

Revised manuscript for submission to Quality of Life Research

Manuscript No: QURE-D-13-00587R1

Article title: Testing gender invariance of the Hospital Anxiety and Depression Scale using the classical approach and Bayesian approach

Authors: Ted C. T. Fong*, MPhil; Rainbow T. H. Ho, PhD

Affiliations: Centre on Behavioral Health and Department of Social Work & Social

Administration, The University of Hong Kong, Hong Kong

Corresponding author: * Ted C. T. Fong, Centre on Behavioral Health, 2/F, The Hong Kong

Jockey Club Building for Interdisciplinary Research, 5 Sassoon Road, Pokfulam, HKSAR

Phone no: 852-2831 5174; E-mail: ttaatt@hku.hk; Fax no: 852-2816 6710

Keywords: Hospital Anxiety and Depression Scale; measurement invariance; maximum likelihood; Bayesian; confirmatory factor analysis; gender

Word count (exc. figures/tables): 2,173; number of tables: 4; number of figures: 0

Testing gender invariance of the Hospital Anxiety and Depression Scale using the classical approach and Bayesian approach

Abstract

Purpose Measurement invariance is an important attribute for the Hospital Anxiety and Depression Scale (HADS). Most of the confirmatory factor analysis (CFA) studies on the HADS adopt the classical maximum likelihood approach. The restrictive assumptions of exact zero cross-loadings and residual correlations in the classical approach can lead to inadequate model fit and biased parameter estimates. The present study adopted both the classical approach and the alternative Bayesian approach to examine the measurement and structural invariance of the HADS across gender.

Methods A Chinese sample of 326 males and 427 females was used to examine the two-factor model of the HADS across gender. Configural and scalar invariance of the HADS were evaluated using the classical approach with the robust weighted least square estimator and the Bayesian approach with zero-mean, small variance informative priors for cross-loadings and residual correlations.

Results Acceptable and excellent model fits were found for the two-factor model under the classical and Bayesian approaches, respectively. The two-factor model displayed scalar invariance across gender using both approaches. In terms of structural invariance, females

showed a significantly higher mean in the anxiety factor than males under both approaches.

Conclusion The HADS demonstrated measurement invariance across gender and appears to be a well-developed instrument for assessment of anxiety and depression. The Bayesian approach is an alternative and flexible tool that could be used in future invariance studies.

Keywords: Hospital Anxiety and Depression Scale; measurement invariance; maximum likelihood; Bayesian; confirmatory factor analysis; gender

Testing gender invariance of the Hospital Anxiety and Depression Scale using the classical approach and Bayesian approach

Developed by Zigmond and Snaith [1], the Hospital Anxiety and Depression Scale (HADS) is a widely used assessment tool of anxiety and depression. Previous studies indicated satisfactory levels of internal consistency, test-retest reliability, and concurrent validity for the HADS [2,3]. Investigation of measurement invariance is essential for the ongoing assessment of the validity of the HADS. Measurement invariance tests explore increasing levels of measurement invariance: configural invariance (equal factor structures), metric invariance (equal factor loadings), and scalar invariance (equal item intercepts) [4]. A well-fitting configural model is the prerequisite for conducting tests of measurement invariance. Measurement invariance is crucial in ensuring comparability of factor scores across different groups. Without scalar invariance, differences in factor scores across different groups cannot be clearly attributed to differences in the underlying constructs [5].

The mainstream of psychological assessment research is largely based on the classical maximum likelihood (ML) approach. Previous confirmatory factor analysis (CFA) studies [6-8] have examined the metric invariance of the HADS across gender using the classical approach. As these studies did not evaluate the scalar invariance, the measurement invariance for the HADS has yet to be established. Given the known bias for chi-square test

in large sample sizes [9], typical ML-CFA studies disregard significant chi-square test results. In the context of invariance testing, however, researchers [10,11] have warned that the significant chi-square in the configural model may produce distorted results in subsequent chi-square difference tests. Furthermore, the restrictive assumptions of exact zero cross-loadings and residual correlations in the classical approach can lead to inadequate model fit and biased parameter estimates [12,13].

