

Published as: Lai, C-L., Mak, K-K., Cheng, C., Watanabe, H., Nomachi, S., Bahar, N., Young, K.S., Ko, H-C., Kim, D. & Griffiths, M.D. (2015). Measurement invariance of Internet Addiction Test among Hong Kong, Japanese, and Malaysian adolescents: An item response analysis. *Cyberpsychology, Behavior and Social Networking*, 18, 609-617. DOI: 10.1089/cyber.2015.0069

Abstract

There has been increased research examining the psychometric properties on the Internet Addiction Test across different ages and populations. This population-based study examined the psychometric properties using Confirmatory Factor Analysis and measurement invariance using Item Response Theory (IRT) of the IAT in adolescents from three Asian countries. In the Asian Adolescent Risk Behavior Survey (AARBS), 2,535 secondary school students (55.91% girls) in Grade 7 to Grade 13 (Mean age = 15.61 years; $SD = 1.56$) from Hong Kong ($n = 844$), Japan ($n = 744$), and Malaysia ($n = 947$) completed a survey on their Internet use that incorporated the IAT scale. A nested hierarchy of hypotheses concerning IAT cross-country invariance was tested using multi-group confirmatory factor analysis. Replicating past finding in Hong Kong adolescents, the construct of IAT is best represented by a second-order three-factor structure in Malaysian and Japanese adolescents. Configural, metric, scalar, and partial strict factorial invariance was established across the three samples. No cross-country differences on Internet addiction were detected at latent mean level. This study provided empirical support to the IAT as a reliable and factorially stable instrument, and valid to be used across Asian adolescent populations.

Keywords: measurement invariance, Internet Addiction Test, adolescents

MEASUREMENT INVARIANCE OF INTERNET ADDICTION TEST AMONG HONG KONG, JAPANESE, AND MALAYSIAN ADOLESCENTS

Internet addiction has increasingly become a serious public health concern among the widespread growth of Internet users ¹. The Internet Addiction Test (IAT) was developed by Young as a screening instrument of the disorder ², and later become a widely used assessment of Internet addiction in both Western ³⁻⁴ and Chinese societies ⁵⁻⁷. To adapt its use in different cultures, the IAT has been translated into different languages, including French ⁸, Italian ⁹, Finnish ¹⁰, Korean ¹¹, Malay ¹², and Chinese ¹³. Despite its strong internal consistency, different factor structures have been reported for the construct measured among adult and college students ^{8-10, 14-15}.

In a Chinese population, using a combined exploratory factor analysis-confirmatory factor analysis approach, Chang and Man Law ¹³ successfully replicated an 18-item three-factor structure in a sample of 410 Hong Kong college students. This study was the first to show that the construct of Internet addiction as assessed by IAT was best represented by a second-order structure. Later, Lai and colleagues ¹⁶ successfully replicated this second-order three-factor structure of IAT in a sample of 762 Hong Kong Chinese adolescents. These findings arguably provide indirect evidence for the stability of the construct of IAT from adolescence to adulthood.

In cross-cultural population-based studies, it is a common practice for researchers to directly compare the average or summated score of the scale, and assume that individuals who obtained the same observed score are similar in terms of severity on the construct underlying the research instrument. However, group comparisons of mean differences rely on an established measurement invariance of the research instrument ¹⁷. Therefore, it is necessary to examine the degree to which the IAT assesses the same construct across sociocultural groups, as an extrapolation of its applications in health studies. Nevertheless, whether the factor structure

underlying the IAT construct is equivalent across different sociocultural groups has remained unexplored. These contribute to the importance of the present study, as it examines the factor structure and measures invariance of IAT across three different Asian adolescent populations, thereby allowing the assessment of cross-sociocultural differences with greater confidence.

Comparisons between groups are methodologically valid only after analyzing the equivalence/nonequivalence between test structures across different groups. In other words, the measurement model (relationships between observed indicators and latent variables) do not differ between groups.

With samples from three countries, the present study aimed to extend previous studies to Hong Kong Chinese adolescents, and addressed the following research objectives: (1) to examine the factorial validity of IAT in Japanese and Malaysian adolescents, (2) to examine the measurement invariance of the IAT across Hong Kong, Japanese and Malaysian adolescents, and (3) assuming measurement invariance, to compare the observed and latent means of the IAT between different sociocultural groups.

