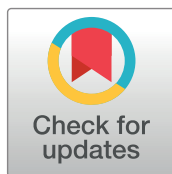




English version



Versión español



Crossmark



OPEN ACCESS

**Citation:** Amaya-Arias AC, Jaimes FA, García VJ. **Limitations of the MBI-HSS in Colombian health personnel psychometric: Evaluation after modification of its response options.** Colomb Méd (Cali);2024; 55(3):e2015818.<http://doi.org/10.25100/cm.v55i3.5818>

**Received:** 21 Oct 2023

**Revised :** 15 Jan 2024

**Accepted :** 25 Sep 2024

**Published :** 30 Sep 2024

**Keywords:**

Burnout, professional; Health Personnel; Psychometrics; MBI-HSS; Factor Analysis, Statistical; Surveys and Questionnaires; Item Response Theory.

**Palabras clave:**

Burnout profesional; Personal sanitario; Psicometría; MBI-HSS; Análisis factorial estadístico; Encuestas y cuestionarios; Teoría de respuesta a los ítems.

ORIGINAL ARTICLE

## Limitations of the MBI-HSS in Colombian health personnel psychometric: Evaluation after modification of its response options.

## Limitaciones del MBI-HSS en personal de salud colombiano: evaluación psicométrica tras la modificación de sus opciones de respuesta.

Ana Carolina Amaya-Arias, <sup>1</sup> Fabián A. Jaimes, <sup>2</sup> Jenny García Valencia <sup>3</sup>

**1** Universidad de Antioquia, Facultad Nacional de Salud Pública, Medellín, Colombia., **2** Universidad de Antioquia, Facultad de Medicina, Departamento de Medicina Interna, Medellín, Colombia., **3** Universidad de Antioquia, Facultad de Medicina, Departamento de Psiquiatría, Medellín, Colombia.

### Abstract

#### Objective:

To establish the construct validity, internal consistency, and difficulty and discrimination of the items in the Maslach Burnout Inventory- Human Services Survey, with modified response options, for measuring burnout in healthcare workers in Colombia.

#### Methods:

A descriptive study of validation of measurement instruments was conducted. With the support of a panel of experts, the response options were modified, and the new version was administered to 377 healthcare workers. Internal consistency was evaluated, and confirmatory factor analysis was conducted to assess the structural construct validity. The structural Equation Modeling model was calculated using the diagonally weighted least squares estimation method. Finally, item discrimination and difficulty thresholds of the response options were calculated using the generalized partial credit model of Item Response Theory.

#### Results:

The emotional exhaustion and personal accomplishment subscales showed good internal consistency ( $\alpha = 0.87$  and  $0.77$ ), while the depersonalization subscale was low ( $0.45$ ). The confirmatory factor analysis indicated an acceptable but not ideal fit ( $NFI$  and  $RFI < 0.95$ ;  $RMSEA > 0.06$ ). Inadequate functioning of the response options and measurement bias were identified.

#### Conclusions:

The previously validated version of the Maslach Burnout Inventory- Human Services Survey and the adapted version in this study have limitations in Colombian healthcare workers, especially in the depersonalization subscale. A thorough revision is suggested to appropriately measure burnout in healthcare personnel in our context.

**Copyright:** © 2024 Universidad del Valle



#### Conflict of Interest:

The research team declares no conflicts of interest in this study.

#### Funding:

His research was funded by the Ministry of Science, Technology, and Innovation of Colombia, under the 2016 call for 757 national doctoral scholarships, and by the Universidad de Antioquia.

#### Data Availability:

The data used in this research are available and can be requested from the corresponding author.

#### Acknowledgments:

The researchers extend special thanks to each of the hospital staff who allowed us access to their facilities and supported the research by responding to the questionnaires, as well as to the scientific and research teams who assisted with logistical aspects and encouraged staff participation.

#### Corresponding author:

Ana C. Amaya-Arias. Universidad de Antioquia, Facultad Nacional de Salud Pública, Medellín, Colombia.  
Email: [acamaya@unal.edu.co](mailto:acamaya@unal.edu.co)

## Resumen

### Objetivo:

Establecer la validez de constructo estructural, la consistencia interna y la dificultad y discriminación de los ítems de la prueba Maslach Burnout Inventory- Human Services Survey, con opciones de respuesta modificadas, para medir el burnout en trabajadores de la salud en Colombia.

### Métodos:

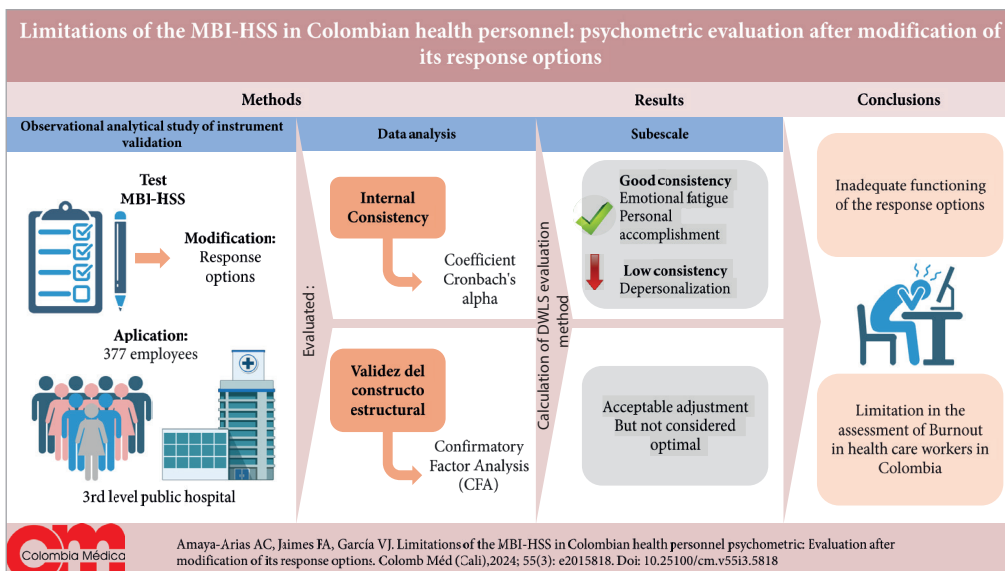
Estudio analítico observacional de validación de instrumentos. Con apoyo de un panel de expertos se modificaron las opciones de respuesta y se aplicó la nueva versión a 377 trabajadores asistenciales. Se evaluó la consistencia interna y se realizó un análisis factorial confirmatorio para la validez de constructo estructural, utilizando el método de estimación de mínimos cuadrados ponderados diagonalmente. Finalmente, se calcularon la discriminación de los ítems y los umbrales de dificultad de las opciones de respuesta haciendo uso del modelo de créditos parciales generalizados.

### Resultados:

Las subescalas fatiga emocional y realización personal mostraron buena consistencia interna ( $\alpha = 0.87$  y  $0.77$ ), mientras que el de la subescala despersonalización fue bajo ( $0.45$ ). El análisis factorial confirmatorio indicó un ajuste aceptable pero no ideal (NFI y RFI  $< 0.95$ ; RMSEA  $> 0.06$ ). Se identificó un inadecuado funcionamiento de las opciones de respuesta y un sesgo en la medición.

### Conclusiones:

La versión validada previamente del Maslach Burnout Inventory- Human Services Survey y la adaptada en este estudio presentan limitaciones en la evaluación del Burnout en los trabajadores de la salud de Colombia, especialmente en la subescala despersonalización. Se sugiere una revisión profunda de este de forma que logre medirse apropiadamente el burnout en personal de salud nuestro contexto.



## Remark

### 1) Why was this study conducted?

In the validation study of the adapted version of the MBI-HSS for Colombia, the authors were unable to confirm satisfactory reliability and validity, and there was a need to eliminate certain items. Additionally, these researchers proposed adjusting the response options as a potential solution to improve the psychometric properties of the instrument. Based on these recommendations, the response scale was modified, and the psychometric characteristics of the test were reevaluated.

### 2) What were the most relevant results of the study?

Even after modifying the response options of the MBI-HSS, limitations were identified in its structural validity, reliability, item discrimination, and response option difficulty. Therefore, the use of this scale to measure burnout among healthcare workers in Colombia is not recommended. If some of its subscales are to be applied, the one that could best measure exhaustion in healthcare personnel is the emotional fatigue subscale.

