UDK: 616.89-008.454-07; DOI: 10.7251/STED0415020S

THE FACTOR STRUCTURE OF THE PATIENT HEALTH QUESTIONNAIRE (PHQ-9) IN A NON-CLINICAL SAMPLE

Siniša Subotić[9], Ivan Knežević[10], Strahinja Dimitrijević[11], Dalibor Miholjčić[12], Savina Šmit[13], Mirjana Karać[14], Jelena Mijatović[15]

The Patient Health Questionnaire – PHQ-9 is a well known self-report measure of nine depression symptoms according to DSM-IV/DSM-V criteria. As the PHQ-9 is not yet validated in most of former Yugoslavian countries, the goal of this research was to determine its factor structure and psychometric properties on a large non-clinical BCS language speaking sample. A total of 1875 participants (61.5% female), the average age of 28.26 (SD=8.32) years, completed the PHQ-9 via an anonymous online survey. WLSMV/DWLS based confirmatory factor analysis (CFA) revealed that two-factor model with correlated (.83) Cognitive/affective and Somatic factors fits the data well ($\chi^2(26)=287.8$, $p<.001$; CFI=0.972, NNFI/TLI=0.961, RMSEA=0.073, 90% CI [0.066, 0.081]), better than single-factor solution ($\chi^2(27)=444.2$, $p<.001$; CFI=0.956, NNFI/TLI=0.941, RMSEA=0.091, 90% CI [0.083, 0.098]). The two-factor model also fits better in other research in non-clinical samples (e.g. palliative care), while unidimensionality is detected in clinical/psychiatric samples. The two-factor model had strong gender invariance (i.e. configural invariance + equal loadings + equal thresholds; $\Delta\chi^2(\Delta df)=9.48(14.75)$, $p=.839$, ΔCFI<.001). However, after adding the equal means constrains, the model remained invariant based on ΔCFI criterion (.008), but became non-invariant based on $\Delta\chi^2(\Delta df)$ criterion (31.63(14.75), $p=.006$). The source of potential invariance was the higher Somatic score for females ($M_{females}=3.32$, $SD_{females}=2.22$, $M_{males}=2.89$, $SD_{males}=2.23$, $t(1873)=4.15$, $p<.001$), but this difference was just at the cutoff between trivial and small effect size ($d=0.197$). Finally, the results of bifactor analysis and good reliability of the whole scale ($\omega=.89$) suggested that using a single PHQ-9 score is probably advisable for most purposes, but researchers should use a bifactor approach to test Cognitive/affective and Somatic domain specific hypotheses. In conclusion, the PHQ-9 exhibits well-fitting latent structure, it has strong gender factor invariance, and good reliability, suggesting good potential for its research purposes use in BCS language. However, its convergent and discriminative validation and norming on clinical samples are pending.

Keywords: *The Patient Health Questionnaire – PHQ-9, Depression, Confirmatory factor analysis (CFA), Bifactor analysis, BCS language*

[9] PIM University & NGO "Persona" Banja Luka, Bosnia and Herzegovina; CEON/CEES, Belgrade, Serbia; Mobile phone: +38765/299-873; Email: sinisasub@gmail.com

[10] Educational advisory center "Auxilium" (within "PRONI" center for youth development), Brčko, Bosnia and Herzegovina; NGO "Persona"& Psychology MA program at PIM University, Banja Luka, Bosnia and Herzegovina; Mobile phone: +38765/841-794; Email: efremhas@gmail.com

[11] University of Banja Luka, Bosnia and Herzegovina; Mobile phone: +38765/887-131; Email: strahinja.dimitrijevic@unibl.rs

[12] NGO "Action Against AIDS" & NGO "Persona" & Psychology MA program at PIM University, Banja Luka, Bosnia and Herzegovina; Mobile phone: +38765/639-089; Email: dalibormiholjcic@gmail.com