Methods

Participants

In the 2012–2013 school year, 2,535 secondary school students in Grades 7 to Grade 13 (aged 12 to 18 years) from Hong Kong ($n = 844$), Japan ($n = 744$), and Malaysia ($n = 947$) participated in the Asian Adolescent Risk Behavior Survey (AARBS). The schools were randomly selected from both rural and urban areas—Kowloon and New Territories of Hong Kong; Western Japan, including Shiga; and Central Western Malaysia, including Selayang. Participants were 55.9% female, and had a mean age of 15.61 years ($SD = 1.56$) (see Table 1).

Participants were asked to complete a 40-minute Internet usage survey that included demographic items, Internet use habits, and Young's Internet Addiction Test (IAT)². The Chinese, Japanese, and Malay versions of IAT were prepared by translating the English version. A forward- and backward-translation procedure was implemented by two independent bilingual translators. Items were translated with consideration to the relevance and comprehensibility to adolescents. For instance, “household chores” in Item 2 was changed to “daily hassles”, and “intimacy with partner” of Item 3 was changed to “activities with companions”. The face validity of the translated questionnaire was then assessed for final approval. Participants in each setting were volunteers whom were approached by teachers in the school classroom. The average response rates were over 90%. Informed consent was obtained and ethics approval was granted from the university ethics committee.

Instruments

Internet Use. Participants were asked about their ownership of personal computer and their frequency and duration of Internet use. They also reported the average number of hours spent online daily during school days and holidays, with responses on a 6-point scale from 1 (once a week or less) to 6 (more than 3 times a day) on frequency of Internet use, and on a 4-point scale (1=no personal or shared computer; 2=shared computer with one sibling; 3=shared computer with more than one siblings; 4=own a personal computer) for the question about computer ownership.

The Internet Addiction Test (IAT). The IAT comprises 20 items rated on a 5-point Likert scale². These items were derived from the DSM-IV-TR¹⁸ diagnostic criteria of pathological gambling, and examines the degree of preoccupation and compulsiveness to go online, and the

impact on life related to Internet usage. A high reliability estimate of IAT was consistently reported in adolescent samples with a Cronbach's alpha over .80 in the previous studies¹⁹⁻²¹.

Data analysis

A series of plausible models identified in the previous studies were evaluated with EQS 6.1. These models include: (1) one-factor model (Model 1), (2) Widyanto, Griffiths, and Brunsten's²² three-factor model (Model 3a); (3) Chang and Man Law's¹³ second-order three-factor model (Model 3b); (4) modified Chang and Man Law's¹³ model with three additional pairs of error covariances (Model 3c); (5) a four-factor model modified from Ng's five-factor model by discarding the single-item factor (Model 4); (6) Widyanto and McMurrans's¹⁴ six-factor model (Model 6a); (7) Ferraro, Caci, D'Amico, and Di Blasi's⁹ six-factor model (Model 6b).

The pattern of factor loadings is shown in Figure 1. Missing answers in any of the IAT items were excluded case-wise, resulting in 79 exclusions in the Hong Kong data, and 171 exclusions in the Malaysia data. The normalized Mardia's coefficient of the Hong Kong, Japan, and Malaysia data was 118.97, 104.02, and 255.43, respectively, indicating that the assumption of multivariate normality was not fulfilled. The Satorra–Bentler scaled χ^2 (SB χ^2) correction²³ was hence used to indicate the overall goodness of fit. By convention, a value of 0.90 or more in the comparative fit index (CFI);²⁴ the normed fit index (NFI) and the non-normed fit index (NNFI);²⁵ and a value less than 0.08 in the root mean square error of approximation Root Mean Square Error of Approximation (RMSEA);²⁶ indicate a well-fitted model²⁷⁻²⁸.