### 3) What do these results contribute?

These findings provide a deeper understanding of the applicability and accuracy of the MBI-HSS in the specific context of our country. They reveal areas for improvement in the instrument, particularly concerning certain subscales and items. Additionally, they highlight the possibility that our understanding and conceptualization of burnout may differ from that originally proposed by Maslach & Jackson, as has been evidenced in other cultures. This suggests the need to review and adapt not only the instrument but also the very concept of burnout for our population. It also invites reflection on the widespread use of this instrument despite its deficiencies in our context, which could imply systematic biases in measurements conducted for monitoring or research purposes. This is a call to action regarding the need for appropriate instrument selection, moving beyond the use of those that are most frequently adopted due to their classic status but may have issues in their adaptations in our context, even from the conceptualization of the construct being measured.

## Introduction

Burnout syndrome is a psychological phenomenon characterized by emotional exhaustion, cynicism or depersonalization, and a diminished sense of personal accomplishment in the workplace <sup>1</sup>. Its presence among healthcare professionals has raised concerns, as it not only affects the well-being and quality of life of those who experience it but may also compromise the quality of care provided and, consequently, patient safety <sup>2,3</sup>.

The Maslach Burnout Inventory (MBI) has been one of the most widely used instruments to assess burnout <sup>4</sup>. This tool has been translated into various languages and used in different countries and occupations, including educators, students, police or military personnel, and healthcare professionals <sup>5-9</sup>.

Despite its widespread use, issues have been identified in its validation across different contexts and cultures. Some studies have confirmed the original three-factor structure proposed by the authors, while others have identified structures ranging from two to five dimensions <sup>5,10</sup>. Additionally, certain questions in the MBI have elicited unfavorable reactions in cultures other than North American, which may introduce measurement biases <sup>11</sup>.

This questionnaire was validated in Spanish, revealing a three-dimensional structure. However, problems were identified with the personal accomplishment dimension, suggesting that it might consist of two aspects: self-competence and an existential component, with the latter having less impact on feelings of exhaustion<sup>5</sup>. In Latin America, Pando Moreno et al.<sup>12</sup> conducted a validation study of the MBI-GS in eight Latin American countries, finding that the factorial structure did not align with the original three-dimensional model. Instead, a two-factor structure emerged, with items related to depersonalization and exhaustion loading onto the first factor and those related to personal accomplishment onto the second.

In Colombia, Córdoba et al.<sup>13</sup>, validated the MBI-HSS among healthcare workers. Although they found good internal consistency for the overall scale, they suggested removing two items to maintain the original three-factor structure. The authors attributed these differences to difficulties in the functioning of the response options.

At the national level, there has been a recognized need to address psychosocial risk factors in the workplace. Resolution 2646 of 2008, issued by the Ministry of Social Protection, establishes the responsibility of Colombian institutions in the prevention, diagnosis, intervention, and control of these factors<sup>14</sup>. This initiative was complemented by Law 1562 of 2012, which modified the General System of Occupational Risks, focusing on the prevention of work-related illnesses and accidents<sup>15</sup>, and Resolution 2764 of 2022, issued by the Ministry of Labor, which adopts a battery of instruments for the evaluation of psychosocial risk factors and provides technical guidelines for the promotion, prevention, and intervention of psychosocial factors and their effects on the working population. These national guidelines underscore the importance of having appropriate tools to assess and address workplace burnout in the Colombian context. Thus, this project emerges as a proactive response to the identified challenges.

Given the importance of measuring burnout among Colombian healthcare workers and the need for instruments with adequate validity and reliability indicators, this project seeks to answer the following question: What are the psychometric properties of the MBI-HSS for measuring burnout among healthcare workers in Colombia after modifying its response options?

## Materials and Methods

For this study, the recommendations of the COSMIN (Consensus-based Standards for the selection of health Measurement Instruments) group were followed for the publication of validation studies of health measurement instruments<sup>16,17</sup>.

### Design

A cross-sectional study was conducted to reevaluate the validity and reliability of the Maslach Burnout Inventory-Human Services Survey (MBI-HSS) in the Colombian healthcare context.

### Setting

This study was conducted in a decentralized, third-level public hospital that provides high-complexity healthcare services. It is located in an intermediate capital city in the southern region of Colombia. Data collection took place between November 1 and December 31, 2022.

### Participants

Convenience sampling was used. The inclusion criteria were: healthcare workers providing clinical care at the hospital where the study was conducted during the data collection period and who voluntarily accepted to participate in the study. Administrative staff were excluded.

The minimum sample size was determined based on Monte Carlo simulations, aiming to establish a robust and reliable structural equation modeling (SEM) framework. For these simulations, a factor loading of 0.5 was adopted for the dimensions of each test, and an expected correlation of 0.3 between subdimensions, based on findings from previous validations<sup>5-9</sup>. It was specified that the diagonally weighted least squares (DWLS) estimation method would be used<sup>18,19</sup>. The following fit indices were calculated: Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI), where values  $\geq 0.95$  are considered optimal; and Standardized Root Mean Squared Residual (SRMR) and Root Mean Square Error of Approximation (RMSEA), which are considered significant with values below 0.06 and 0.08, respectively<sup>20</sup>. Sample sizes of 250, 300, 350, and 400 were tested. The optimal sample size was defined based on the best fit observed in the simulations. It was found that a sample size of 250 already showed excellent fit (CFI = 0.98; TLI = 0.98; RMSEA = 0.015; SRMR = 0.053). Therefore, this was established as the minimum sample size, aiming for a larger sample to minimize the risks of non-convergence and the occurrence of Heywood cases.

### Variables

The main variables of the study were:

1. Burnout: A series of negative responses toward work resulting from prolonged exposure to a highly demanding work environment, characterized by emotional exhaustion, detachment or cynicism, and a diminished sense of personal accomplishment<sup>1</sup>.
2. Demographic Data: Sex, occupation, age, unit or service, type of employment, years of experience in the current occupation, length of service in the unit, type of contract, and average working hours per week.

A detailed description of all study variables is provided in Appendix 1.

### Measurement instruments

MBI-HSS: The Maslach Burnout Inventory - Human Services Survey consists of 22 items that assess professionals' feelings and attitudes toward their work and patients. The response format is an unipolar Likert scale with the following options: never (0), a few times a year or less (1), once a month or less (2), a few times a month (3), once a week (4), a few times a week (5), and every day (6). It includes three subscales:

1. Emotional Fatigue: 9 items assessing physical exhaustion, burnout, and loss of energy.
2. Depersonalization: 5 items evaluating negative aspects of behaviors and attitudes toward others (colleagues and patients).
3. Personal Accomplishment: 8 items measuring emotions such as depression, low morale, avoidance of interpersonal relationships, and low productivity.

In the version validated in Colombia, two items were removed, resulting in a 20-item version with the same three subscales. This version demonstrated intermediate to high internal consistency for the subscales and adequate overall internal consistency<sup>21</sup>.

### Data collection

Data were collected through an online questionnaire using Google Forms, with mandatory responses to avoid incomplete submissions. One of the researchers visited the care units and explained the purpose of the study in small groups, emphasizing the confidentiality of the information and addressing any concerns. Participants who agreed to take part were sent a link to the consent form, which, upon acceptance, automatically redirected them to the questionnaire. Additionally, after the informational sessions, the link was distributed via email and WhatsApp groups for each department.

## Data analysis

**Modification of response options.** A consensus among experts was reached using a modified Delphi methodology, applying the individual aggregation technique<sup>22</sup>. For this purpose, 10 professionals with over 5 years of experience in healthcare or health education were convened (5 medical specialists, 1 general practitioner, 1 specialist nurse, 1 psychometrician, and 2 specialist psychologists). The expert team and their profiles are presented in Appendix 2. This group was asked to review the new response options for the MBI-HSS, which were based on the study by Cañadas & Sánchez<sup>23</sup> on recommended response categories for Likert-type scales. Three modification proposals were sent for evaluation, and each expert rated them on a scale from 1 (Strongly disagree with these response options) to 5 (Strongly agree with these response options). The option with the highest average score among the judges was selected, achieving 100% agreement after consultation with all experts.

**Internal consistency.** Internal consistency was assessed using the inter-item correlation matrix and the correlation of each item with its domain (subscale). Additionally, Cronbach's alpha coefficient was calculated for each subscale of the test, including the alpha value when each item was removed. Cronbach's alpha increases as inter-item correlations grow<sup>24,25</sup>; desirable values for this coefficient range between 0.7 and 0.9<sup>26</sup>. In the inter-item correlation matrix, correlations between 0.2 and 0.5 are considered acceptable<sup>26</sup>. For item-total correlations, values between 0.3 and 0.7 are expected.

