More on the Dimensionality of the GHQ-12: Competitive Confirmatory Models*

Más sobre la dimensionalidad del GHQ-12: modelos confirmatorios competitivos

José M. Tomás[a]
Universidad de Valencia, España
ORCID: http://orcid.org/0000-0002-3424-1668

Pedro Hontangas
Universidad de Valencia, España

Amparo Oliver
Universidad de Valencia, España
ORCID: http://orcid.org/0000-0002-1207-4088

Laura Galiana
Universidad de Valencia, España
ORCID: http://orcid.org/0000-0002-5342-5251

Patricia Sancho
Universidad de Valencia, España
ORCID: http://orcid.org/0000-0002-8624-8757

[a] Correspondence author. Email: tomasjm@uv.es

ABSTRACT

The General Health Questionnaire (GHQ) was designed to measure minor psychiatric morbidity by assessing normal 'healthy' functioning and the appearance of new, distressing symptoms. Among its versions, the 12-item is one of the most used. GHQ-12's validity and reliability have been extensively tested in samples from different populations. In the Spanish version, studies have come to different conclusions, of one, two, and three-factor structures. This research aims to present additional evidence on the factorial validity of the Spanish version of the GHQ-12, using competitive confirmatory models. Three samples of workers (N= 525, 414 and 540) were used to test a set of substantive models previously found in Spanish and international literature. Results showed that multidimensional models had moderate to substantial inter-factor correlations (ranging from 0.29 to 0.76) but not as high as to jeopardize their discriminant validity. The best-fitting models were the original solution by Graetz (1991) and the exploratory three-factor solution offered by Rocha et al. (2011), multidimensional three-factor solutions with correlated factors. The conclusion is that a multidimensional three-factor structure underlies the items in the GHQ-12.

Keywords
Distress; GHQ-12; competitive models; Spanish population; factorial validity.

RESUMEN
El Cuestionario de Salud General (General Health Questionnaire, GHQ) fue diseñado para medir la morbilidad psiquiátrica menor evaluando el funcionamiento "sano" y la aparición de nuevos síntomas molestos. Entre sus versiones, la de 12 ítems es una de las más usadas. La validez y

fiabilidad del GHQ-12 ha sido puesta a prueba en muestras provenientes de distintas poblaciones. En la versión española, los estudios han llegado a distintas conclusiones de estructuras de uno, dos y hasta tres factores. El objetivo de la esta investigación es presentar evidencia adicional de la validez factorial de la versión española del GHQ-12, a través de modelos factoriales competitivos. Tres muestras de trabajadores (N = 525, 414 y 540) fueron utilizada para poner a prueba una serie de modelos sustantivos encontrados previamente en la literatura española e internacional. Los resultados mostraron que los modelos multidimensionales tenían correlaciones inter-factores de moderadas a altas (desde .29 a .76), pero no lo suficiente para poner en peligro la validez discriminante de los modelos. Los modelos que presentaron el mejor ajuste fueron la solución origina de Graetz (1991) y la solución exploratoria de tres factores ofrecida por Rocha et al. (2011), una solución multidimensional de tres factores con correlaciones entre ellos. La conclusión es que hay una estructura multidimensional de tres factores que subyace a los ítems del GHQ-12.

Palabras clave
Distress; GHQ-12; modelos competitivos; población española; validez factorial.

Goldberg (1972) designed the General Health Questionnaire (GHQ) to measure minor psychiatric morbidity by assessing normal 'healthy' functioning and the appearance of new, distressing symptoms (Baksheev, Robinson, Cosgrave, Baker & Yung, 2011). The GHQ is a self-administered screening questionnaire developed in several versions: the original 60-item version (Goldberg, 1972), and shorter versions of 30, 28 and 12 items (Goldberg & Williams, 1988). It is considered "the most widely used instrument for detecting non-psychotic psychiatric cases" (Molina et al., 2006, p. 478). Among all versions, the 12-item and the 30-item versions have been the most used in community samples (Molina et al., 2006), and probably, the GHQ-12 has become the most popular form of the scale because of its relatively good psychometric properties and its brevity (Goldberg & Williams, 1988).

