The Psychologist:
Practice & Research Journal

ORDEM DOS PSICÓLOGOS

Brief assessment of depression: Psychometric properties of the Portuguese version of the Patient Health Questionnaire (PHQ-9)

Tiago Ferreira[1,2], Marlene Sousa[1,2], Liliana Meira[1,2], Carla Cunha[1,2], Anita Santos[1,2], Sara Silva[1,2], Ana Bela Couto[1,2], Pedro Gomes[1], Liliana Costa[1,2], Eunice Barbosa[1,2], Isabel Basto[1,2], João Salgado[1,2]

[1] Instituto Universitário da Maia, Portugal
[2] Centro de Psicologia da Universidade do Porto, Portugal

Corresponding author: João Salgado | jsalgado@ismai.pt

Received:
30 October 2017

Accepted:
27 December 2018

Abstract

Background: The lack of knowledge regarding psychometric characteristics of brief scales for assessing the severity of symptoms of depression for the Portuguese population limits the ability of health professionals to signal clinically relevant situations, as well as the implementation of alternative forms of mental health care. The Patient Health Questionnaire (PHQ-9) is a scale whose usefulness in various professional contexts has been amply demonstrated by research. However, some of its psychometric qualities remain unknown for the Portuguese version.
Goals: The main objective of this study was to analyse the validity of the factorial structure and the psychometric characteristics of the PHQ-9 in two different clinical Portuguese contexts, namely in primary health care units and in a university clinic, also reflecting the similarities and differences found regarding the factorial structures.
Methods: The first study was carried out with 153 patients (83.66% females), with a mean age of 46.10 years (*SD* = 16.79), from eight health units of primary health care (non-probabilistic sampling). All participants underwent an initial screening for the detection of symptoms of depression and suicidal ideation, within the context of a regular family medicine appointment or nursing visit. Participants who scored positively for one of these symptoms were invited to participate in the study. The PHQ-9 and the Work and Social Adjustment Scale (WSAS) were used and exploratory and semi-confirmatory factor analyses of the PHQ-9 were performed. In the second study, 106 adults (70.78% females) participated in a psychological counselling session at a university clinic, with a mean age of 32.47 years (*SD* = 11.01), also constituting a non-probabilistic sample. As instruments, the PHQ-9, the Beck-II Depression Inventory and the State-Trait Anxiety Inventory were used, which participants completed before the beginning of the psychological treatment. As in the first study, exploratory and semi-confirmatory factor analyses of the PHQ-9 were conducted.
Results: The results of the two studies revealed a factorial structure consisting of three factors. In the first study, the three factors were: (1) depressed mood and anhedonia; (2) self-devaluation and suicidal ideation; and (3) fatigue, psychomotor agitation/retardation, and changes in appetite, sleep problems or difficulty in concentration. In the second study, the three factors that were found were: (1) sleep problems and fatigue; (2) changes in appetite and difficulty in concentration, psychomotor agitation/retardation, and suicidal ideation; and (3) depressed mood, anhedonia, and self-devaluation. The factorial structure found in both studies revealed a good fit. The PHQ-9 also presented good internal consistency, convergent validity and criterion validity in both studies. The cut-off point of 9 emerged as an indicator of major depression.
Discussion: The results of this study enhance the use of this scale in different contexts of health care provision in Portugal, since it was possible to determine the cut-off point for the detection of cases with major depression, as well as the criteria for delimiting severity degrees of depression-related symptomatology.

Keywords: PHQ-9, Depression, Psychometric qualities, Primary health care, University clinic

Introduction

Considering its high and increasing worldwide prevalence, depression is currently one of the most serious mental health problems in the world (World Health Organization [WHO], 2017). Between 2005 and 2015, the prevalence of depression increased by 18.4%, affecting 322 million people worldwide (WHO, 2017). In Portugal, major depression has a lifetime prevalence of 16.7%, corresponding to the third highest rate in Europe (Almeida et al., 2013). The literature has consistently demonstrated the association of this problem with other psychological and medical problems, as well as the loss of quality of life, labour and economic productivity, and the increased risk of suicide (Richards, 2011). The direct and indirect repercussions of depression have prompted different government institutions (e.g., the National Institute for Health and Clinical Excellence [NICE] to adopt prevention programs as well as treatment models (e.g., stepped-care model) with the aim of reducing its impact in the short and long term (NICE, 2009). Although a large number of people seek help in the primary health care contexts, the difficulties of a rigorous and timely evaluation by professionals constitute an important obstacle to the treatment of this mental disorder (Martin, Rief, Klaiberg, & Braehler, 2006). In this sense, the mobilization of strategies to speed up the process of evaluation and treatment of clinical cases in these contexts is very important (NICE, 2009).

Numerous formal assessment tools have been developed to aid in the diagnosis of depression (Pettersson, Boström, Gustavsson, & Ekselius, 2015). Self-administered questionnaires have been considered useful tools in clinical screening and decision making (Martin et al., 2006). One of these instruments is the Patient Health Questionnaire (PHQ-9) developed by Kroenke, Spitzer and Williams (2001) to evaluate the severity of depression-related symptomatology according to the criteria of the Diagnostic and Statistical Manual of Mental Disorders (DSM) IV-TR (American Psychiatric Association, 2006).

The PHQ-9 has been widely used in both clinical practice and research (Manea, Gilbody, & McMillan, 2012). In the original study (Kroenke et al., 2001), this instrument revealed good internal consistency (α = .86 to .89), test-retest validity (r = .84), and convergent validity with the MOS Short-form General Health Questionnaire (Stewart, Hays, & Ware, 1988). The same study (Kroenke et al, 2001) found evidence to support the following criteria for evaluating the severity of depression-related symptomatology: 1 to 4, "*minimal*"; 5 to 9, "*slight*"; 10 to 14, "*moderated*"; 15-19, "*moderately severe*"; 20 to 27, "*severe*." A score of 10 presented high sensitivity and specificity in that study and was indicated as the cut-off point for major depression. After that initial study, the PHQ-9 has been the target of several validations in different populations (e.g., university students: Adewuya, Ola, & Afolabi, 2006; epilepsy: Fiest et al., 2014); health care contexts (e.g., primary care context: Chen et al., 2010) and countries (e.g. Australia: Arroll et al., 2010; Mexico: Arrieta et al., 2017). The results obtained in the different studies show that this questionnaire has good psychometric qualities, which makes it a valid measure to evaluate symptoms of depression (Kroenke et al., 2001). In fact, good internal consistency as well as good sensitivity to the treatment and to the changes in the severity of the symptoms of depression have been reported (e.g., Löwe, Schenkel, Carney-Doebbeling, & Göbel, 2006). The meta-analysis conducted by Manea et al. (2012) revealed that the studies have identified cut-off points for the identification of major depression between 7 and 15. Cut-off point 11 seems to present better sensitivity and specificity, although there are no significant differences in the diagnostic properties for the cut-off points located between 8 and 11 (Manea et al., 2012). In Portugal, a study with university students (Monteiro, Torres, Pereira, Albuquerque, & Morgadinho, 2013) observed a good internal consistency (α = .86) and a good convergent validity with the Beck Depression Inventory (BDI, Beck, Steer, & Brown, 1996) and with the Hospital Anxiety and Depression Scale (HADS; Zigmond & Snaith, 1983). In another study conducted with 63 women suffering from breast cancer (Torres, Monteiro, Pereira, & Albuquerque, 2016), the PHQ-9 also showed good internal consistency (α = .86), test-retest validity (ICC = .87), and a convergent validity with HADS (Pais-Ribeiro et al., 2007; Zigmond & Snaith,

1983) and EORTC QLQ-C30 (Aaronson et al., 1993; Pais-Ribeiro, Pinto, & Santos, 2008).