**Structural validity.** Given that the original structure of the instrument is known, a Confirmatory Factor Analysis (CFA) was conducted to test the theoretical model representing the original structure of the instrument using a SEM. The model tested considered the interrelationship between the three subscales, as proposed by the original authors. The following fit indices were calculated: Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Normed Fit Index (NFI), Relative Fit Index (RFI), Goodness of Fit Index (GFI), and Adjusted Goodness of Fit Index (AGFI). Values equal to or greater than 0.95 were considered indicators of good fit. Additionally, the Root Mean Square Error of Approximation (RMSEA) and Standardized Root Mean Squared Residual (SRMR) were calculated, with values below 0.06 and 0.08, respectively, indicating good fit<sup>20</sup>. The model was estimated using the diagonally weighted least squares (DWLS) method, which has shown better performance in confirmatory factor analyses with ordinal response items<sup>27</sup> and uses polychoric correlations for its calculation<sup>28,29</sup>.

**Item functioning.** Item functioning was evaluated using the generalized partial credit model, an extension of the Rasch model, which is used for analyzing items with polytomous response options<sup>30</sup>. Unidimensionality within each subscale was assumed, as each subscale theoretically assesses a different dimension of the overall construct.

Item discrimination (a) and difficulty thresholds for response options (b) were calculated. This analysis evaluates the probability of each response option being selected by participants based on their level of the measured attribute or trait<sup>31</sup>. Discrimination values between 0.65 and 2.5 indicate adequate discrimination<sup>30,32</sup>. Additionally, the functioning of response options was assessed using category characteristic curves.

Item Information Functions (IIFs) were plotted to evaluate the quality of the test items in terms of their ability to measure individuals' traits accurately<sup>31</sup>. Test Information Functions (TIFs) were also generated for each subscale, showing the relationship between individuals' trait levels and the amount of information provided by the subscale as a whole at that trait level<sup>31</sup>.

**Bias control.** Data collection was conducted by the lead researcher, who is specialized in the application of measurement instruments. Throughout the process, the anonymity and confidentiality of the information were emphasized to minimize biases related to social desirability or acquiescence.



**Table 1.** Correlations between items of the MBI-HSS, subscale by subscale (Spearman's Rho)

Emotional fatigue									
Items	1	2	3	6	8	13	14	16	20
1	1								
2	0.57	1							
3	0.49	0.48	1						
6	0.26	0.28	0.38	1					
8	0.64	0.49	0.53	0.39	1				
13	0.44	0.3	0.41	0.31	0.55	1			
14	0.54	0.48	0.48	0.33	0.6	0.41	1		
16	0.42	0.26	0.4	0.47	0.42	0.36	0.34	1	
20	0.45	0.38	0.49	0.36	0.54	0.49	0.4	0.4	1

Personal Accomplishment								
Items	4	7	9	12	17	18	19	21
4	1							
7	0.33	1						
9	0.21	0.36	1					
12	0.31	0.24	0.31	1				
17	0.24	0.32	0.35	0.24	1			
18	0.31	0.3	0.38	0.31	0.48	1		
19	0.3	0.35	0.33	0.35	0.22	0.35	1	
21	0.23	0.29	0.28	0.19	0.31	0.29	0.3	1

Depersonalization					
Items	5	10	11	15	22
5	1				
10	0.23	1			
11	0.12	0.38	1		
15	0.2	0.17	0.15	1	
22	0.22	0.19	0.34	0.14	1

**Ethical considerations.** This project was approved by the ethics and research committees of the participating hospital (Act 006-002, 07/28/2020) and the National School of Public Health at the Universidad de Antioquia (Act No. 228, 02/21/2020), in compliance with regulations for research involving human subjects. Informed consent was obtained and applied digitally. No cases of consent withdrawal occurred during the study.

## Results

### Sample description.

A total of 377 healthcare workers from the hospital completed the questionnaire. Of these, 78% were women and 22% were men, with a median age of 36 years (IQR= 29-44; min= 20; max= 66). Nursing assistants represented 41.6%, followed by other care professionals (22.3%), nurses (16.4%), general practitioners (12.7%), and specialists (6.9%). Regarding their work areas, 26.3% worked in hospitalization, 26.0% in intensive care, 20.4% in emergency care, 13.3% in surgery, 4.7% in imaging, and 9.3% in other units.

Regarding employment conditions, 65.7% were unionized, 24.2% had permanent contracts, and the rest varied between interns, service contracts, and mixed arrangements. The majority (67.0%) had more than 5 years of experience, and 32.0% had worked at the hospital for more than 5 years. The rest were distributed among those with 1-5 years of experience and those with less than one year.

### Modification of response options

The response options presented to the expert panel were as follows:

1. Never, almost never, sometimes, almost always, and always.
2. Never, almost never, rarely, sometimes, almost always, and always.
3. Never, rarely, occasionally, frequently, and always.

**Table 2.** Cronbach's alpha for the MBI-HSS subscales.

Items	Sign	Item-test corr.	Item-rest corr	Interitem corr	Alpha variation
<b>Emotional Fatigue (EF)</b>					
1	+	0.76	0.69	0.36	0.85
2	+	0.67	0.58	0.39	0.86
3	+	0.74	0.65	0.36	0.85
6	+	0.54	0.43	0.42	0.87
8	+	0.83	0.76	0.34	0.84
13	+	0.67	0.57	0.38	0.86
14	+	0.74	0.65	0.36	0.85
16	+	0.61	0.5	0.4	0.86
20	+	0.72	0.64	0.37	0.85
<b>Alpha subscale</b>					0.87
<b>Personal Accomplishment (PA)</b>					
4	+	0.56	0.43	0.24	0.75
7	+	0.61	0.47	0.22	0.75
9	+	0.68	0.54	0.21	0.74
12	+	0.54	0.39	0.23	0.76
17	+	0.63	0.48	0.22	0.75
18	+	0.72	0.59	0.2	0.73
19	+	0.62	0.49	0.22	0.74
21	+	0.58	0.41	0.22	0.76
<b>Alpha subscale</b>					0.77
<b>Depersonalization (D)</b>					
5	+	0.45	0.2	0.18	0.42
10	+	0.63	0.28	0.12	0.36
11	+	0.65	0.32	0.11	0.32
15	+	0.52	0.13	0.19	0.48
22	+	0.52	0.28	0.16	0.38
<b>Alpha subscale</b>					0.45

The first proposal received the highest average rating of 4.2, with 100% of the judges agreeing on this option. Thus, the new version was defined with a unipolar Likert-type response scale of 5 options, ranging from “never” to “always.”

### Internal consistency

Table 1 shows the inter-item correlations of the MBI-HSS. In the emotional fatigue subscale, correlations did not exceed 0.7. In the personal accomplishment subscale, most correlations ranged between 0.2 and 0.4, except for items 12 and 21, which had a correlation of 0.18. In the depersonalization subscale, several correlations were below 0.2, suggesting potential inconsistencies in the measurement of these items.

Table 2 details the item-total correlations, Cronbach's alpha for each subscale, and its variation when items were removed. The emotional fatigue subscale showed robust internal consistency with an alpha of 0.87 and moderate to high item-subscale correlations. The personal accomplishment subscale had an alpha of 0.77, indicating good consistency, and moderate item-subscale correlations. On the other hand, the depersonalization subscale had an alpha of 0.45, reflecting insufficient internal consistency. Although item-subscale correlations were moderate, inter-item correlations were low. Excluding item 15, which had the lowest correlation with other items in the subscale, increased the alpha, but it did not reach a satisfactory level of internal consistency.

### Structural validity

Figure 1 shows the results of the CFA applied to the MBI-HSS. Although the fit was acceptable, it was not considered optimal. The fit indices were: CFI= 0.964; TLI= 0.960; NFI= 0.913; RFI= 0.903; GFI= 0.951; AGFI= 0.940; RMSEA= 0.043 (90% CI: 0.071-0.085); SRMR= 0.071, and  $p= 0.000$ .



