

Measurement Equivalency of Web Service Quality Instruments: A Test on Chinese and African American Consumers

Juan (Gloria)

Abstract

In this study, we reviewed the quality of customer service (E-S-QUAL) model and the quality of service recovery (E-RecS-QUAL) model and applied them to the African American and Chinese cultural settings. Utilizing LISREL 8.72 program, we conducted measurement equivalence test and found that the e-service quality measurement can be generalized to different cultures, and the equivalence of measurement between two cultures considered here is partially supported. In addition, t-tests were conducted on each dimension of web service quality in order to compare the mean differences between the perceptions of Chinese and African American consumers. It was found that Chinese consumers rate web service quality at a significantly lower level than African American consumers on all seven dimensions. Methodological and practical implications of the study are discussed.

Key words/phrases: web-service quality, African American consumers, Chinese consumers, measurement equivalence, cross-cultural study.

Introduction

In the context of cross-cultural research, various cultural differences in consumer behavior have been revealed (e.g., Aaker and Williams 1998; Klein, Ettenson, and Morris 1998), raising the larger issue of the validity of applying marketing-related inferences developed in the United States to marketing efforts in other cultures, especially non-Western cultures. Since the US-based multinationals, such as Coca-Cola, McDonald's, IBM, and others, derive a significant portion of their revenues from international operations, the development of effective marketing strategies that are sensitive to cross-national cultural differences would seem to be of considerable importance for success in the global marketplace (Gurhan-Canli and Maheswaran 2000).

A primary concern, and what arguably should be the first step in such cross-cultural research, is to confirm measurement invariance: “whether or not, under different conditions of observing and studying phenomena, measurement operations yield measures of the same attribute” (Horn and McArdle 1992, p. 117). More specifically, in cross-cultural studies, it is essential to test whether the instruments used to measure the relevant constructs in one culture can also be applied to other cultures before any cultural comparisons are conducted. Otherwise, one cannot be certain whether differences in consumer behaviors reflect true differences between countries on the underlying construct or arise out of measurement biases (Steenkamp and Baumgartner 1998).

However, although such calls for measurement invariance testing in cross-cultural studies have been heard for almost two decades, it is still very common to see many cross-cultural studies take it for granted that the instruments developed in the U.S. are universal and draw conclusions on consumers from different cultures based on the same instruments without conducting cross-cultural measurement invariance testing (Steenkamp and Baumgartner 1998). The lack of robust measurement frameworks across cultures has severely limited and even biased the development of theory-based empirical work (Maheswaran and Shavitt 2000).

The purpose of this study is to examine the measurement generalizability and invariance of the quality of customer service (E-S-QUAL) model and the quality of service recovery (E-RecS-QUAL) model between Chinese and African American consumers. This study intends to build a foundation for cross-cultural studies on web service quality and more generally, to promote the cause of measurement invariance testing in cross-cultural studies.

Web Service Quality Models

Recognizing the need for a comprehensive conceptualization of online shopping service quality, frameworks have been proposed and scales developed in recent years (Wang and Tang 2003; Parasuraman et al 2005). Of these, the two models described below have been extensively tested in the U.S.

E-S-QUAL model: Based on the conceptualization of Zeithaml et al (2002), a 22-item E-S-QUAL scale was developed by Parasuraman et al (2005). This scale consists of four dimensions: efficiency, fulfillment, system availability, and privacy which are considered to be critical to e-service quality.

E-RecS-QUAL model: The importance of service recovery efforts for addressing service failures which are more likely to occur in case of e-retailing has led Parasuraman et al (2005) to develop an accompanying model to E-S-QUAL. The E-RecS-QUAL model consists of three dimensions – responsiveness, compensation, and contact which are measured through an 11-item scale. These two scales were tested and validated in the U.S. through a large-scale study undertaken by Parasuraman et al (2005). To our knowledge, no other published study has employed these models in research focusing on web service quality.

The Rationale for the Selection of Two Cultures

There are several reasons for selecting these two cultures for the present study. First, insight into Asian consumers is critically important to the global marketplace as Asian consumers move from traditional economies to contemporary mainstream consumerism (Farrell et al., 2006; Ho et al., 1997; McKinsey Group, 2006). Many recent practitioner-based articles discuss the seemingly endless business opportunities in the Chinese marketplace (cf Garner, 2005) or have conducted marketplace research on Chinese consumers (cf Boston Consulting

Group, 2007). However, research at the nexus of culture and web service quality perception and research that is theoretically grounded in traditional consumer behavior literature, particularly in Asian contexts, has been sparse.

