE) usando el método de mínimos cuadrados no ponderados con base en la matriz de correlaciones policóricas, e implementando el Análisis Paralelo (Timmerman & Lorenzo-Seva, 2011) para determinar el número de factores.

Además del supuesto de unidimensionalidad, el Modelo de Respuesta Graduada (MRG; Samejima, 1968) también asume que, dado un nivel fijo del rasgo a medir, las respuestas a los ítems son independientes entre sí (independencia local). Para corroborar este supuesto, se calculó el índice X^2_{LD} (Chen & Thissen, 1997) para cada par de ítems. Si dicho índice es mayor que 10, se indica que el supuesto no se cumple.

Por último, se exploró la presencia de DIF según sexo con el test de Wald modificado (Cai, 2012; Cai, Thissen, & du Toit, 2011; Langer, 2008). Este se aplicó a cada uno de los diez ítems que componen la ECE, tomando al resto como anclaje. Un resultado estadísticamente significativo indica que el ítem posiblemente exhiba DIF, es decir, que mida el rasgo de forma distinta entre los grupos.

Luego de verificar el cumplimiento de los supuestos, se implementó el MRG (Samejima, 1968), que es el más utilizado en la actualidad para respuesta politómica ordenada, utilizando el método de estimación de máxima verosimilitud (Bock & Aitkin, 1981). El ajuste global de los datos al MRG se calculó con el estadístico M_2 (Maydeu-Olivares & Joe, 2005, 2006) y el índice RMSEA asociado, aceptándose un ajuste óptimo si este último es menor o igual a .05. El ajuste de cada ítem al modelo se evaluó con el índice $S-\chi^2$ (Orlando & Thissen, 2000, 2003), considerando que ajusta si $p > .01$ (Toland, 2014).

Mediante la implementación del MRG, se obtuvieron los parámetros a (discriminación) y b_m (de umbral o de localización) de cada ítem.

Mientras que el parámetro a permite conocer qué tan bien el ítem distingue entre participantes con diferentes niveles de AEA, los parámetros b_m indican la dificultad de cada categoría de respuesta. Entonces, si el parámetro de inclinación a es moderado ($> .65$), alto (> 1.35) o muy alto (> 1.70 ; Baker, 1985) y las distancias entre los parámetros b_m de un mismo ítem son amplias, sus categorías de respuesta son eficaces para discriminar entre participantes con diferentes niveles de AEA. Las *Curvas Características de las Categorías del Ítem* (CCCI) grafican, para cada nivel de AEA, la probabilidad de seleccionar cada una de las categorías de respuesta.

Por otro lado, se hallaron evidencias de confiabilidad en el marco de la TRI, tanto para el ítem como para el test. La *Función de Información del Ítem* (FII) indica la precisión de determinado ítem en la medición de cada nivel de AEA, y su sumatoria de forma la *Función de Información del Test* (FIT), que informa la confiabilidad de la escala según el nivel de AEA. Finalmente, la *Confiabilidad Marginal* (CM) permite conocer la fiabilidad global de la escala.

En cuanto a los programas estadísticos utilizados, el AFE fue realizado con el programa FACTOR versión 10.5 (Ferrando & Lorenzo-Seva, 2017), mientras que los análisis de independencia local, DIF y modelización con TRI con el programa IRTPRO 4.2 (Cai et al., 2011).

Resultados

Cumplimiento de supuestos

Inicialmente se analizó la unidimensionalidad, y mediante la implementación del análisis paralelo sobre las respuestas a la ECE se concluyó que una única dimensión explicó una variabilidad significativa de los ítems (49.11%) y el análisis paralelo señaló una estructura de un solo factor. Entonces, el supuesto de unidimensionalidad pudo darse por satisfecho, dado que los valores del estadístico de Bartlett ($p < .01$) y del test KMO (0.948) indicaron que la matriz de correlaciones podía ser factorizada.

Posteriormente, los resultados del DIF por sexo, la prueba basada en el χ^2 de Wald indicó que la totalidad de los ítems se encuentran libres de DIF uniforme y no uniforme por sexo ($p > .05$; Tabla 1).

