

Measuring Internet Retail Service Quality Using Confirmatory Factor Analysis

D.O.I - 10.51201/Jusst12605

<http://doi.org/10.51201/Jusst12605>

Jyoti Kumari, Assistant professor

School Of Management Sciences, Lucknow

Abstract:

In the early years of online retailing, having an online presence and low prices were believed to be key drivers of success. More recently, electronic service quality has become essential as an online marketing strategy. Online stores provide higher service quality to create online customer loyalty, improve customer satisfaction and keep a lasting competitive advantage. Despite its recognized significance, there are few comprehensive empirical researches that analyze the quality of Internet retail services. Therefore, there is always a scope of conducting study on the development of e-service quality scales. This paper demonstrates the development of a new scale having new measures for measuring customer satisfaction and patronage intention in e-retailing. For the purpose, Confirmatory factor Analysis was used using AMOS 16. After analysis, the measures like website design, fulfillment, personalization, customer service and ratings & reviews are identified the crucial measures for determining the satisfaction level and patronage intentions of online customers. At the end of this paper, limitations and future scope of the study was given.

Keywords- E-Service Quality, Customer Satisfaction, Patronage Intention, EFA, CFA, AMOS

Introduction and Background

With the emergence of internet retailing, service quality has been considered as an important tool in measuring the success or failure of e-commerce companies in the online world (Santos, 2003; Yang, 2001; Zeithaml, Parasuraman, & Malhotra, 2002). A number of existing e-service quality studies have attempted to identify the elements that characterize the perception of service quality among customers and to create models that outline the variations between consumer perceptions and the actual experience of service. (Janda, Trocchia, & Gwinner, 2002; Zeithaml et al., 2002). . However, Wolfinbarger and Gilly (2003) argue that the consistency of the measurement service

should go beyond the interface of the website. This is because the online shopping experience of a customer involves everything from information gathering, product assessment, decision-making, order, shipment, returns and customer support. It is obvious that measures to evaluate only websites cannot be adequate to assess the quality of service across different stages of delivery of online retail services. This also corresponds to Parasuraman, Zeithaml, and Malhotra's (2005); Henceforth, who state that the purpose of developing e-service quality scales is to quantify customers' entire experience of the service received rather than generating information for website designers. To date, relatively few systematic empirical studies have analyzed the efficiency of Internet retail services. Such studies include Barnes and Vidgen (2002); Janda et al. (2002); Loiacono, Watson, and Goodhue (2002); Wolfinbarger and Gilly (2003); and Yoo & Donthu (2001). However, many of these studies do not include all aspects of service quality (see, e.g., Boshoff, 2006). Some scales have also been developed by several researchers, but they have their disadvantages and cannot be used in all forms of industry. Therefore, the development and conceptualization of e-service efficacy has emerged as a prominent concern.

Due to the importance of service quality in the success of e-retailers, external validation of e-service quality measures through replication is extremely important. Replications not only help to assess the reliability and validity of newly established measurement methods, but also help in establishing the scope and limitations of their generalization to other contexts. (Hubbard, Vetter, & Little, 1998) It is therefore important to test and determine the validity of newly developed scale in a cross-national context in order to recognize the drawbacks it may have with respect to its generalizability.

Research method

As this research paper concentrates on validating the newly developed scale therefore a new scale having 51-items was developed in this study. To measure customer satisfaction in the online environment, the measures from (Anderson & Sullivan, 1993) and for patronage intention the measures from Shim and Drake, 1990, Dolbec and Chebat, 2013; Vlachos and Vrechopoulos, 2012) was adopted. For collecting responses, Uttar Pradesh was selected as population of the study because of the most populous state as per census 2011. Proportionate stratified random sampling was used. This population was divided into four stratus i.e. .Eastern, Western, Central

and Bundelkhand region. From each stratum, responses were collected on random basis by distributing questionnaires. Around 610 questionnaires were distributed in which 598 questionnaires were received back. After initial filtration, rejection of incomplete 158 questionnaires and finally 440 responses were considered for further analysis.

Table 1.DEMOGRAPHIC PROFILE OF RESPONDENTS

		Name of the city where respondent belong to				Total
		Ghaziabad	Jhansi	Lucknow	Varanasi	
Age of the respondents	18-25	12	12	13	14	51
	26-35	29	32	35	31	127
	36-45	42	29	31	25	127
	46-55	17	25	13	28	83
	above 55	10	12	18	12	52
Gender of the respondents	Male	53	59	51	56	219
	Female	57	51	59	54	221
Marital status of respondents	Married	51	46	57	43	197
	Single	52	63	48	65	228
	Separated	7	1	5	2	15
Education level of respondents	up to 12 th	1	24	3	6	34
	Graduate	28	41	22	31	122
	Postgraduate	70	31	56	47	204
	Others	11	14	29	26	80
Employment status of respondents	Govt. job	12	10	24	13	59
	Private job	47	46	33	39	165
	Business	21	15	19	22	77