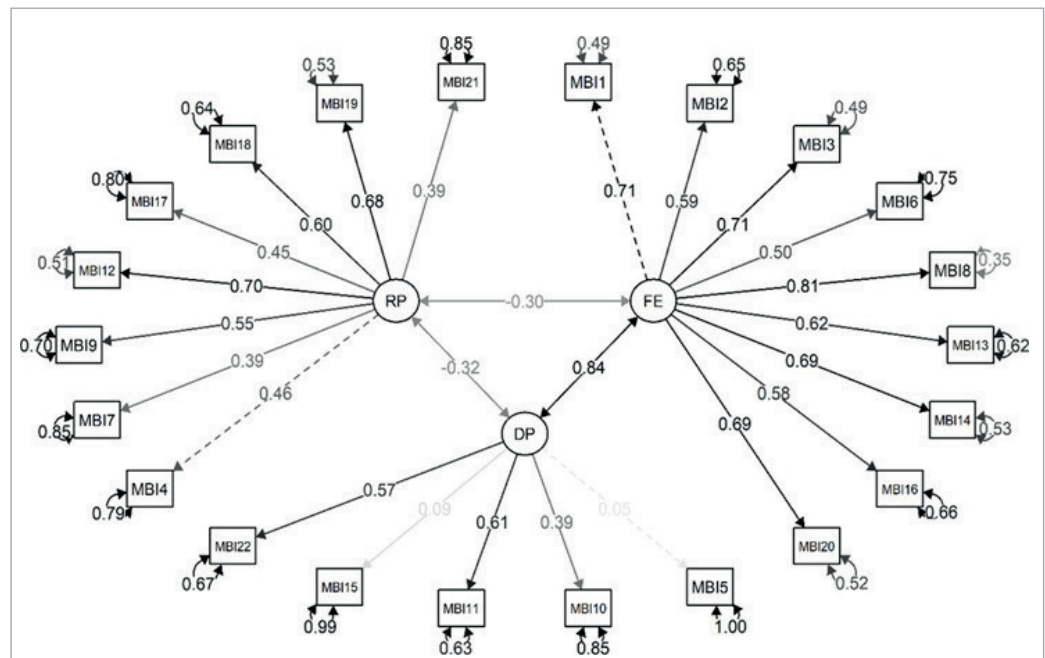


Figure 1. Structural equation model from the confirmatory factor analysis of the MBI-HSS

Figure 1 reveals a direct and strong relationship between emotional fatigue and depersonalization (0.84). In contrast, the relationship between emotional fatigue and personal accomplishment was inverse and weak (-0.30). The connection between personal accomplishment and depersonalization was indirect, weak, and non-significant ( $p \geq 0.05$ ), suggesting that the personal accomplishment subscale does not have a relevant correlation with the other two. It is important to note that the factor loading items in the depersonalization subscale ranged from very low to moderate. Items 5 and 15, in particular, were not significant and had unique measurement errors of 1.0 and 0.99, respectively, indicating that 0% and only 1% of their variance was due to the latent factor. This suggests that these items may not be reliable indicators of the depersonalization construct<sup>33</sup>.

Given these findings, an alternative model was explored, excluding the personal accomplishment subscale and items 5 and 15 from the depersonalization subscale (SEM Figure A in Appendix 3). This modification yielded superior fit indices: CFI= 0.999; TLI= 0.999; NFI= 0.982; RFI= 0.977; GFI= 0.989; AGFI= 0.984; RMSEA= 0.000 (90% CI: 0.000-0.033); SRMR= 0.048, and  $p=0.477$ . With this adjustment, the model demonstrated excellent fit, and all proposed relationships in the model were significant. Additionally, the relationship between the two remaining subscales was direct and very strong (0.84).

### Item functioning

Table 3 presents the results of item discrimination and difficulty thresholds for the response options in the emotional fatigue and depersonalization subscales. Most items showed discrimination values within the expected range, with the exception of item 8, which slightly exceeded the limit but did not significantly affect the subscale. However, certain items, such as 4 and 17, showed overlapping and reversed thresholds in their response options, which was more evident in the Item Characteristic Curves (Appendix 4 and and55).

In the emotional fatigue subscale, response thresholds tended to cluster at high trait values, with overlapping and reversed thresholds in items such as 6 and 16. The responses “almost never” and “almost always” were rarely selected, with “never” being the predominant option. On the other hand, in the personal accomplishment subscale, the most frequent responses

**Table 3.** Item discrimination and difficulty (thresholds) of response options for the MBI-HSS

Items	Discrimination (a)	b1 (2 vs 1)	b2 (3 vs 2)	b3 (4 vs 3)	b4 (5 vs 4)
<b>Emotional Fatigue</b>					
1	2.36	-0.79	-0.48	1.88	2.39
2	1.42	-1.59	-1.51	1.38	2.25
3	1.35	-0.1	0.38	2.04	3.15
6	0.71	1.57	1.52	4.73	1.53
8	2.71	-0.1	0.11	1.73	1.95
13	1.08	0.7	0.46	3.34	2.15
14	1.31	-0.08	-0.48	2.29	1.98
16	0.78	0.66	0.86	4.16	2.35
20	1.44	0.72	0.64	2.72	2.08
<b>Personal Accomplishment</b>					
4	1.03	-1.33	-3.74	-2.07	-0.19
7	1.08	-1.14	-2.18	-2.18	0.27
9	1.18	-1.19	-2.41	-1.32	0.22
12	0.84	-2.02	-3	-1.61	0.31
17	1.34	-0.81	-2.45	-1.47	0.61
18	1.65	-1.19	-2.25	-0.75	0.58
19	1.16	-2.14	-2.22	-1.84	-0.79
21	0.67	-0.57	-3.3	-1.74	0.45

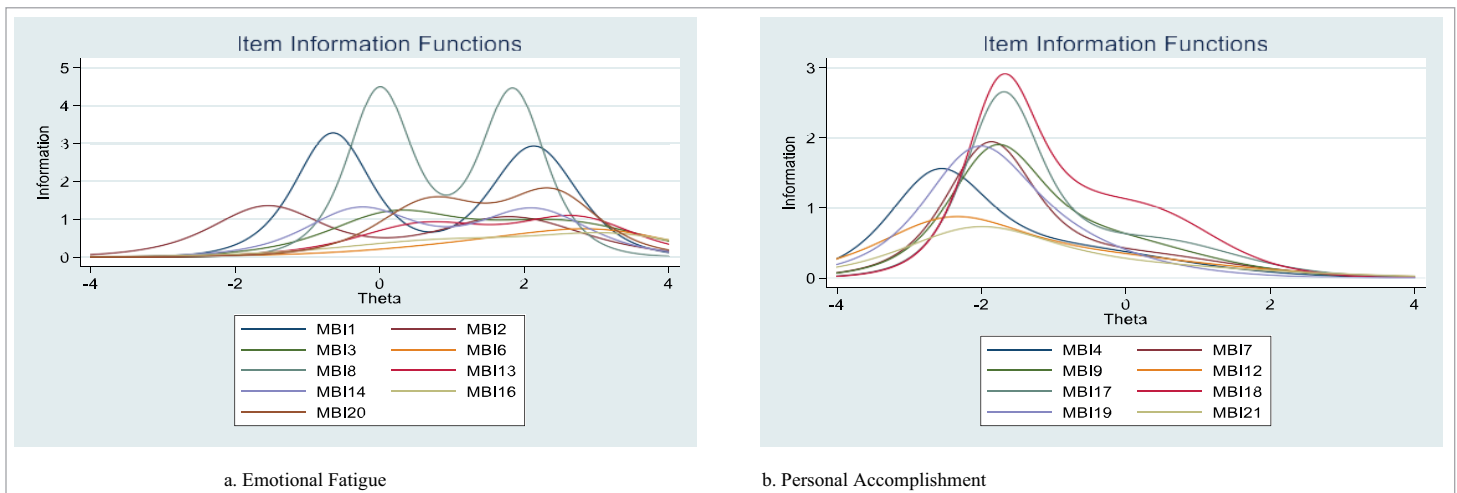
were “always” and “almost always.” The options “never,” “almost never,” and “sometimes” for items 4, 17, 19, and 21 showed overlaps without a clear gradient. The depersonalization subscale did not yield discrimination, and difficulty results due to convergence issues, possibly due to marked response tendencies. A frequency analysis revealed a preference for the options “never” and “almost never,” especially in items 5, 15, and 22, suggesting a measurement bias, possibly due to social desirability.