Tabla 1.
Análisis de Funcionamiento Diferencial del Ítem entre Mujeres y Hombres

Ítem	χ^2_{Total} (gl 5)	p	χ^2_a (gl 1)	p	χ^2_b (gl 4)	p
1. Los exámenes me producen una tensión excesiva.	4.0	.55	0.0	.98	4.0	.41
2. Creo que me esfuerzo mucho para lo poco que consigo.	1.4	.92	0.5	.48	0.9	.92
3. Me siento bajo de ánimo, como triste, sin motivo aparente.	0.9	.97	0.7	.41	0.3	.99
4. Hay días que no duermo bien a causa del estudio.	4.4	.49	0.1	.73	4.3	.37
5. Tengo dolor de cabeza y otras molestias que afectan a mi rendimiento académico.	2.9	.72	1.0	.31	1.8	.77
6. Hay días que noto más la fatiga y me falta energía para concentrarme.	4.2	.52	0.2	.68	4.1	.40
7. Me siento emocionalmente agotado por mis estudios.	5.2	.39	0.7	.41	4.6	.34
8. Me siento cansado al final de la jornada de estudio.	3.4	.64	1.0	.33	2.4	.66
9. Estudiar pensando en los exámenes me produce estrés.	3.9	.57	0.3	.60	3.6	.47
10. Me falta tiempo y me siento desbordado por los estudios.	4.0	.55	0.2	.63	3.8	.44

Nota: El χ^2_{Total} se refiere al test global, el χ^2_a al test para DIF no uniforme y el χ^2_b al test para DIF uniforme. Se consideran significativos los $p < .05$.

Para terminar, con respecto a la independencia local, el estadístico X^2_{LD} tomó valores para cada par de ítems entre -0.7 y 9.7, menores al valor límite de 10. Por lo tanto, se consideró que los pares de ítems son lo suficientemente independientes para implementar el MRG.

Análisis principal

En cuanto al ajuste del modelo, se apreció que el ajuste global ($M_2 = 1288.17$; $gl = 710$; $p < .001$; $RMSEA = .04$) y el ajuste de cada ítem ($S-\chi^2$ asociados a un $p > .01$) al MRG resultaron adecuados, lo que hace viable su implementación (Tabla 2).

Tabla 2.
Ajuste de los ítems al Modelo de Respuesta Graduada

Ítem	χ^2	$g.l.$	p
1	94.42	88	0.3001
2	101.36	92	0.2364
3	99.43	86	0.1524
4	121.76	92	0.0206
5	74.95	84	0.7498
6	71.76	65	0.2633
7	72.53	64	0.2169
8	68.07	76	0.7302
9	67.50	67	0.4610
10	52.10	73	0.9695

Con respecto a los parámetros de discriminación a , éstos tuvieron valores que representan una capacidad moderada/alta de las categorías de respuesta para distinguir entre participantes con diferentes niveles de AEA (≥ 1.16), siendo los ítems del seis, siete y nueve los que presentan niveles más elevados de discriminación (> 2.00) y, en consecuencia, brindan más información del AEA. Por otro lado, la distancia entre los b_m fue amplia, indicando que fue posible discriminar entre distintos niveles de AEA con cada categoría de respuesta (Tabla 3).

En cuanto a la interpretación individual de la CCCI de algunos ítems, el segundo (*Creo que me esfuerzo mucho para lo poco que consigo*) tiene un nivel moderado de discriminación entre los participantes ($a = 1.20$), que va desde niveles bajos hasta muy altos de AEA (los b_m van desde -1.56 hasta 2.77 en una escala estandarizada del rasgo), pero es el de menor capacidad discriminativa entre los 10 ítems. No obstante, discrimina menos entre participantes que tengan niveles muy bajos de AEA (Figura 1).

Por otro lado, el ítem siete (*Me siento emocionalmente agotado por mis estudios*) tiene la mejor capacidad de discriminación entre los participantes ($a = 2.68$) desde niveles bajos hasta niveles altos de AEA (los b_m van desde -1.21 hasta 1.69 en una escala estandarizada del rasgo). Empero, discrimina menos entre participantes que tengan niveles extremos de AEA (muy altos o muy bajos; Figura 2).