In addition, we choose to study African American consumers as well in order to establish the generalizability of the web service quality scales introduced by Parasuraman et al (2005). Prior literature has examined the “digital divide” that exists between the African American and majority (Caucasians) populations in the U.S. Empirical evidence establishing a lower internet usage rate among African Americans (Hoffman et al 2004; Jackson et al 2004; Smith 2005) has prompted scholars to propose a relationship between race and attitude toward internet usage (Porter and Donthu 2006). The present study intends to validate the E-S-QUAL and the E-RecS-QUAL scales by testing them among the African American consumers. Evaluation of the E-S-QUAL and the E-RecS-QUAL models in the vastly different Chinese and African American cultural settings, will severely test their generalizability. Likewise, by comparing the model fit, factor loadings, correlations, and residuals, any perceptual differences between Chinese and African American cultures will be revealed. These findings will be valuable to the consumer behavior literature.

Objectives and Hypotheses

The objectives of this study are to test the external validity (generalizability) of E-S-QUAL model and the E-RecS-QUAL model in different cultural settings and to gain insights about how web service quality factors differ across cultures.

In addition, at the end of our work, our goal is to identify a single instrument that measures these seven constructs across multiple markets and obviates the need for an assessment

of the measurement model on a culture-by-culture basis. Toward this end, the following five hypotheses relating to measurement invariance were tested via structural equation modeling. The hypotheses progress from least restrictive to most restrictive, in order to test the applicability of the web service quality scales in the Chinese and African American cultural contexts. The sequential ordering of such measurement invariance tests was suggested by Joreskog and Dorbom (1999).

Finally, after the hypotheses regarding the measurement model are tested, we also compare the mean differences between the two samples on the seven web service quality constructs. As this study is the first one to our knowledge to use a stable measurement scale to examine the web service quality perceptions of Chinese and African American consumers simultaneously, no specific hypotheses were advanced regarding the directionality of mean differences.

Hypothesis 1: The constructs of E-S-QUAL model and the E-RecS-QUAL model exist cross-culturally: the same number of web service quality constructs emerge for consumers in both cultures, and the scale variables load on the same constructs for both cultures.

Hypothesis 2: The loadings between the indicators and the latent constructs are the same for consumers across both the cultures (LX=IN).

Hypothesis 3: The variances of latent variables are equal between the two cultures (PH=IN).

Hypothesis 4: The correlations between latent variables are equivalent in different cultural groups (PH=IN).

Hypothesis 5: The correlations of error between any two indicators in the web service quality perception model are invariant across cultures (TD=IN).

Methodology

Measurement of the Constructs in the Model

The E-S-QUAL model and the E-RecS-QUAL model items (Parasuraman et al 2005) corresponding to the seven web service quality perception constructs in the model are utilized without revision. More specifically, 22-item E-S-QUAL scale consists of four dimensions: efficiency, fulfillment, system availability, and privacy. In addition, an 11-item scale was developed to measure the quality of service recovery offered by the website, which includes responsiveness, compensation, and contact. Respondents indicated their level of agreement with each of the original 33 items on a 5-point, Likert-type scale, with 1 indicating “strongly disagree” and 5 indicating “strongly agree.” [Subjects were asked to recall their most recent online shopping experience and rate the corresponding website service quality on the 33 items originally developed in the models.](#)

Questionnaire Translation and Validation

The E-S-QUAL model and the E-RecS-QUAL model items were initially translated into Chinese by one of authors, then back-translated into English by a native Chinese speaker fluent in English. In order to avoid bias, the original English version of the questionnaire was not shown to the back translator. These two English version questionnaires—the original version and the version back-translated from the initial Chinese version—were compared by a native English speaker. The Chinese version was modified slightly based on this comparison and sent to a small sample of Chinese consumers via e-mail. These consumers suggested some improvements in both word usage and sentence structure. Thus, the final Chinese version of the questionnaire went through several iterations with a variety of native speakers for final refinement.

Data Collection

[Convenient sampling method was used in both cultures. More specifically, in the U.S.,](#) English-version questionnaires were distributed at an East Coast University, to African American

students enrolled in business courses. These students as well as their family members and friends constituted the sample. Data collection in China occurred in Beijing, the capital city of China. Questionnaires were distributed to students enrolled in a comprehensive four-year University with an emphasis on economics and business. After questionnaires were collected in China, they were airmailed to the U.S.