Figures 2 and 3 show the Item Information Functions (IIFs) and Test Information Functions (TIFs) for both subscales. In the emotional fatigue subscale, several items had low discrimination ability, showing better discrimination at high trait levels. In the personal accomplishment subscale, items tended to discriminate better at low trait levels. Both subscales lacked good discrimination at intermediate trait levels. The TIF figures support these findings, highlighting limitations in the reliability of the test.

## Discusión

The primary objective of this study was to determine the psychometric properties of the MBI-HSS after modifying its response options, addressing the suggestion made by the team that initially validated the instrument in Colombia <sup>21</sup>. With the support of expert judges, a unipolar Likert-type response scale with five options ranging from “never” to “always” was defined. These options are widely used in questionnaires where responses imply a frequency of occurrence.

When conducting validity and reliability analyses, it was found that while the fit indices were acceptable, they were not optimal. This, combined with the low factor loadings and high measurement errors in two items of the depersonalization subscale, as well as the weak and non-significant relationship of the personal accomplishment subscale with the other two subscales, suggests that the model does not adequately fit the data and may have deficiencies in the theoretical model's adequacy. This is further supported by the fact that testing an alternative model without the personal accomplishment subscale and without items 5 and 15 resulted in an excellent fit, indicating that the data represent an underlying model different from the one originally proposed.



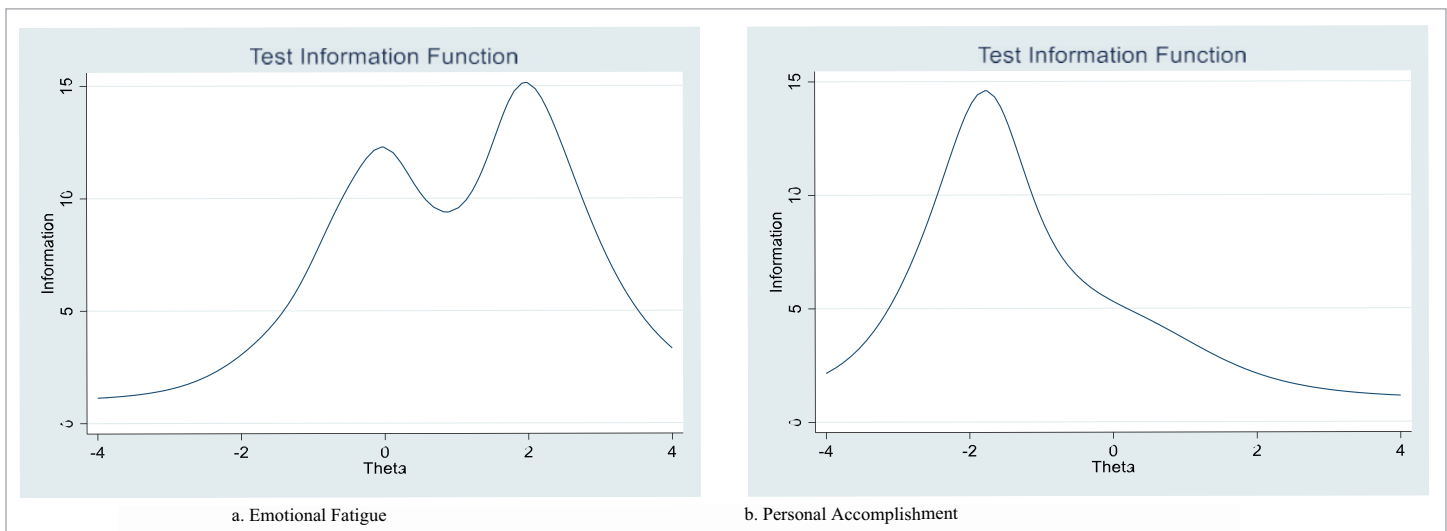
**Figure 2.** Item information functions for the Emotional Fatigue and Personal Accomplishment subscales of the MBI-HSS

On the other hand, internal consistency revealed differences in the performance of the subscales. While the emotional fatigue subscale showed a high alpha of 0.87, the personal accomplishment subscale had a moderate value of 0.77, and the depersonalization subscale had a low value of 0.45. The correlations between items in the latter subscale were also low, indicating variability in the functioning of the subscales. This suggests that the way depersonalization is measured may not be reliable in our context.

This issue was previously highlighted by Kristensen et al.<sup>11</sup>. In their study, they conducted a pilot test of this questionnaire with 70 healthcare workers. Participants noted the questions they found difficult to answer or had comments on. The results revealed that most negative comments were directed at questions related to depersonalization. Items 5 (“I feel I treat some patients as if they were objects”) and 15 (“I don’t really care what happens to some patients”) generated negative reactions, even anger, among participants, which seems to also be happening in our context, as these items show measurement bias. Additionally, some questions about personal accomplishment were criticized for being “too American,” such as “I feel I am positively influencing other people’s lives through my work” and “I have accomplished many worthwhile things in this job.” These findings align with the results of the study by Pando Moreno et al.<sup>12</sup> in Latin America, where the personal accomplishment subscale did not correlate with the other two and was placed in a separate dimension.

These concerns underscore the need to adapt the MBI-HSS to other languages, as the depersonalization and personal accomplishment subscales may present significant challenges due to cultural differences. For example, the perception of the doctor-patient relationship versus the doctor-client relationship may vary across cultures. Moreover, aspects considered important for a sense of personal accomplishment are also influenced by cultural context and may differ.

In the Colombian validation study, the CFA did not find a fit of the model to the original structure, so the researchers decided to conduct an exploratory factor analysis (EFA) and run a new model with a reorganization of items derived from the EFA. When performing a new CFA with this reorganization, they found an adequate but not optimal fit, which, combined with the low internal consistency of the depersonalization and personal accomplishment subscales, did not allow the research team to confidently assert that the measurement of burnout with this adapted version was entirely valid and reliable. Due to this, they recommended revising the response options, which they believed could be causing confusion among respondents due to their number, lack of familiarity in this culture, and some options not aligning well with



**Figure 3.** Test information functions for the Emotional Fatigue and Personal Accomplishment subscales of the MBI-HSS

the questions. For example, responding to “I easily understand how my patients feel” with a frequency scale that includes options like “Once a month or less” or “Once a week” does not seem coherent in our language and is not realistic for someone to recall if this happens a few times a year, once a month, or once a week <sup>13</sup>.

In addition to the validation in Colombia, other Latin American countries have also conducted validation exercises for the MBI-HSS. In Argentina, the authors modified the wording of some items: in the CFA with the original structure, they did not find an adequate fit, so they removed item 12, which improved the fit and internal consistency of the subscales, with values between 0.7 and 0.9 <sup>34</sup>. In Mexico, a validation study with psychologists found an adequate fit of the model with the original structure, and internal consistencies similar to ours: high for personal accomplishment (0.81) and emotional fatigue (0.86) and low for depersonalization (0.53) <sup>35</sup>. In the study conducted in Chile <sup>36</sup>, the authors found that items 12, 13, 14, 20, and 21 had low factor loadings or loaded onto parallel dimensions, so they decided to remove them and test the fit with a three-factor model excluding these items. With these modifications, they found an adequate but not optimal fit (CFI= 0.91, RMSEA= 0.063, GFI= 0.93, and AGFI= 0.91).

Similarly, in Peru, items 1, 16, and 21 were removed to achieve an adequate fit of the model to the proposed three factors (CFI= 0.98, RMSEA= 0.046) <sup>37</sup>. These authors highlight the fact that in different countries, some items are removed, the structure of the scales is modified, and yet the test is still considered valid and reliable. However, this compromises comparability with the original version and across multicenter studies. As they rightly state, “inferring the validity of the MBI-HSS compromises the ethics of the researcher and the inefficacy of using an instrument intended to identify a problem that requires clinical and psychosocial attention” <sup>37</sup>. This is a critical point, as evaluating something as significant as burnout with an instrument whose validity and reliability have not been optimally verified could affect measurement accuracy, generate systematic research errors, and lead to incorrect decision-making.

Since this study modified the response options of the MBI-HSS, it was important to evaluate the functioning of the new proposed options. Thus, item difficulty and discrimination analyses were conducted. It was found that in the depersonalization subscale, the skewness was so pronounced in some items (5, 15, and 22) that it was not possible to determine the discrimination and difficulty of these items, supporting the arguments of Kristensen et al. <sup>11</sup>. It seems that these types of questions generate such an emotional conflict that respondents either do not truly feel this way or find it difficult to accept that they might have reached a state where they see and treat their patients as “objects”.