[13] Psychology BA program at PIM University, Banja Luka, Bosnia and Herzegovina; Mobile phone: +38765/169-372; Email: savinasmit@yahoo.com

[14] Psychology BA program at PIM University, Banja Luka, Bosnia and Herzegovina; Mobile phone: +38766/681-831; Email: mirka.80@hotmail.com

[15] Psychology BA program at PIM University, Banja Luka, Bosnia and Herzegovina; Mobile phone: +38765/968-148; Email: jmijatovic88@hotmail.com

Introduction

Depressive disorders are common mental disorders, affecting more than 350 million people of all ages worldwide (World Health Organization [WHO], 2012). They are a leading cause of disability adjusted life years (DALYs), and a second leading cause of years lived with disability (YLDs) (Ferrari et al., 2013). While having short and valid screening tools for depression is certainly not enough to achieve better outcomes, it is arguably still an important starting point (Gilbody, Sheldon, & House, 2008; Williams, Pignone, Ramirez, & Stellato, 2002). The Patient Health Questionnaire (PHQ-9; Kroenke & Spitzer, 2002; Kroenke, Spitzer, & Williams, 2001) is one of the most known, public domain self-report depression screening tools, which is, according to the most recent meta-analysis (Moriarty, Gilbody, McMillan, & Manea, 2015), "widely used in both clinical and research settings" (p. 1). It shows good diagnostic properties in various settings, but best in primary care (Gilbody, et al., 2008; National Collaborating Centre for Mental Health, 2010; Moriarty et al., 2015).

However, the PHQ-9 is not yet validated, or even widely used in most of the former Yugoslavian countries (with a partial exception of Slovenia; see, e.g., Kozel et al., 2012) and psychometric properties and factor structure of the PHQ-9 in this area are unknown. Thus, the goal of this research was to determine the factor structure and psychometric properties of the PHQ-9 in a non-clinical sample, as the first step in a potential facilitation of its wider local use and as a prequel for its further validation in clinical settings. Previous research on general population (e.g., Germany: Martin, Rief, Klaiberg, & Braehler, 2006; China: Yu, Tam, Wong, Lam, & Stewart, 2012) demonstrated favorable results. We limited this research on Bosnian-Croatian-Serbian (BCS) language speaking parts of the former Yugoslavia.

From other research, based on the results of the exploratory procedures, we know that in primary care and psychiatric/clinical samples the PHQ-9 appears to be unidimensional (Cameron, Crawford, Lawton, & Reid, 2008; Hansson, Chotai, Nordstöm, & Bodlund, 2009; Huang, Chung, Kroenke, Delucchi, & Spitzer, 2006). Scarce confirmatory findings in general population (Chinese sample; Yu et al., 2012) suggest one factor as well. In contrast, there are both exploratory and confirmatory findings on spinal cord injury and palliative care patients which suggest or confirm a two-factor solution, with (highly) correlated Cognitive/affective and Somatic factors (Chilcot et al., 2013; Kalpakjian et al., 2009; Krause, Bombardier, & Carter, 2008; Richardson & Richards, 2008). It is possible that the Somatic factor emerged due to the specific nature of these medical conditions. It is also plausible that the difference in the number of factors is an artifact of the usage of exploratory and confirmatory procedures.

Previous research also confirms that the PHQ-9 is largely invariant between culture, gender, and age groups, with only slight hints of the possible invariance/differential item functioning (Huang et al., 2006; Kalpakjian et al., 2009; Thibodeau & Asmundson, 2014; Yu et al., 2012).

According to all of this, we would expect to obtain a satisfactory fit of either one- or two-factor solutions, hopefully with an invariance of the confirmed factor structure.

