

RESEARCH ARTICLE (ORIGINAL) 8

Psychometric properties of the 28-Item General Health Questionnaire in nurses: a proposal with 24 items

Propriedades psicométricas do Questionário Geral de Saúde 28 itens em enfermeiros: proposta com 24 itens

Propiedades psicométricas del Cuestionario de Salud General de 28 ítems en enfermeros: propuesta con 24 ítems

Paulo Seabra^{1,2}
 <https://orcid.org/0000-0001-8296-1021>
Manuel Luís Capelas³
 <https://orcid.org/0000-0002-8993-6854>
Joaquim Oliveira Lopes^{1,2}
 <https://orcid.org/0000-0003-2571-7078>
Mariana Calado⁴
 <https://orcid.org/0000-0003-3409-6058>
Ezequiel Pessoa¹
 <https://orcid.org/0000-0002-0756-2693>
¹ Lisbon Nursing School, Lisbon, Portugal² Nursing Research, Innovation and Development Centre of Lisbon (CIDNUR), Lisbon, Portugal³ Universidade Católica Portuguesa, Health Sciences Institute, Center for Interdisciplinary Research in Health, Lisbon, Portugal⁴ Hospital CUF Descobertas, Lisbon, Portugal**Corresponding author**

Paulo Seabra

E-mail: pauloseabra@escl.pt

Received: 12.08.20

Accepted: 27.02.21

Abstract

Background: The psychometric properties of measurement instruments should be assessed in different populations.

Objective: To analyze the psychometric properties of the 28-Item General Health Questionnaire (GHQ-28) in nurses.

Methodology: Methodological validation study of measurement instruments. Analysis of psychometric properties, exploratory factor analysis (EFA) by the principal component analysis method, and Receiver Operating Characteristic (ROC) curve, with a nonprobability sample of 1,264 nurses.

Results: EFA suggests keeping four factors and deleting four items in 3 dimensions (items 3, 16, 21, and 26). The variance explained by the 4 factors was 61.5%, and Cronbach's alpha was 0.93. The general health scores did not change significantly after deleting the 4 items. The comparison between the 24-item version and the 28-item version revealed an area under the curve (AUC) of 0.996 ($p < 0.001$) and a cutoff point of 20.5, with a 96.3% sensitivity and a 98.4% specificity.

Conclusion: The reassessment of the psychometric properties of GHQ-28 in nurses suggests a reduction to 24 items.

Keywords: psychometrics; general health questionnaire; mental health; nursing

Resumo

Enquadramento: A avaliação das propriedades psicométricas dos instrumentos de medida, quando aplicados em diferentes populações, é essencial.

Objetivo: Analisar as propriedades psicométricas do Questionário de Saúde Geral 28 itens (GHQ28) em enfermeiros.

Metodologia: Estudo metodológico de validação de instrumentos de medida. Análise das propriedades psicométricas, análise fatorial exploratória (AFE) pelo método dos componentes principais, e curva *Receiver Operating Characteristic* (ROC), com uma amostra não probabilística de 1264 enfermeiros.

Resultados: A AFE propõe a manutenção de 4 fatores e eliminação de 4 itens em 3 dimensões (item 3, 16, 21, 26). A variância explicada pelos 4 fatores foi 61,5% e o alfa de Cronbach 0,93. Os resultados de saúde geral não sofreram variabilidade significativa após se retirarem os 4 itens. A curva ROC, comparando a versão 24 itens com a de 28, revelou uma área sob a curva (AUC) de 0,996 ($p < 0,001$) e ponto de corte 20,5, com uma sensibilidade de 96,3% e especificidade de 98,4%.

Conclusão: A reavaliação das propriedades psicométricas do GHQ28 em enfermeiros, sugere redução para 24 itens.

Palavras-chave: psicometria; questionário geral de saúde; saúde mental; enfermagem

Resumen

Marco contextual: La evaluación de las propiedades psicométricas de los instrumentos de medición, cuando se aplican en diferentes poblaciones, es esencial.

Objetivo: Analizar las propiedades psicométricas del Cuestionario de Salud General de 28 ítems (GHQ28) en enfermeros.

Metodología: Estudio metodológico sobre la validación de instrumentos de medida. Análisis de las propiedades psicométricas, análisis factorial exploratorio (AFE) mediante el método de componentes principales y curva Receiver Operating Characteristic (ROC), con una muestra no probabilística de 1264 enfermeros.