The Bayesian approach is an alternative estimation approach recently incorporated in the context of structural equation modeling [14,15]. This approach relaxes the restrictive assumptions of exact-zero cross-loadings and residual correlations in the classical approach via the use of zero-mean, small variance priors [14]. The informative priors specify credible values and uncertainty for the parameters based on prior beliefs from existing knowledge and substantive theories. This approach does not rely on large-sample normal theory and is likely to produce robust results in the cases of strongly skewed variables or small sample sizes. A recent study [16] applied the Bayesian approach to evaluate the factor structure of the HADS and found empirical evidences for the two-factor structure. The present study aimed to further examine the measurement and structural invariance of the HADS across gender using the alternative Bayesian approach. To evaluate the feasibility of using the Bayesian approach to conduct invariance testing, conventional invariance testing was performed using the classical approach as a comparison.

Method

Participants and measure

This study was based on 763 participants recruited under simple random sampling in Hong Kong in accordance with ethical guidelines of local research ethics committee. The participants provided written informed consents and completed the HADS. The study sample included 326 males (mean age = 42.6 years, $SD = 11.5$, inter-quartile range = 35-51) and 427 females (mean age = 41.1 years, $SD = 11.4$, inter-quartile range = 33-50). The majority of the male and female samples was married (60.1% and 69.3%) and had completed secondary school (51.3% and 50.0%). The two samples showed no significant age difference (mean difference = 1.5, $t_{745} = 1.782$, $p = .075 > .05$, Cohen $d = .065$).

The Chinese version of the 14-item HADS [16,17] inquires the two domains of anxiety (7 items) and depression (7 items) during the past week on a 4-point response format. In the present study, adequate levels of Cronbach's alphas were found for anxiety and depression in the male ($\alpha = .86$ and $.72$) and female ($\alpha = .86$ and $.74$) samples.

Data analysis

To assess the factorial validity and the gender invariance of the HADS, multigroup confirmatory factor analysis was performed in Mplus 7.1 [18] using the classical and

Bayesian approaches. The robust weighted least square estimator was used for estimation in the classical approach. This estimator has been found to produce unbiased parameters and standard errors for analysis of the 4-point ordinal response items [19]. Model fit of the two-factor model was assessed across gender using the following criteria on the approximate fit indices [20]: comparative fit index (CFI) ≥ 0.95 , Tucker-Lewis index (TLI) ≥ 0.95 and root mean square error of approximation (RMSEA) ≤ 0.06 .

Measurement invariance across gender was evaluated via tests of configural invariance (invariance of factor structure) and scalar invariance (invariance of factor loadings and thresholds). Invariance of structural parameters was then examined for the factor variance, covariance, and factor means. The invariance tests were evaluated using chi-square difference tests to compare nested models through the DIFFTEST option. A significant chi-square difference test suggested model noninvariance where partial invariance model would be considered by relaxing the noninvariant item based on modification indices.

In the Bayesian approach, the two-factor model was first evaluated across gender using the Bayesian estimator. Model estimation was carried out in two independent Markov chain Monte Carlo chains with the Gibbs sampler [21,14,22] in 80,000 iterations. The potential scale reduction (PSR) factor [23] was used with a PSR value below 1.1 suggesting model convergence. The use of zero-mean, small-variance informative priors, which specify approximate zeros with a 95% limit of -.2 to .2, was investigated for cross-loadings and

residual correlations [14]. Analyses were performed on the standardized scores and factor variances were set at one so that the scale of priors corresponded to standardized loadings.

To assess the measurement invariance across gender, models on configural invariance (invariance of factor structure) and scalar invariance (invariance of factor loadings and item intercepts) were evaluated and compared. Model fit was evaluated using posterior predictive p value and its associated 95% credibility interval [14]. A low posterior predictive p value ($p < .05$) and positive 95% lower limit suggest a poor model fit, while a posterior predictive p value of around .5 and a symmetric credibility interval centering close to zero indicate an excellent model fit. The deviance information criterion (DIC) and Bayesian information criterion (BIC), which balance model likelihood with a penalty for model complexity [23], were used to compare the competing invariance models. Model with a smaller value of DIC or BIC was preferred.

