

Dimensionality of the 9-item Utrecht Work Engagement Scale revisited: A Bayesian structural equation modeling approach

Ted C. T. FONG¹ and Rainbow T. H. HO^{1,2}

¹Centre on Behavioral Health, The University of Hong Kong, Hong Kong and ²Department of Social Work & Social Administration, The University of Hong Kong, Hong Kong

Abstract: Dimensionality of the 9-item Utrecht Work Engagement Scale revisited: A Bayesian structural equation modeling approach: Ted C. T. FONG, et al. Centre on Behavioral Health, The University of Hong Kong, Hong Kong—Objectives: The aim of this study was to reexamine the dimensionality of the widely used 9-item Utrecht Work Engagement Scale using the maximum likelihood (ML) approach and Bayesian structural equation modeling (BSEM) approach. **Methods:** Three measurement models (1-factor, 3-factor, and bi-factor models) were evaluated in two split samples of 1,112 health-care workers using confirmatory factor analysis and BSEM, which specified small-variance informative priors for cross-loadings and residual covariances. Model fit and comparisons were evaluated by posterior predictive p-value (PPP), deviance information criterion, and Bayesian information criterion (BIC). **Results:** None of the three ML-based models showed an adequate fit to the data. The use of informative priors for cross-loadings did not improve the PPP for the models. The 1-factor BSEM model with approximately zero residual covariances displayed a good fit (PPP>0.10) to both samples and a substantially lower BIC than its 3-factor and bi-factor counterparts. **Conclusions:** The BSEM results demonstrate empirical support for the 1-factor model as a parsimonious and reasonable representation of work engagement. (J Occup Health 2015; 57: 353–358)

Key words: Bayesian, Factor structure, Informative priors, Maximum likelihood, Work engagement

The concept of work engagement, defined as “a positive, fulfilling, work-related state of mind that is characterized by vigor, dedication, and absorption¹⁾”, is an active research topic in current applied psychology. A high level of work engagement represents posi-

tively oriented psychological capacities in the workplace such as high energy levels, mental resilience, enthusiasm, strong connection to the environment and being engrossed by one’s work²⁾. The 9-item Utrecht Work Engagement Scale (UWES-9)^{1,3)} is a concise and widely used measure of work engagement across countries^{4,5)}. Though previous studies have demonstrated acceptable reliability and convergent validity for the scale^{6,7)}, there remain unresolved issues about the scale’s dimensionality.

The UWES-9 was originally hypothesized to assess three aspects of work engagement: vigor (3 items), dedication (3 items) and absorption (3 items). On one hand, results from previous studies^{3, 4, 8)} reveal a better fit for a 3-factor model than a 1-factor model, which appears to support interpretation with the three subscale scores. On the other hand, very high correlations ($r \geq 0.90$) were consistently found among the three factors, suggesting potential model redundancy. In view of the inadequate fit for the 1-factor model and lack of discriminant validity for the 3-factor model, de Bruin and Henn⁹⁾ examined a bi-factor model as an alternative factor structure for the UWES-9¹⁰⁾. The bi-factor model, which specified a general work engagement factor and two specific factors on dedication and absorption¹⁰⁾, provided a superior fit to the 1-factor and 3-factor models. The general factor was found to be a dominant factor that accounts for significant portions of the variance in the UWES-9 items.

A number of methodological problems are noteworthy in the previous studies on the UWES-9 based on the traditional maximum-likelihood (ML) approach. The first problem relates to inappropriate evaluation of model fit. Researchers frequently ignore significant χ^2 tests of exact fit by claiming that the χ^2 test is oversensitive to trivial misspecification at large sample sizes. Despite the high power of the χ^2 test to detect model misfit, significant χ^2 tests do not routinely imply trivial misspecification^{11,12)}. The use

Received Feb 24, 2015; Accepted Mar 30, 2015

Published online in J-STAGE May 8, 2014

Correspondence to: T. C. T. Fong, Centre on Behavioral Health, 2/F, Hong Kong Jockey Club Building for Interdisciplinary Research, 5 Sassoon Road, Pokfulam, Hong Kong (e-mail: ttaatt@hku.hk)

of approximate fit indices in evaluating and justifying model fit has been a contentious issue. Second, the ML approach fixes all cross-loadings and residual covariances at zero. This assumption may be unrealistic and overly restrictive¹³⁾ and could lead to inadequate model fit and biased parameter estimates¹⁴⁾. To locate the source of misfit, model diagnostics are performed based on modification indices to estimate a particular cross-loading or residual covariance one at a time. However, such a practice often lacks theoretical justification and likely capitalizes on idiosyncratic features of the sample.