Resultados: La AFE propuso el mantenimiento de 4 factores y la eliminación de 4 ítems en 3 dimensiones (ítem 3, 16, 21, 26). La varianza explicada por los 4 factores fue del 61,5% y el alfa de Cronbach del 0,93. Los resultados de salud general no experimentaron una variabilidad significativa tras eliminar los 4 ítems. La curva ROC, que compara la versión de 24 ítems con la de 28, mostró un área bajo la curva (AUC) de 0,996 ($p < 0,001$) y un punto de corte de 20,5, con una sensibilidad del 96,3% y una especificidad del 98,4%.

Conclusión: La reevaluación de las propiedades psicométricas del GHQ28 en enfermeros sugiere una reducción a 24 ítems.

Palabras clave: psicometría; cuestionario de salud general; personal de salud; salud mental; enfermería



How to cite this article: Seabra, P., Capelas, M. L., Lopes, J. O., Calado, M., & Pessoa, E. (2021). Psychometric properties of the 28-Item General Health Questionnaire in nurses: a proposal with 24 items. *Revista de Enfermagem Referência*, 5(6), e20136. <https://doi.org/10.12707/RV20136>



Introduction

Health assessment offers a wide variety of possibilities based on the several needs associated with the multidimensionality and complexity of the concept. One of the dimensions is mental health, with subjective factors that are difficult to monitor. Mental health can be defined as a “state of well-being in which the individual realizes his or her own abilities, can cope with the normal stresses of life, can work productively and fruitfully, and is able to make a contribution to his or her community” (World Health Organization [WHO], 2007, p. 1), that is, a state that is not contrary to that of mental illness but rather influenced by day-to-day factors, and contributes to coping with stressful situations.

In Portugal, studies with small samples of health professionals and difficult to compare due to the use of different instruments have shown high levels of burnout (25%-60% of nurses; Marôco et al., 2016) and mental suffering (15%-61.5% suffer from anxiety, and 14%-18% suffer from depression; Gomes & Oliveira, 2013). Therefore, many instruments have been developed to assess mental health or clarify a possible diagnosis of mental illness (Pais-Ribeiro, 2011; Ali et al., 2016). These instruments become internationally accepted for cross-sectional indicators of specific symptomatology, but not necessarily pathological symptomatology, such as the 28-Item General Health Questionnaire (GHQ-28). Jackson (2007), as cited in Pais-Ribeiro et al. (2015), reported that the GHQ-28 is one of the most widely used mental health screening questionnaires, particularly for its suitability to any type of population. However, it is still little used with health professionals. Thus, this study aims to analyze the psychometric properties of GHQ-28 in Portuguese nurses.

Background

GHQ-28 was developed by Goldberg and Hillier (1979) to assess the ability to carry out some activities and cope with stressful phenomena. It is not intended to assess stable traits but rather changes in functioning. Therefore, it is an instrument to assess mental health and psychological well-being (Pais-Ribeiro & Antunes, 2003). It consists of four subscales (Somatic symptoms, Anxiety and insomnia, Social dysfunction, and Severe depression), with seven items each. The items are rated on a 4-point (0-3) Likert-type scale. Through the sum of items, the scores for the subscales are 0-21 and the scores of the total scale are 0-84. Higher scores indicate worse mental health. Concerning the cutoff points, validation studies in Portugal have found cutoff points of 4/5 for the subscales and 23/24 for the total scale, with higher scores indicating the need for the person to be clinically assessed by a health professional. Its dimensions refer to symptomatology rather than to diagnoses, although the scores may be relevant and independent in each subscale (Pais-Ribeiro et al., 2015).

GHQ-28 is a scale used internationally, and it currently has shorter versions (12 or 20 items). Several studies have been conducted in Portugal for its validation in different

populations. In 2003, a study conducted with 30 people with tuberculosis and 30 healthy people obtained a Cronbach's alpha of 0.94 (Pais-Ribeiro & Antunes, 2003). In 2011, a study with 171 people in a non-clinical situation obtained a Cronbach's alpha of 0.92 (Monteiro, 2011). In 2015, a study with 384 people who used general health services and social services obtained a Cronbach's alpha of 0.94 (Pais-Ribeiro et al., 2015). The 12-item version was validated for Portugal with a population of 790 secondary school students, obtaining an internal consistency of 0.91 (Laranjeira, 2008). The reliability values are acceptable in all studies (Marôco, 2018).

The studies for validation of the scale in health professionals are still scarce. In India, a study with 448 health professionals found that 41.1% of them had signs of impaired mental health (Vinod et al., 2017), although no measure of internal consistency was indicated in the study. In Japan, a study with professionals who emigrated from Indonesia found an internal consistency of 0.89 (Sato et al., 2016), with 22.5% of them having a lower perception of their mental health.