	Student	17	26	16	21	80
	Housewife	13	13	18	15	59
Family annual income of respondents	Below 2	4	14	12	11	41
	4-6	27	34	28	28	117
	6-10	49	48	38	46	181
	10-15	13	6	14	15	48
	16 & above	17	8	18	10	53
Preferred device for online shopping	Smart phone	97	103	98	93	391
	Laptop/Desktop	13	7	12	17	49

Method of analysis

As in this study a new scale was developed. For identifying the reliability of the new developed scale, exploratory factor analysis was performed initially and after doing so some iteration, the developed scale was redefined and then confirmatory factor analysis (CFA) was used to assess its unidimensionality, reliability, and validity. Gerbing and Anderson (1988) argue that because factors obtained via EFA are defined as the ‘weighted sum of all observed variables’, they do not represent the theoretical constructs underlying each set of indicators. Therefore, CFA was used for the same reference.

The research started by looking at each of the dimensions separately, as suggested by Garver and Mentzer (1999), to determine whether each of the items loaded on the dimension it was intended to be, and to evaluate the unidimensionality of the constructs. All the items loaded quite well on the appropriate dimensions. Cronbach’s alpha value greater than 0.6, with values from 0.7 – 0.8 considered to be acceptable for scale reliability (Field, 2009). Cronbach’s alpha value greater than 0.6, with values from 0.7 – 0.8 considered to be acceptable for scale reliability (Field, 2009). In view of this, the research proceeded to the next stage: performing CFA by analyzing all the measurements together.

Validity Analysis: The degree to which data collection methods accurately measure what they were meant to measure is known as validity (Saunders and Thornhill, 2003). Validity can be defined as the ability of a measurement (including questions or items) to avoid systemic or non-random mistakes (Hair et al. 1998). Validity therefore depends on how correctly and well the measure's content reflects the definition of the study we are evaluating.

To satisfy the validity procedure, we carried out following validity checks:

Content Validity: The validity of the content defines how well the measure's content corresponds to the conceptual description of the phenomenon under study (Nunnally 1978). The data collection instrument has been developed after extensive literature review and content validity was ensured by a thorough review and expert opinions including academicians and practitioners.

Construct Validity: Validity of Construct determines how well a test or experiment lives up to its arguments. It refers to whether a variable's operational definition actually represents a concept's true theoretical meaning.

There are two ways of ensuring this validity:

➤ **Convergent validity:** Convergent validity explains that a high proportion of variation in general should be converged or shared by the objects that are measures of a particular construct (Hair, 2006). Anderson and J.C. Gerbing and, D.W. (1991) advocates that convergent validity is assessed in the measurement model by deciding whether or not the objects on a scale converge or load together on a single build. In other words, convergent validity is the degree of convergence seen when two attempts are made to measure the same construct through maximally different methods. If there is no convergence, then either the theory used in the analysis must be evaluated or measure purification must be carried out by removing the items.

➤ **Discriminant Validity:** The degree to which measurements of two different structures are comparatively distinctive from each other (Campbell & Fiske, 1959) is calculated by discriminant validity. Discriminant validity assesses the degree of variance between a definition and its indicators and another definition and its indicators. It means that items from one scale should not load or converge too closely with items from a different scale and that different latent

variables which correlate too highly may indeed be measuring the same construct rather than different constructs (Garver and Mentzer, 1999).

According to Hair(1998) the data must meet the threshold values to ensure the validity and reliability. These are achieved by measuring Composite Reliability (CR), Average Variance Expected (AVE), Maximum Shared Variance (MSV) and Average Shared Variance (ASV).

- ✓ To achieve proper reliability, value of CR shall be >0.70 .
- ✓ For Convergent reliability to be achieved, $CR > AVE$ and $AVE > 0.50$.
- ✓ For Discriminant validity, $MSV < AVE$ and $ASV < AVE$. AVE should be higher than 0.5

Table.2. all the constructs the value of AVE, CR and MSV were under acceptable range hence no issues regarding validity was observed.

Table 2 Result of validity test

	CR	AVE	MSV	MaxR(H)	SATIS	WD	FUL	RR	PS	CS	PI
SATIS	0.815	0.595	0.120	0.819	0.772						
WD	0.944	0.736	0.060	0.946	0.099	0.858					
FUL	0.907	0.765	0.254	0.937	0.102	0.128	0.875				
RR	0.820	0.608	0.097	0.897	0.183	0.245	0.130	0.780			
PS	0.840	0.648	0.173	0.916	0.152	0.141	0.197	0.187	0.805		
CS	0.966	0.905	0.097	0.970	0.236	0.095	0.159	0.312	0.131	0.951	
PI	0.954	0.726	0.254	0.975	0.346	0.099	0.504	0.147	0.416	0.252	0.852

After checking validity issues,a first order CFA was then constructed to check whether the selected measures provides a good fit to the data by identifying the relationship between the

observed and unobserved variables and the constructs (Child, 1990). In the present study CFA was performed on 51 items using AMOS 16.

The below table 1.3 shows the acceptable value of each indices and also the observed value calculated in this study.