Given the limitations of the ML approach, Muthén and Asparouhov¹³⁾ proposed Bayesian structural equation modeling (BSEM) as an alternative modeling approach. This pioneering approach relaxes the restrictive assumptions of the ML approach via the use of zero-mean, small variance informative priors¹³⁾. Specification of informative priors can better reflect prior knowledge and substantive theories by taking into account the plausible uncertainty over the approximately zero cross-loadings and residual covariances¹³⁾. With reference to recent applications of the BSEM approach^{15, 16)}, the present study aimed to provide new insights on the latent structure of the UWES-9. This was done by reexamining the one-factor, three-factor, and bi-factor models under both the ML and BSEM approaches.

Subjects and Methods

Participants

The participants were 1,112 Chinese adults working in the health-care service sector in Hong Kong. The participants provided written informed consents

and completed a self-report questionnaire in Chinese. Ethical approval was obtained from the local institutional review board. The majority of the participants were female (81.7%), had a secondary education level (58.6%), and were middle-aged (52.8%), with an age range from 41 to 55. All participants had at least a year's work experience. The sample comprised support workers (53.8%), professional workers (16.8%), administrative workers (14.1%) and medical workers (11.3%). The present sample was randomly split into two samples, with Sample 1 being used for primary analysis and Sample 2 being used for cross-validation.

Measure

Work engagement was assessed by the 9-item Utrecht Work Engagement Scale (UWES-9)¹⁷⁾. The UWES-9 was hypothesized to measure work engagement in three dimensions: vigor (3 items), dedication (3 items) and absorption (3 items). The items are scored on a 7-point Likert scale ranging from 0 ("never") to 6 ("every day"). Previous studies reported good reliability for the UWES-9 total score (median Cronbach's $\alpha=0.92$) and subscale scores (median $\alpha=0.77-0.85$)³⁾. Table 1 presents the descriptive statistics of the UWES-9 items for the two samples. The items were positively and significantly correlated ($r=0.19-0.68$, $p<0.01$) and displayed minor skewness or kurtosis (≤ 0.5).

Data analysis

The 1-factor, 3-factor, and bi-factor models of the UWES-9 were examined using ML-based confirmatory factor analysis (CFA) and BSEM using Mplus version 7.2¹⁸⁾. The one-factor model specifies a work

Table 1. Descriptive statistics and correlations of the UWES-9 items in the two samples

Item	1	2	3	4	5	6	7	8	9	M	SD	S	K
1		0.67	0.57	0.39	0.38	0.19	0.45	0.43	0.35	3.89	1.36	-0.30	0.03
2	0.60		0.67	0.53	0.50	0.37	0.57	0.56	0.44	3.89	1.41	-0.24	-0.35
3	0.52	0.65		0.50	0.50	0.25	0.64	0.59	0.37	4.36	1.25	-0.35	-0.35
4	0.48	0.41	0.46		0.48	0.32	0.50	0.47	0.43	3.39	1.39	-0.21	0.10
5	0.39	0.45	0.50	0.43		0.41	0.48	0.53	0.43	3.38	1.56	-0.08	-0.47
6	0.20	0.28	0.25	0.36	0.36		0.35	0.36	0.43	2.26	1.61	0.32	-0.52
7	0.52	0.54	0.60	0.49	0.45	0.29		0.68	0.41	3.94	1.51	-0.38	-0.32
8	0.46	0.56	0.61	0.47	0.50	0.39	0.64		0.59	3.53	1.49	-0.22	-0.21
9	0.30	0.35	0.31	0.46	0.38	0.35	0.40	0.57		3.10	1.56	-0.13	-0.40
M	3.82	3.92	4.33	3.35	3.27	2.24	3.87	3.49	3.10				
SD	1.32	1.42	1.30	1.46	1.59	1.58	1.53	1.52	1.60				
S	-0.23	-0.30	-0.48	-0.20	0.04	0.46	-0.39	-0.13	-0.08				
K	-0.07	-0.24	-0.06	-0.03	-0.58	-0.19	-0.21	-0.34	-0.45				

M=mean; SD=standard deviation; S=skewness; K=kurtosis. Descriptive statistics for Sample 1 (N=556) are displayed on the lower diagonal, while those for Sample 2 (N=556) displayed on the upper diagonal. All correlations are statistically significant at $p<0.01$.

