

Cross-Group Comparisons: A Cautionary Note

Boonghee Yoo

St. Cloud State University

ABSTRACT

This article alerts researchers to the importance of factorial invariance in comparative studies. Cross-group or cross-national comparisons, without a clear understanding of factorial structures, can result in misleading conclusions regarding compared groups. The types and process of invariance test are discussed. Then, as an empirical illustration, American consumer ethnocentrism toward Japanese products is examined across gender and age. © 2002 Wiley Periodicals, Inc.

In marketing research, cross-group differences are widely compared because groups (e.g., gender, age, culture, ethnic origin, city, and nation) are considered important criteria for major market segmentation and consumer classification. However, from a methodological viewpoint, mean group comparisons assume measurement invariance in which the factorial structure is equivalent across groups. Minimally, mean comparisons require scalar invariance or strong factorial invariance (i.e., cross-group invariance of every variable's factor loading and intercept; Meredith, 1993). Marketing and consumer researchers, however, have paid little attention to factorial invariance, taking the underlying methodological assumptions for granted. To the extent that invariance is weak, the credibility of study findings is questionable.

The purpose of the present article is to discuss and illustrate the importance of the measurement invariance and structural invariance of

measures across groups of consumers. Cross-group, cross-cultural, or cross-national comparisons, without a clear understanding of factorial structures, may produce misleading conclusions regarding compared groups.

CROSS-GROUP COMPARISONS AND FACTORIAL INVARIANCE

Cross-group factorial invariance assesses whether specific parameters of the measurement or structural model are equivalent across groups of demographic types and/or nations. There are two types of invariance tests: the measurement invariance test and the structural invariance test. The structural invariance test assesses whether a proposed structural model is equivalent across groups. On the other hand, the measurement invariance test assesses whether a measurement model is equivalent across groups.

Full measurement invariance refers to the overall equality of covariance structure across groups [i.e., $H_0: \Sigma_1 = \Sigma_2 = \dots = \Sigma_g$, where g is the number of groups]. Failure to reject this null hypothesis furnishes sufficient evidence of invariance. But rejection of the hypothesis leads to further testing of the equivalence in the factor loadings [i.e., $H_0: \Lambda_1 = \Lambda_2 = \dots = \Lambda_g$] and error variances [i.e., $H_0: \Theta_1 = \Theta_2 = \dots = \Theta_g$] of the indicators, as well as the factor variance-covariances [i.e., $H_0: \Phi_1 = \Phi_2 = \dots = \Phi_g$] (Bollen, 1989; Jöreskog & Sörbom, 1993).

Furthermore, there exist three levels of factorial invariance for a measurement model, which are specialized conditions of measurement invariance because in every level of factorial invariance the factor variance-covariances can be freely correlated across groups (Meredith, 1993; Steenkamp & Baumgartner, 1998). First, *metric invariance* or weak factorial invariance (i.e., invariance of factor loadings) is a minimal requirement for testing mean differences in individual observed variables across groups. This level of invariance, which achieves equal metrics, indicates that members in different groups interpret and respond to measures in an equivalent manner. In other words, metric invariance is evidence of measuring the same construct in multiple groups. Second, *scalar invariance* or strong factorial invariance (i.e., invariance of intercepts of the manifest variables in addition to metric invariance) is required for further testing of mean differences of the latent variables across groups. Meredith (1993) insists that scalar invariance is required for comparative validity in cross-group studies and the lack of scalar invariance can make mean group comparisons hazardous. Third, *strict factorial invariance* (i.e., invariance of unique variances in addition to scalar invariance) is a condition under which composite variances and covariances can be meaningfully compared. But

this highest level of factorial invariance is not considered necessary for the measurement invariance.

Invariance Testing Process

To detect invariance, a chi-square fit difference test is conducted between an unconstrained model and a constrained model. In the unconstrained model of measurement invariance, every parameter is specified to vary across groups, whereas in the constrained model of measurement invariance, factor loadings, error variances, and factor variance-covariances are increasingly fixed to be equal across groups (Bollen, 1989; Jöreskog & Sörbom, 1993). If the chi-square test exhibits no significant difference between the models, there exists cross-group invariance for the parameters. In a scalar invariance test, both factor loadings and intercepts are constrained to be the same in the constrained model (Meredith, 1993).

