International Journal of
Organizational Innovation

# THE PRACTICAL APPROACH FOR TESTING MEASUREMENT INVARIANCE ACROSS GENDER ISSUE FOR THE EMPATHY ITEMS IN SERVQUAL SCALE

Dr. Ming Gao, Professor
School of Economics and Management
Fuzhou University, Fujian, China
gaoming65@163.com

Ling-Chuan Huang, Doctor Candidate
School of Economics and Management
Fuzhou University, Fujian, China
linjan709@gmail.com

Dr. Yu-Jia Hu, Assistant Professor
Department of Distribution Management
Shu Te University, Taiwan, R.O.C.
h090805@gmail.com

## Abstract

Measurement invariance issues have caused more and more attention as some studies suggested that measurement invariance did cause a number of significant differences for the contents on the questionnaires. As typical requirements for measurement invariance were difficult to reach, scholars suggested some practical approaches to study measurement invariance issues. This research was to utilize the practical statistic approach to examine the measurement invariance issues for the part of Empathy in SERVQUAL Scale for retail chain stores business in Taiwan across gender. The sample in this research was selected as customers from four retail chain stores in Taiwan, resulting in 200 individual surveys for analysis. The results indicated the Taiwan version of Empathy 3-item scale achieved invariant across gender groups. Finally, this research generated the proposals for retail chain stores business in Taiwan and recommended future scholar studies.

Key Words: Marketing, Statistics, Service Quality, SERVQUAL, Empathy, Measurement Invariance, Retail Chain Stores, Taiwan

## Research Background and Theoretical Foundation

Deming (1981) and Garvin (1987) claimed the service quality was the awareness of person satisfaction for counterpart the customers' demand. Garvin (1987) suggested the consumer's perception of service quality was recognized by biased judgment which was different to the perception of tangible product management. Since service quality has been described as the critical factor to sustain the proficiency for business development, the SERVQUAL Scales has become the most popular mechanism for evaluating the perception of service quality. The SERVQUAL Scales was formed by Parasuraman, Zeithaml, and Berry (1988). This instrument tries to discover the variation between anticipation and consciousness on a number of specific areas in service quality. The SERVQUAL Scales provided five dimensions to assessing the perception of service quality, such as tangible, responsiveness, reliability, empathy, and assurance. Bolton and Drew (1991) expanded a model to exploring the influences of a service changes on consumer attitudes about service quality. The results indicated service changes were indentified to strongly impact consumer evaluations if service quality through their effect on consumer awareness of current concert and disconfirmation.

Even though this questionnaire has applied generally to perform studies for service quality in industries or countries, very few studies have examined the gender issues of the measurement invariance for SERVQUAL Scale across gender. In addition, numerous researchers (Baird 1976, Blus-tain 2000, Canary and Hause 1993) have suggested theories for gender differences in numerous facets on psychology and manner, such as social talents, communications, work approaches and so on.

People also have paid much attention on the issue of increasing woman buying power for those years, and the studies for gender differences in consumer behavior have become important topic for current academy researcher and marketing administrators in business. Research by Meyers-Levy and Maheswarm (1991), and Mitchell and Walsh (2004) revealed gender differences will lead to different buyer behaviors. Mitchell and Walsh (2004) also claimed gender differences resulted in different decision making processes. These theories and studies suggested that there was the significant relationship between gender and service quality. Hu (2014) conducted a study to examine the measurement invariance across gender in the version of the part of Tangibles in SERV- QUAL Scale for retail chain stores business in Taiwan. The results indicated the Taiwan version of Tangible 5-item scale only achieved partial measurement invariance. However, in another study, Hu, Liu, Su and Huang (2016) conducted a research to examine the measurement invariance across gender in the version of the part of Responsiveness in SERVQUAL Scale. The results indicated the Responsiveness 5-item scale achieved strict measurement invariance across gender. These results explained the measurement invariance across gender for SERVQUAL Scale maybe differently depends on individual items. Current researches need to develop more studies to verify how measurement invari-

ance across gender issues effect the SERVQUAL Scale.

The proposed methods for assessing measurement invariance across groups differ from researchers to researchers (Vandenberg & Lance, 2000). Byrne (2010) suggested the tests for multi-group invariance were: (a) factor loadings, (b) factor covariances, and (c) structural regression paths. Jöreskog (1971) recommended the first step to measure the equality of covariance structure was the test of null hypothesis. The groups were considered to have equivalent covariance structures if $H_0$ cannot be rejected. Jöreskog (1971) also suggested the group data should perform the signal group tests first as group data were considered to have equal covariance indexes. Thus, conducting a configurable test was necessary to verify if males and females shared the similar indicators for model fit. If this occurred, configurable invariance would exist and follow up invariance analyses were unnecessary.