Results

Sample Characteristics

In the United States, a total of 228 valid questionnaires were collected. Over half of the respondents (55.9%) were full-time students. The sample is slightly skewed in terms of gender, with over 60% being female. A significant majority of the respondents (78.1%) are under the age of 30. In the other cultural context, data were collected from a convenience sample of 237 Chinese students. The sample is comparable to the U.S. sample in terms of gender (60.3% female), but younger, with an average age of 20.2 years. Among 237 valid questionnaires collected, only 147 (55.47%) cases were utilized in the following analyses, since the others lacked internet shopping experience.

Measurement Model Analyses

Using LISREL 8.72, model fit was assessed through Chi-square tests, error levels (RMSEA and RMR), and multiple fit indices such as IFI, CFI, NNFI and GFI. All hypotheses were tested using the structural equation modeling approach.

Web Service Quality Measurement Model Fit Testing. The original E-S-QUAL model and the E-RecS-QUAL model scales (comprising 33 items together) were tested using the data obtained from the Chinese and African American samples. The LISREL results for the two samples are summarized in Table 1. All of the indices suggest a good model fit. For the African

American sample, although the Chi-square statistic (992.89) was significant at 474 degree of freedom, the ratio of chi-square to degrees of freedom (2.09) was below the 3.0 threshold value. RMSEA and RMR were as low as .076 and .062, respectively, and all of the fit indices were around .95 (CFI, IFI, and NNFI) (Figure 1).

Insert Table 1 and Figure 1 about Here

Likewise, LISREL results in Table 1 also showed good model fit for the Chinese cultures, suggesting that the web service quality model is valid in this culture as well. All of the path loadings were significant and the majority of them were above .50 (Figure 2). Instrument reliabilities (Cronbach's alpha) for each factor in each culture were calculated and reported in Table 3. All of them were in the acceptable range from .69 to .94. Therefore, H1 was supported.

Insert Figures 2 and Table 2 about Here

Model Invariance Testing. After establishing the distinctive presence of the seven web service quality constructs in our African American and Chinese samples, we proceeded to sequentially test the progressively restrictive assumptions to examine whether the relationships among these constructs were constant across consumers in different cultures.

First, a two group comparison test was conducted without constraining any paths (Table 3, model 1). Although the Chi-Square (2992.21) was significant at 948 degree of freedom, the fit indices, such as IFI, NNFI, and CFI, were around .91, and the RMSEA and RMR were as low as .090 and .061. Therefore, this two-group model was accepted as a baseline model in this study.

The following model tests with all kinds of restrictions will be compared with this baseline model. In addition, the good fit of the baseline model also supported H1.

Insert Table 3 about Here

Next step was to constrain the LX matrix (Model 2), thereby restricting the factor loadings of the measurement items to be equal between the two cultures. Model 2 exhibits a good fit to data. In fact, the ratio of the Chi-Square and degree of freedom, the fit indices, and the error (RMSEA) were almost identical to the results from model 1. Moreover, the Chi-Square which changed by 51.29 from model 1 to model 2 was not significant at 33 degrees of freedom change, which means that model 2 was not significantly worse than model 1. Therefore, model 2 was accepted, and H2 was fully supported.

As for H3 and H4, in addition to LX path constrains, the PH matrix (correlations among factors and factor variance) was set as invariant between the samples from the two cultures (model 3). Although the goodness of fit indices were acceptable, the Chi-Square changed by 92.02 from model 2 to model 3, which was significant ($p < .01$) at 21 degrees of freedom change. Model 3 was rejected since the fit was significantly worse than in the case of model 2.

Then we tested model 4, by releasing some of the PH path constraints according to modification indices of model 3. As seen from the results presented in Table 4, model 4 showed a good fit to data in terms of the ratio of Chi-Square to the degrees of freedom, fit indices, and RMSEA. Moreover, the Chi-Square changed by 33.00 from model 2 to model 4, which was not significant at 17 degrees of freedom change. Model 4 was accepted since it was not significantly worse than model 2. From the results of the tests of models 3 and 4, H3 and H4 received partial support.

Finally, H5 was tested by constraining TD paths (correlations among the errors of indicators) into invariance between two cultures (model 5). Model 5 did not fit well according to fit indices, RMSEA, and RMR, and it was significantly worse ($p < .01$) than model 4 (Chi-Square change=259.59, df change = 33). Therefore, model 5 was rejected, and H5 was not supported.

In summary, we can conclude that the measurement of E-S-QUAL model and E-RecS-QUAL model is valid and consistent between different cultures - at least in the two cultures considered in this study. Although the error correlation (H5) was not invariant in our test, according to Steenkamp and Baumgartner (1998), such degree of variation is acceptable in a cross-cultural study, especially for tests involving comparison of means. As a general conclusion, we can utilize this measurement instrument to analyze web service quality perceptions in other cultures and to compare web service quality perceptions among different cultures. Therefore, mean comparison tests on web service quality constructs were conducted between African American and Chinese samples in the following section.