Provided that any of the posited models were acceptable in the subgroups, analyses of hierarchical multi-sample confirmatory factor analyses (CFA) across countries were performed. First, configural invariance was established across subgroups by imposing the best-fit model previously obtained separately to each subgroup. This model (Model A) was used as the baseline models to which nested models with equality constraints could be directly contrasted. RMSEA

less than 0.05, complemented with CFI greater than .90 was used to evaluate the configural model fit²⁹. Following sequence of increasingly more restrictive tests of invariance across groups was implemented: (i) first-order factor loadings (Model B), (ii) second-order factor loadings (Model C), (iii) intercepts of observed items (Model D), (iv) intercepts of the latent first-order factors (Model E), (v) the residual variances and covariances of the observed items (Model F), and (vi) disturbances of the latent first-order factors³⁰. If configural invariance did not exist, the analyses would stop at that immediate level. If any measurement variance were found on any level, the Lagrange Multiplier Test (LM Test) was conducted to search for the cross-group equality constraint that most contributed to the misfit. The “full” forms of invariance would be relaxed to obtain partial measurement invariance. Since the corrected scaled difference χ^2 (SDCS) test²³ has been criticized to be too sensitive to large sample size and non-normality^{29, 31}, Chen’s³¹ suggestion that a difference of larger than .01 in CFI and 0.015 in RMSEA were adopted as the criteria for the test of measurement invariance. In the case of incompatible conclusion, the CFI would be chosen as the main criterion.

To evaluate the influence of measurement invariance on cross-group mean comparisons, observed and latent mean differences for each latent construct were computed. Latent mean values were fixed to 0 for Hong Kong adolescents in Hong Kong VS Japan; and Hong Kong VS Malaysia comparison, and Japanese adolescents in Japan VS Malaysia comparison. Finally, statistical significance associated with differences between the latent means was determined on the basis of the z -statistic, and t -statistic for the observed means.

Results

Confirmatory Factor Analysis in Japan and Malaysia data

Similar to the CFA results in Hong Kong adolescents reported elsewhere, Model 3c fitted the data the best in Japanese and Malaysian adolescents, as supported by the smallest SB χ^2 to df

ratio and AIC value (see Table 2). The Cronbach's alphas of the total IAT, Withdrawal and Social Problems (W & SP), Time Management and Performance (TM & P), and Reality Substitute (RS) were .85, .84, .68, and .92, respectively.

Measurement invariance across countries

As shown in Table 3, the configural invariance models all provided a good model fit, providing the baseline model for subsequent comparison of the more restrictive invariance models.

Hong Kong vs. Japan samples. After constraining the first-order factor loadings, no substantial differences between the equal first-order factor loadings model and the baseline model were observed. Following this, both first-order and second-order factor loadings were constrained to be equal, non-significant SDCS in addition with the small change in the fit indices provided evidence of second-order metric invariance of Model 3c across countries. The scalar invariance of the IAT was tested by imposing equality constraint on the intercepts of observed items, and first-order factors. The equal item intercept model and equal first-order factor intercept model did not provide fit significantly better than the less constrained models.

This was followed by imposed equality constrain on all factor loadings, item and factor intercepts, and residual variances and covariances of all observed items. Although the result of the SDCS test was again not significant change in CFI exceeded the cutoff criterion. The LM Test statistics revealed that residual variances of Item 10 “block disturbing thoughts about life with soothing thoughts of the internet” operated differentially between HK and JP adolescents after Bonferroni correction for the number of parameters tested. Allowing this parameter to vary freely resulted in a comparable model fit relative to the factor intercepts invariant model on the basis of CFI and RMSEA indices.

Hong Kong vs. Malaysia samples. The equal first-order factor loadings model was compared with the baseline model, and reported no substantial decrease in CFI and RMSEA albeit the significant differences in SDCS test. Once again, the second-order factor loadings, item, and first-order factor intercepts of Model 3c were all considered to be invariant across the groups as evidenced small change in the fit indices relative to the less restrictive models. Nonetheless, Δ CFI of the model with equality constraint on all residual variances and covariances of the observed items has exceeded the cutoff value. Results of LMTest with Bonferroni correction suggested that six residual variances (Item 13, 15, 16, 17, 18 and 20) and two residual covariances (between Item 1 and 2, and Item 16 and 17) were different across the groups. Freeing the eight analogous pairs of residual variances and covariances of IAT items resulted in no substantial difference in the supplementary fit indices.