Previous research has found inconsistent results regarding the factorial structure of the PHQ-9. Some studies have observed single-factor structures (Cameron, Crawford, Lawton, & Reid, 2008), while others have obtained two-factor structures with diverse distributions of items by factors (Chilcot et al., 2013). Elhai et al. (2012) carried out a study that confirmed a 1-factorial structure and three structures constituted by 2 different factors. In that study, the 2-factors structures were more adjusted than the one-factor structure, and the one that separated somatic factors (sleep, fatigue, appetite and concentration problems, agitation/retardation) from non-somatic factors (anhedonia, depressed mood, self-devaluation, suicide ideation) proved to be the most appropriate solution. However, an important limitation in some of these previous exploratory studies has been the use of less appropriate methodologies based, for example, on principal components analysis (Cameron et al., 2008). In addition, studies based on confirmatory factor analysis have sought to identify the most accurate exploratory factorial solution and confirm its replicability (Elhai et al., 2012). Those studies did not, however, allow the evaluation of alternative factorial structures to those identified by the exploratory studies having the aforementioned limitations. Therefore, there is the need to explore better the factorial structure of the PHQ-9. Previous investigation has suggested different cut-off points to define the presence of depression (Manea et al., 2012).

In addition to these inconsistencies found in international studies, there is the fact that the factorial structure and the criteria for delimiting degrees of severity of depression-related symptomatology remain unknown for the Portuguese version of the PHQ-9. This article reports two studies that explored the validity of the factorial structure as well as the psychometric characteristics of the PHQ-9 in two different Portuguese clinical contexts: primary health care units (Study 1) and a university clinic (Study 2). It aimed to examine whether the factorial structure and psychometric qualities of this questionnaire differ according to the clinical context analysed, as pointed out by some studies (e.g., Manea et al., 2012).

Study 1: Psychometric qualities of the PHQ-9 in primary health care units

Methods

Participants

A sample of 153 adults (M_{age}=46.10, DP_{age}=16.79, Min_{age}=18, Max_{age}=75) participated in this study, of which 128 (83,66%) were female. This sample was collected as part of the *"STOP Depression: Stepped care treatments and digital solutions for depression and suicide prevention in primary care"*, financed by the European Economic Area Grants (PT06) asked review" pilot study on the efficacy of a stepped-care model (Van Straten, Hill, Richards, & Cuijpers, 2014) for the treatment of depression in primary health care. This is a convenience sample, since the participants of eight health units from the Agrupamento de Centros de Saúde do Porto Ocidental (ACeS Porto Ocidental) and 18 health professionals collaborated in this study. Participants had to meet the following inclusion criteria: (a) age between 18 and 75 years old; (b) a positive response to at least one of the three questions of the initial screening; (c) diagnosed with depression. In terms of education, 65 (43.05%) of the 153 participants completed primary education, 35 (35.76%) had secondary education and 32 (21.19%) had university education. Regarding occupation, 53 participants (34.67%) were employed, 39 (26%) were unemployed, 37 (24.67%) were retired, 13 (6.67%) were students and 9 (6%) were self-employed.

Measurements

Patient Health Questionnaire (PHQ-9)

The PHQ-9 (Kroenke et al., 2001) consists of 9 items that assess the severity of depression-related symptomatology on a 4-points Likert type scale (from 0 "*Not at all*" to 3 "*Nearly every day*"). The Portuguese version of the PHQ-9 was made available by Pfizer (www.phqscreeners.com).

Work and Social Adjustment Scale (WSAS)

The WSAS (Mundt, Marks, Shear, & Greist, 2002) consists of 5 items that assess the degree of deterioration of social and occupational adjustment on a 9-points Likert type scale (from 0 "*Not at all*", to 8 "*Very severely*"). In the original study, WSAS revealed good internal consistency under various conditions of use (α = .79 to .94), as well as a good test-retest correlation (*r* = .73). In our study, the WSAS revealed adequate internal consistency (α = .87, *99%CI BCa* [.81, .91], $\Omega_{categorical}$ = .89, *99%CI BCa* [.83, .94]).

Procedures

The study was conducted in 8 health care units of the ACeS Porto Ocidental with the collaboration of 18 health professionals (physicians and nurses). These health professionals were instructed to implement in their regular appointments a short screening procedure for all their patients between the ages of 18 and 75. This screening procedure aimed to evaluate the presence of depressed mood, anhedonia and suicidal ideation using three questions: (1) "*During the last month did you feel, for the most part of the time, particularly down, depressed or without hope?*"; (2) "*During the last month did you feel less pleasure or interest in doing things that previously gave you pleasure and/or had an interest in doing?*"; (3) "*In the last few weeks, did you think that life was not worth it, that it would be better to fall asleep and not wake up or that it would be better if you were not alive?*" This screening procedure was implemented with 771 patients. The PHQ-9, WSAS and/or a semi-structured interview, to respectively assess the severity of depression, the deterioration of personal, social or occupational adjustment and the risk of suicide were applied to patients who scored positively on screening (*n* = 153). Data regarding these 153 participants was used for this first study. Based on the severity of the symptoms and on the deterioration of the personal, social or occupational adjustment, as well as on the risk of suicide, these participants were then referred for low or high intensity treatments or referenced for specialized mental services (NICE, 2009).

The study received a positive opinion from the Ethics Committee and obtained authorization from the Comissão Nacional de Proteção de Dados. All participants gave informed consent prior to admission to the study.

Data analysis

Data analysis follows an exploratory and semi-confirmatory orientation through robust methods to accommodate possible limitations resulting from sample size. The univariate and multivariate normality analyses were performed, respectively, through the symmetry and kurtosis evaluation with the psych package and the MVN package for R (R Core Team, 2017). In this study, there were slight deviations of univariate normality and deviations of multivariate normality. The exploratory analysis procedures were implemented in the polychoric correlation matrix between the scores obtained for each of the items (e.g., Garrido, Abad, & Ponsoda, 2013).

Traditional methods tend to underestimate the number of factors in situations where the number of items is low and the factors reveal high correlations (Keith, Caemmerer, & Reynolds, 2016). The exploratory graph analysis (EGA; Golino & Demetriou, 2017; Golino & Epskamp, 2017) was recently developed to overcome these limitations. Within EGA, a graph refers to a network consisting of vertices (scale items), and the links between the vertices represent their associations. A gaussian graphic model was used to estimate the strength of each association. The extended bayesian information criterion (Foygel & Drton, 2010) was used to select the most parsimonious network while controlling for type I errors. Finally, the Fruchterman-Reingold algorithm (1991) was used to graphically represent the network of associations between items. Once the network of the associations between the items was estimated, it was then explored in order to identify subgroups of items that are most closely linked. These subgroups, called communities (Fortunato, 2010), correspond to factors in traditional exploratory factor analysis and were identified through the walktrap algorithm (Pons & Latapy, 2005). Vertices with similar structural characteristics were aggregated into subgroups that constitute the communities present in the network.

The EGA (Golino & Epskamp, 2017) and bootnet (Epskamp, Borsboom, & Fried, 2018) packages for R were respectively used for the exploratory analysis of the PHQ-9 network and to assess their validity and stability in this analysis. Once the communities were identified, a semi-confirmatory factorial analysis was performed. A rotation matrix was partially specified based on the information from the exploratory graph analysis. This matrix specified the number of communities (factors) in the EGA, as well as the items belonging to each of the communities. Subsequently, a matrix of polychoric correlations was estimated. The suitability of this matrix for the factorial analysis was evaluated through the determinant and sphericity tests of Bartlett and Kaiser-Meyer-Olkin (KMO). A robust form of the weighted least squares method (Yang-Wallentin, Jöreskog, & Luo, 2010) was applied on the asymptotic variance/covariance matrix, in order to extract the factors related to the PHQ-9. A procrustean rotation was then implemented in order to approximate the extracted factors to the rotation matrix initially specified. An oblique version of the procrustean rotation was implemented in order to allow the interaction between the factors. The adjustment of the post-rotation factor solution to the semi-specified solution was evaluated by the Tucker (1951) congruence coefficient between the two solutions.