However, Marsh and Hocevar (1985), and Byrne, Shavelson, and Muthén (1989) argued the strict invariance requirement was typically hard to satisfy. If the testing results cannot meet the strict invariance requirements, a small portion of non-invariant items will not affect the comparisons for all groups in invariant measurement level, and the researchers were not necessary to conduct individual item test. Moreover, Cheung and Rensvold (2002) suggested other moderate statistic methods to test the between-group invariance of CFA models, such as the statistics of △CFI, △Gamma hat, and △McDonald's NCI criteria were utilized. Therefore, this research expanded previous research

and applied practical statistic approach to evaluate the measurement invariance across gender for the component of Empathy in SERVQUAL Scale in chain restaurant business of Taiwan.

Research Purposes and Hypotheses

Following the theory concept as stated, the purposes and the significance for this research were: (a) to examine the measurement invariance for the part of Empathy in SERVQUAL Scale across gender in retail chain stores business of Taiwan, (b) to generate the proposals for managerial applications of retail chain stores business, and (c) to suggest areas for future scholarly inquiry. Following this concept, the researcher proposed one hypothesis as follows.

**Hypothesis :** Assuming all three empathy items are statistic invariant across gender.

Methodology

Following Five Likert Scale: strongly disagree, disagree, neutral, agree, strongly agree, the three items of Empathy in this research were little modified from the SERVQUAL Scales (Parasuraman et al, 1988). The three items were: (a). Service person won't ignore my request even though in busy status. b). Service person put the interests of customers first., c). Service person understands the needs of customers. Previous studies by Hu (2011, 2012, and 2013) have demonstrated the validity and reliability for this scale were reasonable and acceptable. The customers from four retail chain stores in the Kaohsiung city of south Taiwan attended this research. The pretest was conducted with Item Analysis in 50

samples. The researcher applied the method of random sampling. Each store randomly invited volunteer customers who shopping in stores to participate the questionnaire survey. The total number of valid responses was 200 (not including 15 invalid response), providing an adjusted response rate of 93%. This research applied the moderate and practical statistic method suggested by Cheung and Rensvold (2002), and test the between-group invariance of CFA models, such as the statistics values of △CFI and △Gamma hat to examine the measurement invariance issues for the three items of Empathy from the SERVQUAL Scales across gender.

Analysis of Results

This research applied the Structure Equating Modeling (SEM) by Analysis of Moment Structure (AMOS) software to test the model structure and hypothesis. The factor analysis for three empathy items for male, female and all groups were showed as Figure One, Figure Two and Figure Three. The results of multi- sample analysis for the unconstrained and the three constrained models were listed as Table one. The unconstrained model showed an acceptable baseline model for both males and females.

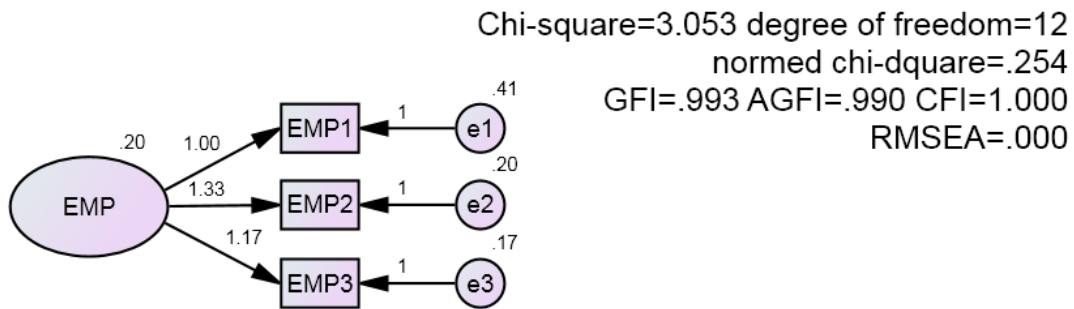Figure 1. Factor Analysis for Empathy Items for Male Group



Chi-square=3.053 degree of freedom=12
normed chi-dquare=.254
GFI=.993 AGFI=.990 CFI=1.000
RMSEA=.000