Japan vs. Malaysia samples. On the basis of the practical fit indices, first-order factor loadings, second-order factor loadings, item intercepts, and first-order factor intercepts were all considered to be invariant across the groups, given the trivial change in the more restrictive model. It should also be noted that although Δ RMSEA has slightly exceeded cutoff, the respective change in CFI, that is a critical criterion, was less than .01. The equal item intercepts model is thereby considered tenable. Similarly, when analogous pairs of residual variances and covariances were constrained to equal, substantial degradation in fit was revealed. Releasing equality constraint on the nine residual variances (Item 10, 12, 13, 15, 16, 17, 18, 19 and 20) and one pair of residual covariances (between Item 1 and 2) suggested by the LM Test as the major source of misfit, and resulted in a substantial improvement; changes in CFI and RMSEA were consistently small in the modified mode.

Tests of Latent and Observed Mean Differences across Countries

The full-scale invariance established above allowed comparison of latent factor mean differences across the subgroups. The Wald test suggested that significant lower latent means in Japan than Hong Kong adolescents on RS and higher means on TM & P (see Table 3). Consistently, MY adolescents obtained significant higher means on W & SP than HK adolescents in *t* test. However, significant higher observed mean score on TM & P but not on W & SP were obtained. The comparison between Japan and Malaysia adolescents revealed significant difference on all three first-order factors of the IAT. Higher latent means on W & SP and RS, and lower latent means on TM & P were revealed. When observed scores were compared, only the mean of TM & P remained significant. The equality constraints on the intercept of first-order factors were then additionally imposed. When there was no significant latent mean difference across the three samples, significant observed mean difference between Malaysia and Japan sample was found.

Discussion

The large increase in internet use over the last two decades has generated an increased need to use research instruments with individuals across different cultures. Nevertheless, whether the factor structure underlying the IAT construct is equivalent across different sociocultural groups has remained untested until this study. With this in mind, the present study examined the factor structure of the IAT in Malaysian and Japanese adolescents. Additionally, we evaluated the measurement invariance of IAT across three groups of Asian adolescents (from Hong Kong, Malaysian, and Japanese), thereby allowing the assessment of cross-sociocultural differences with greater confidence. Finally, mean differences of the IAT composite, and factor scores from the CFA-based and traditional *t*-test approach was assessed.

With a confirmatory approach, it was found that the construct of the IAT is best represented by a second-order three-factor structure in Malaysian and Japanese adolescents,

which replicated an earlier finding in Hong Kong adolescents¹⁶. This provides evidence for the stability of the construct of Internet addiction as assessed by IAT. It also suggests that the original 20-item IAT – initially designed as unitary dimensional – should be re-organized into 18 items underlying three symptom dimensions including withdrawal and social problems, time management and performance, and reality substitute, that are in turn derived from a higher-order construct “Internet Addiction”. This hierarchical factor structure lends support for the use of both total and factor score of IAT. Whilst the former may be used as a criterion to determine severity of the problem, the latter aids understanding about symptom presentation of the individual.

When the IAT is used in practice, each of its items is assumed to make an equal and important contribution to the three latent constructs. No matter that the factor loadings and/or the intercepts contribute differentially to the factor score across countries, results of comparisons of mean scores and regression coefficients across countries would be ill-advised. The more important findings of this study are that measurement invariance of the IAT was sufficiently upheld. More specifically, factor form, first- and second-order factor loadings, latent factor, and observed item intercepts of IAT were invariant across these three countries. However, residual variances of some items were not invariant across groups. These residual variances reflect portions of variance of these items that were not attributable to the factors of IAT. A cross-group inequality might be the result of both nonsystematic error variance and method variance that cast differential effects on the individual’s item response. Future attention could be placed on the interpretation of Item 10 for Japanese adolescents, and Item 13, 15, 16, 17, 18, and 20 for Malaysian adolescents, in which residual variance inequality were consistently reported. Researchers and practitioners are urged to interpret cross-sociocultural differences corresponding to these non-invariant items with caution.

In relation to the mean differences, even though all invariant item residual variances were fixed to be equal, the impact of measurement error could still be observed. Except the contrast between Hong Kong and Japan sample, the conclusions from the latent means and observed means differences were not entirely consistent. The two methods disagreed particularly in the comparison between Malaysian and Japanese adolescents. Although both statistical analytical methods tend to reflect the difference of the groups on their levels of the underlying construct, the latent mean comparison was commented to be more precise and powerful because measurement errors are adjusted²⁹.