The items in the emotional fatigue and personal accomplishment subscales showed adequate discrimination values but significant overlaps in response options and underused or skewed options in some items (6, 16, 4, 17, 19, and 21). Overall, the scores help to discriminate individuals with high levels of exhaustion but are less effective for those with moderate levels.

Given these results, we cannot confirm that the modifications to the response options in this study resolved the measurement issues of the test. Therefore, considering the findings from other countries, the MBI-HSS, as currently designed, is not recommended for assessing burnout in our healthcare population.

These results imply that our understanding and conceptualization of burnout may differ from that originally proposed by Maslach & Jackson, as evidenced in other cultures. This suggests the need to review and adapt not only the instrument but also the very concept of burnout for our population. These findings invite reflection on the widespread use of this instrument despite its deficiencies in Latin American contexts, which could lead to systematic biases in measurements used for monitoring or research on this construct in our region. It is a call to action for the appropriate selection of instruments, moving beyond those that are frequently adopted due to their classic status but may have issues in their adaptations in our context, even in the conceptualization of the construct being measured.

### Limitations and strengths

Although the sample was obtained through convenience sampling at a single institution, the findings reflect limitations of the MBI-HSS in the Colombian context, consistent with previous studies in the region. This suggests that the identified issues may not be exclusive to this sample, providing relevant evidence to reconsider its use in similar populations<sup>38</sup>.

As strengths, this study stands out from other structural validation studies of the MBI-HSS due to the use of the diagonally weighted least squares (DWLS) estimation method. This method has been shown to offer superior parameters when dealing with categorical variables, as it employs polychoric correlations instead of Pearson correlations for estimation<sup>29</sup>. When comparing the results of CFAs obtained with ML and MLM methods to those of this study, it is evident that the DWLS method provides optimal fit indices. This suggests that the estimates made with DWLS are more accurate and that the factor loadings or relationships identified between variables align better with the true nature of the data. Additionally, a preliminary simulation was conducted to determine the sample size, ensuring a sample with the necessary statistical power.

### Conclusions

Even after modifying the response options of the MBI-HSS, limitations in structural validity, reliability, item discrimination, and response option difficulty were identified in the validation conducted among healthcare workers in our country. If some of its subscales are to be applied, the one that could best measure exhaustion among healthcare personnel is the emotional fatigue subscale.

### References

1. Maslach C, Schaufeli WB, Leiter MP. Job burnout. *Annual review of psychology*. 2001;52(1):397-422. Doi: 10.1146/annurev.psych.52.1.397
2. Al-Ghunaim TA, Johnson J, Biyani CS, Alshahrani KM, Dunning A, O'Connor DB. Surgeon burnout, impact on patient safety and professionalism: A systematic review and meta-analysis. *Am J Surg*. 2022; 224(1, Part A): 228-238. Doi: 10.1016/j.amjsurg.2021.12.027.

3. Jun J, Ojemeni MM, Kalamani R, Tong J, Crecelius ML. Relationship between nurse burnout, patient and organizational outcomes: Systematic review. *Internat J Nursing Studies*. 2021; 119: 103933. Doi: 10.1016/j.ijnurstu.2021.103933
4. Maslach C, Jackson SE. The measurement of experienced burnout. *J Organizational Behavior*. 1981; 2(2): 99-113.
5. Gil-Monte PR. Factorial validity of the Maslach Burnout Inventory (MBI-HSS) among Spanish professionals. *Rev Saúde Pública*. 2005; 39(1): 1-8. doi: 10.1590/s0034-89102005000100001
6. Gil-Monte PR, Zuñiga-Caballero LC. Validez factorial del "Cuestionario para la Evaluación del Síndrome de Quemarse por el Trabajo" (CESQT) en una muestra de médicos mexicanos. 2010; 9(1): 169-178. Doi: 10.11144/Javeriana.upsy9-1.vfce
7. Portoghese I, Leiter MP, Maslach C, Galletta M, Porru F, D'Aloja E, et al. Measuring Burnout among university students: factorial validity, invariance, and latent profiles of the Italian version of the Maslach Burnout Inventory Student Survey (MBI-SS). *Frontiers Psychol*. 2018; 9: 2105. Doi: 10.3389/fpsyg.2018.02105
8. Matejic B, Milenovic M, Kisic Tepavcevic D, Simic D, Pekmezovic T, Worley JA. Psychometric properties of the Serbian version of the Maslach burnout inventory-human services survey: a validation study among anesthesiologists from Belgrade Teaching Hospitals. *Scient World J*. 2015; 2015: 903597. Doi: 10.1155/2015/903597
9. Khani MH, Mohammadi M, Anvari F, Farsi M. Students School Burnout Inventory: Development, Validation, and Reliability of Scale. *Internat J School Health*. 2018; 5(1): p. 1-6. Doi: 10.5812/intjsh.57594
10. Bria M, Spânu F, Baban A, Dumitrascu DL. Maslach burnout inventory-general survey: factorial validity and invariance among Romanian healthcare professionals. *Burn Res*. 2014;1(3):103-111. Doi: 10.1016/j.burn.2014.09.001
11. Kristensen TS, Borritz M, Villadsen E, Christensen KB. The Copenhagen Burnout Inventory: A new tool for the assessment of burnout. *Work Stress*. 2005;19(3):192-207. Doi: 10.1080/02678370500297720
12. Pando MM, Aranda BC, López PMdR. Validez factorial del Maslach Burnout Inventory-General Survey en ocho países latinoamericanos. *Ciencia Trabajo*. 2015; 17(52): 28-31. Doi: 10.4067/S0718-24492015000100006
13. Córdoba L, Tamayo JA, González MA, Martínez MI, Rosales A, Barbato SH. Adaptation and validation of the Maslach Burnout inventory-human services survey in Cali, Colombia. *Colomb Méd*. 2011;42(3):286-293. Doi: 10.25100/cm.v42i3.874
14. Ministerio de Salud. Resolución 2646: Por la cual se establecen disposiciones y se definen responsabilidades para la identificación, evaluación, prevención, intervención y monitoreo permanente de la exposición a factores de riesgo psicosocial en el trabajo y para la determinación del origen de las patologías causadas por el estrés ocupacional. Bogotá: Ministerio de Salud; 2008.
15. Congreso de Colombia. Ley 1562 Por la cual se modifica el Sistema de Riesgos Laborales y se dictan otras disposiciones en materia de Salud Ocupacional. Bogotá: Congreso de Colombia; 2012.
16. COSMIN. Checklist for Assessing study qualities. Accessed: agosto 7 2018; s.f. <https://www.cosmin.nl/tools/checklists-assessing-methodological-study-qualities/>
17. Mokkink L, Terwee C, Patrick D, Alonso J, Stratford P, Knol D. International consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient reported outcomes: results of the COSMIN study. *J Clin Epidemiol*. 2010; 63(7): 737-45. Doi: 10.1016/j.jclinepi.2010.02.006
18. Ortiz MS, Fernández-Pera M. Modelo de ecuaciones estructurales: una guía para ciencias médicas y ciencias de la salud. *Terapia Psicológica*. 2018; 36(1): 51-57. Doi: 10.4067/s0718-48082017000300047