Figure 2. Factor Analysis for Empathy Items for Female Group



Chi-square=3.053 degree of freedom=12
normed chi-dquare=.254
GFI=.993 AGFI=.990 CFI=1.000
RMSEA=.000
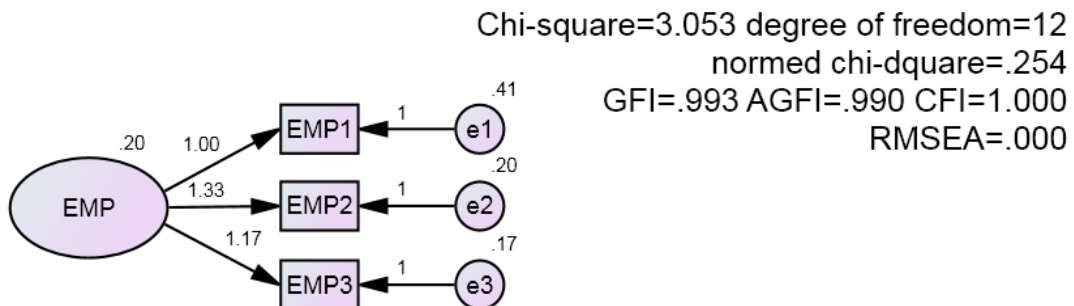
Figure 3. Factor Analysis for Empathy Items for Group



Chi-square=3.053  degree of freedom=12
normed chi-dquare=.254
GFI=.993  AGFI=.990  CFI=1.000
RMSEA=.000
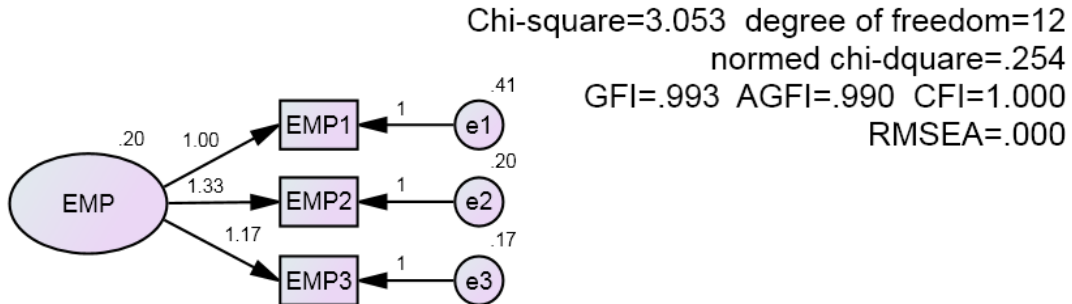
Table 1. Model Fit Indexes for Unconstrained and Constrained Model

| Model | χ2 | DF | P | χ2/DF | GFI | AGFI |
|---|---|---|---|---|---|---|
| Unconstrained | .000 | 0 | | | 1.000 | |
| Measurement weights | .030 | 4 | 1.000 | .007 | 1.000 | 1.000 |
| Structural covariances | .912 | 6 | .989 | .152 | .998 | .994 |
| Measurement residuals | 3.053 | 12 | .995 | .254 | .993 | .990 |

Table 2. Baseline Comparisons for CFI Test

| Model | NFI Delta1 | RFI rho1 | IFI Delta2 | TLI rho2 | CFI |
|---|---|---|---|---|---|
| Unconstrained | 1.000 | | 1.000 | | 1.000 |
| Measurement weights | 1.000 | 1.000 | 1.018 | 1.041 | 1.000 |
| Structural co variances | .996 | .994 | 1.023 | 1.035 | 1.000 |
| Measurement residuals | .987 | .990 | 1.041 | 1.031 | 1.000 |
| Saturated model | 1.000 | | 1.000 | | 1.000 |
| Independence model | .000 | .000 | .000 | .000 | .000 |

Table 3. Nested Model Comparisons for TLI Tests

| Model | DF | CMIN | P | NFI Delta-1 | IFI Delta-2 | RFI rho-1 | TLI rho2 |
|---|---|---|---|---|---|---|---|
| Measurement weights | 4 | 0.030 | 1.000 | .000 | .000 | | |
| Structural co variances | 2 | .882 | .643 | .004 | .004 | .006 | .006 |
| Measurement residuals | 6 | 2.140 | .906 | .009 | .010 | .004 | .004 |

The multi-sample analysis also showed the indexes of model fit for three constrained models (measure weights, structural co variances, and measurement residuals) across gender, and these indexes indicated the three constrained models were accepted. The researcher listed the testing values for CFI in Table 2. The results in Table 2 indicated all △CFI values were less than .01, and all △TLI values in Table 3 were less than .01(P>.05). Both △CFI and △TLI values indicted the items achieve practical statistic invariant across the test groups.