In a recent study, the GHQ-28 was applied to a population of nurses from the national care network to assess its reliability (internal consistency of 0.94) and construct validity (KMO = 0.946 and Bartlett's test of sphericity with $p < 0.001$). The study found that 60.6% of the participants had a negative perception of their mental health, 71.6% perceived somatic symptoms, 76% had significant anxiety, 94.1% had some form of social dysfunction, and 22.2% had symptoms of severe depression (Seabra et al., 2019). Due to the lack of studies in Portugal on this topic, the psychometric properties and structure of this scale should be further explored (Marôco, 2018) to contribute to its adaptation to this population.

Methodology

A methodological study of validation of measurement instruments (Marôco, 2018; Hair et al., 2009) was conducted with a convenience sample of 1,264 nurses from a universe of 69,486 nurses enrolled in the *Ordem dos Enfermeiros* (Portuguese Nursing and Midwifery Regulator, OE). Between April and July 2017, the OE made a link available on its website. Nurses were asked to reply by email. Using this link, participants could complete a self-applied questionnaire for sociodemographic, professional, and occupational characterization and the GHQ-28 to assess their mental health (Pais-Ribeiro & Antunes, 2003). The need to fill out the GHQ in full reduced the non-response rate.

Data were then extracted from the database in Excel format for Windows. Statistical analysis was performed using IBM SPSS Statistics software, version 25 (SPSS Inc., Chicago, IL, USA) for Windows. The item relational structure was analyzed using an exploratory factor analysis (EFA) on the correlation matrix, with extraction of the factors through the principal components method, followed by Varimax rotation, which is a similar method to that used in previous studies with the same instrument (Pais-Ribeiro et al., 2015). The validity of the EFA was assessed using correlation and partial correlation matrices, Kaiser-Meyer-Olkin measure (KMO),

and Bartlett's test of sphericity. The following mean KMO values of model adequacy were used: medium between 0.7 and 0.8, good between 0.8 and 0.9, and excellent between 0.9 and 1.0. The p -value in Bartlett's test of sphericity should be < 0.05 (Marôco, 2018; Hair et al., 2009). The common factors retained were those with an eigenvalue greater than 1, which is in line with the scree plot and the percentage of variance retained (Marôco, 2018). Regarding factor loadings, in case of cross-loading in more than one factor, the minimum loading should be greater than 0.3 (for samples larger than 350 participants), and the difference between them should be greater than 0.1 (Marôco, 2018), which was associated with a qualitative evaluation when it met these criteria and loaded on more than one factor. The scale's internal consistency was assessed using Cronbach's alpha, which indicates the extent to which the several items in the instrument, measuring the same construct, produce similar scores (Oliveira, 2014). The psychometric and diagnostic sensitivity was also assessed as suggested by Marôco (2018). The ROC curve and the Youden Index were used to determine the cutoff points and the sensitivity and specificity values of the subscales and the total scale, using the comparative evaluation with GHQ-28 as the only standard.

Authorization was requested and obtained to use the GHQ-28 from the authors of the validation of the scale for the Portuguese population. The study was approved by the Ethics Committee of the Health Sciences Institute of the *Universidade Católica Portuguesa* (Opinion no. 23/2017) and is already published (Seabra et al., 2019). This study makes secondary use of the data.

Results

The sample's sociodemographic characterization revealed that 83.2% (1052) of the participants were women, with a median age of 36 years (IQR = 14), a minimum age of 22, and a maximum age of 64 years; 87.7% (1108) of them lived with someone, and 55.0% (695) had people dependent

on them. At a professional level, 60.0% (759) worked in hospitals and 22.2% (280) in primary health care; 58.5% (740) worked in shifts, with a median of 40 (IQR = 15) working hours per week; the mean length of professional experience was 15.2 (± 9.5) years.

EFA was preceded by the analysis of the factorability of the GHQ-28 correlation matrix. Visual inspection of the correlation matrix revealed a substantial number of correlations higher than 0.30, suggesting an item cluster structure. The analysis of partial correlations (anti-image matrix), with values lower than 0.7, suggested that the items, individually, can be explained by the other items on the factor (Hair et al., 2009). Bartlett's test of sphericity revealed statistical significance for the correlation matrix ($p < 0.001$) and a KMO value of 0.946 that indicates an excellent factorability of the correlation matrix, according to the criteria defined by Marôco (2018). Thus, the validity of the EFA for the correlation matrix of the GHQ-28 was confirmed. Then, the EFA was performed using the principal components method, followed by a Varimax rotation to extract the factors. The decision on the number of factors to be extracted was made based on the eigenvalues-greater-than-one rule and the scree plot (Marôco, 2018), establishing a relational structure of the GHQ-28 items explained by four latent factors.