Tabla 3.
Parámetros estimados de los ítems de la ECE

Ítem	a (e.e.)	b_1 (e.e.)	b_2 (e.e.)	b_3 (e.e.)	b_4 (e.e.)	b_{prom}
1	1.45 (0.13)	-2.06 (0.18)	-1.00 (0.11)	0.60 (0.09)	1.82 (0.15)	-0.16
2	1.20 (0.11)	-1.56 (0.16)	-0.27 (0.09)	1.24 (0.13)	2.77 (0.25)	0.55
3	1.44 (0.13)	-0.73 (0.10)	0.06 (0.08)	1.32 (0.12)	2.37 (0.20)	0.76
4	1.16 (0.11)	-2.07 (0.20)	-0.91 (0.12)	0.43 (0.10)	2.26 (0.21)	-0.07
5	1.57 (0.13)	-1.10 (0.11)	-0.26 (0.08)	0.82 (0.09)	2.07 (0.17)	0.38
6	2.49 (0.20)	-1.73 (0.12)	-0.80 (0.07)	0.30 (0.06)	1.65 (0.11)	-0.15
7	2.68 (0.21)	-1.21 (0.09)	-0.35 (0.06)	0.61 (0.06)	1.69 (0.11)	0.19
8	1.96 (0.16)	-1.92 (0.14)	-0.91 (0.09)	0.20 (0.07)	1.55 (0.12)	-0.27
9	2.44 (0.20)	-1.56 (0.11)	-0.69 (0.07)	0.20 (0.06)	1.02 (0.08)	-0.26
10	2.16 (0.17)	-1.65 (0.12)	-0.72 (0.08)	0.41 (0.07)	1.33 (0.10)	-0.16

Nota. a : parámetro de discriminación calculado con el MRG; b_1 , b_2 , b_3 y b_4 : parámetros de umbral calculados con el MRG; e.e.: error estándar; b_{prom} : promedio de los parámetros de umbral b .