Results

The classical approach

Table 1 shows the goodness of fit indices of the HADS models under the classical approach. The two-factor model showed an acceptable fit (CFI and TLI $> .95$ and RMSEA $\sim .06$) to the male and female subsamples. The configural invariance model fitted adequately in the multigroup CFA (CFI and TLI $> .95$ and RMSEA $\sim .06$), supporting

equality of factor structure across gender. Table 2 displays the solution for the configural model using the classical approach. For both subsamples, all of the 14 items had major factor loadings greater than .50 and cross-loadings fixed at zero. The two factors were significantly and strongly correlated ($r = .80 - .82, p < .05$).

[Insert Table 1 about here]

[Insert Table 2 about here]

The model on scalar invariance showed a satisfactory fit to the data (CFI and TLI $> .97$ and RMSEA $< .06$). Compared to the configural invariance model, the chi-square difference test was nonsignificant ($MD\Delta\chi^2 = 48.77, df = 38, p = .11$), supporting equality of factor loadings and item thresholds across gender. In terms of structural invariance, the male and female subsamples showed invariant factor variances and covariance. The two subsamples showed partial latent mean invariance on the depression factor (mean difference = 0.02, $t = 0.26, p = .80, Cohen d = 0.02$). However, the female subsample showed a significantly higher latent mean in anxiety than the male subsample (mean difference = 0.18, $t = 2.79, p < .05, Cohen d = 0.23$).

The Bayesian approach

Table 3 shows the Bayesian CFA results for the two-factor model for the male and female subsamples. Zero-mean, small-variance informative priors were specified for

cross-loadings and residual correlations in the Bayesian approach. The two-factor model with the informative priors converged (PSR = 1.03 and 1.02) and provided excellent fits (posterior predictive p values close to .5 and 95% intervals centering at zero) to the two subsamples, respectively. Similarly, the configural invariance model converged (PSR = 1.08) and showed an excellent fit to the data (posterior predictive p values close to .5 and 95% intervals centering at zero), supporting equality of factor structure across gender.

[Insert Table 3 about here]

Table 4 displays the solution for the configural model using the Bayesian approach. For both subsamples, all proposed major factor loadings were statistically significant and substantial (range = .46 - .80) while none of the cross-loadings (range = -.05 to .08) were statistically significant. Only one out of the 91 residual correlations were statistically significant in the male subsample (range = -.14 to .14) and female subsample (range = -.18 to .14). Significant and strong correlations were found between the two factors ($r = .63 - .64$, $p < .05$) in both subsamples.

[Insert Table 4 about here]

The Bayesian model on scalar invariance converged (PSR = 1.06) and displayed a good fit to the data (posterior predictive p values = .34 and a negative 95% lower limit). Compared to the configural invariance model, the scalar invariance model showed a lower DIC and BIC, supporting equality of factor loadings and intercepts across gender.

For the structural parameters, the factor variances and covariance were invariant across gender. Though the latent depression mean was not significantly different across gender (Cohen $d = 0.04$, $SD = 0.08$, $p = .33$), the latent anxiety mean was significantly higher for the female subsample than the male subsample (Cohen $d = 0.22$, $SD = 0.08$, $p < .05$).

Discussion

The present study examined the factorial validity and measurement invariance of the HADS across gender using the classical approach and the alternative Bayesian approach. Regarding the factorial validity of the HADS, the two-factor model showed an approximate model fit in terms of acceptable goodness-of-fit indices under the classical approach. Similar to previous psychometric studies on the HADS [8,16,7,6,17], the chi-square test of exact fit was found to be significant. The significant chi-square test could be attributed to its oversensitivity bias for large sample size or the underlying assumption of exact zero cross-loadings and residual correlations. As shown in the Bayesian model results, the minor model misfits in the classical model may lie in the omission of trivial cross-loadings and residual correlations among the items.