The analysis of the factor matrix showed that items 3, 16, 21, and 26 had high factor loadings (item-factor correlation), almost overlapping, on more than one factor and with differences < 0.10 between them (Table 1). According to Marôco (2018), this situation can occur if an item is explained simultaneously by more than one factor, which does not contribute to the orthogonality (non-correlation) of the scale's dimensions. In this case, the deletion of these items is considered. Before this decision was taken, additional analysis was carried out with Quartimax and Equimax orthogonal rotation methods to define a simple structure in which each variable was associated with a single factor (Hair et al., 2009). However, the analysis with these methods did not provide a solution to this key problem.

Table 1

Rotated factor matrix based on 28 items (summary)

ITEMS	Factor loadings				KMO
	Factor 1	Factor 2	Factor 3	Factor 4	
GHQ (28 items included)					0.946
3 – Have you been feeling run down and out of sort?	0.501			0.508	
16 – Have you been taking longer over the things you do?			0.494	0.437	
21 – Have you been able to enjoy your normal day-to-day activities?	0.405		0.476		
26 – Have you found at times you couldn't do anything because your nerves	0.401	0.491			
Eigenvalue	11.122	2.557	1.930	1.105	
Explained variance	39.7%	9.1%	6.9%	3.9%	
Explained cumulative variance	39.7%	48.8%	55.7%	59.7%	

Note. GHQ = General Health Questionnaire; KMO = Kaiser-Meyer-Olkin. The factor loadings are in bold, keeping the item-factor correlation values above 0.32 (Tabachnick & Fidel, 2007).

Then, each item was deleted successively (3, 16, 21, and 26, in this order), and each factor matrix was assessed in each step. The results of the EFA with items 27, 26, and 25 continued to show high cross-loadings, with differences < 0.10, for which reason these items were deleted from the scale. Subsequently, the analysis of the factorability of the correlation matrix of the GHQ with 24 items revealed a Bartlett's test of sphericity with statistical significance ($p < 0.001$) and a KMO = 0.931, which indicates an excellent

factorability of the correlation matrix (Marôco, 2018) and confirms the validity of the EFA for the item correlation matrix. Then, the EFA was conducted by the principal components method, followed by Varimax rotation to extract the factors. The most stable relational structure of the scale with 24 items was explained by four latent factors, which were extracted based on the factor loadings of each item, the communalities, the eigenvalue > 1, the covariances, and the scree plot (Marôco, 2018; Table 2).

Table 2

Rotated factor matrix based on 24 items

ITEMS	Factor loadings				h ²
	Factor 1	Factor 2	Factor 3	Factor 4	
1- Have you been feeling perfectly well and in good health?	0.366		0.320	0.519	0.513
2- Have you been feeling in need of a good tonic?	0.306			0.406	0.337
4- Have you felt that you are ill?	0.414			0.563	0.612
5- Have you been getting any pains in your head?				0.816	0.733
6- Have you been getting a feeling of tightness or pressure in your head?				0.797	0.723
7- Have you been having hot or cold spells?				0.479	0.377
8- Have you lost much sleep over worry?	0.663			0.331	0.587
9- After falling asleep wake up several times?	0.557			0.365	0.457
10- Have you felt constantly under strain?	0.771				0.700
11- Have you been getting edgy and bad-tempered?	0.701				0.643
12- Have you been getting scared or panicky without a good reason?	0.597	0.337			0.536
13- Have you found getting everything on top of you?	0.745				0.684
14- Have you been feeling nervous and strung-up all the time?	0.766				0.743
15- Have you been managing to keep yourself busy and occupied?			0.535		0.394
17- Have you felt on the whole you were doing things well?			0.773		0.638
18- Have you been satisfied with the way you've carried out your task?			0.777		0.668
19- Have you felt that you are playing a useful part in things?			0.751		0.607
20- Have you felt capable of making decisions about things?			0.733		0.603
22- Have you been thinking of yourself as worthless person?		0.592	0.343		0.591
23- Have you felt that life is entirely hopeless?		0.772			0.720
24- Have you felt that life isn't worth living?		0.797			0.740
25- Have you thought of the possibility that you might end your life?		0.824			0.695
27- Have you found yourself wishing you were dead and away from it all?		0.803			0.722
28- Have you found that the idea of taking your own life kept coming into your mind?		0.847			0.743
Eigenvalue	9.391	2.451	1.861	1.062	
Explained variance	39.1%	10.2 %	7.8 %	4.4 %	
Explained cumulative variance	39.1%	49.3%	57.1%	61.5%	

Note. h² = Communalities. The factor loadings are in bold, keeping the item-factor correlation values above 0.32 (Tabachnick & Fidel, 2007).