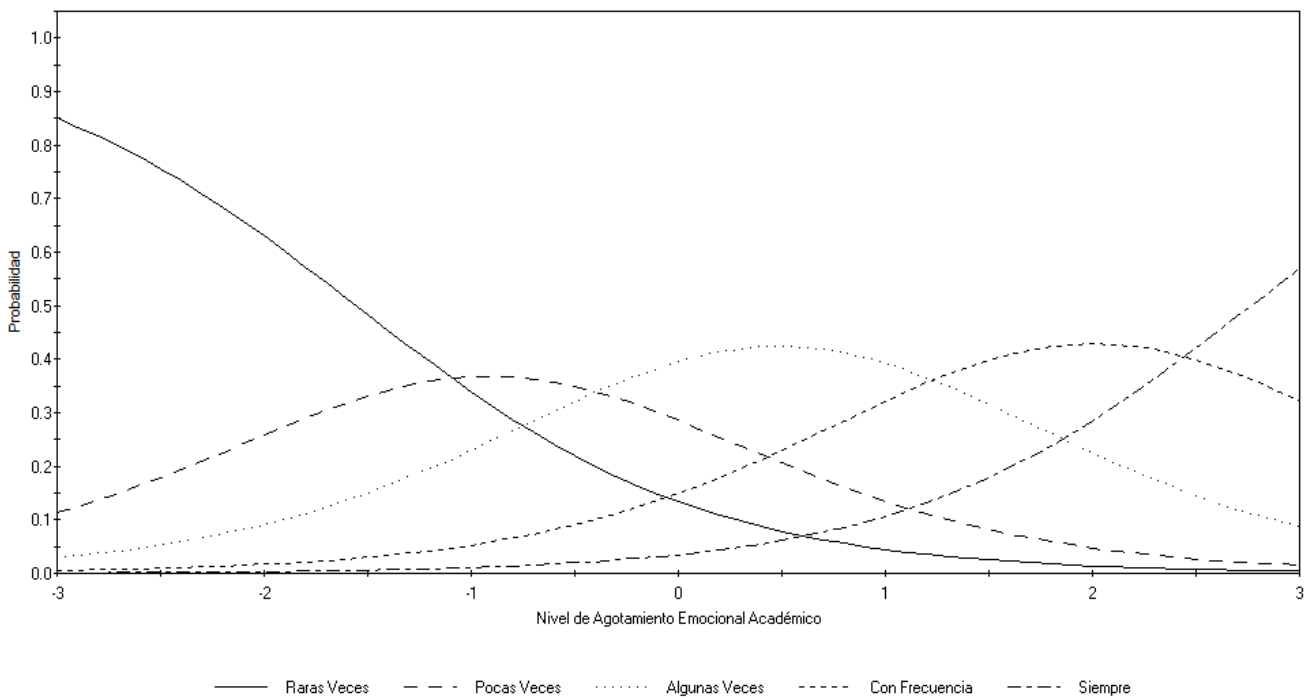


Figura 1. Curvas Características de las Categorías de Respuesta del Ítem 2

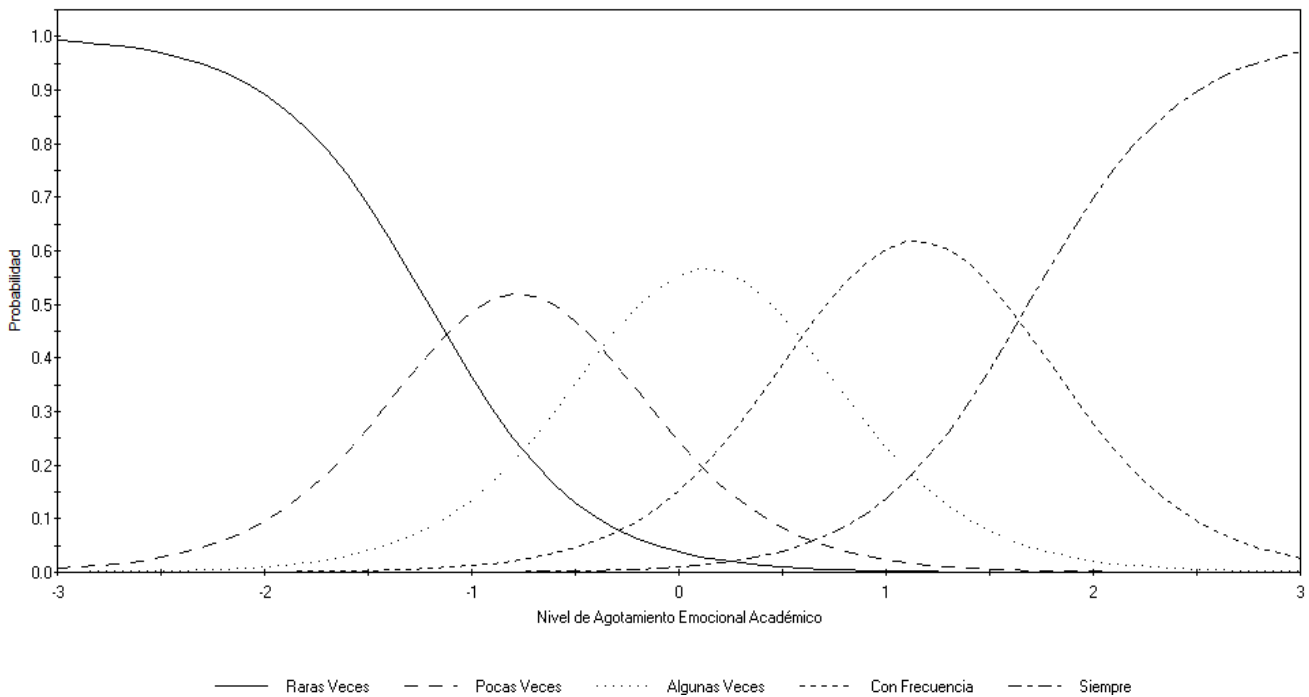


Figura 2. Curvas Características de las Categorías de Respuesta del Ítem 7

La confiabilidad desde la TRI fue estimada con la FII y la FIT para determinados niveles de la ECE. La escala es muy informativa desde niveles bajos hasta a altos del rasgo, disminuyendo en los extremos (Tabla 4). Los últimos cinco ítems (6 al 10) son los que aportan mayor información al test.

Tabla 4.
Funciones de Información de los Ítems y del test

	θ -2.4	-1.6	-0.8	0	0.8	1.6	2.4
Ítem							
1	0.52	0.62	0.61	0.59	0.62	0.60	0.45
2	0.29	0.40	0.44	0.44	0.43	0.43	0.42
3	0.16	0.36	0.57	0.63	0.62	0.63	0.57
4	0.35	0.40	0.42	0.41	0.40	0.38	0.35
5	0.25	0.55	0.74	0.76	0.75	0.72	0.60
6	0.83	1.70	1.74	1.64	1.43	1.60	0.71
7	0.27	1.39	2.00	1.94	1.93	1.90	0.81
8	0.79	1.10	1.12	1.10	1.01	1.00	0.52
9	0.61	1.59	1.73	1.73	1.70	0.95	0.19
10	0.65	1.29	1.36	1.29	1.36	1.11	0.38
Test	5.71	10.42	11.73	11.53	11.25	10.32	6.00
e.e.	0.42	0.31	0.29	0.29	0.30	0.31	0.41

Nota: θ : nivel de agotamiento emocional académico; Test: nivel de información de la Escala de Agotamiento Emocional Académico para cada nivel de Agotamiento Emocional Académico; e.e.: error estándar.