In the unconstrained model of structural invariance, the structural path parameters (i.e., Γ and B matrices), which recognize the relationships between exogenous and endogenous latent variables, are freely estimated across groups, whereas the invariant parameters of the measures discovered in the measurement invariance test (to be equal across groups) are fixed. In the constrained model, on the other hand, the structural path parameters are specified to be equal.

Partial Invariance Test

Although a full invariance model is desirable, such a strict invariant model is regarded as practically impossible and scientifically unrealistic (Steenkamp & Baumgartner, 1998). But the failure of full invariance does not prevent some reasonable interpretation of the invariance tests. Thus, as a pragmatic compromise, researchers recommend partial invariance, in which a subset of parameters is constrained to be equal, whereas the rest of the parameters are allowed to vary, achieving a satisfactory fit (Byrne, Shavelson, & Muthén, 1989). A partial invariance test is useful for identifying the sources of inequivalence. In a sequential test, one parameter at a time is specified to be invariant across groups, and a chi-square fit difference test with the unconstrained model is then performed. The series of tests will identify cross-group invariant parameters, upon which a partial invariant model can be developed. There is no agreement, however, on the proportion of the parameters in the model that should be invariant so that the entire model is considered reasonably invariant, but, in general, most of the parameters should be invariant. Under the concept of partial invariance, constructs and structural paths can be compared and interpreted meaningfully across groups even if some parameters are not invariant.

AN EMPIRICAL ILLUSTRATION: CONSUMER ETHNOCENTRISM

Obviously, the objective of this empirical study was to illustrate the testing of measurement as well as structural invariance in a real-life context centered on consumer ethnocentrism toward Japanese products. The study involved four gender and age groups of U.S. consumers: female adults, male adults, female students, and male students. The focal constructs included consumer ethnocentrism, perceived quality, purchase intention, and actual ownership of Japanese products. The relationships among the constructs are well established in the literature. Specifically, consumer ethnocentrism negatively influences perceived quality judgments and purchase intentions regarding imported products, and quality judgments positively affect the purchase intention, which positively influences the actual ownership of foreign products (e.g., Shimp & Sharma, 1987; Yoo & Donthu, 2001).

Although consumer ethnocentrism studies recognize gender and age as important variables, researchers have investigated consumer ethnocentrism and their consequent constructs without gender and age invariance tests. This tradition must be based on the assumption of measurement and structural invariance as well as scalar invariance across gender and age. In this study, this assumption is tested. Accordingly, the relevant null hypotheses are:

- H₀:** (a) The measurement model of consumer ethnocentrism and its consequent constructs is invariant across gender and age (full measurement invariance).
- H₀:** (b) The structural model of consumer ethnocentrism and its consequent constructs is invariant across gender and age (full structural invariance).
- H₀:** (c) The factorial loadings and intercepts of the variables measuring consumer ethnocentrism and its consequent constructs are invariant across gender and age (full scalar invariance).

Sample

Consumers from the Midwest region of the United States participated in this study. A similar number of completed responses was collected for each gender and age group: 90 female adults, 121 male adults, 96 female students, and 109 male students. College students were selected to represent younger people, and nonstudent adults were selected to represent older people. The four groups were significantly different, for example, in age (35.2 years for female adults, 33.3 for male adults, 21.6 for female students, and 23.4 for male students; F value = 64.21, $p < .0001$) and personal income (\$30,830 for female adults, \$33,420 for male

adults, \$8,490 for female students, and \$12,500 for male students; F value = 83.38, $p < .0001$).

Measures

The Shimp and Sharma (1987) 17-item consumer ethnocentric tendency scale, CETSCALE, was used to measure consumer ethnocentrism. The scale has been well validated in various sample types and countries. The Klein, Ettenson, and Morris (1998) scales were used to measure perceived quality and purchase intention of Japanese products, which were measured by six items each. For the above three measures, participants expressed their agreement with statements for the constructs with the use of 5-point Likert scales anchored with 5 = "strongly agree" and 1 = "strongly disagree." Similar to The Klein et al. (1998) work, ownership of Japanese products was assessed as the number of products the participant owned. The products asked about were televisions, cars, stereos, cameras, camcorders, and VCRs. The value of the ownership of Japanese products could range from 0 (i.e., owns no product) to 6 (i.e., owns all six products). Data showed significant ownership differences across gender and age. The F statistic of ANOVA was 6.59 ($p < .001$): female adults, male adults, female students, and male students owned 1.64, 2.09, 1.77, and 2.23 Japanese products on average, respectively.