Method

Sample and procedure

The sample comprised 1875 participants (61.5% female). The average age was 28.26 (*SD*=8.32) years, with no differences in age between the female and male participants ($t(1873)=0.99$, $p=.69$, $d=0.05$). The participants completed the PHQ-9 as a part of several anonymous online LimeSurvey (LimeSurvey Project Team/Carsten Schmitz, 2012) surveys, conducted for various other purposes over a course of mid 2013 to early 2015. Due to the strict anonymity and differences between surveys regarding the participants' location measurement, it can only be estimated with some margin of error that 60% of the participants were from Bosnia and Herzegovina, 23% from Serbia, 10% from Croatia, and 7% were from a diaspora or it was impossible to determine their location.

Instrument

PHQ-9 (Kroenke & Spitzer, 2002; Kroenke et al., 2001). This questionnaire is a measure of nine depression symptoms (one item each) according to the DSM-IV(-TR) (American Psychiatric Association, 2000) criteria. But the PHQ-9 also conforms to and is endorsed by (Moriarty, et al., 2015) a new DSM-V (American Psychiatric Association, 2013). The PHQ-9 items are 4-point symptom frequency estimates, ranging from 0 ("not at all") to 3 ("every day"). Values of 5, 10, 15, and 20 were originally proposed as cutoffs for mild, moderate, moderately severe, and severe depression, respectively (Kroenke & Spitzer, 2002; Kroenke et al., 2001), with a cutoff of 10 being most often reported and used, for which it has a pooled sensitivity of .78 (95% CI [.70, .84]) and pooled specificity of .87 (95% CI [.84, .90]) (Moriarty, et al., 2015).

Statistical analyses

In order to compare one- and two-factor PHQ-9 solutions we conducted a confirmatory factor analysis (CFA) using lavaan program for R (Rosseel, 2012). As the PHQ-9 items are four-point ordered polytomous items, we used conceptually appropriate Weighted Least Squares Means and Variance Adjusted (WLSMV) estimator (Beauducel, & Herzberg, 2006) to compute robust standard errors and a mean- and variance-adjusted test statistic, with Diagonally Weighted Least Squares (DWLS) as an estimator of the model parameters. When determining model fit, we relied on the following fit indices (see, e.g., Hooper, Coughlan, & Mullen, 2008), which are compatible with WLSMV/DWLS estimators: the Comparative Fit Index (CFI; CFI≥0.95 indicates good fit), the Non-Normed Fit Index (NNFI, also known as the Tucker-Lewis Index, TLI; NNFI/TLI≥0.95 indicates good fit), and the Root Mean Square Error of Approximation (RMSEA; RMSEA<0.06 indicates good fit, with a stringent upper limit of RMSEA<0.07).

In order to determine if there is a difference between the CFA and exploratory factor analysis (EFA) in regards to the number of factors, we also applied two robust procedures for determining the number of factors in EFA. The first procedure is a version of parallel analysis (PA) proposed by Timmerman and Lorenzo-Seva (2011), based on polychoric correlations and minimum rank factor analysis (MRFA; ten Berge, & Kiers, 1991) method of extraction. Both mean and 95 percentile criteria were consulted. The second procedure that we relied upon is the "HULL" method, proposed by Lorenzo-Seva, Timmerman, and Kiers (2011), which, similarly, to the CFA, seeks to determine the model with the best balance between fit and number of parameters. Both the PA and HULL were calculated using a FACTOR program (Lorenzo-Seva & Ferrando, 2006).

We were also interested in evaluating measurement invariance, which means that we wanted to determine if a person's scores on a questionnaire do not significantly depend on that person's group membership (Wu, Li, & Zumbo, 2007). We were specifically interested in gender invariance (note that despite the sample size, our sample was not balanced enough for the age groups and country invariance tests). The measurement invariance testing consists of several progressively more constrained model comparisons (Hirschfeld & von Brachel, 2014; Wu et al., 2007). We tested the following invariance steps, appropriate for the models based on ordered polytomous items (Hirschfeld & von Brachel, 2014): 1) configural invariance, which determines whether or not the same factor model specifications hold for all groups, 2) weak invariance, which adds a constraint of equal loadings across groups, 3) strong invariance, which also adds thresholds constraints, and 4) strong invariance + equal factor means constraints.