Mean Differences in Constructs between the Two Cultures

To examine mean differences in the web service quality perception subscales across cultures, we first averaged the items comprising each construct for each respondent. Then t-tests were conducted on the seven subscale values between the two cultural groups. Significant differences were found between two cultures on all seven constructs of web service quality, at the customary .05 level. Further examinations of mean differences via planned contrasts showed that Chinese consumers had significantly lower perceptions on all seven dimensions of the web service quality than the African American consumers (Table 4 and Figure 3). Constructs such as efficiency, privacy, and system exhibit larger differences than the other variables.

Insert Table 4 and Figure 3 about Here

Discussion

The E-S-QUAL model and E-RecS-QUAL model are recently developed models that were employed for measuring consumers' web service quality perceptions in the U.S. Our study extends them by validating the seven dimensions and their scales in other cultures. It was concluded that the E-S-QUAL model and E-RecS-QUAL model are robust across different cultures, and that the instruments can be used in other cultures.

More specifically, we showed that the same variables load on the same constructs in both the cultures tested (H1) and that the factor loadings can be constrained to be equal between these two cultures (H2). On the other hand, our study partially supported the hypotheses of the equivalence of factor variance and the correlations among factors (H3 and H4), but did not support the hypothesis of error correlations being equal (H5) between two cultures, which means some degree of variation exists in terms of factor variance, the correlations among factors, and error correlations. However, according to Steenkamp and Baumgartner (1998), such a degree of variation is acceptable in a cross-cultural study. Therefore, our work suggests that the *measurement* of the web service quality perception constructs can be generalized across these cultures.

Our research is arguably amongst the early work to generalize the E-S-QUAL model and E-RecS-QUAL model to different cultural settings, to test the invariance of measurement across different cultures, and to use a single web service quality perception measurement scale with good psychometric properties in comparisons between Chinese and African American

consumers. Looking at mean differences, we found that there were significant differences between these two cultures tested on all seven web service quality scales, and that the Chinese consumers consistently perceive web service quality to be lower than their African American counterparts across all seven dimensions. Plausible explanations include the nascence of the internet retailing activity in China and the relatively lower levels of service standards (CNNIC 2007).

This study makes important theoretical, methodological, and practical contributions. First, our efforts provide theoretical support to the general finding in the literature on web service quality, that consumers in different cultures do have significantly different web service quality perceptions. This finding makes another call for breaking the “American Box” and for extending the domestic web service quality research to global markets.

Methodologically, this study utilized students in other countries instead of sampling international students studying in the U.S., which avoids acculturation concerns. While students may not be representative of the entire populations, they are considered to be appropriate sample for a study such as this, for they are in the vanguard of the internet adopters in both cultures (Hoffman et al 2004; CNNIC 2007).

Our final contribution is practitioner-related: because service quality is a fundamental issue for every company, this research is valuable for practitioners who are involved in online international business or trying to expand their business to the cultures included here. Our work shows that these seven web service quality constructs exist across cultures, and that the instruments developed by Parasuraman et al (2005) can be used to measure web service quality in other cultures. Moreover, since significant differences were found across cultures on all seven

dimensions, assuming that “business is business everywhere” or, “price perception is the same everywhere” could potentially lead to disastrous consequences for an internet retailer.

In addition, it is also very important to the practitioners in the international business arena to understand that Chinese consumers have relatively low perceptions of the web service quality. There is an opportunity here for the experienced retailers such as Amazon, to elevate the quality of experience for Chinese consumers and to capture a lion’s share of the market at this early stage.

Limitations and Future Research

This study examined the generalizability of the E-S-QUAL and E-RecS-QUAL models across two cultures, using convenient samples of students. In future studies, these models should be tested in other important markets such as Europe and India, with distinctive cultures and established e-retailing merchants. Future studies could also examine the influence of cultural factors (such as individualism-collectivism and long term-short term orientation) on the service quality perceptions, satisfaction and loyalty of internet shoppers.

References

- Aaker, J. L. and Williams, P. (1998), “Empathy versus Pride: The Influence of Emotional Appeals Across Cultures,” *Journal of Consumer Research*, 25 (December), 241-61.
- Boston Consulting Group (2007), “Winning the hearts and minds of China’s consumers”, accessed at http://www.bcg.com/publications/files/Winning_Chinas_Consumers_Sept_2007.pdf, [23 October 2007].
- CNNIC (2007), “Statistical Survey Report on the Internet Development in China”, (Accessed online).
- Farrell, D., Gersch, U. A., and Stephenson, E. (2006), “The Value of China's Emerging Middle Class,” *McKinsey Quarterly*, Special Edition, 60-69.