Data Analyses

Confirmatory factor analysis (CFA), based on a partial disaggregation approach, was used on the items. The partial disaggregation approach compromises the most aggregate approach, in which all items are summed to form one composite for a construct, and the most disaggregate approach, in which each item is used as an independent indicator for a construct (see Sweeney, Soutar, & Johnson, 1999). The most aggregate approach lacks item distinction, so it loses information, whereas the most disaggregate approach is often unmanageable because of the great number of errors and parameters to estimate. In contrast, the partial disaggregation approach maintains the advantage of the multiple measure approach and solves the "small sample size" problem. In this study, the full disaggregation approach produces a low item to sample ratio ranging from 3.9 to 5.3 for the items of the three constructs (i.e., consumer ethnocentrism, and perceived quality and purchase intention of Japanese products). The full measurement model of each single construct, in which all individual items were included, was acceptable for every group and the reliability was high. This indicated that measurement error should not be a problem. Thus, the items for each construct were divided into three groups and summed to form three aggregate indicators for the construct. As a result, nine composite indicators were generated.

The fit of the measurement model based on the covariance matrix of the nine composite indicators was satisfactory for each of the four demographic groups. The chi-square fit index for the 24 degrees of freedom ranged from 22.20 to 36.17, resulting in low chi-square to degrees-of-freedom ratios between 0.93 and 1.51. Other fit indexes were also acceptable: CFI ranged from 0.92 to 0.96; CFI and TLI were higher than 0.97; RMSEA ranged from 0 to 0.075; and RSMR ranged from 0.018 to 0.066. In addition, the scale composite reliability and the average variance extracted for each construct were reasonable (Fornell & Larker, 1981). The composite reliability, that is, internal consistency reliability, ranged from 0.85 to 0.97. The average variance extracted for each construct ranged from 0.65 to 0.91, exceeding the acceptable level of 0.50. These satisfactory indices show that the measurement model has an excellent fit for each demographic group and that it can be used as the baseline model to investigate further invariance (Byrne et al., 1989).

Measurement Invariance across Gender and Age

The unconstrained model, in which the factor loadings of the three constructs were relaxed to vary across gender and age, yielded a reasonable fit to the data ($\chi^2_{df=69} = 114.43$, CFI = 0.99, TLI = 0.99, GFI = 0.95, SRMR = 0.025, and RMSEA = 0.043). This unconstrained model was first compared with the fully constrained model ($\chi^2_{df=132} = 222.59$), which expected the overall equality of covariance structure across groups including factor loadings, error variances, and factor variance-covariances. The chi-square difference between the two models was significant ($\Delta \chi^2 = 108.16$ with $\Delta df = 63$, $p < .001$). Thus, the assumption of full measurement invariance was rejected. This result led to sequentially testing the invariance of factor loadings, error variances, and factor variance-covariances to detect the sources of inequivalence. First, the invariance of factor loadings was supported ($\Delta \chi^2 = 13.22$ with $\Delta df = 18$, $p > .75$). Next, the invariance of factor loadings and error variances were tested but not supported ($\Delta \chi^2 = 63.90$ with $\Delta df = 45$, $p < .001$). Accordingly, the invariance of factor loadings as well as factor variance-covariances were tested but not supported ($\Delta \chi^2 = 57.57$ with $\Delta df = 36$, $p < .05$). In summary, there was evidence only for the existence of metric invariance. Thus, the research tradition of assuming full measurement invariance was partially supported.