GHQ-12's validity and reliability have been extensively tested in samples from different populations. The original English version has deserved much psychometric attention (e.g., Mann et al., 2011; Smith, Fallowfield, Stark,

Velikova, & Jenkins, 2010; Smith, Oluboyede, West, Hewison, & House, 2013). The scale has also been translated into and validated in languages different from English: French (Salama-Younes, Montazeri, Ismaïl, & Roncin, 2009), German (Romppel, Brachler, Roth & Glaesmer, 2013), Dutch (Cornelius, Groothoff, Van der Klink, & Brouwer, 2013), Turkish (Krespi et al., 2010), or Portuguese (Fernandes & Vasconcelos-Raposo, 2012) among others.

Regardless of accumulated evidence, the factorial validity of the GHQ-12 is still under controversy. Originally developed as a one-factor screening instrument, indeed some studies found one underlying factor (Banks et al., 1980), but multifactorial solutions were more likely, either with two factors (for example, Andrich & Van Schoubroeck, 1989; Politi, Piccinelli & Wilkinson, 1994; Schmitz, Kruse & Tress, 1999; Villa, Zuluaga-Arboleda & Restrepo-Roldán, 2013) or three factors (Graetz, 1991; Urzúa, Caqueo-Urízar, Bargsted & Irarrázaval, 2015). To further complicate things, a closer look at these multidimensional structures shows that the structure could be confounded by wording effects associated with negatively worded items. Consequently, several models that tested for the GHQ-12 factor structure adding these wording effects started to appear along literature, and some have found support for a unidimensional structure with wording effects associated to negatively worded items (Smith et al., 2010; Smith et al., 2013; Solís-Cámara, Meda-Lara, Moreno-Jiménez & Juárez-Rodríguez, 2016).

Concerning the Spanish version of the scale, there is also a certain amount of accumulated evidence on its psychometric properties, and specifically on its factor structure. Some of this evidence comes from exploratory factor analyses. López-Castedo and Fernández (2005), for example, studied a non-probabilistic sample of 1930 Spanish adolescents and found support for a two-factor solution of anxiety and social dysfunction that conjointly explained 46.8% of the variance. Sánchez-López and Dresch (2008), in turn, found a multidimensional structure of three correlated factors in another non-probabilistic sample of the general Spanish

population. They were named successful coping, which included all positively worded items, self-esteem (items 6, 9, 10, and 11), and stress (items 2, 5, and 9). The factors were moderately inter-correlated, with a more significant value for factors two and three. Finally, Rocha, Pérez, Rodríguez-Sanz, Borrell and Obiols (2011) studied the psychometric properties of the GHQ-12, including its factorial validity, on a Spanish representative sample (29476 participants). They used exploratory factor analyses forced to one, two, and three-factor solutions that respectively explained 67%, 82%, and 91% of the variance. They concluded that the unidimensional solution adequately represented the observed scores and, consequently, gave normative values to be used in the Spanish population.

More convincing results are offered when confirmatory factor analyses (CFA) are employed, and this has been the case in some studies (González-Romá, Lloret & Espejo, 1993). González-Romá et al. (1993) studied two-factor structures (one factor and two factors of anxiety and depression) in two non-probabilistic samples of 167 and 112 workers and found poor model fit for both structures, but slightly better for the multidimensional. Padrón, Galán, Durbán, Gandarillas and Rodríguez-Artalejo (2012) tested four confirmatory models: unidimensional model, two-factor (positive and negative items) structure, a three-factor model found by Graetz (1991), and the three-factor solution explored in their data. Their results showed that the three-factor structure was the best fitting model, but with high inter-factor correlations (ranging between 0.72 and 0.84). Finally, Aguado et al. (2012) were the first ones to model method effects into the Spanish version of the GHQ-12. They concluded that a one-factor model including method effects among negatively worded items better represented the data of their postpartum women sample.

Taking all these into account, this research aims to present additional evidence on the factorial validity of the GHQ-12 in its Spanish version. A set of substantive models that have been previously found in Spanish and international literature, including method effects associated with negatively worded items, will be tested. This complete set of models will be systematically tested in three independent samples of workers.