- Garner, J. (2005), *The Rise of the Chinese Consumer*, John Wiley & Sons Ltd.
- Gürhan-Canli, Z. and Maheswaran, D. (2000), "Cultural Variations in Country of Origin Effects," *Journal of Marketing Research*, 37 (3), 309-317.
- Ho, F. N., Ong, B.S., and Lee, S. (1997), "A Multicultural Comparison of Shopping Patterns among Asian Consumers," *Journal of Marketing Theory & Practice*, 5 (Winter), 42-51.
- Hoffman, D. L., Novak, T. P., and Alladi Venkatesh (2004), "Has The Internet become Indispensable," *Communications of the ACM*, 47 (7), 37-42,
- Horn, J. L. and McArdle, J. J. (1992), "A Practical and Theoretical Guide to Measurement Invariance in Aging Research," *Experimental Aging Research*, 18 (Fall-Winter), 117-44.
- Jöreskog, K. G. and Sörbom, D. (1999), *LISREL 8 User's Reference Guide*. Lincolnwood, IL: Scientific Software International.
- Klein, J. G., Ettenson, R., and 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 (1), 89-100.
- Malhotra, N.K., Ulgado, F.M., Agarwal, J., Shainesh, G., and Wu. L. (2005), "Dimensions of Service Quality in Developed and Developing Economies: Multi-Country Cross-Cultural comparisons", *International Marketing Review*, 22 (3), 256-78.
- McKinsey Group (2006), "From 'Made in China' to 'Sold in China': The Rise of the Chinese Urban Consumer. http://www.mckinsey.com/mgi/publications/china_consumer/, 15 October 2007.
- Parasuraman, A., Zeithaml, V.A., and Malhotra, A. (2005), "E-S-QUAL: A Multiple-Item Scale for Assessing Electronic Service Quality", *Journal of Service Research*, 7 (3), 213-33.
- Porter, C. E. and Donthu, N. (2006), "Using the Technology Acceptance Model to Explain How Attitudes Determine Internet Usage: The Role of Perceived Access Barriers and Demographics," *Journal of Business Research*, 59 (9), 999-1007,
- Smith, D. T. (2005), "Developing Self-Employment Among African Americans: The Impact of Household Social Resources on African American Entrepreneurship," *Economic Development Quarterly*, 19 (4), 346-55.
- Steenkamp, J. E. M. and Baumgartner, H. (1998), "Assessing Measurement Invariance in Cross-National Consumer Research," *Journal of Consumer Research*, 25, 78-90.
- Wang, Y-S. and Tang, T-I (2003), "Assessing Customer Perceptions of Website Service Quality in Digital Marketing Environments", *Journal of End User Computing*, 15 (3), 14-31

Tables and Figures

Table 1
Fit Indices for Chinese and African American Consumers

	Chi-Square	DF	Ratio	Sig.	RMS EA	RM R	CFI	IFI	NN FI	H1 Decision
AA	992.89	474	2.09	.000	.076	.062	.95	.95	.94	Accept
China	1012.64	474	2.14	.000	.079	.078	.90	.90	.89	Accept

Table 2
Reliability (Cronbach's Alpha) of Web Service Quality Subscales by Culture

	Efficiency	System	Fulfill	Privacy	Response	Compensation	Contact
AA	.89	.88	.94	.90	.88	.73	.83
China	.77	.69	.80	.71	.80	.77	.80