Conclusions and Suggestions

After examining the significant statistic values including the values for CFI and TLI, all △CFI values were less than .01, and all △TLI values were less than .01(P>.05). The results supported that the hypothesis should not be rejected. The results revealed this Taiwan version of the Empathy 3-item scales from the SERVQUAL Scales achieved measurement invariance across the gender. The results were consistent with the study by Huang, Huang, Hung, and Hu (2017) that indicated the Empathy 3- item scale achieved strict measurement invariance across gender. However, the results did not consist with the views of Meyers-Levy and Maheswarm (1991), and Mitchell and Walsh (2004) suggested gender differences will lead to dissimilar buyer actions. Although this model was fully measurement invariance, Milfont and Fischer (2010) claimed full measurement invariance is questionable to hold in reality.

Based on the research by Hu (2014), the part of Tangibles in SERVQUAL Questionnaire was partial measurement invariance across the gender under strict testing approach. Actually, there are many different approaches for testing measurement invariance issues. There are no agreements on the methods or sequences for researchers. Those studies revealed the researcher should keep sensible manner to conduct research for multiple items in questionnaire for the issues of measurement invariance across group, based on the suitable tactic. And this study suggested the part of Empathy in SERVQUAL Scale was measurement invariance across genders. Scholars may apply these items for conducting studies for service industry, specially exploring the service quality issues for the retail chain store business in Taiwan.

References

Baird, J. E., Jr. (1976). Sex differences in group communication: A review of relevant research. *Quarterly Journal of Speech, 62*, 179-192.

Blustain, S. (2000). The new gender wars. *Psychology Today, 33*, 1-5.

Bolton, R. N., & Drew, J. H. (1991). A longitudinal analysis of the impact of service changes on consumer attitudes. *Journal of Marketing, 55*(Jan), 1-9.

Byrne, B. M. (2010). *Structural Equation Modeling With AMOS: Basic Concepts, Applications, and Programming*, 2 Ed, Routledge, Taylor & Francis Group, NJ. 197-200.

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*, 456–466.

Canary, J. D., & Hause, S. K. (1993). Is there any reason to research sex differences in communication? *Communication Quarterly, 41*(2), 129-144.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*, 233-255.

Deming, W. E. (1981). Improvement of quality and productivity through action by management. *National Productivity Review, 11*, 12-22.

Garvin, D. A. (1987). Competing on the eight dimensions of quality. *Harvard Business Review, 65*(Nov-Dec), 101-109.

Hu, Y. J., Liu, F. M., Su, L. C., & Huang, J. H. (2016). Examining the measurement invariance across gender in the version of the part of responsiveness in SERVQUAL scale in retail chain stores business of Taiwan. *International Journal of Organizational Innovation, 9*(1), 1-7.

Hu, Y. J. (2014). Assessing measurement invariance across gender in the version of the part of tangibles in Servqual scale for retail chain stores business in Taiwan. *International Journal of Organizational Innovation, 7*(1), 181-187.

Hu, Y. J. (2013). Using structural equation modeling to evaluate a consumer behavior model: the case of retail chain stores in Taiwan. *International Journal of Organizational Innovation, 6*(1), 121-127.

Hu, Y. J. (2012). The moderating effect of brand equity and the mediating effect of marketing mix strategy on the relationship between service quality and customer loyalty: The case of retail chain stores in

Taiwan. *International Journal of Organizational Innovation, 5*(1), 155-162.

Hu, Y. J. (2011). How Brand Equity, Marketing Mix Strategy and Service Quality Affect Customer Loyalty: The Case of Retail Chain Stores in Taiwan. *International Journal of Organizational Innovation, 4*(1), 59-73.

Huang, L. C., Huang, Y. C., Hung, C. Y., & Hu, Y. J (2017). The study of the empathy items in SERVQUAL scale for measurement invariance across gender, *10*(1), 238-245.

Jöreskog, K.G. (1971). Simultaneous factor analysis in several populations. *Psychometrika, 36*, 409–426.

Marsh, H. W., & Hocevar, D. (1985). Application of confirmatory factor analysis to the study of self-concept: First - and higher order factor models and their invariance across groups. *Psychological Bulletin, 97*, 562–582.

Meyers-Levy, J., & Maheswaran, D. (1991). Exploring Differences in Males' and Females' Processing Strategies. *Journal of Consumer Research, 18*(1), 63-70.

Milfont, T. L., & Fischer, R. (2010). Testing measurement invariance across groups: Applications in cross-cultural research. *International Journal of Psychological, 3*(1), 110-121.

Mitchell, V. W., & Walsh, G. (2004). Gender differences in German consumer decision making Styles. *Journal of Consumer Behavior, 3*(4), 331-346.

Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1988). Communication and control processes in the delivery of service process. *Journal of Marketing, 52*(1), 36-58.

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.