engagement factor, and the three-factor model assumes three factors, vigor, dedication and absorption. The bi-factor model specifies a general factor that loads on all items and two specific factors that each load on three items⁹. The general and specific factors are uncorrelated with each other. The present study did not estimate the full bi-factor model (1 general factor + 3 specific factors for vigor, dedication, and absorption), as the extra specific factor would be likely fully accounted for by the general factor. The ML-based CFA models were conducted using a robust maximum likelihood estimator, and model fit was evaluated based on a χ^2 test of exact fit and two fit indices¹⁹, namely, comparative fit index (CFI) \geq 0.95 and root mean square error of approximation (RMSEA) \leq 0.06. Factor loadings greater than 0.30 were taken as practically significant. Over 93.3% (N=519) of the participants provided complete responses for the UWES-9 items in both samples. The missing data were handled by full-information maximum likelihood estimation. McDonald's coefficient ω , which denotes the proportion of observed variance of the measured items explained by the factor, was used as a measure of composite reliability²⁰.

The BSEM models were estimated using the Bayes estimator with a series of prior specifications for the cross-loadings and residual correlations for the standardized item scores. The BSEM analysis in this study was carried out with reference to Appendix 1 of the recent paper by Asparouhov, Muthén, and Morin²¹. First, BSEM models specified diffuse priors for the hypothesized factor loadings and did not specify informative priors for the cross-loadings and residual covariances. Next, we specified small-variance informative priors for the cross-loadings, choosing prior variances of 0.01 in line with Muthén and Asparouhov¹³. Finally, informative Inverse Wishart (dD,d) priors were added for the residual covariances. A starting value of d=100 was recommended for the informative priors with a sample size near 500²¹ and D referred to the residual variances of the Bayesian CFA models. Two independent Markov chain Monte Carlo chains were used for BSEM estimation using the Gibbs sampler^{22, 23}. Model convergence was monitored by potential scale reduction factor²⁴ and posterior parameter trace plots.

Model fit was evaluated using the posterior predictive *p*-value (PPP), associated 95% confidence interval and number of iterations needed for convergence¹³. A PPP<0.05 and a positive 95% lower limit imply a poor model fit. Sensitivity analysis was performed by varying the informative priors for cross-loadings (variances=0.001, 0.01, 0.05 and 0.1) and residual covariances (d=100, 200, 300 and 400), with the aim of arriving at BSEM models with good model identi-

fication (fast convergence), a PPP>0.05 and reasonable confidence interval limits. The deviance information criterion (DIC)²⁵ or Bayesian information criterion (BIC)²⁶ was used for comparison of BSEM models with different or the same specifications of informative priors, respectively. Both the DIC and BIC avoid model over-fitting by imposing a model complexity penalty based on the estimated and actual number of parameters, respectively. Models with a lower information criterion (a difference of 10 or above) were favored.

Results

ML-CFA model results

Table 2 reports the results of the ML-CFA models for the UWES-9. In both samples, all three models were rejected by the χ^2 test ($p=0.000$), and the fit indices failed to meet the suggested cutoff (CFI<0.95 and RMSEA>0.06). Given the lack of theoretical justification and the possibility of capitalizing on chance features of the sample, model respecification was not carried out in ML-based models using model modification indices. Instead, we turned to BSEM diagnostic analysis to locate the source of model misfit.

BSEM results

Table 3 presents the fit statistics of the BSEM results with different priors. All three BSEM models without informative priors were rejected by the data (PPP=0.000) with a high 95% lower PP limit in both samples. Specification of cross-loading priors (variances=0.01) led to a lower DIC than the Bayesian CFA models and shifted the 95% PP limits closer to zero. However, the PPP for the BSEM with cross-loadings remained at 0.000 in both samples. All BSEM models with informative residual covariance priors (d=300 in Sample 1 and d=200 in Sample 2) consistently provided an adequate fit to the data, with PPP=0.113–0.193 and a negative 95% lower PP limit.

Table 2. Results of three ML-CFA models for the UWES-9

Model	Sample	#	χ^2	<i>p</i>	CFI	RMSEA
1-factor	1	27	131.7	0.000	0.903	0.084
	2	27	177.3	0.000	0.884	0.100
3-factor	1	30	93.5	0.000	0.936	0.072
	2	30	138.7	0.000	0.912	0.093
Bi-factor	1	33	85.4	0.000	0.941	0.074
	2	33	87.4	0.000	0.949	0.075

N=556; ML=maximum likelihood; CFA=confirmatory factor analysis; #=number of free parameters; χ^2 =chi-square value; CFI=comparative fit index; RMSEA=root mean square error of approximation.