The measurement error inequality entailed may contraindicate group comparison on the observed means with the use of traditional *t* test. This may also account for the incompatible findings of the two statistical comparison methods. On the basis of latent means comparison, no significant differences were obtained on the total IAT score between Hong Kong, Japanese, and Malaysian adolescents. For latent means, “withdrawal and social problems” as it relates to Malaysian adolescents, and “time management and performance” as it relates to Japanese adolescents were significantly higher in Hong Kong adolescents; “withdrawal and social problems” as it relates to Malaysian adolescents and “reality substitute” in relating to Hong Kong adolescents in addition appeared to be higher in Japanese adolescents. Finally, Malaysian adolescents on average scored higher on “time management and performance” than Japanese adolescents. These findings, when taken together, suggest that the three countries studied did not differ with respect to the overall severity on Internet addiction, yet were prone to different domains of symptoms associating with Internet addiction.

The present study may not be able to establish the strict factorial invariance of IAT across countries. Nevertheless, this last step is generally considered to be difficult to achieve and a very stringent test of equivalence. It is therefore still arguable whether or not this step is necessary¹⁷.

^{29, 32-33}. With the evidence of scalar invariance and partial strict factorial invariance, this study provided evidence for the equality of meaning of the construct Internet addiction in the three Asian countries casting confidence for future use of the IAT in a cross-national fashion. From a practical standpoint, the present study also demonstrated the critical effects of item measurement errors on observed group (country; i.e., cultural backgrounds in our case) mean differences, and reminds future researchers to be careful that latent mean scores should also be taken into account alongside the observed mean scores. The present authors believe that test adaptation should not be limited to linguistic analyses of translated instrument, but should also involve psychometric analyses of the inferences derived from the score to assure the equivalence to the original test by the measurement invariance.

Limitations and future studies

First, with the aim of investigating cross-sociocultural difference on Internet addiction, only Hong Kong Chinese were recruited. Consequently, the results are not generalizable to other Chinese populations. Furthermore, this study only focused on internal validity of IAT. Other criteria for an effective and useful assessment, namely predictive, concurrent, and content validity were not examined. As discussed above, mixed findings concerning the factor structure of IAT in Caucasian adults were highlighted in previous validation studies. This calls the generalization of the current model to non-Asian adolescents into question. To claim the IAT as a “culture neutral” instrument, future studies are needed to explore whether non-Asian adolescents conceptualize Internet addiction or problematic Internet use as the same way as their Asian counterparts. Similarly, the measurement invariance of the IAT in adults is also an issue that is yet to be studied and requires further exploration.

In conclusion, the results of the present study strongly support the reliability and factorial validity of the IAT construct. The findings also take a further step in understanding of the

internal validity of IAT, namely to confirm its hierarchical three-factor structure to other non-Chinese Asian samples. In addition, with the employment of multi-group CFA, measurement structure of the IAT was shown to be invariant across Hong Kong, Malaysia, and Japanese adolescents. The major implication is that IAT generally measures the same construct across different Asian countries in adolescents. These findings lay a foundation for future cross-sociocultural investigations of Internet addiction. Meaningful comparisons of statistics such as means and regression coefficients could only be made if the scale (i.e., IAT) is a widely adapted assessment and comparable across different groups. The existence of non-invariant measurement errors of some items suggest that merely interpreting observed mean difference of IAT could lead to erroneous inference. However, after partialing out variance attributable to measurement error, no cross-country comparisons were found to result in significantly different levels of Internet addiction severity.

References

1. Block JJ. Issues for DSM-V: Internet Addiction. *American Journal of Psychiatry*. 2008;165(3):306-7.
2. Young KS. (1998) *Caught in the Net: How to recognize the signs of Internet addiction and a winning strategy for recovery*. New York: John Wiley & Sons.
3. Durkee T, Kaess M, Carli V, et al. Prevalence of pathological Internet use among adolescents in Europe: Demographic and social factors. *Addiction*. 2012;107:2210-22.
4. Johansson A, Götestam KG. Internet addiction: Characteristics of a questionnaire and prevalence in Norwegian youth (12–18 years). *Scandinavian Journal of Psychology*. 2004;45(3):223-9.
5. Cao F, Su L. Internet addiction among Chinese adolescents: Prevalence and psychological features. *Child: Care, Health and Development*. 2007;33(3):275-81. Epub 2007/04/19.
6. Chou C, Hsiao M-C. Internet addiction, usage, gratification, and pleasure experience: The Taiwan college students' case. *Computers & Education*. 2000;35(1):65-80.
7. Zhou Y, Lin F-C, Du Y-S, et al. Gray matter abnormalities in Internet addiction: A voxel-based morphometry study. *European Journal of Radiology*. 2011;79(1):92-5.
8. Khazaal Y, Billieux J, Thorens G, et al. French validation of the Internet Addiction Test. *CyberPsychology & Behavior*. 2008;11(6):703-6.
9. Ferraro G, Caci B, D'Amico A, Blasi MD. Internet addiction disorder: An Italian study. *CyberPsychology & Behavior*. 2007;10(2):170-5.
10. Korkeila J, Kaarlas S, Jääskeläinen M, Vahlberg T, Taiminen T. Attached to the web — harmful use of the Internet and its correlates. *European Psychiatry*. 2010;25(4):236-41.