The global adjustment of the factorial structure was evaluated through indexes selected on bases on the results of previous studies that considered ordinal data (Hutchinson & Olmos, 1998) and samples of similar size to that used in this study (Sivo, Fan, Witta, & Willse, 2006). Consistent with the previous proposals, indexes of different classes were also considered (Jaccard & Wan, 1996). These procedures for the semi-confirmatory factorial analysis were performed with the FACTOR software (Lorenzo-Seva & Ferrando, 2013). The categorical omega and Cronbach's alpha were calculated using the MBESS package for software R. The convergent validity between the PHQ-9 and the WSAS was explored using the Pearson correlation coefficient calculated between the total scores of the two instruments. This coefficient and the respective confidence intervals were calculated with the psychometric package for software R. To explore the criterion validity of the PHQ-9, the scores reported by the participants in the pre-treatment evaluation were stratified according to points defined by the previous investigation (Kroenke et al., 2001). Through the JASP software, a univariate analysis of variance was then used to compare the different strata relative to the total scores in the WSAS.

Results

Exploratory analysis of the PHQ-9 graph

This questionnaire can have a minimum score of 0 and a maximum score of 27. Scores ranged from 0 to 25 (*M* = 11.12, *99%CI BCa* [10.29, 12.00], *SD* = 5.20, *Variance* = 27.04) (Table 1). No substantial floor or ceiling effects were observed on the different items. As shown in Figure 1.a, three communities were identified (*Mdn* = 3.0, *SD* = 0.92, *95%CI BCa* [2.97, 3.03]). Community 1 includes: depressed mood and anhedonia. Community 2 includes: sleep problems, fatigue, changes in appetite, difficulty in concentration and retardation. Finally, community 3 includes: self-devaluation and suicidal ideation.

Semi-confirmatory factorial analysis of the PHQ-9

The correlation matrix was adequate for factor analysis, with the following results: low to moderate correlation coefficients (*M* = 0.31, *SD* = 0.13, *Variance* = 0.02, *Min* = 0.05, *Max* = 0.63); appropriate determinant (0.15); significant Bartlett sphericity test ($\chi^2_{(36)}$ = 283.70, p < .001); and appropriate KMO test (.76, *99%CI BCa* [.67, .84]). Table 2 presents the semi-confirmatory factorial solution. Overall, PHQ-9 items revealed a low-to-average but adequate commonality (MacCallum, Widaman, Zhang, & Hong, 1999) (*M* = 0.56, *SD* = 0.15, *Variance* = 0.02, *Min* = 0.25, *Max* = 1.00), except for item 3. All items showed adequate factor loadings, greater than .35 (Costello & Osborne, 2005). After oblique procrustean rotation, the factors explained 13.45%, 15.28% and 13.62% of the variance, respectively. The congruence index of items and factors is presented in Table 3. Item 4 showed a congruence slightly below the reference value (Lorenzo-Seva & ten Berge, 2006). The remaining items, the factors and the overall solution

presented optimal congruence with the pre-specified rotation matrix. Adjustment indices for the factorial solution also showed good adjustment (Table 4). The PHQ-9 as well as Factor 1 revealed adequate internal consistency (α = .77, *99%CI BCa* [.64, .85], $\Omega_{categorical}$ = .75, *99%CI BCa* [.61, .84]). Factor 2 (α = .62, *99%CI BCa* [.46, .73], $\Omega_{categorical}$ = .63, *99%CI BCa* [.45, .74]) and Factor 3 (α = .61, *99%CI BCa* [.38, .77], $\Omega_{categorical}$ = .69, *99%CI BCa* [.42, .85]) showed lower internal consistencies.

Table 1. *Descriptive characteristics of the PHQ-9 items*

Study		Item	*M* (*99%CI BCa*)	*SD*	*% Min*	*% Max*
Primary health care	**1**	**Anhedonia**	1.490 (1.294, 1.693)	0.981	15.686	19.608
	2	**Depressed mood**	1.601 (1.418, 1.784)	0.906	8.497	20.261
	3	**Sleep Problems**	1.667 (1.425, 1.908)	1.170	2,569	35.294
	4	**Fatigue**	1.882 (1.667, 2.085)	1.019	9.804	36.601
	5	**Appetite changes**	1.301 (1.065, 1.530)	1.107	30.719	19.608
	6	**Self-devaluation**	0.941 (0.725, 1.157)	1.102	49.020	14.379
	7	**Difficulty in concentration**	1.118 (0.915, 1.350)	1.038	35.948	12.418
	8	**Psychomotor retardation or agitation**	0.850 (0.647, 1.072)	0.958	47.059	7.190
	9	**Suicide ideation**	0.275 (0.150, 0.425)	0.681	81.699	3.922
University clinic	**1**	**Anhedonia**	1.722 (1.463, 1.981)	1.084	14.815	33.333
	2	**Depressed mood**	1.731 (1.472, 1.972)	1.073	14.815	32.407
	3	**Sleep Problems**	1.704 (1.445, 1.954)	1.035	12.037	30.556
	4	**Fatigue**	1.694 (1.463, 1.935)	0.990	12.037	25.926
	5	**Appetite changes**	1.287 (1.028, 1.556)	1.103	30.556	19.444
	6	**Self-devaluation**	1.463 (1.184, 1.741)	1.106	24.074	24.074
	7	**Difficulty in concentration**	1.315 (1.065, 1.556)	1.038	25.926	16.667
	8	**Psychomotor retardation or agitation**	1.000 (0.750, 1.250)	1.050	42.593	12.037
	9	**Suicide ideation**	0.611 (0.407, 0.833)	0.874	58.333	6.481

Note. M = mean. 99%CI BCa = bias corrected and accelerated confidence intervals for 99% of confidence. SD = standard deviation. % Min = percentage of responses in the minimum value of the Likert scale. % Max = percentage of responses in the maximum value of the Liker scale.

Table 2. *Factorial Solutions for the PHQ-9*

Study	Item	Factor Loadings (99%CI BCa)			Communalities
		Factor 1	**Factor 2**	**Factor 3**	
Primary health care	1	**0.803 (0.637, 2.810)**	-0.068 (-2.945, 0.138)	0.128 (-0.149, 0.565)	.696
	2	**0.705 (0.544, 1.928)**	0.004 (-2.380, 0.113)	0.124 (-0.153, 0.362)	.679
	3	0.097 (-0.345, 0.546)	**0.350 (-0.264, 0.693)**	0.100 (-0.148, 0.445)	.246
	4	0.295 (-1.595, 0.797)	**0.542 (0.054, 1.569)**	-0.252 (-0.698, -0.062)	.624
	5	0.006 (-0.645, 0.638)	**0.357 (-0.149, 2.939)**	-0.049 (-0.297, 0.372)	.503
	6	0.014 (-0.249, 0.279)	0.019 (-0.535, 0.323)	**0.752 (0.546, 1.366)**	.762
	7	-0.085 (-2.653, 0.167)	**0.727 (0.436, 7.245**)	0.037 (-0.209, 0.331)	.559
	8	-0.144 (-1.008, 0.386)	**0.683 (0.372, 2.498)**	0.135 (-0.166, 0.339)	.666
	9	0.112 (-0.134, 0.364)	-0.020 (-0.929, 0.225)	**0.773 (0.471, 1.172)**	.617
University clinic	1	-0.013 (-0.656, 0.318)	**0.955 (0.706, 3.373)**	-0.099 (-2.552, 0.429)	.916
	2	0.058 (-0.209, 0.316)	**0.906 (0.704, 3.169)**	-0.039 (-3.014, 0.293)	1.000
	3	-0.021 (-0.358, 0.628)	0.118 (-0.252, 0.621)	**0.583 (-0.409, 9.652)**	.510
	4	0.097 (-0.296, 0.551)	0.317 (-0.235, 0.812)	**0.434 (-0.452, 6.558)**	.678
	5	**0.720 (0.296, 1.161)**	-0.276 (-2.370, 0.101)	0.333 (-0.154, 1.449)	.617
	6	-0.109 (-0.948, 0.162)	**0.738 (-0.170, 1.474)**	0.254 (-0.431, 1.751)	.700
	7	**0.603 (0.356, 1.821)**	0.294 (-0.068, 0.689)	-0.159 (-1.684, 0.158)	.667
	8	**0.853 (0.423, 2.325)**	-0.079 (-1.580, 0.281)	-0.083 (-2.526, 0.297)	.665
	9	**0.674 (0.292, 2.337)**	0.086 (-0.345, 0.421)	-0.064 (-2.762, 0.295)	.533