These models showed a substantially lower DIC than previous BSEM models with no informative priors or with cross-loading priors. It is worth noting that the three BSEM models with residual covariances showed comparable PPPs and DICs (difference<10). However, the 1-factor BSEM model provided a substantially lower BIC than the 3-factor and bi-factor BSEM models in both samples.

Table 4 displays the factor loadings of the three well-fitting BSEM models with residual covariances in Sample 1. In the 1-factor model, all 9 items loaded substantially ($\lambda=0.43$ to 0.80) on the overall factor ($\omega=0.88$, 95% C.I.= 0.86 to 0.89). Though 11 out of the 36 specified residual correlations were statistically significant (the 95% C.I. did not cover zero), they were all less than 0.20 with a range of -0.14 to 0.16 . In the 3-factor model, the three factors ($\omega=0.70$ to 0.76 , 95% C.I.= 0.66 to 0.79) showed salient factor loadings ($\lambda=0.44$ to 0.91). However, vigor and dedication were found to be extremely highly correlated ($r=0.98$, 95% C.I.= 0.94 to 0.99), suggesting model redundancy. In the bi-factor model, the overall factor ($\omega=0.89$, 95% C.I.= 0.87 to 0.90) had salient loadings ($\lambda=0.41$ to 0.79) on all 9 items. However, the specific factor of dedication was poorly defined by its indicators ($\lambda=0.06$ to 0.10) and the specific factor of absorption

had low reliability ($\omega=0.39$).

Discussion

The present study performed a systematic examination of the dimensionality of the UWES-9 under the ML and BSEM approaches. Under the traditional ML approach, none of the 1-, 3-, and bi-factor CFA models provided an acceptable fit to the data in terms of the highly significant χ^2 test and the approximate fit indices. The mediocre fit may be attributed to the overly restrictive constraints of exactly zero cross-loadings and residual covariances. Though the ML models could be modified via estimation of cross-loadings or residual covariances, simultaneous estimation of all these parameters is not possible in this approach because of the statistical unidentifiability.

On the other hand, the BSEM approach facilitates simultaneous estimation of all residual covariances via informative priors that permit slight deviation from zero if such additions are warranted by the data. The present study applied BSEM analysis to locate the source of model misfit and identify possible model modifications. Without specifying any informative priors, the BSEM models did not fit the data adequately at all. The poor model fit was consistent with that of the ML-CFA models that both fixed

Table 3. Fit statistics of 1-, 3- and bi-factor BSEM with different priors for the UWES-9

Prior specification	Sample	#	pD	2.5% PP limit	97.5% PP limit	PPP	DIC	BIC
No informative priors								
1-factor model	1	27	26.6	166.5	221.3	0.000	12,126	12,244
	2	27	26.6	222.1	278.1	0.000	11,975	12,092
3-factor model	1	30	30.8	102.5	160.1	0.000	12,065	12,192
	2	30	32.3	157.8	214.8	0.000	11,915	12,040
Bi-factor model	1	33	31.9	103.2	157.0	0.000	12,044	12,216
	2	33	27.3	143.2	199.6	0.000	11,896	12,046
Cross-loading priors								
3-factor model	1	48	32.9	59.0	117.0	0.000	12,022	12,260
	2	48	35.2	49.1	147.6	0.000	11,830	12,069
Bi-factor model	1	45	23.4	16.9	75.6	0.001	11,969	12,197
	2	45	38.5	21.8	78.2	0.000	11,779	11,987
Residual covariance priors								
1-factor model	1	63	39.2	-12.1	49.3	0.113	11,957	12,277
	2	63	41.9	-10.7	51.2	0.124	11,751	12,066
3-factor model	1	66	41.0	-16.1	47.5	0.160	11,955	12,290
	2	66	43.1	-12.7	51.1	0.128	11,752	12,083
Bi-factor model	1	69	42.3	-14.9	48.6	0.149	11,957	12,309
	2	69	44.3	-17.1	45.1	0.193	11,748	12,096

N=556; #=number of free parameters; pD=estimated number of parameters; PP limit=posterior predictive limit; PPP=posterior predictive *p*-value; DIC=deviance information criterion; BIC=Bayesian information criterion.