Statistically significant changes in χ^2 (i.e. $\Delta\chi^2$) for a given difference in degrees of freedom and/or changes in CFI (i.e. ΔCFI) larger than ΔCFI≥.01 (for sample sizes of $N \geq 300$) at each step indicate non-invariance, with the remark that ΔCFI criterion seems to be more robust (Chen, 2007; Hirschfeld & von Brachel, 2014; Wu et al., 2007).

In order to test a post hoc research question regarding the justification of the total questionnaire score usage, we also conducted a confirmatory bifactor analysis, also known as general-specific or nested model analysis (Chen, West, & Sousa, 2006). The bifactor analysis allows researchers to test if there is a general factor responsible for the commonality of the items, but with multiple domain specific factors accounting for the unique influence of the specific domains (Chen et al., 2006). The explanatory power (i.e. explained % of the variance) of general and specific factors can be compared as well.

Results

Confirmatory factor analysis of the PHQ-9

The WLSMV/DWLS based CFA factor loadings (two-factor model) are given in Table 1.

Table 1. *Confirmatory factor analysis loadings for the two-factor model*

Over the last 2 weeks, how often have you been bothered by any of the following problems?	Cognitive/affective	Somatic
1 – Little interest or pleasure in doing things.	.65	
2 – Feeling down, depressed, or hopeless.	.82	
6 – Feeling bad about yourself – or that you are a failure or have let yourself or your family down.	.79	
7 – Trouble concentrating on things, such as reading the newspaper or watching television.	.65	
8 – Moving or speaking so slowly that other people could have noticed? Or the opposite – being so fidgety or restless that you have been moving around a lot more than usual.	.71	
9 – Thoughts that you would be better off dead or of hurting yourself in some way.	.74	
3 – Trouble falling or staying asleep, or sleeping too much.		.68
4 – Feeling tired or having little energy.		.78
5 – Poor appetite or overeating.		.63

Notes: All factor loadings are standardized item loadings; correlation between the factors is .83; all loadings and correlations are statistically significant at $p<.001$ level.

Results suggested that the two-factor model with correlated Cognitive/affective and Somatic factors fits the data well ($\chi^2(26)=287.8$, $p<.001$; CFI=0.972, NNFI/TLI=0.961, RMSEA=0.073, 90% CI [0.066, 0.081]). The two-factor model fit was better than the single-factor solution ($\chi^2(27)=444.2$, $p<.001$; CFI=0.956, NNFI/TLI=0.941, RMSEA=0.091, 90% CI [0.083, 0.098]).

Both factors had good reliabilities (McDonald's ω; Zinbarg, Revelle, Yovel, & Li, 2005): $\omega_{Cognitive/affective}=.86$, $\omega_{Somatic}=.74$.

A difference in the suggested number of factors between the CFA and EFA

While the CFA supported the two-factor model, both PA (according to both mean and 95 percentile criteria) and HULL procedures clearly favored a single factor. In the single-factor MRFA based EFA solution, factor loadings were between .57 and .82, and the factor explained 74% of the common variance.

Gender invariance of the two-factor model

Due to space limitations, we present gender invariance tests for the best fitting model (i.e. two-factor model) in an abbreviated form. Two factors had 'strong' gender invariance (i.e. a step number three, described previously: 'configural invariance' + equal loadings + equal thresholds), compared to the baseline configuration model ($\Delta\chi^2(\Delta df)$=9.48(14.75), p=.839, ΔCFI<.001). However, after adding the equal means constraints (a step number four), factors remained invariant based on ΔCFI criterion (.008), but became non-invariant based on $\Delta\chi^2(\Delta df)$ criterion (31.63(14.75), p=.006). The source of potential invariance was the higher Somatic factor score for females ($M_{females}$=3.32, $SD_{females}$=2.22, M_{males}=2.89, SD_{males}=2.23, t(1873)=4.15, p<.001), but this difference was just at the Cohen's (1992) cutoff (0.20) between trivial and small effect size (d=0.197).