Under the Bayesian approach, the present study supported a well-fitting two-factor model with approximately zero cross-loadings and residual correlations. All the cross-loadings and residual correlations fell within the hypothesized 95% credibility limits

of -0.2 to 0.2. Comparing the configural model results of the two approaches, the major factor loadings and factor correlation were in general lower in the Bayesian model than in the classical model. That the two-factor model showed an excellent fit in the Bayesian model but only an approximate fit in the classical model appears to lend support to a greater precision for the Bayesian parameter estimates. The discrepancy could reflect an overestimation of the factor loadings and factor correlation in the classical approach due to its restrictive assumption of exact zero cross-loadings and residual correlations.

Concerning the gender invariance of the HADS, the measurement invariance was supported in the classical approach by the non-significant chi-square different test between the scalar invariance and configural invariance models. The scalar invariance model appeared to show slightly better fit in terms of the approximate fit indices. In the Bayesian approach, the scalar invariance model was preferred with a good model fit and greater model parsimony in terms of the lower information criteria. Overall, results from both approaches supports the HADS as a well-developed measurement instrument for anxiety and depression across gender. The scalar invariance denotes equivalent scale calibration and response tendencies which facilitate meaningful comparison of structural parameters across gender. Both approaches revealed invariance of factor variance/covariance for the HADS and partial latent mean invariance with a significantly higher mean of anxiety for the female sample. Future studies can examine the measurement invariance of the HADS across

cultural contexts to elucidate the cross-cultural comparability.

In conclusion, the present study demonstrated the feasibility of using the Bayesian estimator for invariance testing in psychometric research. Via the use of informative priors to permit slight deviation from the hypothesized zeros for model parameters, the Bayesian approach, which explicitly quantifies the degree of subjectivity and precision, could be a realistic reflection of theoretical and substantive knowledge on factor modeling. Besides testing for exact measurement invariance, the Bayesian approach allows a test of approximate measurement invariance via the use of zero-mean, small variance priors for parameter differences between groups [24,25]. This framework provides an effective and convenient approach for detecting non-invariant parameters particularly in the case of multiple groups or time points. The degree of non-invariance can be assessed for multiple parameters in a single step where a large number of non-invariant parameters may exist. Despite the potential complexities in Bayesian model specification [26,27], further use of the approach is recommended in future psychometric research with gradual improvement in software implementation, accumulation of practical experience, and development of formal guidelines [28].

References

1. Zigmond, A. S., & Snaith, R. P. (1983). The Hospital Anxiety and Depression Scale. *Acta Psychiatrica Scandinavica*, 67(6), 361-370.
2. Bjelland, I., Dahl, A. A., Haug, T. T., & Neckelmann, D. (2002). The validity of the Hospital Anxiety and Depression Scale: An updated literature review. *Journal of Psychosomatic Research*, 52(2), 69-77, doi:10.1016/s0022-3999(01)00296-3.
3. Herrmann, C. (1997). International experiences with the hospital anxiety and depression scale - A review of validation data and clinical results. *Journal of Psychosomatic Research*, 42(1), 17-41, doi:10.1016/s0022-3999(96)00216-4.
4. Vandenberg, R. J., & Lance, C. E. (2000). A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research. *Organizational Research Methods*, 3(1), 4-70, doi:10.1177/109442810031002.
5. Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review*, 18(4), 210-222, doi:10.1016/j.hrmr.2008.03.003.
6. Hunt-Shanks, T., Blanchard, C., Reid, R., Fortier, M., & Cappelli, M. (2010). A psychometric evaluation of the Hospital Anxiety and Depression Scale in cardiac