Table 4. Factor loadings of 1-, 3- and bi-factor BSEM with residual covariances for the UWES-9 (Sample 1)

Item	1-factor	3-factor			Bi-factor		
	WE	Vig	Ded	Abs	WE	Ded	Abs
1	0.67 [†]	0.70 [†]			0.68 [†]		
2	0.75 [†]	0.78 [†]			0.76 [†]		
5	0.64 [†]	0.63 [†]			0.64 [†]		
3	0.77 [†]		0.78 [†]		0.79 [†]	0.10	
4	0.64 [†]		0.63 [†]		0.64 [†]	0.06	
7	0.76 [†]		0.76 [†]		0.76 [†]	0.06	
6	0.43 [†]			0.44 [†]	0.41 [†]		0.18 [†]
8	0.80 [†]			0.91 [†]	0.78 [†]		0.31 [†]
9	0.56 [†]			0.61 [†]	0.52 [†]		0.52 [†]

N=556; WE=work engagement; Vig=vigor; Ded=dedication; Abs=absorption. Factor loadings were freely estimated using diffuse priors. Daggers indicate that the 95% credibility interval does not contain zero.

cross-loadings and residual covariances exactly at zero. Despite the improvement in the 95% PP limits and DIC to some extent, the BSEM models with cross-loading priors were still rejected by the data. This implies that the model misfit is unlikely to be attributed to the absence of cross-loadings.

The BSEM models with residual covariance priors showed a good PPP, a negative 95% lower PP limit and a substantially lower DIC than the previous models. Despite equivalent PPPs and DICs for the BSEM models with residual covariances, the substantially lower BIC strongly favors the 1-factor model over the other two models. The 11 residual correlations that were found to be statistically significant were indeed substantively insignificant (<0.20). In line with previous studies^{3-5, 17}, the exceptionally strong inter-factor correlation highlights excessive overlapping among the factors and absence of discriminant validity for the 3-factor model. Similarly, given the weak factor loadings and poor composite reliability for the specific factors, the bi-factor model was not supported in the present study.

The BSEM results suggest the model misfit is due to minor differences between the model and the data in the form of omitted minor residual covariances. We choose to treat these statistically significant but substantively insignificant parameters as approximately zero and interpret the 1-factor model as a sufficiently good and parsimonious approximation for the data. Instead of interpreting subscale scores that are potentially redundant, the present results demonstrate support for use of the total UWES-9 score as a measure of work engagement^{3, 9}.

Despite the large sample size, the present study was

based on a nonrandom sample of health-care workers. The potential selection bias limits the generalizability of the study results to other worker populations. The self-reported cross-sectional nature of the current study implies the possible existence of common method variance. Future studies that adopt a longitudinal design and incorporate objective measures to elucidate the degree of work engagement and its developmental trajectories are recommended.

In summary, this psychometric study was the first to apply the flexible BSEM approach in reevaluating the dimensionality of the UWES-9. The BSEM results demonstrate empirical support for the overall factor as an adequate and parsimonious representation of work engagement. Future research could investigate the measurement invariance of the UWES-9 across gender or cultural contexts using the Bayesian approach²⁷. This innovative approach allows a test of approximate measurement invariance via zero-mean, small variance informative priors for parameter differences between groups^{28, 29}, thereby providing a useful mean of identifying non-invariance in the case of multiple groups or time points.

Acknowledgment: The authors would like to sincerely thank Dr. Tihomir Asparouhov for his invaluable insights concerning the BSEM analysis results.

References

- 1) Schaufeli WB, Salanova M, Gonzalez-Roma V, Bakker AB. The measurement of engagement and burnout: a two sample confirmatory factor analytic approach. *Journal of Happiness Studies* 2002; 3: 71–92.
- 2) Schaufeli WB, Bakker AB. *Utrecht Work Engagement Scale: Preliminary Manual*. Utrecht: Occupational Health Psychology Unit, Utrecht University; 2003.
- 3) Schaufeli WB, Bakker AB, Salanova M. The measurement of work engagement with a short questionnaire - A cross-national study. *Educational and Psychological Measurement* 2006 ; 66: 701–16.
- 4) Balducci C, Fraccaroli F, Schaufeli WB. Psychometric properties of the Italian version of the Utrecht Work Engagement Scale (UWES-9). *A cross-cultural analysis*. *European Journal of Psychological Assessment* 2010; 26: 143–9.
- 5) Nerstad CGL, Richardsen AM, Martinussen M. Factorial validity of the Utrecht Work Engagement Scale (UWES) across occupational groups in Norway. *Scand J Psychol* 2010; 51: 326–33.
- 6) Littman-Ovadia H, Balducci C. Psychometric properties of the Hebrew version of the Utrecht Work Engagement Scale (UWES-9). *European Journal of Psychological Assessment* 2013; 29: 58–63.
- 7) Halbesleben JRB. A meta-analysis of work engagement: Relationships with burnout, demands, resour-