Table 3
Summary of Invariance Tests

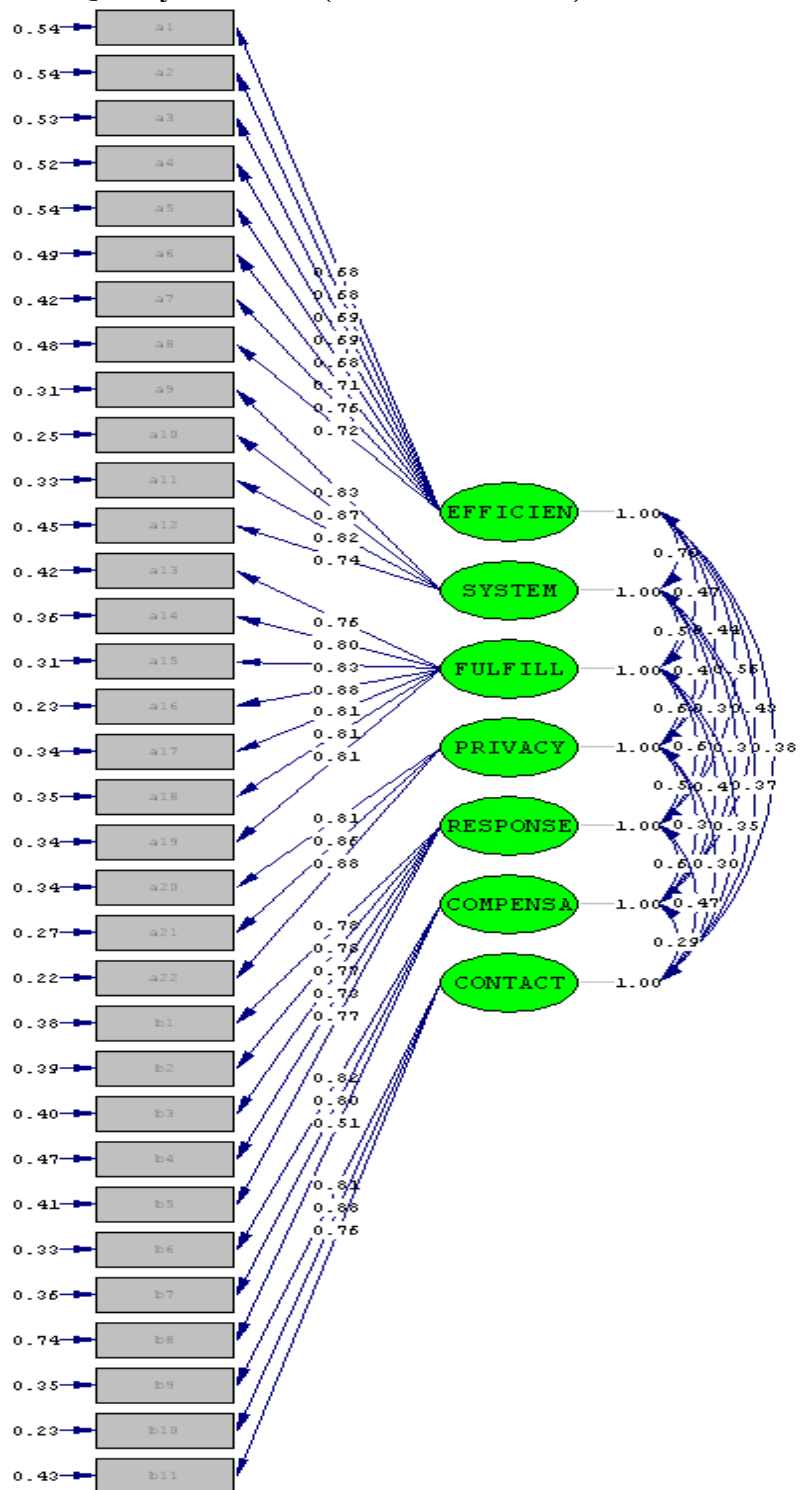
	H1	H2	H3/H4		H5
Model	1	2	3	4	5
Constrain	No	LX=IN	LX=IN PH=IN	LX=IN PH=IN (Partial)	LX=IN PH=IN (Partial) TD=IN
Chi-Square	2992.21	3043.50	3135.52	3076.50	3336.09
DF	948	981	1002	998	1031
Sig.	.000	.000	.000	.000	.000
RMSEA	.090	.090	.091	.090	.094
RMR	.061	.118	.109	.109	.107
CFI	.91	.90	.90	.90	.89
IFI	.91	.90	.90	.90	.89
NNFI	.90	.90	.90	.90	.89
Δ X2	NA	51.29	92.02	33.00	259.59
Δdf	NA	33	21	17	33
Sig.	NA	N.S.	P<.01	N.S.	P<.01
Decision	Y	Y	N	Y	N

Table 4
Web Service Quality Subscale Means by Culture

Web Service Quality	AA	China	Difference	p value
Efficiency	3.86	3.58	.28	.000**
System	3.88	3.59	.29	.000**
Fulfill	3.80	3.60	.20	.006**
Privacy	3.89	3.49	.40	.000**
Response	3.59	3.42	.17	.028*
Compensation	3.29	3.04	.25	.005**
Contact	3.62	3.37	.25	.014*