11. Kim HS. (2000) *Internet addiction*. Seoul: Nanum Press.
12. Ng CG, Isa SM, Hashim AH, Pillai SK, Harbajan Singh MK. Validity of the Malay version of the Internet Addiction Test: A study on a group of medical students in Malaysia. *Asia-Pacific Journal of Public Health*. in press.
13. Chang MK, Man Law SP. Factor structure for Young's Internet Addiction Test: A confirmatory study. *Computers in Human Behavior*. 2008;24(6):2597-619.
14. Widyanto L, McMurrin M. The psychometric properties of the Internet Addiction Test. *CyberPsychology & Behavior*. 2004;7(4):443-50.
15. Jelenchick LA, Becker T, Moreno MA. Assessing the psychometric properties of the Internet Addiction Test (IAT) in US college students. *Psychiatry research*. 2012;196:296-301.
16. Lai CM, Mak KK, Watanabe H, Ang R, Ho RC. Psychometric properties of the Internet Addiction Test in Chinese adolescents. *Journal of Pediatric Psychology*. 2013;38(7):794-807.
17. Schmitt N, Kuljanin G. Measurement invariance: Review of practice and implications. *Human Resource Management Review*. 2008;18(4):210-22.
18. American Psychiatric Association. (2000) *Diagnostic and statistical manual of mental disorders*. Washington, DC: Author.
19. Wang H, Zhou X, Lu C, Wu J, Deng X, Hong L. Problematic Internet use in high school students in Guangdong province, China. *PloS one*. 2011;6(5):e19660.
20. Bayraktar F, Gün Z. Incidence and correlates of Internet usage among adolescents in North Cyprus. *CyberPsychology & Behavior*. 2007;10(2):191-7.
21. Milani L, Osualdella D, Di Blasio P. Quality of interpersonal relationships and problematic Internet use in adolescence. *CyberPsychology & Behavior*. 2009;12(6):681-4.

22. Widyanto L, Griffiths MD, Brunsten V. A psychometric comparison of the Internet Addiction Test, the Internet-Related Problem Scale, and self-diagnosis. *CyberPsychology, Behavior & Social Networking*. 2011;14(3):141-9.
23. Satorra A, Bentler PM. A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*. 2001;66(4):507-14.
24. Bentler PM. Comparative fit indexes in structural models. *Psychological Bulletin*. 1990;107(2):238-46.
25. Bentler PM, Bonett DG. Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*. 1980;88(3):588-606.
26. Steiger JH. Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*. 1990;25(2):173-80.
27. Bentler PM. On the fit of models to covariances and methodology to the bulletin. *Psychological Bulletin*. 1992;112(3):400-4.
28. MacCallum RC, Browne MW, Sugawara HM. Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*. 1996;1(2):130-49.
29. Cheung GW, Rensvold RB. Evaluating goodness-of-fit indexes for testing MI. *Structural Equation Modeling*. 2002;9:235-55.
30. Chen FF, Sousa KH, West SG. Testing measurement invariance of second-order factor models. *Structural Equation Modeling*. 2005;12(3):471-92.
31. Chen FF. Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*. 2007;14(3):464-504.
32. Byrne BM. Testing for the Factorial Validity, Replication, and Invariance of a Measuring Instrument: A Paradigmatic Application Based on the Maslach Burnout Inventory. *Multivariate Behavioral Research*. 1994;29(3):289-311.

33. Vandenberg RJ, Lance CE. A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research. *Organizational Research Methods*. 2000;3(1):4-70.