After the deletion of the four items, the factor matrix minimizes the number of significant factor loadings on each line, causing each variable to be associated with a single factor. This matrix shows the structure of latent factors (item-factor correlations), which is widely known and disseminated in the literature: the first factor is related to Anxiety and insomnia (keeping seven original items), the second factor to Severe depression (it now has six items), the third factor to Social dysfunction (it now has five items), and the fourth factor to Somatic symptoms (it now has six items). The retained factorial model reproduces well the correlational structure observed based on the criterion of commonalities (Hair et al., 2009) and the residual matrix (Marôco, 2018). The commonalities greater than 0.3 demonstrate that the four retained factors are appropriate for describing the latent correlational structure between the items. In turn, the residual matrix

in the model, which results from the difference between the matrix of correlations observed and the matrix of correlations estimated by the model, reveals 79% of residual absolute values below 0.05. A high percentage of residual values (>50%) below 0.05 is an indicator of a factorial model with good fit (Marôco, 2018).

The assessment of the scale's reliability (with 24 items) revealed a Cronbach's alpha of 0.93 for the total scale and the following coefficients for the subscales: Somatic symptoms $\alpha = 0.820$; Anxiety and insomnia $\alpha = 0.890$; Social dysfunction $\alpha = 0.810$; Severe depression $\alpha = 0.900$. These coefficients suggest homogeneity of the items in each of the scale's dimensions. In our proposal (with 24 items), the deletion of items did not have a significant impact on the reliability of the total scale, compared to the original version (28 items), or its subscales, three of them now with fewer items (Table 3).

Table 3

Comparison of the internal consistency coefficients between GHQ-28 and GHQ-24

	Internal consistency (Cronbach's alpha)				
	Total	Somatic symptoms	Anxiety and insomnia	Social dysfunction	Severe depression
GHQ-28	0.940	0.850	0.890	0.842	0.900
GHQ-24	0.930	0.820	0.890	0.810	0.900

Note. GHQ-28 = 28-item General Health Questionnaire; GHQ-24 = 24-item General Health Questionnaire.

Another step that reinforces the analysis of the scale's structure is the inter-factor correlation, which must be < 0.70

and statistically significant, demonstrating its independence. Table 4 shows Spearman's correlation coefficients.

Table 4

Correlation matrix of the factors on the GHQ-24

Factors	Somatic symptoms	Anxiety and insomnia	Social dysfunction	Severe depression
Somatic symptoms	1.000			
Anxiety	0.717*	1.000		
Social dysfunction	0.485*	0.505*	1.000	
Depression	0.428*	0.501*	0.342*	1.000

Nota. * Correlation is significant at 0.01 (2 extremities); GHQ = General Health Questionnaire.

The psychometric sensitivity of the items was also assessed through shape measures (Marôco, 2018). Skewness and kurtosis coefficients were 2.191 and 3.959, respectively, which, according to Kline (2011), indicates that the items can discriminate individuals who are structurally different. Regarding the assessment of the diagnostic/clinical sensitivity, the GHQ-24 maintains the ability to discriminate statistically significant differences in perceived mental health. For example, regarding gender (*men*: $m_e = 20.00$, IQR = 14.75; *women*: $m_e = 24$, IQR = 16; $U = 131039.50$; $p \leq 0.001$) and sports/hobbies (*yes*: $m_e = 21$,

IQR = 14; *no*: $m_e = 26$, IQR = 16; $U = 241145.00$; $p \leq 0.001$). Concerning the subscales (measure adjusted to the number of items) and the same variables, differences were found regarding gender in the Somatic symptoms subscale (*men*: $m_e = 5$, IQR = 5; *women*: $m_e = 7$, IQR = 6; $U = 132447.50$; $p \leq 0.001$), in the Anxiety subscale (*men*: $m_e = 7$, IQR = 7; *women*: $m_e = 9$, IQR = 7; $U = 130420.50$; $p \leq 0.001$), and in the Social dysfunction subscale (*men*: $m_e = 5$, IQR = 2; *women*: $m_e = 6$, IQR = 3; $U = 122997.00$; $p < 0.05$); regarding sports/hobbies, differences were found in the Somatic symptoms subscale

(yes: $m_e = 6$, IQR = 5; no: $m_e = 7$, AIQ = 6; $U = 236115.50$; $p \leq 0.001$), in the Anxiety subscale (yes: $m_e = 8$, IQR = 6; no: $m_e = 10$, IQR = 7,75; $U = 239517.50$; $p \leq 0.001$), in the Social dysfunction subscale (yes: $m_e = 5$, IQR = 2; no: $m_e = 6$, IQR = 3; $U = 228251.50$; $p \leq 0.001$), and in the Severe depression subscale (yes: $m_e = 1$, IQR = 3; no:

$m_e = 1$, IQR = 5; $U = 220532.00$; $p < 0.05$).