19. Li C-H. The performance of ML, DWLS, and ULS estimation with robust corrections in structural equation models with ordinal variables. *Psychological Methods*. 2016; 21(3): 369. DOI: 10.1037/met0000093
20. Schreiber JB, Nora A, Stage FK, Barlow EA, King J. Reporting structural equation modeling and confirmatory factor analysis results: A review. *J Educational Res*. 2006; 99(6): 323-338. Doi: 10.3200/JOER.99.6.323-338
21. Guevara BLM, Ocampo AN. Propiedades psicométricas de confiabilidad y validez del Maslach Burnout Inventory-General Survey. *Rev Interamericana Psicol Ocupac*. 2016; 33(2): 128-142. DOI: 10.21772/ripo.v33n2a04
22. Escobar-Pérez J, Cuervo-Martínez A. Validez de contenido y juicio de expertos: una aproximación a su utilización. *Avances Medición*. 2008; 6: 27-36.
23. Cañadas I, Sánchez A. Categorías de respuesta en escalas tipo Likert. *Psicothema*. 1998; 10(3): 623-631.
24. Martínez R. *Psicometría: Teoría de los test psicológicos y educativos*. Síntesis; 1995.
25. Brown FG. *Principios de la medición en psicología y educación*. Manual Moderno; 1980.
26. De Vet HC, Terwee CB, Mokkink LB, Knol DL. *Measurement in medicine: a practical guide*. Cambridge University Press; 2011.
27. Rhemtulla M, Brosseau-Liard PÉ, Savalei V. When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*. 2012; 17(3): 354. Doi: 10.1037/a0029315
28. Freiberg HA, Stover JB, de la Iglesia G, Fernández LM. Correlaciones policóricas y tetracóricas en estudios factoriales exploratorios y confirmatorios. *Ciencias Psicológicas*. 2013; 7(2): 151-164.
29. Özdemir HF, Toraman Ç, Kutlu Ö. The use of polychoric and Pearson correlation matrices in the determination of construct validity of Likert type scales. *Turkish J Education*. 2019; 8(3): 180-195. Doi: 10.19128/turje.519235
30. de Ayala RJ. *The theory and practice of item response theory*. Guilford Publications; 2013.
31. Embretson SE, Reise SP. *Item response theory*. Psychology Press; 2013.
32. Baker FB. *The basics of item response theory*. ERIC; 2001.
33. Bollen KA, Noble MD. Structural equation models and the quantification of behavior. *Proceedings National Academy Sci*. 2011; 108(supplement 3): 15639-15646. Doi: 10.1073/pnas.1010661108
34. Gilla MA, Belén Giménez S, Moran VE, Olaz FO. Adaptación y validación del Inventario de Burnout de Maslach en profesionales argentinos de la salud mental. *Liberabit*. 2019;25(2):179-193. Doi: 10.24265/liberabit.2019.v25n2.04
35. Lara RMM, Jiménez BM, Muñoz AR, Benadero MEM, Viveros GRO. Análisis factorial confirmatorio del MBI-HSS en una muestra de psicólogos mexicanos. *Psicología Salud*. 2008;18(1):107-116. Doi: 10.25009/pys.v18i1.681
36. Olivares-Faúndez V, Mena-Miranda L, Jélvez-Wilker C, Macía-Sepúlveda F. Validez factorial del Maslach Burnout Inventory human services (MBI-HSS) en profesionales Chilenos. *Universitas Psychologica*. 2014;13(1):145-159. Doi: 10.11144/Javeriana.UPSY13-1.vfmb
37. Calderón-De la Cruz GA, Merino-Soto CS. Análisis de la estructura interna del Maslach Burnout Inventory (Human Service Survey) en médicos peruanos. *Rev Ciencias Salud*. 2020; 18(2): 1-17. Doi: 10.12804/revistas.urosario.edu.co/revsalud/a.9275
38. Babbie ER. *The practice of social research*. Cengage learning; 2020.

## Appendix 1.

### Study variables

Variable	Definition	Level of Measurement	Operational Level
Sex	Biological and physiological characteristics that define males and females.	Qualitative l	1= Female
Age	The length of time a person or other living being has lived since birth.	Quantitative discrete	Years
Age ranges	Age ranges defined according to the WHO classification.	Qualitative l ordinal	1= 18 to 26 2= 27 to 59 3= 60 or more
Care unit	The hospital care unit where the healthcare worker is employed.	Qualitative nominal	1= Hospitalization 2= Hospitalization, gynecology, and delivery room 3= Surgery 4= Surgery 5= Intensive Care Unit (ICU) 6= Other
Occupation	The specific occupational activity performed by the workers at the institution, for which they received formal training.	Qualitative nominal	1= Nursing assistant 2= Nurse 3= General practitioner 4= Specialist physician 5= Other care professions
Weekly working hours	The average weekly time, in hours, that the worker dedicates to their job at the HUN or other institutions.	Qualitative ordinal	1= Less than 40 2= 40 to 59 3= 60 to 79 4= 80 or more
Type of contract	The type of employment contract under which the worker performs their duties in the institution.	Qualitative nominal	1= Unionized 2= Service provision 3= Permanent contract
Work experience	The length of time the worker has been employed in their current occupation since graduation.	Qualitative ordinal	1= Less than 6 months 2= Between 6 months and 1 year 3= Between 1 and 3 years 4= Between 3 and 5 years 5= More than 5 years 4= Mixed contract 5= Intern/Resident
Experience in the unit	The length of time the worker has been employed in their current occupation within the unit where they work at the hospital.	Qualitative ordinal	1= Less than 6 months 2= Between 6 months and 1 year 3= Between 1 and 3 years 4= Between 3 and 5 years 5= More than 5 years
Works at another institution	The employment situation of the worker, indicating whether they also work at another healthcare institution in addition to the hospital.	Qualitative nominal	0= No 1= Yes
MBI-HSS items scores	The responses given by participants to each item of the Maslach Burnout Inventory - Human Services Survey (MBI-HSS).	Qualitative ordinal	To be defined, they will be modified in this study.

## Appendix 2.

### Specialist Physicians

#### •Luz María Gómez

Anesthesiologist

Over 15 years of experience

Medical Advisory Submanager, Colombian Society of Anesthesiology and Resuscitation (S.C.A.R.E.)

#### •Darling Carvajal Duque

Specialist in Critical Care Medicine and Pediatric Intensive Care

Over 15 years of experience

Coordinator, Pediatric Intensive Care Unit, Hernando Moncaleano Perdomo University Hospital, Neiva

#### •Miguel Andrés Bayona Ospina

Pediatric Palliative Care Specialist

Over 10 years of experience

Pediatric Palliative Care Specialist, Hernando Moncaleano Perdomo University Hospital, Neiva

#### •Juan Pablo Zapata Ospina

Physician and Surgeon, Specialist in Psychiatry, MSc in Clinical Epidemiology

Over 10 years of experience

Medical Studies Committee, Alma Mater Hospital de Antioquia

#### •Néstor Daniel Ramírez Borrero

Physician - Master's in Bioethics

Over 15 years of experience

Clinical Bioethicist, Hernando Moncaleano Perdomo University Hospital, Neiva

General Practitioner:

#### •Diana Carolina Caicedo Sánchez

Over 5 years of experience

General Practitioner, Adult Intensive Care Unit, Bucaramanga Emergency Clinic

Specialist Nurse:

#### •Aura María Díaz Arguello

Over 10 years of experience

Head Nurse, Oncology Specialist, Head of Surgery Rooms, National Cancer Institute

Psychometrician:

#### •Claudia Marcela Vélez

Over 15 years of experience

Physician, Specialist, MSc, PhD in Health Policy

Professor, Faculty of Medicine, University of Antioquia

Specialist Psychologists:

•Yudy del Pilar González Gama

Over 5 years of experience

Specialist Psychologist

Professional Verifier, Subdirector of Inspection, Surveillance, and Control of Health

Services - Quality Subdirector

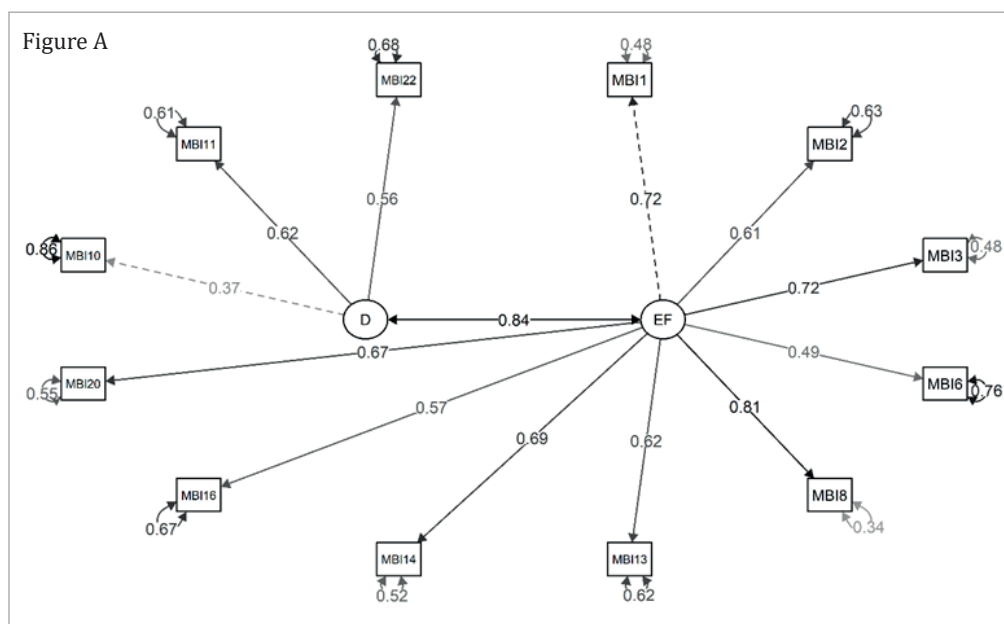
•Freddy Alejandro Barrero G.