Por último, la confiabilidad marginal obtuvo un indicador de .91, siendo su interpretación similar al coeficiente alfa de la TCT. En este caso su obtención es adecuada, dado que la FIT tuvo forma de meseta. No obstante, si el test brindara información sólo en niveles altos o bajos del rasgo (o formar asimétricas de la FIT), el índice de confiabilidad marginal no sería interpretable (Figura 3).

Discusión

La ECE cumplió con el supuesto de unidimensionalidad, lo cual está en concordancia con estudios previos (Dominguez-Lara et al., 2018, 2021; Fontana, 2011; González Ramírez & Landero Hernández, 2007; Herrera Torres et al., 2016; Ramos et al., 2005) dado que se partió de un enfoque basado en la TCT, como el AFE. Con todo, el ajuste de los datos al MRG también fue favorable, lo que indicaría que también desde la TRI tiene carácter unidimensional. Por otro lado, el análisis DIF replicó lo que ocurrió en otro estudio con universitarios peruanos en cuanto al grado de invarianza de la ECE (Dominguez-Lara et al., 2018).

En cuanto al análisis de la ECE bajo el MRG, se evidenció que la escala tiene un nivel de discriminación adecuado desde niveles bajos

hasta altos del rasgo, y teniendo en cuenta que los criterios de valoración que fueron usados son más restrictivos dado que se basan en pruebas de rendimiento (Baker, 1985) y la ECE es una escala

de comportamiento típico, le daría un valor agregado. No obstante, al interior de la ECE existe heterogeneidad con relación a la capacidad discriminativa de los ítems.