Table 3. *Congruence of the factorial solution with the semi-specified rotation matrix for the PHQ-9*

Study	Item / Factor		Congruence *(99%CI BCa)*
Primary health care	1	Anhedonia	.984 (.804, 1.000)
	2	Depressed mood	.985 (.899, 1.000)
	3	Sleep Problems	.929 (.209, 1.000)
	4	Fatigue	.813 (-.551, .996)
	5	Appetite changes	.991 (.897, 1.000)
	6	Self-devaluation	.999 (.997, 1.000)
	7	Difficulty in concentration	.992 (.878, 1.000)
	8	Psychomotor retardation or agitation	.961 (.679, 1.000)
	9	Suicide ideation	.989 (.878, 1.000)
	Factor 1		.943 (.806, .995)
	Factor 2		.957 (.763, .997)
	Factor 3		.949 (.891, .995)
	Total		.935 (.848, .979)
University clinic	1	Anhedonia	.995 (.961, 1.000)
	2	Depressed mood	.997 (.974, 1.000)
	3	Sleep Problems	.979 (.380, 1.000)
	4	Fatigue	.795 (-.066, .998)
	5	Appetite changes	.857 (.450, .996)
	6	Self-devaluation	.937 (-.061, .998)
	7	Difficulty in concentration	.874 (.535, .999)
	8	Psychomotor retardation or agitation	.991 (.956, 1.000)
	9	Suicide ideation	.988 (.883, 1.000)
	Factor 1		.986 (.976, .987)
	Factor 2		.936 (.870, .984)
	Factor 3		.830 (.606, .997)
	Total		.927 (.885, .973)

Table 4. *Adjustment of the factorial solution and internal consistency of the PHQ-9*

		Primary health care	University clinic
Index of Adjustment	RMSCS	$\chi^2_{(12)} = 11.658$, $p = .481$	$\chi^2_{(12)} = 11.512$, $p = .493$
	RMSEA	.000 (.000, .066)	.000 (.000, .095)
	GFI	.986 (.975, 1.000)	1.000 (1.000, 1.000)
	AGFI	.959 (.925, 1.000)	1.000 (1.000, 1.000)
	NNFI	1.000 (.964, 1.000)	1.000 (.981, 1.000)
	CFI	1.000 (.988, 1.000)	1.000 (.994, 1.000)
Internal Consistency	Cronbach's Alfa	.747 (.654, .815)	.869 (.805, .912)
	Omega Categorical	.788 (.661, .857)	.837 (.760, .876)

Note. 99%CI BCa = bias corrected and accelerated confidence intervals for 99% of confidence. RMSCS = *robust mean-scaled chi-square*. RMSEA = *root mean square error of approximation*. GFI = *goodness-of-fit index*. AGFI = *adjusted goodness-of-fit index*. NNFI = *non-normed fit index*. CFI = *comparative fit index*.

Relationship between the PHQ-9 total score and sociodemographic variables

No significant effect of any of the sociodemographic variables was observed on the total scores of the questionnaire, namely gender ($t_{(37,14)} = -0.51$, $M_{dif} = -0.53$, *99%CI BCa* [-3.37, 2.31], $p = .62$) and occupation ($F_{(4, 145)} = 1.31$, $p = .27$, $\eta_p^2 = .04$). There was also no association with age ($r = -.09$, *99%CI BCa* [-.29, .12], $p = .27$) nor with schooling ($\rho = -.02$, $p = .81$).

Convergent validity and criterion validity

The PHQ-9 and the WSAS revealed a moderate-strong association (r = .67, *99%CI BCa* [.54, .77], p < .001). There was a significant effect of the severity of depression-related symptomatology on the impairment of social and professional adjustment ($F_{(4,148)}$ = 27.97, p < .001, η_p^2 = .43). Bonferroni's post-hoc test revealed significant differences between participants with subclinical symptomatology and: (a) moderate symptomatology (M_{dif} = -11.86, *99%CI BCa* [-17.25, -5.24], p < .001), (b) moderately severe symptomatology (M_{dif} = -20.38, *99%CI BCa* [-25.19, -13.40], p < .001), and (c) severe symptomatology (M_{dif} = -26.17, *99%CI BCa* [-32.29, -18.48], p < .001). There were also significant differences between the participants with mild symptoms and: (a) moderate symptomatology (M_{dif} = -5.70, *99%CI BCa* [-10.25, -1.03], p < .001), (b) with moderately severe symptomatology (M_{dif} = -14.22, *99%CI BCa* [-18.87, -8.87], p < .001), and (c) with severe symptomatology (M_{dif} = -20.01, *99%CI BCa* [-25.85, -13.77], p < .001). Finally, significant differences were observed between participants with moderate symptomatology and: (a) moderately severe symptomatology (M_{dif} = -8.52, *99%CI BCa* [-13.83, -2.90], p < .001) and (b) severe (M_{dif} = -14.31, *99%CI BCa* [-20.55, -7.68], p < .001).

Study 2: Psychometric qualities of the PHQ-9 in the context of university clinics

Methods

Participants

A sample of 106 adults (M_{age} = 32.47, SD_{age} = 11.01, Min_{age} = 18, Max_{age} = 62) participated in this study, of which 75 (70.78%) were female. It should be pointed out that the university clinic is open to the community and therefore does not attend exclusively university students; it is also dedicated to the psychological care of different problems in different age groups. This is a convenience sample, since the collection was carried out only in this university clinic, with the participation of 18 years-old patients and older and whose depression and anxiety assessment protocol was completed. Sixty-five (61.32%) of the 106 participants had a university education, 34 (32.08%) secondary education, 5 (4.72%) basic education and 2 (1.89%) primary education. Concerning their occupation, 49 (46.23%) were employed, 28 (26.42%) were students, 15 (14.15%) were working students, 10 (9.43%) were unemployed and 4 (3.77%) had other activities.

Measurements

Patient Health Questionnaire (PHQ-9)

Described in Study 1.

Beck Depression Inventory – II (BDI-II)

BDI-II (Beck et al., 1996; Campos & Gonçalves, 2011) consists of 21 items that evaluate the severity of symptomatology of depression on a 4-point Likert type scale (0 to 3). The BDI-II showed good internal consistency in previous research (α = .83 to .96; Wang & Gorenstein, 2013) and also in the Portuguese version (α = .89, Coelho, Martins & Barros, 2002; α = .90 and .91, Campos & Gonçalves, 2011). A cut-off point of 13 for major depression was estimated for the Portuguese version (Coelho et al., 2002). In this study, BDI-II (M = 22.63, SD = 12.95) showed good internal consistency (α = .93, 99%CI BCa [.90, .95], $\Omega_{categorical}$ = .96, 99%CI BCa [.93, .97]).

State-Trait Anxiety Inventory (STAI)

STAI (Spielberger, 1983; Silva, 2003) consists of 40 items that evaluate the severity of symptomatology of depression on a 4-point scale (1 to 4). STAI showed good internal consistency in previous research (α = .86 to .95, Spielberger, 1983) and also in the Portuguese version (α = .89 to .93, Silva, 2003). In this study (M = 111.65, SD = 22.10), STAI also showed good internal consistency (α = .96, *99%CI BCa* [.94, .97], $\Omega_{categorical}$ = .98, *99%CI BCa* [.89, 1.00]).

Procedures

Data regarding a sample of adult participants under psychotherapeutic treatment in a university clinic was analysed. As part of the psychotherapeutic treatment, the participants comply with an evaluation protocol, prior to initiation and during treatment, which includes evaluation of depression (through PHQ-9 and BDI-II) and anxiety (through STAI) related symptomatology. Pre-treatment evaluation data was used in

this study without involving the intervention of any psychologist. The questionnaires were made available to the participants by the administrative staff of the university clinic and were filled out in an autonomous manner. All participants were informed of the study conditions and gave their informed consent prior to their admission.