Method

Samples and procedure. Three independent samples of Spanish workers were used for this study. The first sample (A) comprised 525 workers, who were gathered through a systematic sampling of all the workers that underwent their annual medical tests at the Valencian Health and Safety Executive (Province of Valencia, Spain). Participants were randomly selected from all the workers attending this health-check during a one-year period. 85.4% were men. Participants' average age was 37 years, SD = 10.84, the minimum age was 16 and the maximum was 64 years old. The second sample (B) was composed of 414 youth employed at the beginning of their work careers. They worked in companies located in 11 Spanish provinces. 57.9% of participants were males. Age ranged from 17 to 34 years old with an average of 22.6 years (SD= 3.85). The third sample (C) was composed of 540 public servants working for two Autonomous Regions in Spain (Valencian Community and Andalucía). 53.81% of them were male with. Age ranged from 22 to 56 years old with an average age of 35 years (SD= 6.41).

Instruments. The General Health Questionnaire (GHQ-12) developed by Goldberg and Williams (1988) was used, with the item content in the Spanish version by Lobo and Muñoz (1996). This questionnaire consists of 12 items, 6 of which are positive and the remaining 6 are negative statements. Items in the three analyzed samples were answered using a 4-point Likert-scale from 0 (better than usual) to 3 (much less than usual). Example items are "Lost much sleep over worry" or Felt you could not overcome your difficulties".

Statistical analyses. Confirmatory Factor Analyses were used to test for the a priori structures in the GHQ-12. These CFA were

estimated with EQS 6.1 (Bentler, 2000-2018). Maximum likelihood with robust Satorra-Bentler corrections on polychoric correlations matrices was the estimation method of choice, given that multivariate normality was not tenable (Mardia multivariate coefficients were 25.6, 22.6, and 95.4, for samples A, B, and C, respectively), and the response format was ordinal (Finney & DiStefano, 2013). Goodness-of-fit for each model was assessed using indexes based on different approaches (Kline, 2015): 1) χ^2 statistic; 2) CFI (Comparative Fit Index), and 3) RMSEA (Root Mean Square Error of Approximation) and its 90% confidence interval. Robust versions of all tests and fit indices have been used. The χ^2 goodness-of-fit statistic is a test of the difference between the observed covariance matrix and the one predicted by the specified model. χ^2 value with a probability value greater than 0.05 indicates good fit; however, this statistic is affected by several limitations and has very restrictive assumptions (dependence on sample size, multivariate normality, use the correct model). Therefore, other indices less affected by sample size and model complexity (Kline, 2015) were used. Values higher than 0.90 for the CFI or lower than 0.08 in RMSEA are considered a reasonable fit (Kline, 2015), although values higher than 0.95 in CFI or lower than 0.05 in RMSEA are more desirable and considered excellent fit (Caycho-Rodríguez et al. 2018). It has also been suggested that the combination of a CFI more than 0.90 with an RMSEA value lower than 0.06 may indicate an extremely good fit (Caycho-Rodríguez et al. 2018). All models found in the literature with a reasonable model fit to represent the underlying structure of the GHQ-12 were specified and tested. Ten completely a priori or strictly confirmatory models were tested:

a) Model 1 was a one-factor model (minor psychiatric morbidity), as found by Banks et al. (1980) or defended in the Spanish version by Rocha et al. (2011). It also served as a baseline (most parsimonious) model.

b) Model 2 contained two correlated factors grouping the positively (items 1, 3, 4, 7, 8, and 12) and negatively worded items (2, 5, 6, 9, 10, and 11), a structure based on Andrich and Van Schoubroeck (1989)

c) Model 3 specified two correlated factors of anxiety/depression (items 1, 2, 3, 6, 7, 10, and 11) and social performance (3, 4, 5, 8, 9, and 12), based on Schmitz et al. (1999).

d) Model 4 was composed of two correlated factors: dysphoria (items 2, 5, 6, 9, 10, 11, and 12) and social dysfunction (items 1, 3, 4, 7, 8, and 12), based on Politi et al. (1994) best fitting model.