Bifactor analysis: Is it justified to use a total PHQ-9 score?

As the two-factor model was supported, and two factors were highly correlated (.83) it was appropriate to pose a question if a total or two factor scores should be used. In order to answer this, we conducted a bifactor analysis, with one general depression factor, comprising all of the items, and two domain specific factors, i.e. Cognitive/affective and Somatic, comprising six and three items respectively. The bifactor model showed an excellent fit (χ^2(18)=158.4, p<.001; CFI=0.985, NNFI/TLI=0.970, RMSEA=0.065, 90% CI [0.055, 0.074]). All the items loaded on the general factor (with loadings ranging from .54 to .77) with much lower, but significant loadings of the corresponding items on Cognitive/affective (loadings from .11 to .44) and Somatic (loadings from .26 to .47) domain specific factors. From this it was calculated that 82.40% of variance was explained by the general factor, and the remaining 17.60% was explained by the two domain specific factors (8.83% and 8.77% respectively). Thus, a vast majority of the PHQ-9 variance is accounted for by the general factor – and a total score.

Table 2. *The PHQ-9 total score distributions*

Participants	Percentages of participants inside the specified cutoffs					
	0	1-4	5-9	10-14	15-19	20-27
Females	2.95	33.71	38.13	15.94	6.76	2.51
Males	7.07	34.67	34.95	14.70	5.96	2.64
Total	4.53	34.08	36.91	15.47	6.45	2.56

In light of this finding, we calculated reliability of whole PHQ-9, which was also good: ω_{Total}=.89. We also calculated distributions for a total PHQ-9 score, which are shown in Table 2. The average PHQ-9 total score for our sample was 6.91 (SD=6.00), and cumulatively, 24.5% of participants had values above the most widely used cutoff of ≥10.

Discussion

Consistent with the findings obtained on spinal cord injury and palliative care patients, our results supported a two-factor PHQ-9 solution (Chilcot et al., 2013; Kalpakjian et al., 2009; Krause et al., 2008; Richardson & Richards, 2008). The model fit was good, with both CFI and NNFI/TLI values well over the proposed cutoffs, and RMSEA just above the upper limit (Hooper et al., 2008). Compared with the reported fit indices in other research which supported the two-factor PHQ-9 structure (i.e. CFI=0.95 and RMSEA=0.08 from Chilcot et al., 2013; RMSEA=0.073 from Krause et al., 2008), our fit was slightly better. Furthermore, an additional comparison of the number of optimal factors between the CFA and EFA approaches is consistent with the possibility that a support for a single- or a two-factor solution in earlier studies might be (at least partly) due to the differences between the exploratory and confirmatory methods, with EFA-based procedures possibly being more likely to reproduce a single factor. A single-factor solution reported in previous studies was mainly a product of the exploratory-based approaches (Cameron et al., 2008; Hansson et al., 2009; Huang et al., 2006). The robust procedures for determining the number of factors to retain in EFA also suggested a single factor on our data, despite the fact that CFA identified two-factors as better fitting. Also note that the only available study that used CFA in the general population (Yu et al., 2012), which confirmed a single-factor solution, did not actually compare it to the two-factor model. Thus, we would advise all other researchers to routinely check for both one and two PHQ-9 factors, regardless of the chosen approach (CFA or EFA).

Probably the main issue that derives from the one- or two-factor structure dilemma is which scoring procedure to use. Specifically, despite the two-factor model having a better fit that a single-factor model on our data, using two factor scores might not be optimal in practice. In the research context, because the factors were very highly correlated, that would pose a serious multicollinearity problem. Furthermore, note that the bifactor solution showed even better fit than the regular correlated two-factor model, and a general factor alone accounted for 82.95% of the variance. Also, the reliability of the whole scale was good, and better than individual specific factor reliabilities (which, granted, is a partial function of higher number of items in the whole scale as opposed to individual factors). Having all this in mind, we argue that the usage of a single/total PHQ-9 score would probably be advisable for most practical purposes, but researchers interested in testing Cognitive/affective and Somatic domain specific hypotheses should use a bifactor approach, as this allows not only to use the general and domain specific factors simultaneously, but it also prevents multicollinearity issues, due to the orthogonality of the general and domain specific factors in a bifactor analysis.