Data analysis

The same analysis procedures described for Study 1 were used to investigate the factorial structure and to implement the semi-confirmatory factorial analysis in the Study 2. Convergent validity was calculated between the total scores of the PHQ-9 and those of BDI-II and STAI. Considering the cut-off point identified for the Portuguese version of BDI-II for major depression, a ROC analysis was performed to identify the most appropriate cut-off point for the PHQ-9. The area under the curve (AUC) was considered low if lower than .70, moderate if between .70 and .90, and elevated if greater than .90 (Streiner & Cairney, 2007). The cut-off point was identified in order to maximize sensitivity and specificity through the Youden method with modifications proposed by Perkins and Schisterman (2006). Precision and accuracy of this cut-off point were also evaluated. The pROC and ROCR packages for R were used.

Results

Exploratory analysis of the PHQ-9 graph

Scores ranged from 0 to 27 (*M* = 12.53, *99%CI BCa* [10.86, 14.10], *SD* = 6.56, *Variance* = 42.962). Table 1 presents the descriptive characteristics for each of the PHQ-9 items. There was no evidence of floor or ceiling effects. Three communities were identified (*Mdn* = 3, *SD* = 1.07, *95%CI BCa* [2.96, 3.04]) (Figure 1.b). Community 1 includes: changes in appetite, difficulty in concentration, psychomotor retardation/agitation, and suicide ideation. Community 2 includes: depressed mood, anhedonia, and self-depreciation. Finally, community 3 includes: changes in sleep and fatigue.

Semi-confirmatory factorial analysis of the PHQ-9

The correlation matrix was adequate for factor analysis, with the following results: moderate to high correlation coefficients (*M* = 0.47, *SD* = 0.12, *Variance* = 0.02, *Min* = 0.27, *Max* = 0.81); appropriate determinant (0.02); significant Bartlett sphericity test ($\chi^2_{(36)} = 0.42$, $p < .001$); and appropriate KMO test (.85, *99%CI BCa* [.80, .91]). Table 2 presents the semi-confirmatory factorial solution. PHQ-9 items revealed a low-to-average but adequate commonality (*M* = 0.70, *SD* = 0.16, *Variance* = 0.03, *Min* = 0.51, *Max* = 1.00). With the exception of item 4, all items presented saturations greater than .51. After oblique procrustean rotation, the factors explained 20.76%, 26.10% and 8.42% of the variance. Regarding the index of congruence of items and factors, factor 3 showed a congruence slightly lower than the reference value (Table 3). The remaining factors, the global solution and the items revealed optimal congruence with the pre-specified rotation matrix. The PHQ-9 adjustment indices for the factorial solution in this sample also showed a good adjustment (Table 4). The PHQ-9 revealed adequate internal consistency, as well as Factor 1 (α = .76, *99%CI BCa* [.63, .85], $\Omega_{categorical}$ = .78, *99%CI BCa* [.63, .63]) and Factor 2 (α = .88, *99%CI BCa* [.80, .92], $\Omega_{categorical}$ = .89, *99%CI BCa* [.82, .93]). Factor 3 (α = .58, *99%CI BCa* [.31, .76], $\Omega_{categorical}$ = .60, *99%CI BCa* [.29, .77]) showed lower internal consistencies.

Convergent validity and cut-off point

The PHQ-9 revealed a high association with BDI-II (r = .79, *99%CI BCa* [.67, .88], $p < .001$) and moderate association with STAI (r = .66, *99%CI BCa* [.46, .78], $p < .001$). ROC analysis revealed high AUC (.92, *99%CI* [.84, .99]) (Figure 2.a). A cut-off point of 9 revealed high sensitivity (.89, *99%CI* [.68, 1.00]) and high specificity (.90, *99%CI* [.81, .98]), as well as high accuracy (.96) and precision (.90) (Figure 2.b).

Relationship between the PHQ-9 total score and sociodemographic variables

No significant effect of any of the sociodemographic variables was observed on the total scores of the

PHQ-9, namely for gender ($t_{(104,00)}$ = 0.32, M_{dif} = 0.45, *99%CI BCa* [-3.26, 4.16], p = .75) and for occupation ($F_{(4,101)}$ = 1.80, p = .13, η_p^2 = .07). There was also no association with age (r = -.02, *99%CI BCa* [-.27, .23], p = .82) nor with schooling (ρ = -.03, p = .67).

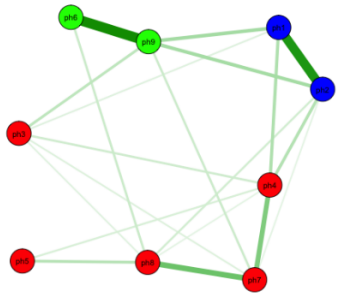

Figure 1.a. *Primary health care study*

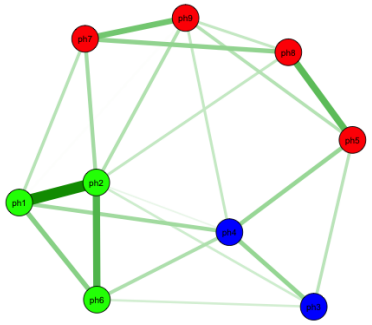

Figure 1.b. *University clinic study*

Figure 1. Exploratory analysis of the PHQ-9 graph. *Figure* 1.a. and Figure 1.b. represent the network of interactions between PHQ-9 items for primary health care units and for university clinics, respectively. The circles represent the items in PHQ-9. PH1 = Anhedonia; PH2 = Depressed mood; PH3 = Sleep problems; PH4 = Fatigue; PH5 = Changes in appetite; PH6 = Self-devaluation; PH7 = Difficulty in concentration; PH8 = Psychomotor retardation or agitation; PH9 = Suicide ideation. The green lines represent the positive associations between the items. The thickness of the lines represents the magnitude of these associations. The colours of the circles represent the communities identified in the exploratory analysis of the PHQ-9 graph.

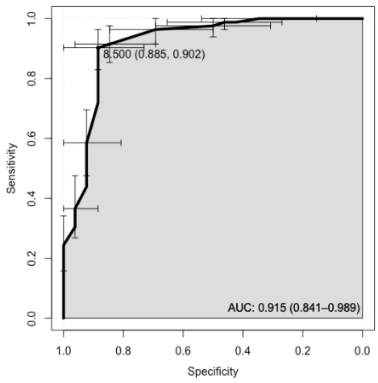

Figure 2.a. *Area under the Curve, Sensitivity, Specificity*

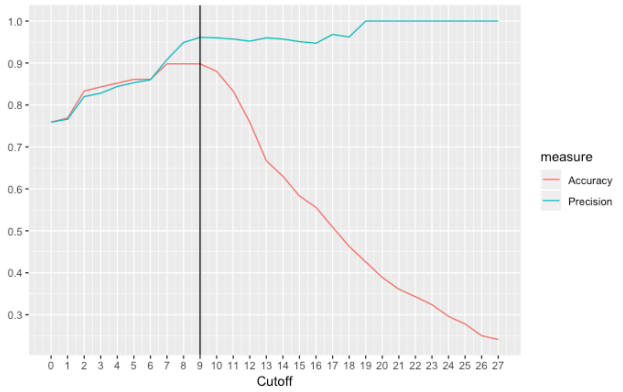

Figure 2.b. *Precision and Accuracy*

Figure 2. ROC analysis for the PHQ-9. *Figure 2.a*. The area under the curve (AUC) is represented by the grey colour. Its value and respective confidence intervals (99%) are displayed at the lower right corner. On the black line is shown the value corresponding to the cut-off point identified by the Youden method with the changes proposed by Perkins and Schisterman (2006), as well as the corresponding values regarding specificity and sensitivity. The narrowest horizontal and vertical lines represent the confidence intervals (99%), relative to specificity and sensitivity, for each of the considered cut-off points. *Figure 2.b*. Precision and accuracy related to the cut-off point determined in the ROC analysis.

Discussion

The aim of this article was to explore the psychometric qualities of the Portuguese version of the PHQ-9 in two different clinical contexts: in primary health care settings and in a university clinic setting. The results revealed that the PHQ-9 is a valid measure for the evaluation of depression in the two mentioned contexts.