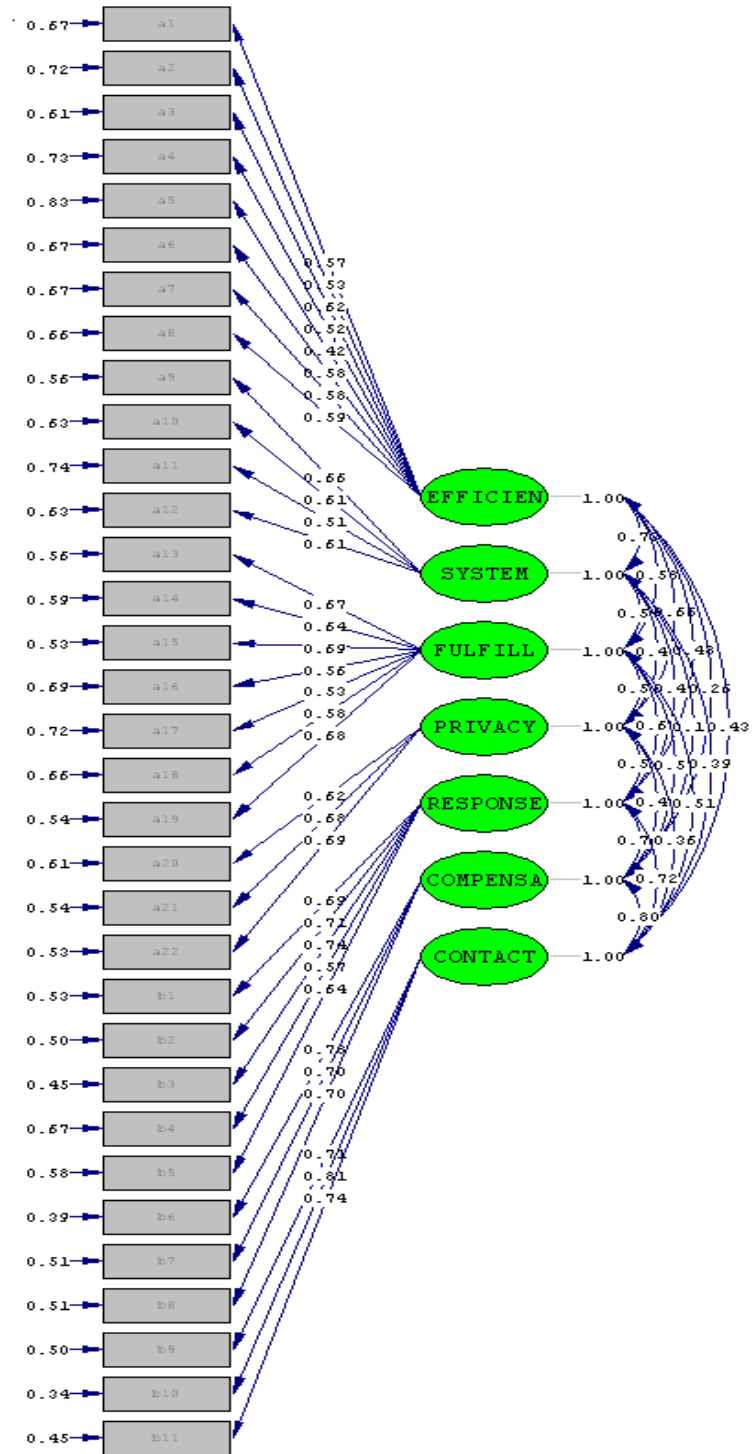
* Significant at .05 level ** Significant at .01 level

Figure 1
Web Service Quality Model Fit (African American)



