

# 測量衡等性方法與流程之文獻回顧

## Review on Measurement Invariance Methods and Procedures

陳文英/真理大學  
Wen-Ing Chen / Aletheia University

真理大學運動知識學院  
運動知識學報 第九期 抽印本

中華民國 一百零一 年 七 月 三十 日

# **Review on Measurement Invariance Methods and Procedures**

**Wen-Ing Chen / Aletheia University**

## **Abstract**

Invariance testing in sport management has been attracting increasing attention in recent years. This paper was to review methods and procedures associated with invariance testing based on past literature. More specifically, this paper centered on the topics including the need for investigating invariance testing, suggested procedures for measurement invariance and issues regarding invariance testing, followed by conclusion. This paper was of great help for researchers dedicated to multiple group research such as cross-cultural studies.

**Keywords:** invariance testing, literature review, structural equation modeling, confirmatory factor analysis

## I. INTRODUCTION

One of the most important assumptions in testing mean differences is that the measurement of the underlying construct is invariant across multiple groups (Byrne, Shavelson, & Muthen, 1989). If researchers ignore this assumption, then there is no assurance whether the score difference is due to real differences on the construct being measured or it is due to the differential functioning of the test across groups (Millsap & Kwok, 2004). Besides, the issues concerning measurement invariance have been of interest to researchers in various research fields, including gerontology, education, individual differences, cross-cultural psychology, developmental psychology, marketing, criminology, sport psychology, and the organizational sciences (Vandenberg & Lance, 2000). Therefore, the issue of measurement invariance warrants researchers' attention. Among various available statistical techniques for investigating whether a given set of items fulfill measurement invariance, confirmatory factor analysis is currently the most important one (Byrne et al., 1989; Vandenberg, 2002; Vandenberg & Lance, 2000). Consequently, the current paper primarily focused on the use of the confirmatory factor analysis. The organization of this paper is as follows: first, the suggested procedures were introduced, followed by the issues associated with the measurement invariance. Finally, conclusions and recommendations were provided.

### **Suggested Procedures for Measurement Invariance**

The recommended procedures for investigating measurement invariance across groups vary from researchers to researchers (Vandenberg & Lance, 2000). Byrne (1998, p. 259) proposed a procedure in which five questions were to be answered. They included (a) is the measurement model group-invariant? (b) is the factorial structure invariant across populations? (c) are certain paths in a specified causal structure invariant across populations? (d) are the latent means of particular constructs in a model different across populations? (e) does the factorial structure of a measuring instrument replicate across independent samples of the same population? It is basically an issue of cross-validation. Another recommended procedure was proposed by Steenkamp and Baumgartner (1998). They started with investigating pool covariance matrices, followed by the examination of the factorial structure of the constructs, latent means, and factor covariance.

Vandenberg and Lance (2000) recommended a set of procedures for investigating measurement invariance using confirmatory factor analysis framework (p. 12); they were as follows:

- Step 1: An omnibus test of the equality of covariance matrices across groups, i.e., a test of the null hypothesis of invariant covariance matrices across different groups.
- Step 2: A test of "configural invariance," (Horn & McArdle, 1992) meaning that a null hypothesis is to be tested in which the same pattern of fixed and free factor loadings is specified for each group. Note that in order for the subsequent tests to be meaningful, configural invariance must be fulfilled.

- Step 3: A test of “metric invariance” (Horn & McArdle, 1992) or a test of null hypothesis in which factor loadings are invariant across groups. (At least partial) metric invariance must be established in order for subsequent tests to be meaningful.
- Step 4: A test of “scalar invariance” (Meredith, 1993; Steenkamp & Baumgartner, 1998) or a test of the null hypothesis where intercepts of regression on the latent variable are invariant across groups.
- Step 5: A test of the null hypothesis that unique variances are invariant across groups. Tests of scalar invariance should be conducted only if (at least partial) metric invariance is met, and tests of invariant uniquenesses should proceed only if (at least partial) metric and scalar invariance has been fulfilled first.
- Step 6: A test of the null hypothesis that factor variances are invariant across groups.
- Step 7: A test of the null hypothesis that factor covariances are invariant across groups. Oftentimes, and Step 6 are combined in an omnibus test of the equality of the latent variables’ variance/covariance matrices across groups.
- Step 8: A test of the null hypothesis of invariant factor means across groups, which often was invoked as a way to test for differences between groups in level on the construct of interest.