First, it should be mentioned that the PHQ-9 showed good internal consistency in the two studies, similar to that observed in previous studies (e.g., Kroenke et al., 2001). In addition, PHQ-9 revealed a moderate association with WSAS in Study 1, as well as a moderate association with STAI and a high association with BDI-II, in Study 2, demonstrating good convergent validity with measures of social and professional functioning evaluation, anxiety and depression, respectively. The

association with measures of anxiety and depression has been widely documented in the literature; so, these results are in line with most studies carried out with different populations and contexts (e.g., Monteiro et al., 2013). On the contrary, the association of this scale with measures of personal, social and professional adjustment is not always found in the literature, and is often low (McKnight, Monfort, Kashdan, Blalock, & Calton, 2016). In this study, however, the severity of depression-related symptoms is associated with the deterioration observed in these dimensions of adjustment, which validates their joint use in clinical practice.

One of the objectives of the study here described was to verify whether the factorial structures produced by the two studies, conducted in different clinical contexts, were similar or different from each other. It was observed that, in the two studies, different factorial structures were found, not only when compared directly with each other, but also when compared with previous studies, as mentioned in the literature (Manea et al., 2012). In Study 1, three communities/factors were identified: (1) depressed mood and anhedonia, that explained 13.45% of the variance; (2) self-devaluation and suicidal ideation, that explained 13.62% of the variance; and (3) sleep problems, appetite and difficulty in concentration, fatigue and psychomotor retardation/agitation, which explained 15.28% of the variance. Similarly, in Study 2, three communities/factors were identified, however composed by different items: (1) sleep problems and fatigue, that explained 8.42% of the variance; (2) changes in appetite and difficulty in concentration, psychomotor agitation/retardation, and suicidal ideation, that explained 20.76% of the variance; and (3) depressed mood, anhedonia and self-devaluation, that explained 26.10% of the variance. As we can see, we found factorial structures that differ according to the context where the PHQ-9 was applied. This result, although it may generate some confusion, is expected considering the literature. Indeed, it is important to emphasize that the diversity of factorial structures observed in this article is consistent not only with previous research on the questionnaire (Elhai et al., 2012), but also with that observed in other instruments for the evaluation of depression (Shafer, 2006) and with the diagnosis of depression itself (Van Loo, De Jonge, Romeijn, Kessler, & Schoevers, 2012). In fact, in previous studies, it has been observed that some psychometric characteristics of the PHQ-9 vary according to the context (Manea et al., 2012), which may have contributed to the diverse observations in the two studies reported here. It has recently been underlined that these variations are somewhat predictable, since the symptoms of depression differ in their sensitivity to various risk factors, their impact on the various dimensions of psychological adjustment, and their association with neurological and genetic components (Fried & Nesse, 2015). The results found in this article seem to point precisely to the influence of the clinical context on the psychometric qualities of the scale, which may be related to important differences in sociodemographic characteristics and risk factors of the two samples used, with relevant expression in the factorial structures found. From a psychometric perspective, these observations suggest that symptoms of depression may not be temporally stable, nor equally adequate indicators of depression (Schmittmann et al., 2013) in different clinical contexts, which may explain the changes in the factorial structure of instruments that assess depression over time (Fried et al., 2016). From the clinical point of view, it suggests that it is important for health professionals to be alert to different symptom configurations, which may reflect different risk factors and vulnerabilities of a given clinical population.

As noted earlier, EGA (Golino & Demetriou, 2017) allows us to produce more detailed factor structures in situations where: (a) the number of participants is less than 500, (b) the number of items is reduced and (c) the correlation between factors is high. Thus, the higher sensitivity of EGA has allowed the identification of three factors in the two studies here presented, unlike previous research that has found factorial structures with one or two factors (Elhai et al., 2012). In fact, the PHQ-9 factorial structures observed in Study 1 and Study 2 differ from the more adjusted factorial structure observed in a meta-analysis that

compared a one-factorial structure and three structures consisting of two factors (Elhai et al., 2012). In that meta-analysis it is reported that the structures constituted by two factors showed to be more adjusted than the one-factorial structure, and that the most adequate was the one that separated somatic factors from non-somatic ones. In such structure, the items related to depressed mood, anhedonia, self-devaluation and suicidal ideation appear in the same (non-somatic) factor, contrary to what happens in Study 1 and Study 2 here reported, in which those items appear separated by two factors. However, it is important to note that the items related to depressed mood and anhedonia emerged now in the same factor, presenting a good internal consistency. These results suggest, therefore, that these two symptom-related items may constitute a valid and rapid measure of depression screening in different clinical contexts, being extremely important for the clinical practice of different health professionals. These data are in line with the items that are included in the PHQ-2 (Kroenke, Spitzer, & Williams, 2003), a reduced version of the PHQ-9. PHQ-2 has shown good psychometric properties as well as a high utility in professional contexts that benefit from brief screening instruments (Arroll et al., 2010). Thus, although they produce distinct factorial structures, the two studies described in this article reinforce the importance of using these two symptoms – anhedonia and depressed mood – as a useful solution for the screening and brief evaluation of depression, which is fundamental for the recognition and detection of this problem in clinical contexts with short contact-time characteristics (Wright, 1994).

The two studies reported in this paper present preliminary evidence supporting a cut-off point to distinguish major depression, consistent with previous research (Manea et al., 2012). In fact, the cut-off point 9 obtained in this article shows high sensitivity, specificity, precision and accuracy, which reinforces its use as a reference value for the diagnosis of major depression disorder (scores higher than 9 suggesting major depression). In addition, the obtained results support the intervals for severity of the depression-related symptoms proposed by previous research (Kroenke et al., 2001), namely: subclinical symptomatology, mild symptomatology, moderate symptomatology, moderately severe symptomatology and severe symptomatology. However, it is important to point out that these results should be confirmed by future studies using as a reference criterion the diagnosis performed by researchers trained in the application of semi-structured interviews for the diagnostic assessment of mental disorders.

It is important to note that the application of the PHQ 9 should not replace a more exhaustive and thorough evaluation performed by the health professional. Although these studies demonstrate that the PHQ-9 provides reliable and valid data for the evaluation of depression, it is highly pertinent that health professionals combine different sources of information so that the formal diagnosis of this problem is as rigorous and supported as possible, in order to avoid false positives and false negatives.

Limitations

Notwithstanding the relevant contributions of this article, some limitations should be considered. Given the limited number of participants in the two reported studies, it is important that future studies use larger samples in confirmatory factual analyses, in order to confirm the factorial structures here observed and to consolidate normative information. It is also important to note that the differences in the data collection procedures of the two studies may have influenced the results obtained (for example, the fact that in one study there was a screening prior to the PHQ-9 completion, which did not occur in the other). In addition, in Study 1, false positives and/or excluded false negatives may have been included because of the screening process, which may have also had an impact on the results obtained. Additionally, this investigation did not evaluate the test-retest fidelity of the PHQ-9 nor the sensitivity to the changes generated by the treatments. Thus, subsequent studies should explore these psychometric characteristics in the Portuguese version of the PHQ-9.

Conclusions

This article appears to be an important contribution to the screening and evaluation of depression in that it enables the signalling of clinically relevant situations,

assisting different health care professionals in decision-making concerning the diagnosis of depression in different contexts, and promoting, ultimately the timely treatment of cases of depression. In addition, it contributes greatly to the research, since it provides data obtained from robust statistical methods that allowed to find more detailed factor structures, surpassing the limitations of the traditional methods. The psychometric qualities of the Portuguese version of the PHQ-9, as well as its cut-off point and the criteria for delimiting its degrees of severity of depression-related symptomatology are now available to aid future research.

Acknowledgments

We would like to thank all the participants, as well as the health professionals who collaborated in the study. We would also like to thank the Agrupamento dos Centros de Saúde (ACeS) do Porto Ocidental, the Administração Central do Sistema de Saúde (ACSS) and the European Economic Area Grants (EEA) for their support throughout the study.