Structural Invariance across Gender and Age

The cross-group tests of the structural model were designed to determine the extent to which a single set of causal parameter estimates could equivalently represent the four demographic groups. Throughout the tests, factor loadings were set to be invariant across gender and age because in the measurement invariance test they were found invariant

across groups. In the structural model, the fourth construct (i.e., ownership of Japanese products) was included. The invariance of four structural paths was examined. The four paths comprised two β paths connecting three endogenous factors (i.e., purchase intention to ownership, β_{12} , and perceived quality to purchase intention, β_{23}) and two γ paths connecting the exogenous factor to two endogenous factors (i.e., consumer ethnocentrism to perceived quality, γ_{31} , and consumer ethnocentrism to purchase intention, γ_{21}).

An unconstrained structural model, in which the four paths were set to vary across gender and age, was first estimated. The model yielded a satisfactory fit ($\chi^2_{df=146} = 168.58$). To test the invariance of all the paths simultaneously, the unconstrained structural model was compared with the constrained model in which all causal paths were fixed to be invariant across groups. The chi-square difference was insignificant ($\Delta\chi^2_{df=12} = 20.24, p > .05$), thus furnishing sufficient evidence for an invariant pattern of causal paths.

Following the preceding, another structural model, in which the factor loadings and structural path parameters were constrained to be invariant whereas other parameters were set to vary across groups, was tested. The overall model fit was satisfactory: $\chi^2_{df=158} = 188.82$, CFI = 0.99, TLI = 0.99, GFI = 0.93, SRMR = 0.063, and RMSEA = 0.044. The significance of an estimate is shown in the estimate's t value, which tests the null hypothesis of the parameter having a zero value. Both invariant and noninvariant estimates were highly significant: t values ranged from 2.39 to 41.51. All four causal paths were significantly equivalent across groups and consistent with expectations. Specifically, β_{12} (0.39, $t = 7.81$) and β_{23} (0.15, $t = 4.11$) were positive while γ_{31} (-0.23, $t = -4.44$) and γ_{21} (-0.74, $t = -17.75$) were negative. In summary, full structural invariance, was supported, but its stability is suspicious because a partially invariant measurement model rather than the fully invariant model was used.

Full Scalar Invariance Test

Analysis of the latent mean structures requires scalar invariance, in which the factor loadings and intercepts of the manifest variables are invariant across groups (Meredith, 1993). As described before, the measurement model of three constructs had demonstrated metric invariance, that is, equivalence of factor loadings satisfying one of the two conditions of scalar invariance. To determine if the second condition could also be satisfied in the three-factor measurement model, the chi-square fit difference test was conducted between the unconstrained model in which the intercepts of the manifest indicators were set to vary across groups and the constrained model in which the intercepts were prescribed to be the same. The factor loadings were kept invariant across gender and age throughout the following tests. The chi-square

difference between the unconstrained and constrained models was not insignificant ($\Delta\chi^2 = 203.56$ with $\Delta df = 27$, $p < .001$), thus furnishing insufficient evidence for full scalar invariance, based on the fully constrained model of intercept parameters.

Partial Scalar Invariance Test

Consequent to the failure of establishing full scalar invariance, a series of invariance tests of the individual intercept parameters was performed to identify the sources of such failure. In each test, one specific intercept was increasingly constrained to be equivalent across groups and other remaining intercepts were relaxed to vary. Then, the partial invariance model was compared with the unconstrained model to test the equivalence of the focal intercepts. The first partial invariance model, in which the intercept of the first variable of purchase intention was held invariant, demonstrated that the intercept was consistent across groups, as indicated in the insignificant chi-square difference ($\Delta\chi^2_{df=3} = 5.6$, $p > .10$). Then, a new constrained model in which the invariant intercept and an additional intercept, that is, the intercept of the second purchase intention variable, were constrained to be the same across groups. This constrained model was compared with the unconstrained model, resulting in an insignificant difference ($\Delta\chi^2_{df=6} = 7.7$, $p > .25$). The series of tests for scalar invariance revealed that six out of the nine intercepts were invariant. Specifically, all three intercepts of purchase intention, the second and third intercepts of perceived quality, and the third of consumer ethnocentrism were identified as invariant across groups. Accordingly, the final model of partial scalar invariance was formed with these six invariant intercepts constrained to be equal and the three non-invariant intercepts relaxed to vary across gender and age. The overall fit of the model was satisfactory ($\chi^2_{df=132} = 169.65$, CFI = 0.99, TLI = 0.99, GFI = 0.93, SRMR = 0.059, and RMSEA = 0.053). The model did not differ from the unconstrained model significantly ($\Delta\chi^2_{df=18} = 26.23$, $p > 0.05$), thus indicating sufficient cross-group equivalence of the six intercepts. This is evidence that there exists a reasonable, but not full, level of scalar invariance.

