

Use of Multi-Group Confirmatory Factor Analysis in Examining Measurement Invariance in Counseling Psychology Research

Hui Xu^{*a}, Terence J. G. Tracey^a

[a] Counseling and Counseling Psychology Department, Arizona State University, Tempe, AZ, USA.

Abstract

The purpose of this article is to introduce the theoretical implications and analytic strategies of measurement invariance. The article is focused on three important invariance conditions, consisting of configural invariance, metric invariance, and scalar invariance. Configural invariance refers to a qualitatively invariant measurement pattern of latent constructs across groups. Metric invariance refers to a quantitatively invariant measurement model of latent constructs across groups. Scale invariance refers to invariant mean levels of latent constructs across groups. While each invariance condition depicts one aspect of the relation between latent constructs with manifest observations, a progressive statistical strategy of measurement invariance was introduced based on multi-group confirmatory factor analysis. The article also provided a case example illustrating how to apply and examine measurement invariance in counseling psychology, with detailed theoretical implications and analytic decision-makings in each step. Application of measurement invariance in measurement comparison across multiple groups (e.g., gender, developmental stages, and national boundaries) was discussed and recommended.

Keywords: measurement invariance, multiple groups, confirmatory factor analysis, configural invariance, metric invariance, scalar invariance

The European Journal of Counselling Psychology, 2017, Vol. 6(1), 75–82, doi:10.5964/ejcop.v6i1.120

Received: 2016-02-28. Accepted: 2016-11-15. Published (VoR): 2017-02-14.

*Corresponding author at: Counseling & Counseling Psychology, MC-0811, Arizona State University, Tempe, AZ, 85287-0811, USA. E-mail: huixu5@asu.edu



This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Measurement invariance has been conceived as an important research topic for assessment in cross-population (e.g., cross-cultural) contexts (Cheung & Rensvold, 2002; Kline, 2015). This issue is crucial in that without measurement invariance, there can be no comparison of scores across groups. The key question of measurement invariance rests on three sequential hypotheses or invariance levels, consisting of: 1. do multiple populations endorse the same measurement pattern for the latent constructs; 2. do multiple populations endorse the same psychological meanings for the latent constructs; 3. do multiple populations endorse the same levels of the latent constructs (Kline, 2015; Vandenberg & Lance, 2000). Different hypotheses have different implications for measurement research and each serves as the foundation for the subsequent hypothesis. However, they could be examined in a systematic manner by using multi-group confirmatory factor analysis (CFA). The purpose of the current article was to introduce the theoretical implications of measurement invariance as well as the corresponding analytic strategies, focusing on the three invariance conditions. One example is used to illustrate the theoretical and analytic points.

Theoretical Implication of Measurement Invariance

Measurement invariance addresses the key question of whether measurement of latent constructs varies across multiple groups, with configural invariance, metric invariance, and scalar invariance as the most common conditions. The most common testing theory in evaluating measurement invariance is Classic Test Theory (CTT), which posits that observations of the latent construct (X) are comprised of true scores of the latent construct (T) and random measurement error (E). Mathematically, the relation of these three measurement components could be simplistically formulated as $X = T + E$ (Lord, Novick, & Birnbaum, 1968; Vandenberg & Lance, 2000). So, the measurement model of latent constructs in CTT could imply three important pieces of measurement information: 1. certain observations could indicate latent constructs; 2. these observations indicate latent constructs in a systematic matter; 3. latent constructs exhibit certain levels based on sample observations. Measurement invariance therefore could be examined on three sequential levels, consisting of configural invariance, metric invariance, and scalar invariance (Kline, 2015; Vandenberg & Lance, 2000).

Configural invariance refers to the condition that the model of latent constructs being indicated by certain observations holds across multiple groups (Abrams et al., 2013; Vandenberg & Lance, 2000). When configural invariance is supported, it indicates that the same latent construct could be indicated by the same manifest observations across groups (Vandenberg & Lance, 2000). However, it does not imply that the relation of latent constructs with manifest observations are equivalent across groups. Therefore, the next step is to examine the more constrained metric invariance.