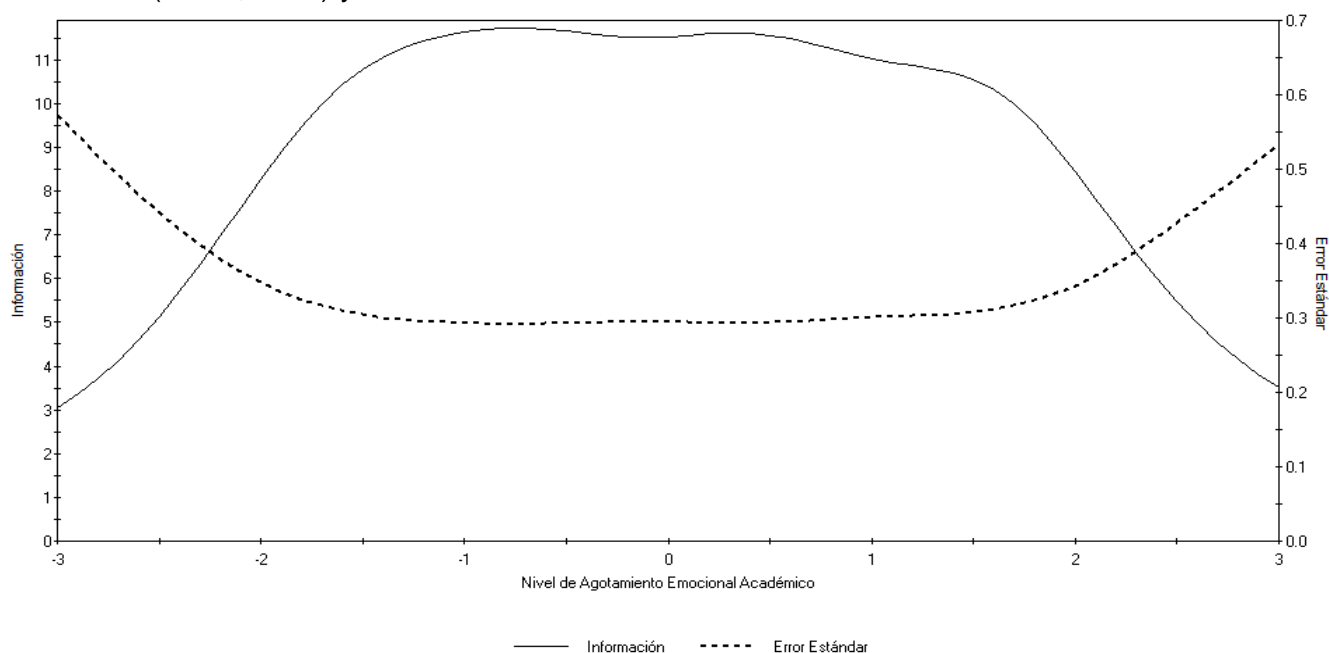


Figura 3. Función de información de la Escala de Cansancio Emocional

Por un lado, los ítems que brindan más información del rasgo son el siete (*Me siento emocionalmente agotado por mis estudios*), seis (*Hay días que noto más la fatiga y me falta energía para concentrarme*) y nueve (*Estudiar pensando en los exámenes me produce estrés*) dado que el parámetro de discriminación (a) es muy alto y las distancias entre los parámetros de localización (b_m) son amplias, lo que indica que las categorías de respuesta resultan eficaces para discriminar entre participantes con diferentes niveles de AEA. En tal sentido, son los que poseen mayor calidad psicométrica para la medición del rasgo, lo que se puede explicar en la medida que recogen los aspectos esenciales del constructo y son los más vinculados a la experiencia académica.

Por otro lado, los ítems con los parámetros (a y b_m) de magnitud menos favorable tienen un contenido más general en cuanto a aspectos negativos, como el pesimismo (ítem 2: *Creo que me esfuerzo mucho para lo poco que consigo*), tristeza (ítem 3: *Me siento bajo de ánimo, como triste, sin motivo aparente*), o poca calidad de sueño (ítem 4: *Hay días que no duermo bien a causa del estudio*). Esto indicaría que algunas

emociones negativas generales (e.g., pesimismo) pueden estar asociadas pero no determinan la experiencia de AEA. Es decir, si el estudiante marca las opciones más elevadas (e.g., *siempre*) no existe la certeza que experimente un elevado nivel de AEA, mientras que una respuesta similar en el ítem seis o siete, sí lo indicaría y sería, por tanto, de mayor importancia para el evaluador.

Con relación a la confiabilidad, la ECE es capaz de informar apropiadamente entre niveles bajos y altos, pero en los extremos (muy altos o muy bajos) de AEA el error de medición crece. En forma general, la capacidad discriminativa, así como el nivel de información alcanzado tuvieron valores adecuados para la totalidad de los ítems que componen la escala, destacando los últimos cinco ítems, y considerando que no es una prueba diagnóstica, es algo favorable.

Existen algunas limitaciones asociadas a los participantes que es necesario mencionar, ya que aspectos como el tipo de muestreo (no probabilístico) y características de la muestra (solo estudiantes de psicología) pudieron afectar los procedimientos realizados, aunque la calidad de indicadores psicométricos permite suponer que las características de la muestra no los afectan

sustancialmente. En cuanto a las limitaciones metodológicas, cuando se midió la independencia local, se halló un grado de dependencia entre los ítems algo cercano al límite de lo deseable. Por ello, puede ser útil en posteriores estudios el análisis de las causas de este nivel de dependencia local. Por otra parte, será deseable la aplicación de otros modelos de la TRI aptos para ítems de respuesta politómica, a los mismos datos y el cálculo de medidas de ajuste relativo. En este caso, se ha implementado en forma directa el MRG por haber sido, en numerosos estudios, el que mejor ha ajustado.

En cuanto a las recomendaciones, dado que en el ámbito aplicado son más importantes los niveles altos de AEA, podrían incorporarse ítems para que la escala pueda ampliar precisión a esos niveles del rasgo. Del mismo modo, conocer que determinados ítems (6, 7 y 9) brindan mayor información del rasgo podría sugerir la validación de una versión ultra breve, aunque para ello es conveniente usar otras estrategias analíticas, sobre todo con relación a criterios externos. Asimismo, si bien las propiedades psicométricas fueron favorables con la TRI, es conveniente replicarlo en estudiantes de otras nacionalidades.

Para concluir, la ECE obtuvo excelentes evidencias de calidad psicométrica mediante la TRI, dados los valores de los parámetros, el ajuste del modelo a los datos, los valores de la función de información de los ítems y del test y el coeficiente de confiabilidad marginal. Las mismas son compatibles con un constructo que no tiene características psicopatológicas y con un instrumento apto para realizar un rastillaje del agotamiento emocional académico, necesario para prevenir el burnout académico.