e) Model 5 specified three substantive correlated dimensions of cope (items 1, 3, 4, and 8), stress (items 2, 5, and 7), and depression (items 6, 9, 10, 11, and 12), and was based on a content analysis made by Martin (1999).

f) Model 6 presented three correlated factors: dysphoria (items 2, 5, 6, and 9), social dysfunction (items 1, 3, 4, 7, 8, and 12), and loss of confidence (items 10 and 11), that is, the factor solution supported in Graetz (1991).

g) Model 7 included three correlated factors, very similar to those in Graetz (1991), but found in the 3-factor exploratory solution by Rocha et al. (2011): dysphoria (2, 5, and 9), social dysfunction (1, 3, 4, 7, 8, and 12), and loss of confidence (6, 9, 10, and 11). It included a cross-loading, as item 9 was considered an indicator of both dysphoria and social dysfunction. Therefore, two additional models were specified, to override the cross-loading:

  a) Model 7a, with the same structure as model 7 but with item 9 loading only on dysphoria.
  b) Model 7b, with model 7 structure but with item 9 loading only on social dysfunction.

h) Additional to the substantive models, Model 8 considered a method factor related to the negatively worded items, together with the trait dimension of minor psychiatric morbidity. This model found support, among others, in the Spanish version of the scale by Aguado et al. (2012).

Results

Goodness-of-fit indexes for the ten models tested are shown in Table 1. A first general look at this table showed that the one-factor solution (model 1) originally proposed by Goldberg for the GHQ-12 was never the best fitting model, and generally speaking, it was always the worst. In a nutshell, a one-factor solution did not adequately represent the variance-covariance matrix among the items in the GHQ-12. Secondly, all models that posited two trait factors (models 2, 3 and 4) also inadequately represented the observed data. In none of the cases their fit indexes reached the stricter cut-off criteria, and most of the time they were well below the more "relaxed" cut-off criteria. Model 5 is the first three-factor model, and again its fit could not be considered enough. The same happened with model 8, which included method effects associated with negatively worded items, and again did not show a relevant increment on its fit compared to other models.

Table 1

Goodness-of-fit indexes for the three samples in the 10 a priori models

Models	Sample A (N= 525)					Sample B (N= 414)					Sample C (N= 540)				
	χ^2	df	CFI	RMSEA	10% CI	χ^2	df	CFI	RMSEA	10% CI	χ^2	df	CFI	RMSEA	10% CI
1	675.2*	54	0.663	0.166	0.155-0.177	261.2*	54	0.825	0.096	0.085-0.108	214*	54	0.907	0.074	0.064-0.085
2	205.4*	53	0.917	0.083	0.071-0.085	179.3*	53	0.893	0.076	0.064-0.088	189.4*	53	0.92	0.069	0.059-0.08
3	643*	52	0.68	0.165	0.153-0.176	262.2*	52	0.823	0.099	0.087-0.111	217.5*	52	0.903	0.077	0.066-0.087
4	188.2*	52	0.926	0.079	0.067-0.091	178.7*	52	0.894	0.077	0.065-0.089	188.5*	52	0.92	0.07	0.059-0.088
5	247.8*	51	0.893	0.096	0.084-0.108	175.5*	51	0.885	0.077	0.064-0.089	157.8*	51	0.938	0.062	0.051-0.073
6	150.1*	51	0.946	0.068	0.056-0.081	103.6*	51	0.956	0.05	0.036-0.064	112.9*	51	0.964	0.047	0.036-0.059
7	106*	50	0.97	0.052	0.038-0.065	118.1*	50	0.942	0.057	0.044-0.071	119.1*	50	0.96	0.051	0.039-0.062
7a	150.6*	51	0.946	0.068	0.056-0.081	143.8*	51	0.922	0.066	0.054-0.079	133.1*	51	0.952	0.055	0.043-0.066
7b	143.5*	51	0.95	0.066	0.053-0.079	148.2*	51	0.918	0.068	0.055-0.081	161.3*	51	0.936	0.063	0.052-0.074
8	184.9*	48	0.926	0.083	0.07-0.095	149.7*	48	0.914	0.072	0.059-0.085	132.3*	48	0.951	0.057	0.046-0.069

Note. df = degrees of freedom; *p < 0.001.