Latent Mean Comparisons Across Gender and Age

The means of the three latent variables were then compared across gender and age based on the partial scalar invariance model. Because the latent means have no origin, absolute means cannot be estimated. The adult male group was selected as the reference group, and every latent mean parameter of the group was fixed to be zero, whereas the latent means of other groups were freely computed in relative terms to the zero means of the adult male group. Statistical significance was based on t values.

The overall invariance of latent means was first tested to evaluate if all the latent means were equal across gender and age. The hypothesis of full latent mean invariance was rejected due to the significant chi-square fit difference ($\Delta\chi^2_{df=6} = 34.42, p < .0001$) between the unconstrained model with varying latent means and the constrained model with equal latent means. This failure of latent mean invariance allowed mean comparisons for individual latent constructs, which were based on their difference from zero of the reference group, that is, the adult male group. The latent means of the adult female group were 0.12 ($t = 1.06$) for purchase intention of Japanese products, -0.13 ($t = -1.82, p < .05$) for perceived quality, and -0.11 ($t = -1.00$) for consumer ethnocentrism. The means of the female student group were 0.25 ($t = 2.36, p < .01$) for purchase intention, -0.07 ($t = -1.12$) for perceived quality, and -0.16 ($t = -1.58$) for consumer ethnocentrism, and those of the male student group were 0.15 ($t = 1.39$) for purchase intention, -0.10 ($t = 1.45$) for perceived quality, and -0.18 ($t = -1.51$) for consumer ethnocentrism. These results indicate that all of the groups were similar in consumer ethnocentrism. Female students had a significantly higher score on purchase intention of Japanese products than male adults, and female adults had a significantly lower score on perceived quality of Japanese products than male adults.

Detailed speculation on the reasons for these mean differences is beyond the scope of this article. One possible explanation is that young women are more materialistic than older men and generally want to possess more products than old men do (Belk, 1985). In addition, women, who are more cautious and risk averse in purchase decisions than men, tend to show lower product-quality evaluations than men because as the gap model of satisfaction predicts, higher expectations of product performance lead to lower satisfaction for the same level of product quality (Parasuraman, Zeithaml, & Berry, 1985).

DISCUSSION

To demonstrate the importance of factorial invariance check in cross-group comparisons, this study examined measurement and structural invariance and latent mean structures of American consumer ethnocentrism and relevant constructs among four demographic groups classified by gender and age. The measurement model achieved a satisfactory fit for each group. But the model demonstrated invariance of factor loadings only (i.e., metric invariance), detecting significant gender and age differences in error variances and factor variance-covariances. More seriously, the invariance did not measure up to scalar invariance (i.e., equivalent factor loadings and intercepts across groups), which is the minimal condition under which latent mean scores can be meaningfully compared across groups. Establishing scalar invariance is important

when studying differences among multiple sociodemographic or multiple cross-national groups. The failure of scalar invariance implies the measurement's fundamental assumptions might have been violated in comparative studies. This failure casts serious doubts on the findings of previous studies that compared mean scores of composite variables for consumer ethnocentric tendencies and relevant constructs across gender, age, or nations.

Generally speaking, the failure of any level of factorial invariance might occur because of instrument, population, or both. In particular, in cross-national or cross-cultural research, both instrument and population differences can be equally important sources of the invariance failure. Complete instrument equivalence is never easy to accomplish because even the meanings of standard scales are often differently interpreted across nations (Riordan & Vandenberg, 1994). Also, a matched sample with identical characteristics and backgrounds, which is required to increase the internal validity of a theory by controlling extraneous variables, is hard to get across nations. Thus, identifying which source is more responsible for the failure of invariance is a challenging task in a cross-national study. But, in the current illustrative study, the instrument used across groups was identical and so population differences (within the United States) were the only plausible source of the invariance failure. This means that previous group comparisons that assumed equivalence of measurement across groups could be erroneous. Previous research examined cross-group mean differences when the groups might have interpreted the measured constructs in a somewhat different manner.