Referencias

- Asociación Médica Mundial (1964). *Declaración de Helsinki*. Helsinki: AMM. Recuperado de: http://www.conamed.gob.mx/prof_salud/pdf/helsinki.pdf
- Ato, M., López, J., & Benavente, A. (2013). Un sistema de clasificación de los diseños de investigación en psicología. *Anales de Psicología*, 29(3), 1038-1059. doi: 10.6018/analesps.29.3.178511
- Attorresi, H. F., Lozzia, G. S., Abal, F. J. P., Galibert, M. S., & Aguerri, M. E. (2009). Teoría de Respuesta al Ítem. Conceptos básicos y aplicaciones para la medición de constructos psicológicos. *Revista Argentina de Clínica Psicológica*, 18(2), 179-188.
- Auné, S. E., Abal, F. J. P., & Attorresi, H. F. (2019). Application of the Graded Response Model to a Scale of Empathic Behavior. *International Journal of Psychological Research*, 12(1), 49-56. doi: 10.21500/20112084.3753
- Auné, S. E., Abal, F. J. P., & Attorresi, H. F. (2022). Modeling of the UCLA Loneliness Scale According to the Multidimensional Item Response Theory. *Current Psychology*, 41(3), 1213-1220. doi: 10.1007/s12144-020-00646-y
- Baker, F. (1985). *The basics of item response theory*. Porstmouth: N. H. Heineman.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. En F. M. Lord, & M. R. Novick (Eds.). *Statistical Theories of Mental Test Scores*. Reading, MA: Addison Wesley.
- Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: application of an EM algorithm. *Psychometrika*, 46(4), 443-459. doi: 10.1007/BF02293801
- Caballero, C., Bresó, E., & González Gutiérrez, O. (2015). Burnout en estudiantes universitarios. *Psicología desde el Caribe*, 32(3), 424-441. doi: 10.14482/psdc.32.3.6217
- Caballero, C., Hederich, C., & Palacio, J. (2010). El burnout académico: delimitación del síndrome y factores asociados con su aparición. *Revista Latinoamericana de Psicología*, 42(1), 131-146.
- Cai, L. (2012). *flexMIRT: Flexible multilevel item factor analysis and test scoring* [Computer software]. Seattle, WA: Vector Psychometric Group, LLC.
- Cai, L., Thissen, D., & du Toit, S. (2011). *IRTPRO user's guide*. Lincolnwood, IL: Scientific Software International.
- Chen, W-H., & Thissen, D. (1997). Local dependence indexes for item pairs using item response theory. *Journal of Educational and Behavioral Statistics*, 22(3), 265-289. doi: 10.3102/10769986022003265
- Colegio de Psicólogos del Perú (2017). *Código de ética y deontología*. Recuperado de: https://www.cpsp.pe/documentos/marco_legal/codigo_de_etica_y_deontologia.pdf
- Dominguez-Lara, S., Fernández-Arata, M., Manrique-Millones, D., Alarcón-Parco, D., & Díaz-Peñaloza, M. (2018). Datos normativos de una escala de agotamiento emocional académico en estudiantes universitarios de psicología de Lima (Perú). *Educación Médica*, 19(3), 246-255. doi: 10.1016/j.edumed.2017.09.002
- Dominguez-Lara, S., Alarcón-Parco, D., Fernández-Arata, M., Tamayo-Agudelo, W., Bernal-Vargas, L., & Campos-Uscanga, Y. (2021). Invarianza factorial de una medida de burnout académico entre estudiantes universitarios de Perú, México y Colombia. *Revista Argentina de Ciencias del Comportamiento* 13(3), 70-86. doi: 10.32348/1852.4206.v13.n3.25946