Metric invariance refers to the condition that the relation of latent constructs with observatory indicators holds across multiple groups (Abrams et al., 2013; Vandenberg & Lance, 2000). When metric invariance is supported, it indicates that the same latent construct could be represented by the same manifest observations in an equivalent manner across groups. In other words, it would indicate that the psychological meanings of the measured latent constructs are equivalent across groups (Vandenberg & Lance, 2000). This step warrants subsequent examination of whether the levels of latent constructs are invariant across groups (Abrams et al., 2013; Vandenberg & Lance, 2000). Without metric invariance, it is senseless to compare the means of latent constructs because they indicate psychologically different constructs.

Scalar invariance refers to the condition that the level of the compared latent construct holds across multiple groups (Abrams et al., 2013; Vandenberg & Lance, 2000). When scalar invariance is supported, it would be suggested that different groups could exhibit the same mean level of the same latent construct (Vandenberg & Lance, 2000). Otherwise, differential mean levels of the same latent construct across groups would be implied.

Measurement invariance serves as important tool to address cross-cultural validity issues raised by the multicultural movement of counseling psychology (Arnett, 2008; Sue & Sue, 1990). It has been widely acknowledged that psychological knowledge garnered from one cultural group cannot automatically generalize to other cultural groups, as different cultures have appreciable differences regarding how individuals should perceive and react with the world (Arnett, 2008; Sue & Sue, 1990). This critique particularly applies to counseling psychology, as the field often investigates culturally sensitive phenomena, such as mental health concerns. Therefore, a careful examination of the cross-cultural measurement invariance is necessary and important as to helping counseling psychologists understand how psychological constructs are perceived cross-culturally.

Analytic Strategies of Measurement Invariance

One of the most common strategies to examine measurement invariance is multi-group confirmatory factor analysis (Kline, 2015). While confirmatory factor analysis examines whether the hypothesized measurement model fits the data well, multi-group confirmatory factor analysis could be used to precisely compare the measurement model across groups. The progressive analytic strategy in multi-group CFA involves three steps, corresponding to the three invariance conditions of configural, metric, and scalar invariance (Kline, 2015; Vandenberg & Lance, 2000).

To examine configural invariance, the same measurement model could be examined separately for each group by using CFA. The fit of the models could be evaluated using the common criteria recommended by Hu and Bentler (1999): robust chi-square, Comparative Fit Index (CFI < .90), Root Mean Square Error of Approximation (RMSEA < .08), and Standardized Root Mean Square Residual (SRMR < .08). If the same measurement model fits the data well across groups, then configural invariance is supported.

The second step of invariance examination is to examine metric invariance by comparing two nested models, consisting of a baseline model and an invariance model. The baseline model allows the factor loadings to be freely estimated across multiple groups. The invariance model constrains the factor loadings to be equivalent across multiple groups. Differences between the two nested models are examined with the chi-square difference test (Muthén & Muthén, 2012) and the Δ CFI (Cheung & Rensvold, 2002; Meade, Johnson, & Braddy, 2008). A non-significant result of the chi-square difference test would indicate that the invariance model is a better representation of the data because it fits the data equivalently relative to the baseline model but has better parsimony (Muthén & Muthén, 2012). In contrast, a significant result of the chi-square difference test would indicate that the baseline model is a better representation of the data, suggesting that the psychological meanings of the latent constructs vary across groups.

However, the chi-square difference test has been revealed to be highly sensitive to the sample size and less sensitive to a lack of invariance than Δ CFI (Cheung & Rensvold, 2002; Meade et al., 2008). Simulation studies comparing multiple goodness-of-fit indices (e.g., chi-square, AIC, RMSEA, and CFI) have recommended Δ CFI as it is independent of model complexity and sample size and a Δ CFI less than .01 indicates invariance (Cheung & Rensvold, 2002; Meade et al., 2008). Meade et al. (2008) has suggested that if Δ CFI indicates invariance and the sample size is greater than 200, any differences between groups are probably trivial and further analyses could proceed, even though the chi-square difference test is significant.