It was clear in Table 1 that both Graetz and Rocha's three-factor models had the best fit through all indexes and samples considered. The goodness-of-fit indexes were very similar in both cases, for model 6, 7 and the derived models 7a and 7b. This was not surprising since both models only varied in one indicator loading. Goodness-of-fit was slightly better for Graezt's solutions in samples B and C, while it was better for Rocha's model in sample A.

Table 2 shows the standardized factor loadings for model 6 estimated in the three samples. All factor loadings were statistically significant and substantial. Finally, Table 3 shows correlations among the factors for multidimensional models. A clear result is that multidimensional models do have moderate to large inter-factor correlations, but not as high as to jeopardize their discriminant validity.

Table 2

Standardized loadings (Samples A, B and C)

Items	A			B			C		
	F1	F2	F3	F1	F2	F3	F1	F2	F3
1	0.519			0.55			0.587		
2		0.617			0.547			0.636	
3	0.64			0.55			0.421		
4	0.676			0.551			0.66		
5		0.663			0.654			0.748	
6		0.51			0.688			0.742	
7	0.669			0.388			0.573		
8	0.739			0.308			0.667		
9		0.86			0.848			0.855	
10			0.895			0.904			0.95
11			0.747			0.706			0.744
12	0.636						0.565		

Table 3

Correlations between factors (Samples A, B and C)

Models	A			B			C		
	r12	r13	r23	r12	r13	r23	r12	r13	r23
1. One trait factor	-	-	-	-	-	-	-	-	-
2. Two factors (Andrich et al., 1989)	0.299	-	-	0.585	-	-	0.708	-	-
3. Two factors (Schmitz et al., 1999)	0.722	-	-	0.948	-	-	0.959	-	-
4. Two factors (Politi et al., 1994)	0.236	-	-	0.549	-	-	0.673	-	-
5. Three factors (Martin, 1999)	0.032	0.239	0.685	0.33	0.661	0.713	0.465	0.626	0.778
6. Three factors (Graetz, 1991)	0.293	0.31	0.768	0.545	0.486	0.61	0.689	0.611	0.76
7. Three factors (Rocha et al., 2011), item 9 in factors 2 and 3	0.114	0.327	0.514	0.404	0.549	0.444	0.643	0.636	0.65
7a. Three factors (Rocha et al., 2011), item 9 in factor 2	0.269	0.308	0.772	0.59	0.502	0.696	0.724	0.633	0.789
7b. Three factors (Rocha et al., 2011), item 9 in factor 3	0.69	0.327	0.648	0.339	0.616	0.727	0.579	0.676	0.734
8. One trait factor and one method factor	-	-	-	-	-	-	-	-	-

Discussion and conclusions

According to all data, the GHQ developed by Goldberg (1972), and specifically its shortest GHQ-12, is one of the most widely used and studied indicator of minor psychiatric disorders. It has been used as a screening tool that can be easily administered in adult and adolescent populations alike (Pena & Caine, 2006), and it is meant to be valid and reliable. Nevertheless, there is not a broad consensus on its structure, the number of dimensions, and items. This holds true across the different languages into which it has been translated. The solution that has received more support along literature is the multidimensional three-factor structure found by Graetz (1991), but the one-factor solution, with method effects associated to negatively worded items, has also had some support (Smith et al., 2013).

Regarding the Spanish version, the cumulated information on its factorial validity is not that large, but as seen in the Introduction, it has also been subject to controversy. Two recent contributions to the existing literature on the

structure of the GHQ-12 are those from Rocha et al. (2011) and Aguado et al. (2012). On the one hand, Rocha et al. (2011) studied an extensive representative sample of the Spanish population (approximately 25000 people), and using EFA found support for a one-factor, two-factor, and three-factor structure, but decided to retain the simplest one-factor structure. They derived population norms for this single dimension to be used in the Spanish population. On the other hand, Aguado et al. (2012) tested several competing models via CFA. Among them, they included the one-factor solution, a one-factor solution with method effects associated with the negatively worded items, and Graetz's (1991) multidimensional model. They found that the best fitting models were the one-factor model with method effects and Graetz's (1991) three-factor solution. Nevertheless, fit indexes for several of the models proposed were extremely close to those of the best fitting models.