Consistent with foreign findings are the average percentages of participants with scores between the questionnaire's cutoffs (i.e. 1-4, 5-9, etc.), which are almost identical to the ones presented by, e.g., Krause and colleagues (2008), with mild exception of slightly fewer people with a score of zero in our sample.

The PHQ-9 showed a strong gender factor invariance, meaning that there were no differences in the factor structure, loadings, and thresholds between females and males, and the Cognitive/affective factor showed no differences in means. Only a small indication of differences in gender means attributable to a Somatic score is consistent with previous findings (Kalpakjian et al., 2009; Thibodeau & Asmundson, 2014). Thus, it can be stated that the PHQ-9 has approximately the same properties for females and males.

In conclusion, the PHQ-9 exhibits well-fitting latent structure, ‘strong’ gender factor invariance, and good reliability. Our results are also largely consistent with the foreign findings. This suggests a good potential for the PHQ-9 research purposes use in the BCS language (with the obvious need for further country- and age-groups specific tests). Convergent and discriminative aspects of the PHQ-9 validations are, however, pending. While an insight in correlations with other relevant measures would undoubtedly be useful, this was beyond the scope and goal of this article and is something that should be done in future research. Norming on clinical samples and predictive validity testing of the PHQ-9 in primary care and clinical settings should, obviously, also be amongst the highest priorities.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4thed., text rev.). Washington, DC: Author.

Beauducel, A., & Herzberg, P. (2006). On the Performance of Maximum Likelihood versus Means and Variance Adjusted Weighted Least Squares estimation in CFA. *Structural Equation Modeling: A Multidisciplinary Journal, 13*(2), 186-203.

Cameron, I. M., 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.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14*(3), 464-504.

Chen, F. F., West, S. G., & Sousa, K. H. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research, 41*(2), 189-225.

Chilcot, J., Rayner, L., Lee, W., Price, A., Goodwin, L., Monroe, B., ... & Hotopf, M. (2013). The factor structure of the PHQ-9 in palliative care. *Journal of Psychosomatic Research, 75*(1), 60-64.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155-159.

Ferrari, A. J., Charlson, F. J., Norman, R. E., Patten, S. B., Freedman, G., Murray, C. J., & ... Whiteford, H. A. (2013). Burden of depressive disorders by country, sex, age, and year: Findings from the Global Burden of Disease Study 2010. *Plos Medicine, 10*(11), 1-12.

Gilbody, S., Sheldon, T., & House, A. (2008). Screening and case-finding instruments for depression: A meta-analysis. *Canadian Medical Association Journal, 178*(8), 997-1003.

Hansson, M., Chotai, J., Nordstöm, A., & Bodlund, O. (2009). Comparison of two self-rating scales to detect depression: HADS and PHQ-9. *British Journal of General Practice, 59*(566), e283-e288.

Hirschfeld, G., & von Brachel, R. (2014). Improving Multiple-Group confirmatory factor analysis in R – A tutorial in measurement invariance with continuous and ordinal indicators. *Practical Assessment, Research & Evaluation, 19*(7), 1-12. Available from http://goo.gl/TtMPu4

Hooper, D., Coughlan, J., Mullen, M. (2008). Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods, 6*(1), 53-60.

Huang, F. Y., Chung, H., Kroenke, K., Delucchi, K. L., & Spitzer, R. L. (2006). Using the patient health questionnaire-9 to measure depression among racially and ethnically diverse primary care patients. *Journal of General Internal Medicine, 21*(6), 547-552.