- patients: Addressing factor structure and gender invariance. *British Journal of Health Psychology*, 15, 97-114, doi:10.1348/135910709x432745.
7. Gale, C. R., Allerhand, M., Sayer, A. A., Cooper, C., Dennison, E. M., Starr, J. M., et al. (2010). The structure of the Hospital Anxiety and Depression Scale in four cohorts of community-based, healthy older people: the HALCYon program. *International Psychogeriatrics*, 22(4), 559-571, doi:10.1017/s1041610210000256.
8. Annunziata, M. A., Muzzatti, B., & Altoe, G. (2011). Defining Hospital Anxiety and Depression Scale (HADS) structure by confirmatory factor analysis: a contribution to validation for oncological settings. *Annals of Oncology*, 22(10), 2330-2333, doi:10.1093/annonc/mdq750.
9. Markland, D. (2007). The golden rule is that there are no golden rules *: A commentary on Paul Barrett's recommendations for reporting model fit in structural equation modelling. *Personality and Individual Differences*, 42(5), 851-858, doi:10.1016/j.paid.2006.09.023.
10. Yuan, K. H., & Bentler, P. M. (2004). On chi-square difference and z tests in mean and covariance structure analysis when the base model is misspecified. *Educational and Psychological Measurement*, 64(5), 737-757, doi:10.1177/0013164404264853.
11. Millsap, R. E. (2007). Structural equation modeling made difficult. *Personality and Individual Differences*, 42(5), 875-881, doi:10.1016/j.paid.2006.09.021.

12. Kolenikov, S. (2011). Biases of parameter estimates in misspecified structural equation models. In T. F. Liao (Ed.), *Sociological Methodology 2011, Vol 41* (Vol. 41, pp. 119-157, Sociological Methodology).
13. Cole, D. A., Ciesla, J. A., & Steiger, J. H. (2007). The insidious effects of failing to include design-driven correlated residuals in latent-variable covariance structure analysis. *Psychological Methods, 12*(4), 381-398, doi:10.1037/1082-989x.12.4.381.
14. Muthen, B., & Asparouhov, T. (2012). Bayesian Structural Equation Modeling: A More Flexible Representation of Substantive Theory. *Psychological Methods, 17*(3), 313-335, doi:10.1037/a0026802.
15. Kaplan, D., & Depaoli, S. (2012). Bayesian structural equation modeling. In R. Hoyle (Ed.), *Handbook of structural equation modeling*. New York, NY: Guilford Press.
16. Fong, T. C. T., & Ho, R. T. H. (2013). Factor analyses of the Hospital Anxiety and Depression Scale: a Bayesian structural equation modeling approach. *Quality of Life Research, 1-7*, doi:10.1007/s11136-013-0429-2.
17. Leung, C. M., Wing, Y. K., Kwong, P. K., Lo, A., & Shum, K. (1999). Validation of the Chinese-Cantonese version of the Hospital Anxiety and Depression Scale and comparison with the Hamilton Rating Scale of Depression. *Acta Psychiatrica Scandinavica, 100*(6), 456-461.
18. Muthen, L. K., & Muthen, B. (1998-2012). *Mplus user's guide* (7th ed.). Los Angeles,

CA: Muthen & Muthen.

19. Schmitt, T. A. (2011). Current Methodological Considerations in Exploratory and Confirmatory Factor Analysis. *Journal of Psychoeducational Assessment*, 29(4), 304-321, doi:10.1177/0734282911406653.
20. Hu, L. T., & Bentler, P. M. (1999). Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives. *Structural Equation Modeling*, 6(1), 1-55, doi:10.1080/10705519909540118.
21. Asparouhov, T., & Muthen, B. (2010). Bayesian analysis of latent variable models using Mplus (Technical report). Los Angeles, CA: Muthen & Muthen.
22. Lee, S. Y., & Song, X. Y. (2012). *Basic and Advanced Bayesian Structural Equation Modeling: With Applications in the Medical and Behavioral Sciences*. Chichester, UK: Wiley.
23. Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis* (2nd ed.). Boca Raton, FL: Chapman & Hall.
24. Verhagen, A. J., & Fox, J. P. (2012). Bayesian tests of measurement invariance. *British Journal of Mathematical and Statistical Psychology*, online first, doi:10.1111/j.2044-8317.2012.02059.x.
25. Muthen, B., & Asparouhov, T. (2013). BSEM Measurement Invariance Analysis. *Mplus Web Notes* (Vol. 17).