Taking both contributions together, the overall conclusion is that the GHQ-12 structure is far from being well established. Rocha et al.'s (2011) results, although coming from an impressive sample, were exploratory and it is pending to confirm if the one-factor model would fit better than the three-factor exploratory solution that explained a 91% of the variance (vs. the 67% in the one-factor solution). With regards to the latest contribution by Aguado et al. (2012), they used CFA to test the main models proposed in the literature. However, the fit among the different solutions was extremely close, especially the three-factor and the method effects' solutions, casting some doubts on GHQ-12 dimensionality. Moreover, their sample was particular, postpartum women, which may lead to sample-specific results. Therefore, these two latest contributions to the factorial validity of the GHQ-12, although extremely important, should be complemented with further analyses of new samples from the Spanish population.

Current results offer evidence on three samples of workers: industrial workers of all ages, young workers entering the labor market, and finally civil servants. Results are consistent among the three samples, as the best fitting models were the original solution by Graetz (1991) and the exploratory three-factor solution offered by Rocha et al. (2011); multidimensional three-factor solutions with correlated factors. It should be borne in mind that both structures are almost equal, as the only difference is in one indicator (item 6). Accordingly, both structures labeled the factors as dysphoria, social dysfunction, and loss of confidence. Additionally, these factors showed discriminant validity, given that correlations ranged from a minimum of 0.29 to a maximum of 0.76, depending on the sample analyzed, and in any case, the interval of confidence included 1. Although results in the three samples showed different values for the correlations among factors, we believe there may be some characteristics on the type of worker surveyed that make the relationship among the factors of the questionnaire vary among samples. In our opinion, if the scale indeed measures three factors with enough discriminant validity, a question should be clearly stated: Are the normative values offered in Rocha et al. (2011) useful in the Spanish population?

A potential limitation of the study is that it only offers evidence on the factorial validity of the scale. Implications of the different factor structures proposed in terms of criterion-related validity would be of interest.

References

Aguado, J., Campbell, A., Ascaso, C., Navarro, P., García-Esteve, L., & Luciano, J. V. (2012). Examining the factor structure and discriminant validity of the 12-item General Health Questionnaire (GHQ-12) among Spanish postpartum women. *Assessment*, *19*(4), 517-525. https://doi.org/10.1177/1073191110388146

Andrich, D., & Van Schoubroeck, L. (1989). The General Health Questionnaire: a psychometric analysis using latent trait theory. *Psychological Medicine*, *19*(2), 469-485. https://doi.org/10.1017/S0033291700012502

Baksheev, G. N., Robinson, J., Cosgrave, E. M., Baker, K., & Yung, A. R. (2011). Validity of the 12-item General Health Questionnaire (GHQ-12) in detecting depressive and anxiety disorders among high school students. *Psychiatry Research*, *187*(1-2), 291-296. https://doi.org/10.1016/j.psychres.2010.10.010

Banks, M. H., Clegg, C. W., Jackson, P. R., Kemp, N. J., Stafford, E. M., & Wall, T. D. (1980). The use of the General Health Questionnaire as an indicator of mental health in occupational studies. *Journal of Occupational Psychology*, *53*(3), 187-194. https://doi.org/10.1111/j.2044-8325.1980.tb00024.x

Bentler, P. M. (2000-2018). *EQS 6.1. for Windows. Structural equations program manual*. Encino, CA: Multivariate Software, Inc. Retrieved from http://www.econ.upf.edu/~satorra/CourseSEMVienna2010/EQSManual.pdf

Caycho-Rodríguez, T., Ventura-León, J., García-Cadena, C. H., Tomás, J. M., Domínguez-Vergara, J., Daniel, L.,…Arias-Gallegos, W. L. (2018). Evidencias psicométricas de una medida breve de resiliencia en adultos mayores peruanos no institucionalizados. *Psychosocial Intervention*, *27*(2), 73-79. https://doi.org/10.5093/pi2018a6