Kalpakjian, C. Z., Toussaint, L. L., Albright, K. J., Bombardier, C. H., Krause, J. K., & Tate, D. G. (2009). Patient Health Questionnaire-9 in spinal cord injury: An examination of factor structure as related to gender. *The Journal of Spinal Cord Medicine, 32*(2), 147-156.

Kozel, D., Zakotnik, J. M., Grum, A. T., Kersnik, J., Pavlič, D. R., Tomori, M. Ž., & Ziherl, S. (2012). Applicability of systematic screening for signs and symptoms of depression in family practice patients in Slovenia. *Zdravniški Vestnik, 81*(12). 838-846.

Krause, J. S., Bombardier, C., & Carter, R. E. (2008). Assessment of depressive symptoms during inpatient rehabilitation for spinal cord injury: Is there an underlying somatic factor when using the PHQ? *Rehabilitation Psychology, 53*(4), 513-520.

Kroenke, K., & Spitzer, R. L. (2002). The PHQ-9: A new depression diagnostic and severity measure. *Psychiatric Annals, 32*(9), 1-7.

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine, 16*(9), 606-613.

LimeSurvey Project Team / Carsten Schmitz. (2012). LimeSurvey: An open source survey tool [Computer Software]. Hamburg, Germany: LimeSurvey Project. Retreived from https://www.limesurvey.org/

Lorenzo-Seva, U., & Ferrando, P. J. (2006). FACTOR: A computer program to fit the exploratory factor analysis model. *Behavior Research Methods, 38*(1), 88-91.

Lorenzo-Seva, U., Timmerman, M. E., & Kiers, A. L. (2011). The Hull method for selecting the number of common factors. *Multivariate Behavioral Research*, *46*, 340-364.

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*(1), 71-77.

Moriarty, A. S., Gilbody, S., McMillan, D., & Manea, L. (2015). Screening and case finding for major depressive disorder using the Patient Health Questionnaire (PHQ-9): A meta-analysis. *General Hospital Psychiatry, 37*(6), 567-576.

National Collaborating Centre for Mental Health. (2010). *Depression in adults with a chronic physical health problem.* Leicester, UK: British Psychological Society.

Richardson, E. J., & Richards, J. S. (2008). Factor structure of the PHQ-9 screen for depression across time since injury among persons with spinal cord injury. *Rehabilitation Psychology, 53*(2), 243-249.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1-36. Retreived from http://goo.gl/l4kgYS

ten Berge, J. M. F., & Kiers, H. A. L. (1991). A numerical approach to the approximate and the exact minimum rank of a covariance matrix. *Psychometrika*, *56*(2), 309-315.

Thibodeau, M. A., & Asmundson, G. J. (2014). The PHQ-9 assesses depression similarly in men and women from the general population. *Personality and Individual Differences*, *56,* 149-153.

Timmerman, M. E., & Lorenzo-Seva, U. (2011). Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychological Methods*, *16*(2), 209-220.

Williams, J. W., Pignone, M., Ramirez, G., & Stellato, C. P. (2002). Identifying depression in primary care: A literature synthesis of case-finding instruments. *General Hospital Psychiatry*, *24*(4), 225-237.

World Health Organization [WHO]. (2012, October). *Depression* [Fact sheet]. Retrieved from http://goo.gl/NRaxhA

Wu, A. D., Li, Z., & Zumbo, B. D. (2007). Decoding the meaning of factorial invariance and updating the practice of multi-group confirmatory factor analysis: A demonstration with TIMSS data. *Practical Assessment Research & Evaluation*, *12*(3), 1-26.

Yu, X., Tam, W. W., Wong, P. T., Lam, T. H., & Stewart, S. M. (2012). The Patient Health Questionnaire-9 for measuring depressive symptoms among the general population in Hong Kong. *Comprehensive Psychiatry, 53*(1), 95-102.

Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α, Revelle's β, and McDonald's ω_H: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika, 70*(1), 123-133.