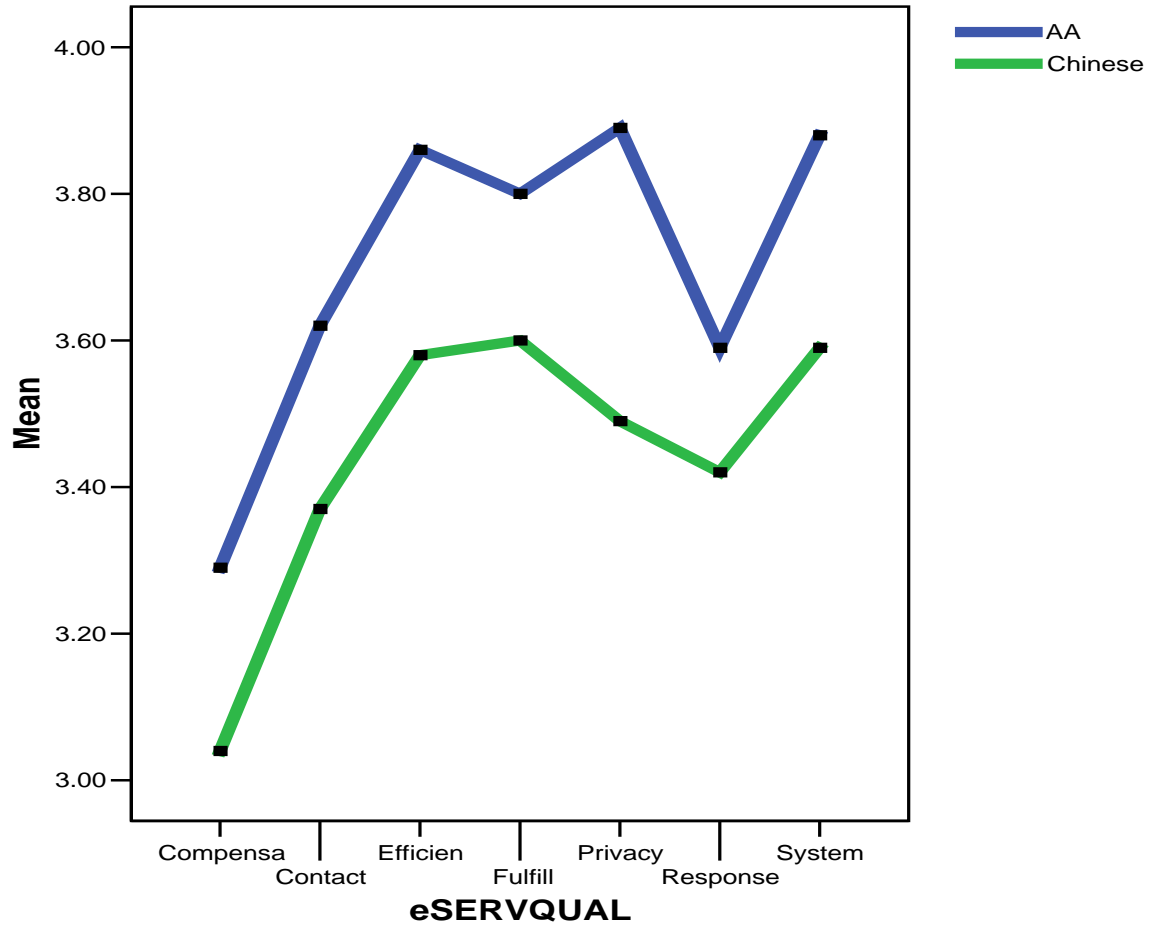
Chi-Square=888.89, df=474, P-value=0.00000, RMSEA=0.076

Figure 2
Web Service Quality Model Fit (CHINA)



Chi-Square=939.41, df=474, P-value=0.00000, RMSEA=0.079

Figure 3
Mean Web Service Quality Subscale Differences between Cultures



**Appendix 1
Instrument Labels**

E-S- QUAL		Label
1.	This site makes it easy to find what I need.	A1
2.	It makes it easy to get anywhere on the site.	A2
3.	It enables me to complete a transaction quickly.	A3
4.	Information at this site is well organized.	A4
5.	It loads its page fast.	A5
6.	This site is simple to use.	A6
7.	This site enables me to get on to it quickly.	A7
8.	This site is well organized.	A8
9.	This site is always available for business.	A9
10.	This site launches and runs right always.	A10
11.	This site does not crash.	A11
12.	Pages at this site do not freeze after I enter order information.	A12
13.	It delivers offers when promised.	A13
14.	This site makes items available for delivery within a suitable time frame.	A14
15.	It quickly delivers what I order.	A15
16.	It sends out the items ordered.	A16
17.	It has in stock the items the company claims to have.	A17
18.	It is truthful about its offerings.	A18
19.	It makes accurate promises about delivery of products.	A19
20.	It protects information about my Web-shopping behavior.	A20
21.	It does not share my personal information with other sites.	A21
22.	This site protects information about my credit card.	A22
E-RecS-QUAL		
1.	It provides me with convenient options for returning items.	B1
2.	This site handles product returns well.	B2
3.	This site offers a meaningful guarantee.	B3
4.	It tells me what to do if my transaction is not processed.	B4
5.	It takes care of problems promptly.	B5
6.	This site compensates me for problems it creates.	B6
7.	It compensates me when the product I ordered does not arrive on time.	B7
8.	It picks up items I want to return from my home or business.	B8
9.	This site provides a telephone number to reach the company.	B9
10.	This site has customer service representatives available online.	B10
11.	It offers the ability to speak to a live person if there is a problem.	B11