In summary, the recommended procedures mentioned in this section are not exhaustive; there are still many similar suggested practices out there. The point here was that there is no consensus among researchers in terms of the number of tests being conducted or the sequence/order of the tests. Although there was an agreement with respect to conducting some specific invariance testing (e.g., constraining factor loadings to be equal across groups to test metric invariance), there was little consensus among the researchers regarding the set of tests that constitutes a thorough examination of invariance testing (Vandenberg & Lance, 2000). Likewise, there was little consensus on the particular sequence of tests that should be conducted, although most researchers agreed that a test of configural invariance is necessary and serves as a baseline model for further tests (Vandenberg & Lance, 2000). Additionally, little consensus exists about what aspects should be tested or in what sequence (Vandenberg & Lance, 2000). The lack of a clear consensus regarding the ordering or sequencing of tests is expected (Marsh, 1994) since the specific invariance tests as well as the ordering of the tests should be and will be determined by the aims and goals of the study (Vandenberg & Lance, 2000), indicating that there may not be a sequence of invariance tests that is universally applicable for all situations. However, Vandenberg and Lance (2000) argued that tests of measurement invariance should precede tests of structural invariance.

### **Issues Regarding the Invariance Testing**

#### *Partial Invariance*

Sometimes when the strict invariance fail to establish, researchers may turn to partial invariance. Partial invariance means that some model parameters are found to vary across populations while others are invariant (Millsap & Kwok, 2004). However, there are two

potential issues concerning the partial invariance: inflated type one error and constraining some parameters for the purpose of identification (Millsap & Kwok, 2004). Additionally, there is no consensus as to the partial invariance. For example, Bagozzi and Edwards (1998) discussed tests of partial invariance only for the tests of metric invariance, whereas Byrne et al. (1989) suggested that partial invariance in reference to metric and structural invariance. Moreover, Steenkamp and Baumgartner (1998) recommended partial invariance tests at each step in their recommended sequence. In conclusion, difference exists in conducting partial invariance, implying that researchers should justify what kind of strategies they used.

#### *SEM Approach VS. IRT Approach*

To test the invariance of an instrument across populations, the most frequently used techniques include structural equation modeling (SEM) approach and item response theory (IRT) approach. The SEM approach has some advantages over the IRT approach. The primary difference between these two approaches lies in the assumptions as to the relationship between the manifest variables and the latent variables. The SEM approach assumes linear relationships between manifest variables and latent variables. The IRT approach, on the other hand, assumes the non-linear relationships between manifest variables and latent variables. The IRT approach's assumption of non-linear relationship comes closer to the real case when data are of dichotomous characteristics; however, the SEM approach would perform better when the scale has multiple levels, e.g. Likert scale. Generally speaking, the SEM approach has better applicability than the IRT approach in social and behavioral sciences because the data format tends to be of Likert scale rather than the dichotomous scale (Chiu, 2003).

#### *Other Issues*

The invariance tests serve as or related with other models. For example, Chan (1998) viewed invariance tests as logical prerequisites to conducting latent growth modeling of longitudinal data. Also, Cole and Maxwell (1985) discussed invariance tests as they related to multitrait-multimethod analyses. Thus, invariance testing has become an important tool that warrants researchers' attention.

Another issue regarding the invariance testing was the inconsistency in the terminology among the researchers. To illustrate, the test of configural invariance was named as a baseline model (Bagozzi & Edwards, 1998), a test of equality of factor structures (Cole & Maxwell, 1985), or of equal number of factors and factor patterns (Taris, Bok, & Meijer, 1998). Additionally, the metric invariance was named variously as a test of equality of scaling units (Cole & Maxwell, 1985), factorial invariance (Horn & McArdle, 1992), factor loading invariance (Marsh, 1994), and so on. Another example was the test of invariant uniquenesses, which is equivalent to invariant error variances (Bagozzi & Edwards, 1998; Steenkamp & Baumgartner, 1998; Taris et al., 1998). Different terminology results in confusion and difficulty in applying this technique. As a result, Vandenberg and Lance (2000) recommended the terminology as common ones for future studies on invariance testing issues based on

technical accuracy and neutrality; they were invariant covariance, configural invariance, metric invariance, scalar invariance, invariant uniquenesses, invariant factor variances, invariant factor covariances, and equal factor means.