It is noteworthy that full metric invariance, where all latent constructs have the same psychological meanings across groups, could be rare in applied settings. In this case, partial metric invariance, where several of the latent constructs exhibit the same psychological meanings across groups, could warrant further examination of scalar invariance on those latent constructs.

The final step of invariance examination is to examine scalar invariance by comparing the means of the latent constructs. Essentially this step could be conducted by using the similar nested-model comparison strategy as introduced in the previous step. In this case, the baseline model would allow the means to be freely estimated across multiple groups. The invariance model would constrain the means to be equivalent across multiple groups. However, different statistical programs could have different default specifications regarding the mean structures. We used Mplus as an example as it is the most commonly used statistical package for structural equation modeling.

In the environment of Mplus, means of latent constructs are freely estimated by default and means for the reference group would be set as zero (Muthén & Muthén, 1998-2010). So a significant mean of a compared group would indicate that this group has a different level of the latent construct relative to the reference group. A practical suggestion regarding this issue is to always check the default setting of the statistical program being used, through reading its manual or help documentation.

One Case Illustration

Measurement invariance provides critical information as to how latent constructs investigated are manifested cross-culturally and therefore it fits closely with the vision and value of multicultural counseling research and practice (Arnett, 2008; Sue & Sue, 1990). The next case example (Xu, Hou, Tracey, & Zhang, 2016) illustrates how to apply and examine measurement invariance step-by-step in one research topic of counseling psychology. The data and method were drawn from the original study and thus only a synopsis was provided. The analytic decision making and theoretical implications of each step were detailed. While the original study examined measurement invariance of all the factors, the current article reanalyzed the data and primarily focused on only one factor for the demonstration purpose.

There has been an emerging proposition in vocational psychology that ambiguity is an inevitable component in the career decision-making process and therefore the ability to handle this ambiguity is critical with respect to career decision outcomes (Xu & Tracey, 2014, 2015a; Xu & Tracey, 2015b). Xu and Tracey (2015b) have proposed and demonstrated that ambiguity tolerance specific to career decision-making (i.e., career decision ambiguity tolerance) is an important construct concerning career decision making and revealed a three-factor structure of preference, tolerance, and aversion in U.S. college students. However, it remains unestablished in terms of the measurement invariance of this construct across cultures. The current example thus examined the measurement invariance of career decision ambiguity tolerance across China and the U.S. college students. It was hypothesized that Chinese students would have a lower mean level of both preference and aversion relative to U.S. students, with no differences on tolerance.

The multi-group CFA was conducted based on comparison between a Chinese sample and a U.S. sample. The Chinese sample consisted of 356 college students recruited from five universities, all of which are located in the urban areas of mainland China. The U.S. sample consisted of 328 undergraduate students recruited from a southwest state university. The detailed demographic information for both groups can be found in Xu et al.'s (2016) original article. The 18-item Career Decision Ambiguity Tolerance Scale (CDAT; Xu & Tracey, 2015b) was used in Xu et al.'s (2016) study to measure people's ambiguity tolerance in career decision making. It contains three subscales, consisting of preference, tolerance, and aversion. The reliability and validity of CDAT have been reported to be satisfactory (Xu & Tracey, 2015b).

Configural Invariance

The configural invariance of the three-factor measurement model between China and the U.S. was first examined (see Table 1 for summarized results).

Table 1

Summary of Model Fit Indices for the Three-Factor Measurement Model

Model	χ^2	df	CFI	RMSEA		SRMR
				Estimate	90% C. I.	
Chinese College (n = 356)						
3-factor original model	462.89	132	.77	.08	[.08, .09]	.09
2-factor final model	127.32	53	.92	.06	[.05, .08]	.06
U.S. College (n = 328)						
3-factor original model	245.70	132	.86	.07	[.06, .08]	.08
2-factor final model	86.07	53	.97	.04	[.03, .06]	.05