- es, and consequences. In: Bakker AB, Leiter MP, editors. *Work engagement: A handbook of essential theory and research*. New York: Psychology Press; 2010. p. 102–17.
- 8) Extremera N, Sanchez-Garcia M, Duran MA, Rey L. Examining the psychometric properties of the Utrecht Work Engagement Scale in two Spanish multi-occupational samples. *International Journal of Selection and Assessment* 2012; 20: 105–10.
 - 9) de Bruin GP, Henn CM. Dimensionality of the 9-item Utrecht Work Engagement Scale (UWES-9). *Psychol Rep* 201; 112: 788–99.
 - 10) Cai L, Yang JS, Hansen M. Generalized full-information item bifactor analysis. *Psychol Methods* 2011;16: 221–48.
 - 11) Barrett P. Structural equation modelling: adjudging model fit. *Pers Individ Differ* 2007; 42: 815–24.
 - 12) McIntosh CN. Improving the evaluation of model fit in confirmatory factor analysis: a commentary on Gundy CM, Fayers PM, Groenvold, M., Petersen, M. Aa., Scott, N.W., Sprangers MAJ, Velikov G, Aaronson NK. (2011). Comparing higher-order models for the EORTC QLQ-C30. *Quality of Life Research*, doi:10.1007/s11136-011-0082-6. *Qual Life Res* 2012; 21: 1619–21.
 - 13) Muthen B, Asparouhov T. Bayesian structural equation modeling: a more flexible representation of substantive theory. *Psychol Methods* 2012; 17: 313–35.
 - 14) Cole DA, Ciesla JA, Steiger JH. The insidious effects of failing to include design-driven correlated residuals in latent-variable covariance structure analysis. *Psychol Methods* 2007; 12: 381–98.
 - 15) Golay P, Reverte I, Rossier J, Favez N, Lecerf T. Further Insights on the French WISC-IV Factor Structure Through Bayesian Structural Equation Modeling. *Psychological Assessment* 2013; 25: 496–508.
 - 16) Fong TCT, Ho RTH. Factor analyses of the hospital anxiety and depression scale: a Bayesian structural equation modeling approach. *Qual Life Res* 2013; 22: 2857–63.
 - 17) Fong TCT, Ng SM. Measuring engagement at work: validation of the Chinese version of the Utrecht Work Engagement Scale. *Int J Behav Med* 2012; 19: 391–7.
 - 18) Muthén LK, Muthén B. *Mplus user's guide*. 7th ed. Los Angeles (CA): Muthen & Muthen; 1998–2013.
 - 19) Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling* 1999; 6: 1–55.
 - 20) Dunn TJ, Baguley T, Brunsden V. From alpha to omega: a practical solution to the pervasive problem of internal consistency estimation. *Br J Psychol* 2014; 105: 399–412.
 - 21) Asparouhov T, Muthén B, Morin AJ. Bayesian Structural Equation Modeling with Cross-Loadings and Residual Covariances: comments on Stromeyer et al. Accepted for publication in *Journal of Management*. 2015.
 - 22) Asparouhov T, Muthen B. Bayesian analysis of latent variable models using Mplus (Technical report). Los Angeles (CA): Muthen & Muthen, 2010.
 - 23) Lee SY, Song XY. *Basic and Advanced Bayesian Structural Equation Modeling: With Applications in the Medical and Behavioral Sciences*. Chichester (UK): Wiley; 2012.
 - 24) Gelman A, Carlin JB, Stern HS, Rubin DB. *Bayesian data analysis*. 2nd ed. Boca Raton (FL): Chapman & Hall; 2004.
 - 25) Vrieze SI. Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychol Methods* 2012; 17: 228–43.
 - 26) Wagenmakers EJ. A practical solution to the pervasive problems of p values. *Psychon Bull Rev* 2007; 14: 779–804.
 - 27) Verhagen AJ, Fox JP. Bayesian tests of measurement invariance. *Br J Math Stat Psychol* 2013; 66: 383–401.
 - 28) Muthen B, Asparouhov T. BSEM Measurement Invariance Analysis. *Mplus Web Notes* 2013; 17: 1–48.
 - 29) Fong TCT, Ho RTH. Testing gender invariance of the hospital anxiety and depression scale using the classical approach and Bayesian approach. *Qual Life Res* 2014; 23: 1421–6.