- Embretson, S. E., & Reise, S. (2000). *Item Response Theory for Psychologists*. Mahwah, NJ: Erlbaum Publishers.
- Ferrando, P. J., & Lorenzo-Seva, U. (2017). Program FACTOR at 10: Origins, development and future directions. *Psicothema*, 29(2), 236-240. doi: 10.7334/psicothema2016.304
- 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. doi: 10.32348/1852.4206.v3.n2.5227
- García, A., Escorcía, C., & Pérez, B. (2017). Síndrome de burnout y sentimiento de autoeficacia en profesores universitarios. *Propósitos y Representaciones*, 5(2), 65-126. doi: 10.20511/pyr2017.v5n2.170
- González Ramírez, M., & Landero Hernández, R. (2007). Escala de Cansancio Emocional (ECE) para estudiantes universitarios: propiedades psicométricas en una muestra de México. *Anales de Psicología*, 23(2), 253-257.
- Hambleton, R. K., & Slater, S. C. (1997). Item Response Theory Models and Testing Practices: Current international status and future directions. *European Journal of Psychological Assessment*, 13(1), 21-28. doi: 10.1027/1015-5759.13.1.21
- Herrera Torres, L. H., Mohamed Mohand, L., & Cepero Espinosa, S. (2016). Cansancio emocional en estudiantes universitarios. *DEDICA. Revista de Educação e Humanidades*, 9, 173-191. doi: 10.30827/dreh.v0i9.6888
- Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine*, 16(9), 606-613. doi: 10.1046/j.1525-1497.2001.016009606.x
- Langer, M. (2008). *A reexamination of Lord's Wald test for differential item functioning using item response theory and modern error estimation* (Unpublished doctoral dissertation). University of North Carolina, Chapel Hill. Recuperado de: https://cdr.lib.unc.edu/concern/parent/fb4948756/file_sets/k930bx38r
- Lazarsfeld, P. F. (1950). The logical and mathematical foundation of latent structure analysis. En S. A. Stoufer et al. (Eds.). *Measurement and Prediction*. Princeton: Princeton University Press.
- Marengo-Escuderos, A., Suárez-Colorado, Y., & Palacio-Sañudo, J. (2017). Burnout académico y síntomas relacionados con problemas de salud mental en universitarios colombianos. *Psychologia*, 11(2), 45-55. doi: 10.21500/19002386.2926
- Maslach, C., & Leiter, M. P. (2016). Understanding the burnout experience: recent research and its implications for psychiatry. *World Psychiatry*, 15(2), 103-111. doi: 10.1002/wps.20311
- Maydeu-Olivares, A., & Joe, H. (2005). Limited- and full-information estimation and goodness-of-fit testing in 2n contingency tables: A unified framework. *Journal of the American Statistical Association*, 100(471), 1009-1020. doi: 10.1198/016214504000002069
- Maydeu-Olivares, A., & Joe, H. (2006). Limited information goodness-of-fit testing in multidimensional contingency tables. *Psychometrika*, 71(4), 713-732. doi: 10.1007/s11336-005-1295-9
- Muñiz Fernández, J. (1997). *Introducción a la teoría de respuesta a los ítems*. Madrid: Ediciones Pirámide.
- Nikodijević, A., Labrović, J. A., & Đoković, A. (2012). Academic Burnout Among Students at Faculty of Organizational Sciences. *Education Management*, 17(64), 47-53. doi: 10.7595/management.fon.2012.0019
- Orlando, M., & Thissen, D. (2000). Likelihood-based item fit indices for dichotomous item response theory models. *Applied Psychological Measurement*, 24(1), 50-64. doi: 10.1177/01466216000241003
- Orlando, M., & Thissen, D. (2003). Further investigation of the performance of S- χ^2 : An item fit index for use with dichotomous item response theory models. *Applied Psychological Measurement*, 27(4), 289-298. doi: 10.1177/0146621603027004004
- Pendergast, L. L., von der Embse, N., Kilgus, S. P., & Eklund, K. R. (2017). Measurement equivalence: A non-technical primer on categorical multi-group confirmatory factor analysis in school psychology. *Journal of School Psychology*, 60, 65-82. doi: 10.1016/j.jsp.2016.11.002
- Ramos, F., Manga, D., & Morán, C. (marzo, 2005). *Escala de cansancio emocional (ECE) para estudiantes universitarios: Propiedades psicométricas y asociación*. Trabajo presentado en 6º Congreso Virtual de Psiquiatría Interpsiquis. Recuperado de: <https://psiquiatria.com/bibliopsiquis/escala-de-cansancio-emocional-ece-para-estudiantes-universitarios-propiedades-psicometricas-y-asociacion/>
- Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen: The Danish Institute for Educational Research.
- Rosales Ricardo, Y., & Rosales Paneque, F. (2013). Burnout estudiantil universitario: Conceptualización y estudio. *Salud Mental*, 36(4), 337-345. doi: 10.17711/SM.0185-3325.2013.041
- Samejima, F. (1968). Estimation of latent ability using a response pattern of graded scores. *Bulletin Research Series*, 1-169. doi: 10.1002/j.2333-8504.1968.tb00153.x.

- Schaufeli, W. B., Martínez, I. M., Pinto, 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
- Timmerman, M. E., & Lorenzo-Seva, U. (2011). Dimensionality Assessment of Ordered Polytomous Items with Parallel Analysis. *Psychological Methods*, 16(2), 209-220. doi: 10.1037/a0023353
- Toland, M. (2014). Practical guide to conducting an item response theory analysis. *The Journal of Early Adolescence*, 34(1), 120-151. doi: 10.1177/0272431613511332
- 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