Funding

This study was funded by the European Economic Area Grants (PT06) mechanism (EEA Grants) - Public Health Initiatives Program – and by the Fundação para a Ciência e Tecnologia and COMPETE 2020, through the grants: PTDC/MHC-PCL/1991/2014 grant; CPUP UID/PSI/00050/2013; FEDER/COMPETE2020 POCI-01-0145-FEDER-007294.

Declaration of Conflicting Interests

The authors declare no conflicts of interest with respect to the research, authorship, and/or publication of this article.

Supplementary material

The supplementary material for this article can be found online here.

References

Aaronson, N., Ahmedzai, S., Bergman, B., Bullinger, A., Cull, A., Duez, N., ...Takeda, F. (1993). The European Organization for Research and Treatment of Cancer QLQ – C30: A quality-of-life instrument for use in international clinical trials in oncology. *Journal of the National Cancer Institute, 85(5)*, 365-376. doi:10.1093/jnci/85.5.365

Adewuya, A. O., Ola, B. A., & Afolabi, O. O. (2006). Validity of the Patient Health Question-naire (PHQ-9) as a screening tool for depression amongst Nigerian university students. *Journal of Affective Disorders, 96(1-2)*, 89-93. doi:10.1016/j.jad.2006.05.021

Almeida, J. A. C., Xavier, A., Cardoso, G., Pereira, A., Gusmão, R., Corrêa, B., ... Silva, J. (2013). *Estudo epidemiológico nacional de saúde mental: 1º relatório*. Lisboa: Faculdade de Ciências Médicas, Universidade Nova de Lisboa.

American Psychiatric Association (2006). DSM-IV-TR: *Manual de Diagnóstico das Perturbações Mentais (4ª Edição)*. Lisbon: Climepsi Editores.

Arrieta, J., Aguerrebere, A., Raviola, G., Flores, H., Elliott, P., Espinosa, A., ... Palazuelos, D. (2017). Validity and utility of the Patient Health Questionnaire (PHQ)-2 and PHQ-9 for screening and diagnosis of depression in rural Chiapas, Mexico: A cross-sectional study. *Journal of Clinical Psychology, 73(9)*, 1076-1090. doi:10.1002/jclp.22390

Arroll, B., Goodyear-Smith, F., Crengle, S., Gunn, J., Kerse, N., Fishman, T., ... Hatcher, S. (2010). Validation of PHQ-2 and PHQ-9 to screen for major depression in the primary care population. *Annals of Family Medicine, 8(4)*, 348 – 353. doi:10.1370/afm.1139.

Beck, A.T., Steer, R.A., & Brown, G. (1996). *BDI-II: Beck Depression Inventory-II Manual*. New York: Psychological Corporation.

Cameron, I. A., Crawford, J. R., Lawton, K., & Reid, I. C. (2008). Psychometric comparison of PHQ-9 and HADS for measuring depression severity in primary care. *British Journal of General Practice, 58(546)*, 32 – 36. doi:10.3399/bjgp08X263794.

Chen, S., Chiu, H., Xu, B., Ma, Y., Jin, T., Wu, A., & Conwell, Y. (2010). Reliability and validity of the PHQ-9 for screening late-life depression in Chinese primary care. International *Journal of Geriatric Psychiatry, 25(11)*, 1127-1133. doi:10.1002/gps.2442

Campos, R. C., & Gonçalves, B. (2011). The Portuguese version of the Beck Depression Inventory-II (BDI-II). *European Journal of Psychological Assessment*, 27, 258-264. http://dx.doi.org/10.1027/1015-5759/a000072

Campos, R. C., & Gonçalves, B. (2011). The Portuguese version of the Beck Depression Inventory-II (BDI-II): Preliminary psychometric data with two nonclinical samples. *European Journal of Psychological Assessment, 27(4)*, 258-264. doi:10.1027/1015-5759/a000072

Chilcot, J., Rayner, L., Lee, W., Price, A., Goodwin, L., Monroe, B., ... Hotopf, A. (2013). *The factor structure of the PHQ-9 in palliative care. Journal of Psychosomatic Research, 75(1)*, 60 – 64. doi: 10.1016/j.jpsychores.2012.12.012

Coelho, R., Martins, A., & Barros, H. (2002). Clinical profiles relating gender and depressive symptoms among adolescents ascertained by the Beck Depression Inventory II. *European Psychiatry, 17(4)*, 222 – 226. https://doi.org/10.1016/S0924-9338(02)00663-6

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical assessment, research & evaluation, 10(7)*, 1-9.

Elhai, J. D., Contractor, A. A., Tamburrino, A., Fine, T. H., Prescott, A. R., Shirley, E., ... Calabrese, J. R. (2012). The factor structure of major depression symptoms: A test of four competing models using the Patient Health Questionnaire-9. *Psychiatry Research, 199(3)*, 169-173. doi:10.1016/j.psychres.2012.05.018

Epskamp, S., Borsboom, D., & Fried, E. I. (2018). Estimating psychological networks and their stability: A tutorial paper. *Behav Res Methods, 50(1)*, 195-212. doi: 10.3758/s13428-017-0862-1.

Fiest, K. A., Patten, S. B., Wiebe, S., Bulloch, A. G., Maxwell, C. J., & Jetté, N. (2014). Vali-dating screening tools for depression in epilepsy. *Epilepsia, 55(10)*, 1642-1650. doi:10.1111/epi.12754.

Fried, E. I., & Nesse, R. A. (2015). Depression sum-scores don't add up: Why analyzing specific depression symptoms is essential. *BMC Medicine, 13(1)*, 72. doi:10.1186/s12916-015-0325-4

Fried, E. I., van Borkulo, C. D., Epskamp, S., Schoevers, R. A., Tuerlinckx, F., & Borsboom, D. (2016). Measuring depression over time... Or not? Lack of unidimensionality and longitudinal measurement invariance in four common rating scales of depression. *Psychological Assessment, 28(11)*, 1354 – 1367. doi:10.1037/pas0000275

Fortunato, S. (2010). Community detection in graphs. *Physics reports, 486(3)*, 75 – 174. doi:10.1016/j.physrep.2009.11.002

Foygel, R., & Drton, A. (2010). Extended Bayesian information criteria for Gaussian graphical models. In J.D. Lafferty, C.K.I. Williams, J. Shawe-Taylor, R.S. Zemel, & A. Culotta (Eds.), *Advances in Neural Information Processing Systems 23 (NIPS 2010)* (pp. 604 – 612). Curran Associates, Inc.

Fruchterman, T., & Reingold, E. (1991). Graph drawing by force-directed placement. *Software: Practice and Experience, 21(11)*, 1129–1164. doi:10.1002/spe.4380211102

Garrido, L. E., Abad, F. J., & Ponsoda, V. (2013). A new look at Horn's parallel analysis with ordinal variables. *Psychological Methods, 18(4)*, 454 – 474. doi:10.1037/a0030005.

Golino, H. F., & Demetriou, A. (2017). Estimating the dimensionality of intelligence like data using Exploratory Graph Analysis. *Intelligence, 62*, 54-70. doi:10.1016/j.intell.2017.02.007

Golino, H. F., & Epskamp, S. (2017). Exploratory graph analysis: A new approach for estimating the number of dimensions in psychological research. *PLoS One, 8;12(6):e0174035*. doi: 10.1371/journal.pone.0174035..

Hutchinson, S. R., & Olmos, A. (1998). Behaviour of descriptive fit indices in confirmatory factor analysis using ordered categorical data. *Structural Equation Modelling, 5*, 344 – 364. doi:10.1080/10705519809540111

Jaccard, J., & Wan, C. K. (1996). *LISREL approaches to interaction effects in multiple regression*. Thousand Oaks, CA: Sage Publications.