26. MacCallum, R. C., Edwards, M. C., & Cai, L. (2012). Hopes and Cautions in Implementing Bayesian Structural Equation Modeling. *Psychological Methods*, *17*(3), 340-345, doi:10.1037/a0027131.
27. Rindskopf, D. (2012). Next Steps in Bayesian Structural Equation Models: Comments on, Variations of, and Extensions to Muthen and Asparouhov (2012). *Psychological Methods*, *17*(3), 336-339, doi:10.1037/a0027130.
28. Muthen, B., & Asparouhov, T. (2012). Rejoinder to MacCallum, Edwards, and Cai (2012) and Rindskopf (2012): Mastering a New Method. *Psychological Methods*, *17*(3), 346-353, doi:10.1037/a0029214.

Table 1 Gender invariance tests on the HADS using classical approach

Step	Invariance Model	χ^2	<i>df</i>	CFI	TLI	RMSEA (90% CI)	$MD\Delta\chi^2$ (<i>df</i>)
0	Male only (<i>N</i> = 326)	165.79*	76	.978	.974	.060 (.048-.073)	
0	Female only (<i>N</i> = 427)	226.66*	76	.970	.964	.068 (.058-.078)	
1	Configural	388.32*	152	.970	.969	.064 (.056-.072)	
2	Scalar	391.45*	190	.978	.979	.053 (.046-.061)	48.77 (38)

Note. χ^2 = chi-square from robust weighted least square estimator; CFI: comparative fit index; TLI: Tucker-Lewis index; RMSEA: root mean square error of approximation; $MD\Delta\chi^2$: Change in χ^2 relative to the model in previous step calculated from Mplus DIFFTEST function. **p* < 0.05.

Table 2 Configural model solution for the HADS using classical approach

Item	Male (N = 326)		Female (N = 427)	
	Anxiety	Depression	Anxiety	Depression
Tense	.789*	0	.760*	0
Frightened	.804*	0	.794*	0
Worrying	.775*	0	.770*	0
Relaxed	.689*	0	.726*	0
Butterflies in stomach	.816*	0	.791*	0
Restless	.696*	0	.713*	0
Panic	.821*	0	.850*	0
Enjoyment as usual	0	.662*	0	.630*
Humor	0	.649*	0	.664*
Cheerful	0	.666*	0	.674*
Slowed down	0	.742*	0	.690*
Disinterest in appearance	0	.547*	0	.629*
Hope for enjoyment	0	.590*	0	.627*
Enjoy a good book/TV	0	.588*	0	.606*
Factor correlation	.818*		.796*	

Note. Bolded values indicate the major loadings. * $p < 0.05$.

Table 3 Gender invariance tests on the HADS using Bayesian approach

Step	Invariance Model	No. free parameters	2.5% PP limit	97.5% PP limit	PP p	DIC	BIC
0	Male only ($N = 326$)	148	-44.2	42.3	.511		
0	Female only ($N = 427$)	148	-44.1	42.4	.498		
1	Configural	296	-62.4	59.6	.517	26822.3	28314.5
2	Scalar	258	-48.2	72.7	.344	26821.9	28091.3

Note. Zero-mean, 0.01 variance informative priors were specified on the cross-loadings and residual correlations at all steps. PP = posterior predictive; DIC = deviance information criterion; BIC = Bayesian information criterion; $*p < 0.05$.

Table 4 Configural model solution for the HADS using Bayesian approach

Item	Male (N = 326)		Female (N = 427)	
	Anxiety	Depression	Anxiety	Depression
Tense	.768*	-.036	.774*	-.049
Frightened	.758*	-.010	.752*	.011
Worrying	.664*	.066	.724*	.033
Relaxed	.636*	.019	.636*	.080
Butterflies in stomach	.625*	.033	.721*	.010
Restless	.672*	.011	.639*	-.016
Panic	.703*	-.002	.798*	-.007
Enjoyment as usual	.017	.588*	-.013	.624*
Humor	-.037	.595*	.010	.475*
Cheerful	.022	.636*	.014	.613*
Slowed down	.078	.531*	.053	.611*
Disinterest in appearance	-.016	.537*	.003	.634*
Hope for enjoyment	-.017	.534*	-.026	.710*
Enjoy a good book/TV	.054	.555*	.036	.462*
Factor correlation	.626*		.636*	

Note. Bolded values indicate the major loadings. * $p < 0.05$.