The analysis of the clinical sensitivity in terms of mental health between the scale with 28 items and 24 items shows that, after adjusting the mean to the number of items, the values are more or less the same (Table 5).

Table 5

Comparison of the clinical sensitivity between GHQ-28 and GHQ-24

	Total Mean (\pm SD) / Median (IQR)	Somatic symptoms Mean (\pm SD) / Median (IQR)	Anxiety and insomnia Mean (\pm SD) / Median (IQR)	Social dysfunction Mean (\pm SD) / Median (IQR)	Severe depression Mean (\pm SD) / Median (IQR)
GHQ-28	30.09 (\pm 13.86) 28 (18)	8.35 (\pm .30) 8 (6)	9.13 (\pm 4.83) 9 (7)	9.23 (\pm 3.13) 4 (4)	3.38 (\pm 4.34) 2 (5)
GHQ-24	25 (\pm 11.92) 24 (16)	6.86 (\pm 3.70) 6 (5)	9.13 (\pm 4.84) 9 (7)	6.32 (\pm 2.28) 6 (2)	2.70 (\pm 3.83) 1 (4)
GHQ-28^a/-24^a	Mean 1.07 / 1.04	Mean 1.19 / 1.14	Mean 1.30 / 1.30	Mean 1.31 / 1.26	Mean 0.48 / 0.45

Note. ^a = Adjusted to the number of items; GHQ-28 = 28-Item General Health Questionnaire; GHQ-24 = 24-Item General Health Questionnaire; *SD* = Standard deviation; IQR = Interquartile range.

Finally, the ROC curve was used in the total scale and the subscales to assess clinical validity taking into account the reference values to interpret the results, that is, the values based on which the participants should be referred to a

more accurate clinical assessment. The analysis revealed areas under the curve (AUC) ranging from 0.993 to 1, with $p < 0.001$ (Table 6).

Table 6

AUC-ROC Curve

Scale	AUC	p
GHQ-24-Total	0.996	
GHQ-24-Somatic symptoms	0.993	
GHQ-24-Anxiety	1.000	<0.001
GHQ-24-Social dysfunction	0.990	
GHQ-24-Depression	0.996	

Note. GHQ-24 = 24-Item General Health Questionnaire; AUC = Area under the curve; p = significance test.

The cutoff points, sensitivities, and specificities of the total scale were determined using the Youden Index. A value of 20.5 was found, corresponding to a sensitivity

of 96.3% and a specificity of 98.4%. Table 7 shows the values for the subscales.

Table 7*Cutoff points, sensitivity, and specificity of GHQ-24*

Scale	Cutoff point	Sensitivity	Specificity
GHQ-24-Total	20.5	96.3%	98.4%
GHQ-24-Somatic symptoms	4.5	96.2%	98.1%
GHQ-24-Anxiety	5.5	100.0%	100.0%
GHQ-24-Social dysfunction	4.5	94.3%	98.6%
GHQ-24-Depression	4.5	95.4%	98.5%

Note. GHQ-24 = 24-Item General Health Questionnaire.

Discussion

The challenge of reevaluating the psychometric properties and factorial validity of a widely used scale nationally and internationally is necessary mainly because its applicability in health professionals is little studied. The sample size allowed further exploring these aspects to ensure the reliability of the statistical analyses and enable more robust conclusions to sustain this proposal. We have reevaluated and maintained the four original factors of the scale: Anxiety and insomnia (keeping seven original items), Severe depression (it now has six items), Social dysfunction (it now has five items), and Somatic symptoms (it now has six items), ensuring that the eigenvalue and the variance explained by factor is based on different recommendations and the statistical significance of factor loadings (Hair et al., 2009; Marôco 2018). The proposal of a new structure and the discussion on the deletion of items and their factorial distribution (item saturation on more than one factor, non-orthogonality of the scale, independence of factors) should follow the statistical recommendations agreed and maintain the theoretical integrity of the constructs (Hair et al., 2009). We believe that this aspect has been achieved by following the recommendations of other shorter versions used at both international and national levels (Laranjeira, 2008; Kashyap & Singh, 2017). In this case, we decided to use the more extensive version due to the specificity of the sample.