As discussed earlier, the concept of partial invariance, which is a compromise for the traditional everything-or-nothing decision rule of invariance, allowed cross-group comparisons of latent means despite incompletely supported scalar invariance. Likewise, structural invariance could be tested on the basis of the partially invariant measurement model in which invariant parameters were fixed to be equal across groups, and noninvariant parameters varied across groups. Specifically, all factor loadings were fixed to be equal because they were revealed to be invariant in the measurement invariance test. Meanwhile, error variances and factor variance-covariances were relaxed because they were not invariant across groups. As reported in detail in the previous section, structural invariance was present among the four American demographic groups, indicating that American consumers do not show any gender or age difference in the consequences of their consumer ethnocentric tendencies. This result implies that if an American consumer is ethnocentric, then this consumer, regardless of his or her age and gender, would show the same level of quality perception, purchase intention, and ownership of Japanese products. This may be strong evidence that any type of sample can be used for theory-testing research on causal relationships.

Although this study was limited to one nation's market (i.e., the United States) and one country of origin (i.e., Japan), the findings provided by the study may have important implications for both consumer ethnocentrism studies and other cross-group comparative studies. Most cross-group comparative studies directly compare differences in mean scores across age, gender, or nations without a measurement invariance test. Researchers' main interest remains at best on metric invariance, although it is not clear in which cases metric invariance is useful for comparative studies. Unfortunately, establishing metric invariance is not a sufficient condition to compare composite means across groups. Nevertheless, scalar invariance, or structural invariance of the focal variables, has rarely been tested despite their essential importance for comparative studies. As the illustrative study in this article found no full scalar invariance among American consumer groups, researchers need to be cautious when comparing means across groups or nations. They should establish full, or at least partial, scalar invariance before doing comparative studies of interest.

REFERENCES

- Belk, R. W. (1985). Materialism: Trait aspects of living in a material world. *Journal of Consumer Research*, 12, 265-280.
- Bollen, K. (1989). *Structural equations with latent variables*. New York: Wiley.
- Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, 105(3), 456-466.
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18, 39-50.
- Jöreskog, K. G., & Sörbom, D. (1993). LISREL 8.02. Chicago: Scientific Software International.
- Klein, J. G., Ettenson, R., & Morris, M. D. (1998). The animosity model of foreign product purchase: An empirical test in the People's Republic of China. *Journal of Marketing*, 62, 89-100.
- Meredith, W. (1993). Measurement invariance, factor analysis, and factorial invariance. *Psychometrika*, 58, 525-543.
- Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1985). A conceptual model of service quality and its implications for future research. *Journal of Marketing*, 49, 41-50.
- Riordan, C. M., & Vandenberg, R. J. (1994). A central question in cross-national research: Do employees of different cultures interpret work-related measures in an equivalent manner? *Journal of Management*, 20, 643-671.
- Shimp, T. A., & Sharma, S. (1987). Consumer ethnocentrism: Construction and validation of the CETSCALE. *Journal of Marketing Research*, 24, 280-289.
- Steenkamp, J. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, 25, 78-90.

Sweeney, J. C., Soutar, G. N., & Johnson, L. W. (1999). The role of perceived risk in the quality–value relationship: A study in a retail environment. *Journal of Retailing*, 75, 77–105.

Yoo, B., & Donthu, N. (2001). Cultural orientation and consumer ethnocentrism. In C. P. Rao (Ed.), *Marketing and multicultural diversity*. Westport, CT: Greenwood.

The author greatly appreciates Edward Rigdon, Les Hayduk, and Roger E. Millsap for their valuable comments on his methodological questions related to earlier drafts, and Rajan Natarajan for his insightful comments and suggestions.

Correspondence regarding this article should be sent to: Boonghee Yoo, Department of Marketing and Business Law, G. R. Herberger College of Business, St. Cloud State University, St. Cloud, MN 56301 (byoo@stcloudstate.edu).

Copyright of *Psychology & Marketing* is the property of John Wiley & Sons, Inc. / Business and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.