Keith, T. Z., Caemmerer, J. A., & Reynolds, A. R. (2016). Comparison of methods for factor extraction for cognitive test-like data: Which overfactor, which underfactor?. *Intelligence, 54*, 37-54. doi:10.1016/j.intell.2015.11.003

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9. *Journal of General Internal Medicine, 16(9)*, 606 – 613. doi:10.1046/j.1525-1497.2001.016009606.x

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2003). The Patient Health Questionnaire-2: Validity of a two-item depression screener. *Medical Care, 41(11)*, 1284-1292. doi:10.1097/01.MLR.0000093487.78664.3C

Lorenzo-Seva, U., & Ferrando, P. J. (2013). Factor 9.2: A comprehensive program for fitting exploratory and semiconfirmatory factor analysis and IRT models. *Applied Psychological Measurement, 37*(6), 497-498. doi: 10.1177/0146621613487794

Lorenzo-Seva, U. & ten Berge, J.A.F. (2006). Tucker's congruence coefficient as a meaningful index of factor similarity. *Methodology, 2*, 57 – 64. doi:10.1027/1614-2241.2.2.57

Löwe, B., Schenkel, I., Carney-Doebbeling, C., & Göbel, C. (2006). Responsiveness of the PHQ-9 to psychopharmacological depression treatment. *Psychosomatics, 47(1)*, 62 – 67. doi:10.1176/appi.psy.47.1.62

MacCallum, R. C., Widaman, K. F., Zhang, S. B., & Hong, S. H. (1999). Sample size in factor analysis. *Psychological Methods, 4(1)*, 84-99. doi:10.1037/1082-989X.4.1.84

Manea, L., Gilbody, S., & McMillan, D. (2012). Optimal cut-off score for diagnosing depression with the Patient Health Questionnaire (PHQ-9): A meta-analysis. *Canadian Medical Association Journal, 184(3)*, E191 – E196. doi:10.1503/cmaj.110829

Martin, A., Rief, W., Klaiberg, A., & Braehler, E. (2006). Validity of the brief Patient Health Questionnaire mood scale (PHQ-9) in the general population. *General Hospital Psychiatry, 28*, 71–77. doi:10.1016/j.genhosppsych.2005.07.003

McKnight, P. E., Monfort, S. S., Kashdan, T. B., Blalock, D. V., & Calton, J. A. (2016). Anxiety symptoms and functional impairment: A systematic review of the correlation between the two measures. *Clinical Psychology Review, 45*, 115 – 130. doi:10.1016/j.cpr.2015.10.005

Monteiro, S., Torres, A., Pereira, A., Albuquerque, E., & Morgadinho, R. (2013). Preliminary validation study of a Portuguese version of the Patient Health Questionnaire (PHQ-9). *European Psychiatry, 28, Supplemment, 1*. doi:10.1016/S0924-9338(13)76982-7

Mundt, J., Marks, I., Shear, A., & Greist, J. (2002). The Work and Social Adjustment Scale: A simple measure of impairment in functioning. *British Journal of Psychiatry, 180*, 461 – 464. doi:10.1192/bjp.180.5.461

National Institute for Health and Clinical Excellence (2009). *Depression in adults: The treatment and management of depression in adults*. NICE Clinical Guideline 90. London: NICE.

Pais-Ribeiro, J., Pinto, C., & Santos, C. (2008). Validation study of the Portuguese version QLC-C30- V.3. *Psicologia, Saúde & Doenças, 9(1)*, 89-102.

Pais-Ribeiro, J., Silva, I., Ferreira, T., Martins, A., Meneses, R., & Baltar, A. (2007). Validation study of a Portuguese version of the Hospital Anxiety and Depression Scale. *Psychology, Health & Medicine, 12(2)*, 225-237. doi:10.1080/13548500500524088

Perkins, N. J., & Schisterman, E. F. (2006). The inconsistency of "optimal" cutpoints obtained using two criteria based on the Receiver Operating Characteristic Curve. *American Journal of Epidemiology, 163(7)*, 670–675. doi:10.1093/aje/kwj063

Pettersson, A., Boström, K. B., Gustavsson, P., & Ekselius, L. (2015). Which instruments to support diagnosis of depression have sufficient accuracy? A systematic review. *Nordic Journal of Psychiatry, 69(7), 497-508*. doi: 10.3109/08039488.2015.1008568

Pons, P., & Latapy, A. (2005). Computing communities in large networks using random walks. In *Computer and Information Sciences-ISCIS 2005* (pp. 284 – 293). Springer Berlin Heidelberg.

R Core Team (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/. [Computer software]

Richards, D. (2011). Prevalence and clinical course of depression: A review. *Clinical Psychology Review, 31*, 1117–1125. doi:10.1016/j.cpr.2011.07.004

Schmittmann, V. D., Cramer, A. O., Waldorp, L. J., Epskamp, S., Kievit, R. A., & Borsboom, D. (2013). Deconstructing the construct: A network perspective on psychological phenomena. *New Ideas in Psychology, 31(1)*, 43-53. doi:10.1016/j.newideapsych.2011.02.007

Shafer, A. B. (2006), Meta-analysis of the factor structures of four depression questionnaires: Beck, CES-D, Hamilton, and Zung. *Journal of Clinical Psychology, 62*, 123-146. doi:10.1002/jclp.20213

Silva, D. R. (2003). O Inventário de Estado-Traço de Ansiedade (S.T.A.I.). In A. A. Gonçalves, A. R. Simões, L. S. Almeida, & C. Machado (Eds.), *Avaliação Psicológica: Instrumentos validados para a população portuguesa (vol. 1, pp. 45-63)*. Coimbra: Quarteto.

Sivo, S. A., Fan, X., Witta, E. L., & Willse, J. T. (2006). The search for "optimal" cutoff properties: Fit index criteria in structural equation modelling. *The Journal of Experimental Education, 74(3)*, 267-288. https://doi.org/10.3200/JEXE.74.3.267-288

Spielberger, C. D. (1983). *Manual for the State-Trait Anxiety Inventory (STAI).* Palo Alto, CA: Consulting Psychologists Press.

Stewart, A. L., Hays, R. D., & Ware, J. E. (1988). The MOS short-form general health survey: Reliability and validity in a patient population. *Medical Care, 26(7)*, 724 – 735.

Streiner, D. L., & Cairney, J. (2007). What's under the ROC? An introduction to receiver operating characteristics curves. Canadian *Journal of Psychiatry, 52*, 121–128. doi:10.1177/070674370705200210

Torres, A., Monteiro, S., Pereira, A., & Albuquerque, E. (2016). Reliability and validity of the PHQ-9 in Portuguese women with breast cancer. In *2nd International Conference on Health and Health Psychology Reliability*. doi:10.15405/epsbs.2016.07.02.38

Tucker, L. R. (1951). A method for synthesis of factor analysis studies. *Personnel Research Section Report, 984*. Washington, D. C.: Department of the Army.

Van Loo, H. A., De Jonge, P., Romeijn, J. W., Kessler, R. C., & Schoevers, R. A. (2012). Data-driven subtypes of major depressive disorder: A systematic review. *BMC Medicine, 10(1)*, 156. doi:10.1186/1741-7015-10-156

Van Straten, A., Hill, J., Richards, D. A., & Cuijpers, P. (2014). Stepped care treatment delivery for depression: A systematic review and meta-analysis. *Psychological Medicine, 45*, 1-16. doi:10.1017/S0033291714000701

Wang, Y. P., & Gorenstein, C. (2013). Psychometric properties of the Beck Depression Inventory-II: A comprehensive review. *Revista Brasileira de Psiquiatria, 35(4)*, 416-431. doi:10.1590/1516-4446-2012-1048

World Health Organization (2017). *Depression and other common mental disorders: Global health estimates*. From http://apps.who.int/iris/bitstream/10665/254610/1/WHO-MSD-MER-2017.2-eng.pdf?ua=1

Wright, A. F. (1994). Should general practitioners be testing for depression? *British Journal of General Practice, 144*, 132–135. doi: 10.4324/9780203408582_chapter_20

Yang-Wallentin, F., Jöreskog, K. G., & Luo, H. (2010). Confirmatory factor analysis of ordinal variables with misspecified models. *Structural Equation Modelling, 17(3)*, 392 – 423. doi:10.1080/10705511.2010.489003

Zigmond, A. S., & Snaith, R. P. (1983). The Hospital Anxiety and Depression Scale. *Acta Psychiatrica Scandinavia, 67*, 361-370. doi:10.1111/j.1600-0447.1983.tb09716.x.