Cornelius, B. L. R., Groothoff, J. W., Van der Klink, J. J. , & Brouwer, S. (2013). The performance of the K10, K6 and GHQ-12 to screen for present state DSM-IV disorders among disability claimants. *BMC Public Health*, *13*, 128. https://doi.org/10.1186/1471-2458-13-128

Fernandes, H. M., & Vasconcelos-Raposo, J. (2012). Factorial validity and invariance of the GHQ-12 among clinical and nonclinical samples. *Assessment*, *20*(2), 219-229. https://doi.org/10.1177/1073191112465768.

Finney, S. J., & DiStefano, C. (2013). Non-normal and categorical data in SEM. In G. R Hancock & R. O. Mueller (Eds.), *Structural Equation Modeling: A second course* (pp. 269-314). Greenwich, CO: Information Age Publishing. Retrieved from https://bayanbox.ir/view/3083173369357234507/Finney-DiStefano-non-normal-and-categorical-data-in-structural-equation-modeling.pdf

Goldberg, D. (1972). *The Detection of Psychiatric Illness by Questionnaire*. Oxford: Oxford University Press.

Goldberg, D., & Williams, P. (1988). *A user's guide to the General Health Questionnaire*. Windsor, UK: NFER-Nelson.

González-Romá, V., Lloret, S., & Espejo, B. (1993). Comparación de dos modelos de medida del cuestionario de salud general (GHQ) [Comparison of two measurement models of the General Health Questionnaire (GHQ)]. *Psicológica*, *14*, 259-268.

Graetz, B. (1991). Multidimensional properties of the General Health Questionnaire. *Social Psychiatry and Psychiatric Epidemiology*, *26*(3), 132–138. https://doi.org/10.1007/BF00782952

Kline, R. B. (2015). *Principles and Practice of Structural Equation Modeling*. New York, NY: Guilford.

Krespi, M. R., Hill, J., Holcombe, C., Clark, L., Fisher, J., & Salmon, P. (2010). The accuracy of HADS and GHQ-12 in detecting psychiatric morbidity in breast cancer patients. *Turkish Journal of Psychiatry*, *21*(1), 49-59.

Lobo A., & Muñoz P. E. (1996). Versiones en lengua española validadas. In D. Goldberg, & P. Williams (Eds.), *Cuestionario de Salud General GHQ (General Health Questionnaire). Guía para el usuario de las distintas versiones* (p. 115). Barcelona, Spain: Editorial Masson.

López-Castedo, A., & Fernandez, L. (2005). Psychometric properties of the Spanish version of the 12-item General Health Questionnaire in adolescents. *Perceptual and Motor Skills*, *100*(3), 676-680. https://doi.org/10.2466/pms.100.3.676-680

Mann, R. E., Paglia-Boak, A., Adlaf, E. M., Beitchman, J., Wolfe, D., Wekerle, C.,… Rehm, J. (2011). Estimating the

prevalence of anxiety and mood disorders in an adolescent general population: An evaluation of the GHQ12. *International Journal of Mental Health and Addiction*, 9, 410-420. https://doi.org/10.1007/s11469-011-9334-5

Martin, A. J. (1999). Assessing the multidimensionality of the 12-Item General Health Questionnaire. *Psychological Reports*, *84*(3), 927-935. https://doi.org/10.2466/pr0.1999.84.3.927

Molina, J. D., Andrade-Rosa, C., González-Parra, S., Blasco-Fontecilla, H., Real, M. A., & Pintor, C. (2006). The factor structure of the General Health Questionnaire (GHQ): a scaled version for general practice in Spain. *European Psychiatry*, *21*(7), 478–486. https://doi.org/10.1016/j.eurpsy.2006.03.002

Padrón, A., Galán, I., Durbán, M., Gandarillas, A., & Rodríguez-Artalejo, F. (2012). Confirmatory factor analysis of the General Health Questionnaire (GHQ-12) in Spanish adolescents. *Quality of Life Reesearch*, *21*(7), 1291-1298. https://doi.org/10.1007/s11136-011-0038-x.