As can be seen by the value of RMSEA (.08), SRMR (.09), and CFI (.77), the three-factor structure derived from the U.S. college students was not a good representation of the CDAT structure in Chinese college students. It was found that the factor of tolerance in the original structure had poor loadings on the last three items. Therefore, configural invariance was not supported on the whole three-factor measurement model. Given this, only the two factors of preference and aversion were examined with respect to metric invariance and the tolerance factor was dropped. We examined the structure of this two-factor model in the two cross-cultural samples. As can be seen by the values of RMSEA (.06 and .04), SRMR (.06 and .05), and CFI (.92 and .97), the two-factor structure fit the data well in the Chinese sample as well as fitting in the U.S. sample. Thus, the factors of preference and aversion were supported in both Chinese college students and U.S. college students, which allowed for the following metric and scalar invariance examination.

Metric Invariance

While the metric invariance of CDAT-preference and CDAT-aversion were examined in the original study, the present case only focused on CDAT-aversion in the subsequent examinations of metric and scalar invariance between Chinese and U.S. college students for the purpose of brevity. As can be seen by the values of CFI (.93 and .99), RMSEA (.09 and .03), and SRMR (.04 and .03), the measurement model of CDAT-aversion in Chinese and U.S. college students both fit the data adequately. Table 2 summarized the results of model-data fit for the two nested models examined for metric invariance.

Table 2

Summary of Multi-Group Comparisons for Metric Invariance of CDAT-Aversion

Model	χ^2	df	CFI	RMSEA		SRMR
				Estimate	90% C. I.	
Model a. Loadings freely estimated	60.91	23	.94	.07	[.05, .09]	.04
Model b. Factor loadings invariant	87.84	28	.91	.08	[.06, .10]	.09

We first specified the baseline model (Model a), where all factor loadings were freely estimated. As can be seen by the values of CFI (.94), RMSEA (.07), and SRMR (.04), Model a fit the data adequately. Based on Model a, we specified the invariant model (Model b), where all factor loadings across Chinese and U.S. college students

were constrained to be equal. As can be seen by the values of CFI (.91), RMSEA (.08), and SRMR (.09), Model b fit the data adequately as well.

The corrected chi-square difference test indicated that Model b was significantly worse in fit than Model a, scaled $\Delta\chi^2 (5, N = 684) = 26.79, p < .05$. In addition, the Δ CFI was $> .01$, indicating that the two models could have substantial difference in practical terms. Therefore, it was indicated that the metric invariance of CDAT-aversion was not supported between Chinese and U.S. college students, suggesting that the factor of aversion could have different psychological meanings across China and the U.S. college students.

Scalar Invariance

As the metric invariance of CDAT-aversion was not supported, comparison of the mean of CDAT-aversion across Chinese and U.S. college students became senseless. However, we still proceeded to examine scalar invariance only for the purpose of demonstration.

While Chinese students were set as the reference (i.e., zero), it was found that CDAT-aversion was not significantly different from zero (unstandardized estimate = $-.02, p > .05$). The results, thus, suggested that Chinese and U.S. college students tend to experience an equivalent level of anxiety and avoidance tendency in career decision making. Again, it should be cautioned that such an interpretation relies on metric invariance, which was not supported in this case.

Conclusion

The illustrational study demonstrated that Chinese and U.S. college students perceive ambiguity tolerance differently at both the qualitative and quantitative levels. As can be seen by the configural results, Chinese students do not consistently endorse tolerance as an approach in interacting with ambiguity in career decision making, while U.S. students perceive tolerance as an approach independent from preference and aversion. It was additionally revealed by the scalar results that students in both cultural contexts endorse a similar level of aversion associated with ambiguity in career decision making. The illustrational case thus spoke well to the potential variance of measurement psychometrics across cultures.

Over the past three decades, counseling psychology as a field has witnessed an arising need to investigate and examine counseling theory in its multicultural and international context (Arnett, 2008; Sue & Sue, 1990). Meanwhile, there has been a noticeable shift of research methodology to quantitative methods in counseling psychology (e.g., Frazier, Tix, & Barron, 2004; Quintana & Maxwell, 1999). The quantitative method of measurement invariance introduced in this article reflects a promising fusion of the two trends, which makes precise examination of cross-cultural comparison plausible.