The factors previously determined were confirmed, which allows assessing the practical significance of the results, that is, key variables were identified, which, in this case, are the dimensions (Hair et al., 2009). Data analysis showed that items 3, 16, 21, and 26 loaded on two factors: 1 and 4, 3 and 4, 1 and 3, 1 and 2, respectively, reflecting a conceptual approximation and an interaction between theoretical constructs (psychological conditions) and common points, namely between somatic symptoms and anxiety, between social dysfunction and somatic symptoms, between anxiety and social dysfunction and, finally, between depression and anxiety, which is in line with the literature (Lu et al., 2019). We know that they are symptoms or expressions of individuals' mental state that overlap and influence each other, even because

their objective and subjective indicators are often similar. The scale's domains have been organized based on the literature (Pais-Ribeiro et al., 2015), with the ability to point out independent clinical indications, verified by the independence of the correlations between constructs (between 0.342 and 0.717). As previously identified, a greater correlation was found between the Anxiety and insomnia subscale and the Somatic symptoms subscale (Pais-Ribeiro & Antunes, 2003). Thus, these items were deleted after the exploration of orthogonal and non-orthogonal solutions. Although the deleted items are linked to the theoretical constructs of their sub-dimensions, the reliability results, when compared to the GHQ-28, reveal that these constructs can be measured without the deleted items, as shown in the results of the total scale and subscales in terms of the adjusted mean to the number of items and in terms of their clinical sensitivity in relation to some sociodemographic or socioprofessional variables (Hair et al., 2009; Mâroco, 2018; Pais-Ribeiro et al., 2015). A methodological study is justified to validate this short 24-item version by comparing it with standard instruments of analysis of the overall outcome and sub-dimensions, such as the Short Form-36, a health assessment questionnaire used by Pais-Ribeiro and Antunes (2003), and others clinically relevant to the constructs. It should also be noted that items 3 and 26 had previously been identified as having high loadings on more than one factor, and their discrimination was questioned (Pais-Ribeiro et al., 2015).

For the operationality of the 28-item scale, the literature proposes a total score of 23/24 as a border value for the presence of a case to be referred to further clinical assessment and above 4/5 in the subscales (with seven items each; Pais-Ribeiro et al., 2015). These results indicate the possibility of adjusting the number of items in the total scale and the subscales, with the latter including five to seven items. The scale's total score will range from 0 to 72, the subscales with five items from 0 to 15, the subscales with six items from 0 and 18, and the one that remains with seven items from 0 to 21. In this process, the analysis was carried out mainly considering the statistical data, avoiding the variability provided by the various methods for factor extraction, rotation, and scoring. Each method

produces different outcomes, so we avoided experimenting with different methods until the results matched some previous convictions of the researchers' group, as mentioned in the literature (Oliveira, 2014). A more psychometric and less theoretical-clinical option was made (Pais-Ribeiro et al., 2015) that is reliable because the qualitative analysis kept the items in the factors with higher loading.

Regarding practical applicability, the ROC curve analysis showed that GHQ-24 maintains excellent psychometric characteristics compared to GHQ-28. It will be necessary to perform the criterion-oriented validation of each dimension against gold standard instruments in the future. The reassessment of psychometric properties would benefit from using another instrument with a similar theoretical construct to reinforce the results with a concurrent validation analysis, which can be a limitation of this study. The participant selection process is also a limitation of the study because, although the access is through the OE website, we could not guarantee that the participants were the only ones accessing it.

Conclusion

The analysis of the psychometric properties of GHQ-28 suggests a reduction in the number of items from 28 to 24 (deleting items 3, 16, 21, and 26, which were explained simultaneously by more than one factor). It maintains a (not correlated) orthogonal factorial structure that is theoretically significant and easy to interpret in an instrument that produces a single score resulting from the sum of its dimensions.

The analysis of the GHQ-24 confirmed the good internal consistency of both the scale and its dimensions, as well as the homogeneous and significant contribution of the items to assessing the concept and its dimensions. The ability of the scale's structure to discriminate different individuals with different perceptions is also confirmed, reflecting its sensitivity for assessing symptoms of mental health and psychological well-being. The reduction in the number of items reduces the participants' response time without losing reliability in the results.

This study reinforces the need to regularly reassess the properties of an international instrument, especially when used in a different and specific population. In the future, this study should be conducted with a population of health professionals other than nurses to expand the validation process among health professionals.

Author contributions

Conceptualization: Seabra, P., Lopes, J., Calado, M.

Data curation: Seabra, P., Lopes, J., Calado, M., Capelas, M., Pessoa, E.

Methodology: Seabra, P., Lopes, J., Calado, M., Capelas, M., Pessoa, E.

Writing – original draft: Seabra, P., Lopes, J., Calado, M., Capelas, M., Pessoa, E.