FAKTORSKA STRUKTURA UPITNIKA PACIJENTOVOG ZDRAVLJA (PHQ-9) NA NEKLINIČKOM UZORKU

Upitnik pacijentovog zdravlja (eng. The Patient Health Questionnaire) – PHQ-9 je poznati upitnik koji mjeri devet simptoma depresije po DSM-IV/DSM-V kriterijumima. Pošto PHQ-9 nije validiran u većini zemalja bivše Jugoslavije, cilj ovog istraživanja bio je ispitivanje njegove faktorske strukture i psihometrijskih svojstava na velikom nekliničkom uzorku sa BHS govornog područja. Ukupno 1875 ispitanika (61.5% žena), prosječnog uzrasta od 28.26 (SD=8.32) godina, ispunilo je anonimni online PHQ-9 upitnik. Konfirmativna faktorska analiza (CFA) zasnovana na WLSMV/DWLS pokazala je da dvofaktorski model sa koreliranim (.83) kognitivnim/afektivnim i somatskim faktorima dobro fituje podatke ($\chi^2(26)=287.8$, $p<.001$; CFI=0.972, NNFI/TLI=0.961, RMSEA=0.073, 90% CI [0.066, 0.081]), bolje i u odnosu na jednofaktorsko rješenje ($\chi^2(27)=444.2$, $p<.001$; CFI=0.956, NNFI/TLI=0.941, RMSEA=0.091, 90% CI [0.083, 0.098]). Dvofaktorski model bolje fituje i u drugim istraživanjima na nekliničkim uzorcima (npr. palijativna njega), dok se na kliničkim/psihijatrijskim uzorcima tipično detektuje jednodimenzionalnost. Dvofaktorski model pokazao je snažnu polnu invarijantnost (konfiguracijska invarijantnost + jednaka zasićenja + jednaki pragovi; $\Delta\chi^2(\Delta df)=9.48(14.75)$, $p=.839$, ΔCFI<.001). Međutim, nakon fiksiranja aritmetičkih sredina, model je ostao invarijantan po ΔCFI kriterijumu (.008), ali nije po $\Delta\chi^2(\Delta df)$ kriterijumu (31.63(14.75), $p=.006$). Izvor potencijalnog gubitka invarijantnosti bio je viši skor žena na somatskom faktoru ($M_{žene}=3.32$, $SD_{žene}=2.22$, $M_{muškarci}=2.89$, $SD_{muškarci}=2.23$, $t(1873)=4.15$, $p<.001$), ali ova razlika je bila na samoj granici između trivijalnog i niskog intenziteta efekta ($d=0.197$). Konačno, rezultati bifaktorske analize i dobra pouzdanost upitnika u cjelini (ω=.89), sugerisali su na to da bi upotreba jedinstvenog PHQ-9 skora bila vjerovatno prikladna za većinu potreba, ali da istraživači mogu koristiti bifaktorski pristup kako bi testirali specifične hipoteze u vezi s kognitivnim/afektivnim i somatskim specifičnim domenima. Zaključujemo da PHQ-9 pokazuje dobro fitujuću latentnu strukturu, snažnu polnu faktorsku invarijantnost i dobru pouzdanost, što ukazuje na potencijal za njegovu upotrebu za istraživačke svrhe na BHS govornom području. Međutim, upitnik tek treba konvergentno i diskriminativno validirati i normirati na kliničkim uzorcima.

Ključne riječi: *Upitnik pacijentovog zdravlja – PHQ-9, depresija, konfirmativna faktorska analiza (CFA), bifaktorska analiza, BHS jezik*

Recommended citation:

Subotić, S., Knežević, I., Dimitrijević, S., Miholjčić, D., Šmit, S., Karać, M., & Mijatović, J. (2015). The factor structure of the Patient Health Questionnaire (PHQ-9) in a non-clinical sample. In S. Subotić (Ed.), *STED 2015 Conference Proceedings – Psychology Section* (pp. 20-28). Banja Luka, B&H: University for Business Engineering and Management.