Overall, measurement invariance is an important research topic when researchers hope to investigate whether the measurement model of the latent constructs (including pattern and magnitude) holds across groups and whether the mean levels of the latent constructs hold across groups. An equivalent measurement model serves as a prerequisite for subsequent mean comparison, as it is illogical to compare means when means indicate different psychological constructs. While the present paper used a cross-cultural study to demonstrate a flexible and powerful statistical examination of measurement invariance, this analytic approach could be eventually applied

in any multi-group comparison, such as comparison across gender, developmental stages, social economic status, and national boundaries.

Funding

The authors have no funding to report.

Competing Interests

The authors have declared that no competing interests exist.

Acknowledgments

The authors have no support to report.

References

- Abrams, M. D., Ómarsdóttir, A. O., Björnsdóttir, M. D., Einarsdóttir, S., Martin, C., Carr, A., . . . Rector, C. (2013). Measurement invariance of the Career Indecision Profile: United States and Iceland. *Journal of Career Assessment, 21*, 469-482. doi:10.1177/1069072712475181
- Arnett, J. J. (2008). The neglected 95%: Why American psychology needs to become less American. *The American Psychologist, 63*, 602-614. doi:10.1037/0003-066X.63.7.602
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*, 233-255. doi:10.1207/S15328007SEM0902_5
- Frazier, P. A., Tix, A. P., & Barron, K. E. (2004). Testing moderator and mediator effects in counseling psychology research. *Journal of Counseling Psychology, 51*, 115-134. doi:10.1037/0022-0167.51.1.115
- 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*, 1-55. doi:10.1080/10705519909540118
- Kline, R. B. (2015). *Principles and practice of structural equation modeling*. New York, NY, USA: Guilford.
- Lord, F. M., Novick, M. R., & Birnbaum, A. (1968). *Statistical theories of mental test scores*. Reading, MA, USA: Addison-Wesley.
- Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *The Journal of Applied Psychology, 93*, 568-592. doi:10.1037/0021-9010.93.3.568
- Muthén, L. K., & Muthén, B. O. (1998-2010). *Mplus User's Guide: Statistical analysis with latent variables* (6th ed.). Los Angeles, CA, USA: Muthén & Muthén.
- Muthén, L. K., & Muthén, B. O. (2012). Chi-square difference testing using the Satorra-Bentler Scaled Chi-Square. Retrieved from <http://www.statmodel.com/chidiff.shtml>
- Quintana, S. M., & Maxwell, S. E. (1999). Implications of recent developments in structural equation modeling for counseling psychology. *The Counseling Psychologist, 27*, 485-527. doi:10.1177/0011000099274002

- Sue, D. W., & Sue, D. (1990). *Counseling the culturally different: Theory and practice*. New York, NY, USA: John Wiley & Sons.
- 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, 4-70. doi:10.1177/109442810031002
- Xu, H., Hou, Z.-J., Tracey, T. J., & Zhang, X. (2016). Variations of career decision ambiguity tolerance between China and the United States and between high school and college. *Journal of Vocational Behavior*, 93, 120-128. doi:10.1016/j.jvb.2016.01.007
- Xu, H., & Tracey, T. J. (2014). The role of ambiguity tolerance in career decision making. *Journal of Vocational Behavior*, 85, 18-26. doi:10.1016/j.jvb.2014.04.001
- Xu, H., & Tracey, T. J. (2015a). Ambiguity tolerance with career indecision: An examination of the mediation effect of career decision-making self-efficacy. *Journal of Career Assessment*, 23, 519-532. doi:10.1177/1069072714553073
- Xu, H., & Tracey, T. J. (2015b). Career Decision Ambiguity Tolerance Scale: Construction and initial validations. *Journal of Vocational Behavior*, 88, 1-9. doi:10.1016/j.jvb.2015.01.006