Over 5 years of experience

Specialist Psychologist

Specialized Professional - Quality and Humanization Area

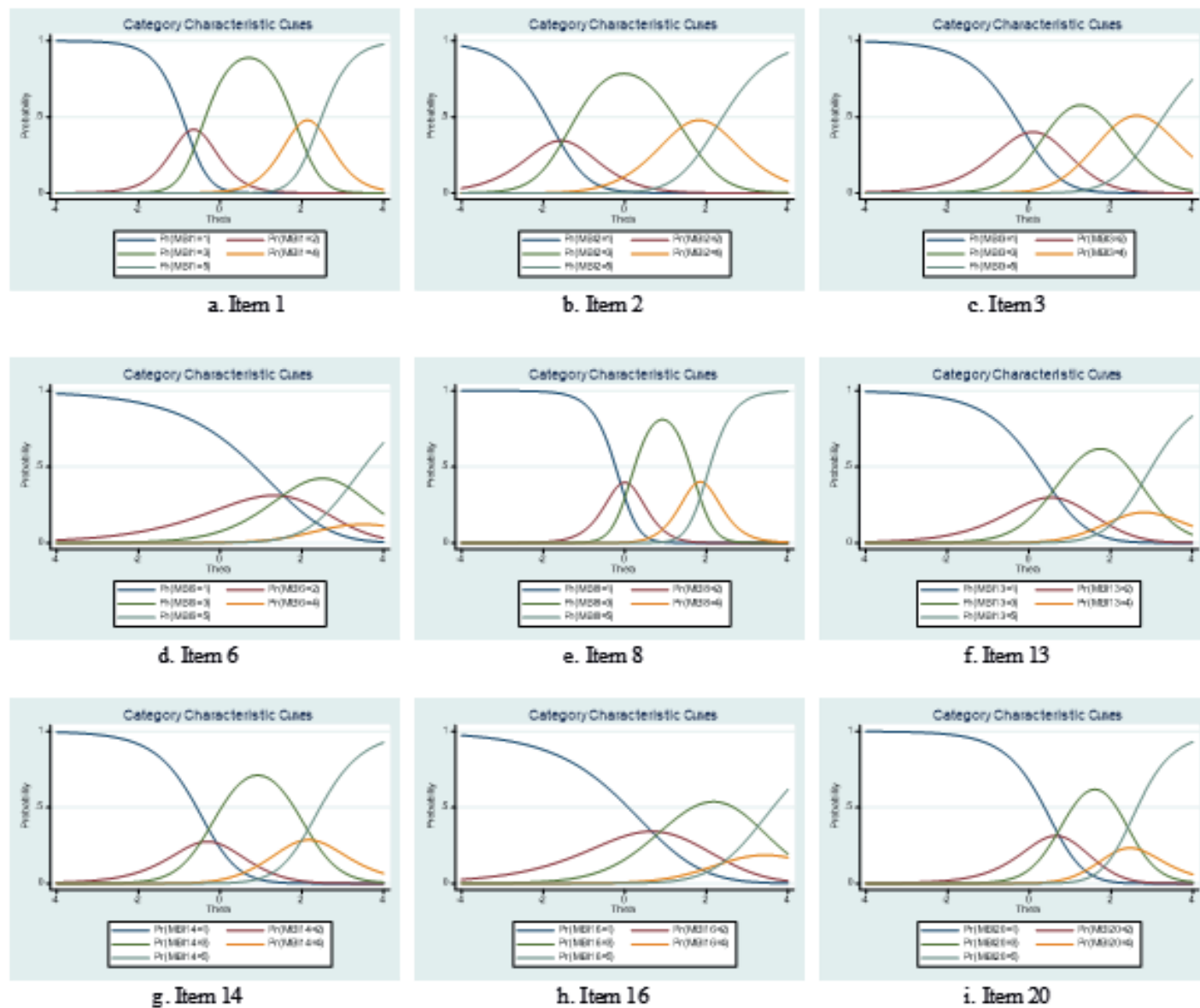
### Appendix 3.



Alternative structural equation model for the MBI-HSS

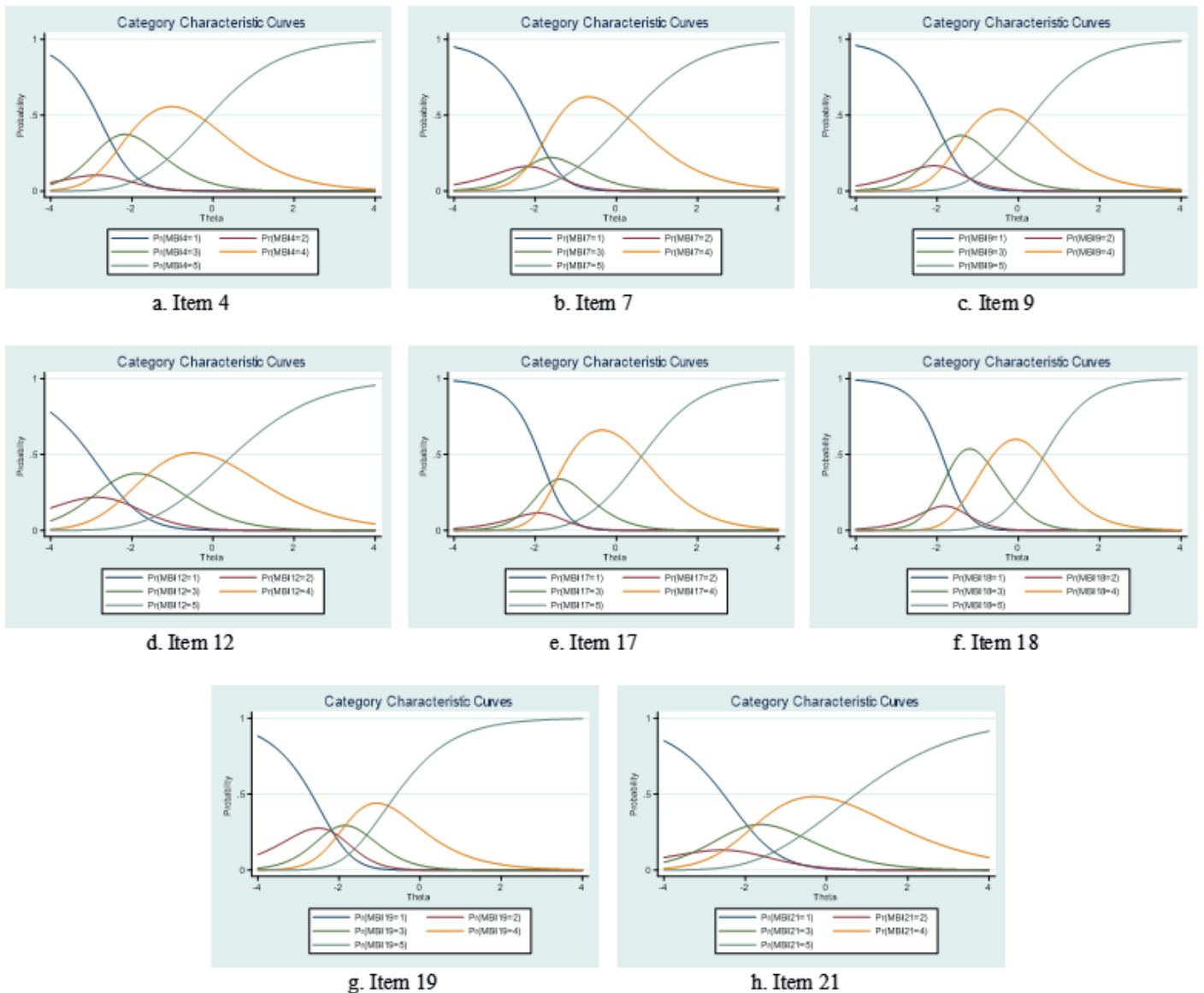
## Appendix 4.

Figure B



Item characteristic curves for the Emotional Fatigue subscale.

Figure B



Item characteristic curves for the Personal Accomplishment subscale.