Writing – review & editing: Seabra, P., Lopes, J., Calado, M., Capelas, M., Pessoa, E.

References

- Ali, G.-C., Ryan, G., & De Silva, M. J. (2016). Validated screening tools for common mental disorders in low and middle income countries: A systematic review. *PLoS ONE*, *11*(6), e0156939. <https://doi.org/10.1371/journal.pone.0156939>
- Goldberg, D., & Hillier, V. (1979). A scaled version of the General Health Questionnaire. *Psychology Medicine*, *9*(1), 139-145. <https://doi.org/10.1017/S0033291700021644>
- Gomes, R. K., & Oliveira, V. B. (2013). Depressão, ansiedade e suporte social em profissionais de enfermagem. *Boletim de Psicologia*, *68*(138), 23-33. http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S0006-59432013000100004
- Hair, J. F., Jr., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2009). *Análise multivariada de dados* (6th ed.). Bookman.
- Kashyap, G. C., & Singh, S. K. (2017). Reliability and validity of general health questionnaire (GHQ-12) for male tannery workers: A study carried out in Kanpur, India. *BMC Psychiatry*, *17*, 102. <https://doi.org/10.1186/s12888-017-1253-y>
- Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). The Guilford Press.
- Laranjeira, C. A. (2008). General health questionnaire - 12 items: Adaptation study to the Portuguese population. *Epidemiology and Psychiatric Sciences*, *7*(2), 148-151. <https://doi.org/10.1017/S1121189X00002840>
- Lu, Y., Li, J., & Liu, Y. (2019). Depression as a mediator of quality of life in patients with neuropathic pain: A cross-sectional study. *Journal of Advanced Nursing*, *75*(11), 2719-2726. <https://doi.org/10.1111/jan.14111>
- Marôco, J. (2018). *Análise estatística com o SPSS Statistics 25* (7th ed.). Report Number.
- Marôco, J., Marôco, A. L., Leite, E., Bastos, C., Vazão, M. J., & Campos J. (2016). Burnout in Portuguese healthcare professionals: An analysis at the national level. *Acta Médica Portuguesa*, *29*(1), 24-30. <https://doi.org/10.20344/amp.6460>
- Monteiro, A. P. (2011). Assessment of the factor structure and reliability of the Portuguese version of the General Health Questionnaire-28 among adults. *Journal of Mental Health*, *20*(1), 15-20. <https://doi.org/10.3109/09638237.2010.492414>
- Oliveira, A. G. (2014). *Bioestatística decodificada: Bioestatística, epidemiologia e investigação* (2nd ed.). Lidel.
- Pais-Ribeiro, J. L. (2011). *Inventário de saúde mental*. Placebo.
- Pais-Ribeiro, J. L., & Antunes, S. (2003). Contribuição para o estudo de adaptação do Questionário de Saúde Geral de 28 itens. *Revista Portuguesa de Psicossomática*, *5*(1), 37-45. <https://www.redalyc.org/articulo.oa?id=28750105>
- Pais-Ribeiro, J. L., Silva, M., Abrantes, C., Coelho, M., & Nunes, J. (2015). Ulterior validação do Questionário de Saúde Geral de Goldberg de 28 itens. *Psicologia, Saúde e Doenças*, *16*(3), 278-286. http://www.scielo.mec.pt/scielo.php?script=sci_abstract&pid=S1645-00862015000300001&lng=en&nrm=iso
- Sato, F., Hayakawa, K., & Kamide, K. (2016). Investigation of mental health in Indonesian health workers immigrating to Japan under the Economic Partnership Agreement. *Nursing and Health Sciences*, *18*, 342-349. <https://doi.org/10.1111/nhs.12275>
- Seabra, P., Lopes, J., Calado, M. E., & Capelas, M. L. (2019). A national survey of the nurses' mental health - The case of Portugal. *Nursing Forum*, *54*(3), 425-433. <https://doi.org/10.1111/nuf.12350>
- Tabachnick, B., & Fidel, L. (2007). *Using multivariate statistics* (5th

ed.). Pearson/Allyn & Bacon.

- Vinod, A., Choudhari, S. B., Waghmare, S., Rajshri, Dewani, K., & Moni, S. (2017). Study of psychiatric morbidity among health professionals of different groups using the General Health Questionnaire-28: A cross-sectional study. *International Journal of Indian Psychology*, 4(2). <https://doi.org/10.25215/0402.186>
- World Health Organization. (2007). *Strengthening mental health promotion*. Geneva, WHO Fact sheet. <https://mindyourmindproject.org/wp-content/uploads/2014/11/WHO-Statement-on-Mental-Health-Promotion.pdf>