The frequency of the tests being discussed varied across researchers. Among all tests, metric invariance has been discussed most frequently, followed by tests of configural invariance and of invariant uniqueness across groups. On the contrary, a test of scalar invariance has been discussed least frequently, followed by the test of equal factor means. The reason may be that mean structures is a relatively recent developed area in the structural equation modeling literature. Therefore, the infrequently discussed tests do not imply that some authors omitted some important aspects of invariance testing that others did discuss (Vandenberg & Lance, 2000).

## II. Conclusions

The multiple-group SEM is basically an extension of the basic SEM analysis by adding a categorical variable such as gender and social-economic status. It can be of extensive applicability in a variety of situations. However, the basic assumptions have to be met under the generic SEM framework. Besides, researchers should pay attention to the logic of conducting the invariance testing. For instance, invariance testing across groups assumes well fitting models (Byrne et al., 1989, p. 456). Additionally, if configural invariance fails to fulfill across groups, then further tests are unwarranted due to the fact that observed measures represent different constructs within each group (Vandenberg & Lance, 2000). As a consequence, researchers should be cognizant of what is supposed to be conducted.

Although there are various recommended procedures regarding invariance testing proposed by different researchers; there is no consensus among researchers in terms of the number of tests being conducted or the sequence/order of the tests. The lack of a clear consensus regarding the ordering or sequencing of tests might be expected since the aims and goals of the study determine the specific invariance tests as well as the ordering of the tests. As noted by Vandenberg and Lance (2000), tests of measurement invariance should precede tests of structural invariance.

Other issues over the invariance testing need to be addressed. For the partial invariance issue, there is no consensus on the steps or procedures. Researchers should make wise judgment on choosing the strategy and provide appropriate justifications. The inconsistency in the terminology over the invariance testing among researchers is another issue. Recommendations regarding the terminological issue have been proposed by Vandenberg and Lance (2002). Finally, the frequency of different tests in invariance testing being conducted varies greatly because the mean structures is a relatively new topic in the structural equation modeling area. Infrequent tests do not imply that they are of less importance. Instead, applied researchers should avoid adopting everything from what have been done in previous study without wise judgment. Rather, adopt and utilize what is appropriate to answer your own research questions in your study.

### III. References

- Bagozzi, R. P., & Edwards, J. R. (1998). A general approach for representing constructs in organizational research. *Organizational Research Methods, 1*, 45-87.
- Byrne, B. M., Shavelson, R. J., & Muthen, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105*(3), 456-466.
- Chan, D. (1998). The conceptualization and analysis of change over time: An integrative approach incorporating longitudinal mean and covariance structures analysis (LMACS) and multiple indicator latent growth modeling (MLGM). *Organizational Research Methods, 1*, 421-483.
- Chiu, H. C. (2003). *Principles and practice of structural equation modeling with LISREL*. Taiwan, Taipei: Yeh-Yeh Publication.
- Cole, D. A., & Maxwell, S. E. (1985). Multitrait-multimethod comparisons across populations: A confirmatory factor analytic approach. *Multivariate Behavioral Research, 20*, 389-417.
- Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research, 18*, 117-144.
- Marsh, H. W. (1994). Confirmatory factor analysis models of factorial invariance: A multifaceted approach. *Structural Equation Modeling, 1*, 5-34.
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*, 525-543.
- Millsap, R. E., & Kwok, O-M. (2004). Evaluating the impact of partial factorial invariance on selection in two populations. *Psychological Methods, 9*(1), 93-115.
- Steenkamp, J. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research, 25*, 78-90.
- Taris, T. W., Bok, I. A., & Meijer, Z. Y. (1998). Assessing stability and change of psychometric properties of multi-item concepts across different situations: A general approach. *Journal of Psychology, 132*, 301-316.
- Vandenberg, R. J. (2002). Toward a further understanding of and improvement in measurement invariance methods and procedures. *Organizational Research Methods, 5*(2), 139-158.
- Vandenberg, R. J., & Lance, C. E. (2002). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 2*, 4-69.

## 測量衡等性方法與流程之文獻回顧

陳文英/真理大學

### 摘 要

近年來衡等性測驗的應用在運動管理逐漸受到重視。本文主要回顧過去相關文獻探討衡等性測驗的方法與流程。明確的說，本文主要包含下列議題：探討衡等性測驗的迫切性、衡等性測驗的建議流程與相關議題，最後本文提出結論。本文將有助於致力多群體研究的研究者，如跨文化研究議題等。

**關鍵詞：**衡等性測驗、文獻回顧、結構方程模式、驗證性因素分析