Pena, J. B., & Caine, E. D. (2006). Screening as an approach for adolescent suicide prevention. *Suicide and Life-Threatening Behavior*, *36*(6), 614-637. http://dx.doi.org/10.1521/suli.2006.36.6.614

Politi, P. L., Piccinelli, M., & Wilkinson, G. (1994). Reliability, validity and factor structure of the 12-item General Health Questionnaire among young males in Italy. *Acta Psychiatrica Scandinavica*, *90*(6), 432-437. https://doi.org/10.1111/j.1600-0447.1994.tb01620.x

Rocha, K. B., Pérez, K., Rodríguez-Sanz, M., Borrell, C., & Obiols, J. E. (2011). Propiedades psicométricas y valores normativos del General Health Questionnaire (GHQ-12) en población general española [Psychometric properties and normative values for the General Health Questionnaire (GHQ-12) in Spanish general population]. *International Journal of Clinical and Health Psychology*, *11*(1), 125-139.

Romppel, M., Brachler, E., Roth, M., & Glaesmer, H. (2013). What is the General Health Questionnaire-12 assessing? Dimensionality and psychometric properties of the General Health Questionnaire-12 in a large scale German population sample. *Comprehensive Psychiatry*, *54*(4), 406-413. https://doi.org/10.1016/j.comppsych.2012.10.010

Salama-Younes, M., Montazeri, L., Ismaïl, A, & Roncin, C. (2009). Factor structure and internal consistency of the 12-item General Health Questionnaire (GHQ-12) and the Subjective Vitality Scale (VS), and the relationship between them: A study from France. *Health and Quality of Life Outcomes*, *7*(22), 1-6. https://doi.org/10.1186/1477-7525-7-22

Sánchez-López, M. P., & Dresch, V. (2008). The 12-Item General Health Questionnaire (GHQ-12): Reliability, external validity and factor structure in the Spanish population. *Psicothema*, *20*(4), 839-843.

Schmitz, N., Kruse, J., & Tress, W. (1999). Psychometric properties of the General Health Questionnaire (GHQ-12) in a German primary care sample. *Acta Psychiatrica Scandinavica*, *100*(6), 462-468. https://doi.org/10.1111/j.1600-0447.1999.tb10898.x

Smith, A. B., Fallowfield, L. J., Stark, D. P., Velikova, G., & Jenkins, V. (2010). A Rasch and confirmatory factor analysis of the General Health Questionnaire (GHQ)-12. *Health and Quality of Life Outcomes*, *8*(45), 1-10. https://doi.org/10.1186/1477-7525-8-45

Smith, A. B., Oluboyede, Y., West, R., Hewison, J., & House, A. O. (2013). The factor structure of the GHQ-12: the interaction between item phrasing, variance and levels of distress. *Quality of Life Research*, *22*(1), 145-152. https://doi.org/10.1007/s11136-012-0133-7

Solís-Cámara, P., Meda-Lara, R. M., Moreno-Jiménez, B., & Juárez-Rodríguez, P. (2016).

Estructura factorial del cuestionario de salud general GHQ-12 en población general de México. *Salud & Sociedad*, *7*(1), 62-76. http://doi.org/10.22199/S07187475. 2016.0001.00004

Urzúa, A., Caqueo-Urízar, A., Bargsted, M., & Irarrázaval, M. (2015). ¿Afecta la forma de puntuación la estructura factorial del GHQ-12? Estudio exploratorio en estudiantes iberoamericanos. *Cadernos de Saúde Pública*, *31*(6), 1305-1312. http://dx. doi.org/10.1590/0102-311X00122913

Villa G., I. C., Zuluaga Arboleda, C. & Restrepo Roldán, L. F. (2013). Propiedades psicométricas del Cuestionario de Salud General de Goldberg GHQ-12 en una institución hospitalaria de la ciudad de Medellín. *Avances en Psicología Latinoamericana*, *31*(3), 532-545.

Notes

*   Research article.