Table 3: Fit Indices and Acceptable Cut-off Points		
Statistics	Interpretation	Observed Values
RMSEA	Below 0.08 is acceptable, below 0.05 shows a good fit, model rejected if above 0.1 (Browne & Cudeck, 1993);	.071
	Between 0.08 to 0.10 is mediocre fit (MacCallum et al., 1996);	
	Should be close to or below 0.07 (Steiger, 2007)	
RMR	Should be as low as possible (Tabachnick & Fidell, 2012);	.070
	Not preferred due to its sensitivity to scale level (Hooper et al., 2008)	
	Should be close to or below 0.08 (Hu & Bentler, 1999);	
NFI	Above 0.90 means a good fit (Bentler & Bonnet, 1980);	.865
	Should be above 0.95 (Hu & Bentler, 1999);	
	Up to 0.80 is acceptable (Hooper et al., 2008)	
NNFI-TLI	Should be close to or above 0.95, between 0.90 and	.888

	0.95 is acceptable, model error if below 0.90 (Hu & Bentler, 1998, 1999)	
CFI	Close to or above 0.95 shows a good fit, 0.90-0.95 is acceptable, model error if below 0.90 (Hu and Bentler, 1998, 1999)	.906
IFI	0.90 and above shows a good fit (Marsh & Hau, 1996)	.907
RFI	0.90 and above shows a good fit (Marsh & Hau, 1996)	.839
GFI	Should be above 0.95 (Miles & Shevlin, 1998);	.865
	0.90 and above shows a good fit (Hooper et al., 2008)	
AGFI	0.90 and above shows a good fit (Hooper et al., 2008)	.827
PNFI	There is not a certain limit but can be as low as 0.50 (Hooper et al., 2008; Mulaik et al., 1989);	.727
PGFI	There is not a certain limit but can be as low as 0.50 (Hooper et al., 2008; Mulaik et al., 1989);	.675

The above indices states that there is significant relationship among the selected measures and shows a good fit for the selected area of study.

Discussion and conclusions

As per the study, Website design, Fulfillment, Rating & Reviews, Personalization and Customer Service are identified as the major dimensions of e-service quality in the field of e-retailing. All the dimensions (Website design, Fulfillment, Rating & Reviews, Personalization and Customer Service) have significant impact on customer satisfaction. Rating and reviews appears the most crucial dimension which creates impact on customer satisfaction. In the districts of Uttar Pradesh online customers firstly see the ratings and reviews given on the preferred product then move forward to do payment and give order of the product. It means ratings & reviews create influence

on purchase decision of online customers. The services provided by online stores in terms of exchanging and returning the items from the desired place and quick resolution of their queries etc makes customers satisfied and happy. Website design, Fulfillment, Personalization and Customer Service are identified as the major determinants of patronage intention in the field of e-retailing which help in determining the intention of online customers to visit that particular or same store again and again. Customer services like easy return and exchange options pick up from the given address and quick query solving etc. play important role in patronizing the behavior of online buyers. Other crucial factor personalization means customers intend to repeat the buying process from the same online store when they feel special and online store gives information about the search category of products and also for inculcating a sense of uniqueness that they are special and given special attention by the online store.

Managerial implications

The multidimensional context, specific conceptualization of e-service quality will allow, service providers to have a more detailed and clear understanding of the various aspects of the construct. The developed instrument can be applied by service providers for monitoring of service quality as per the perceptions of customers. The developed scale will also provide directions to service providers as to which particular dimension needs more surveillance in terms of their significance. As per the importance of these constructs, the service providers can take decisions regarding investment of resources

Limitations and future scope of the study:

The collected data is from the cities of Uttar Pradesh, i.e. Lucknow, Ghaziabad, Varanasi and Jhansi. While these cities are believed to be the region's true representatives they may still have some cultural and behavioural diversity within themselves, which may serve as a study constraint. Only 4 districts of U.P. were selected for collecting the responses because of involvement of time and money. Industry assistance was also sought to acquire a list of buyers from some websites, but the response was negative. Future research could be conducted to explore some of other factors apart from the variables recommended in this research because other variables may influence customer understanding and preference that could be gained by adding more service variables in the study,. The assessment of the quality of service of online

stores may vary from one location to another location due to various demographic, webographic or infrastructural differences, etc. Researchers may take samples from other domains and can identify the variations or similarities in the findings.

References:

- Anderson, E. W., & Sullivan, M. W. (1993). The antecedents and consequences of customer satisfaction for firms. *Marketing science*, 12(2), 125-143.
- Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological bulletin*, 103(3), 411.
- Anderson, J. C., & Gerbing, D. W. (1991). Predicting the performance of measures in a confirmatory factor analysis with a pretest assessment of their substantive validities. *Journal of applied psychology*, 76(5), 732.
- Anderson, R. E., & Srinivasan, S. S. (2003). E-satisfaction and e-loyalty: A contingency framework. *Psychology & marketing*, 20(2), 123-138.
- Barnes, S.J., & Vidgen, R.T. (2002). An integrative approach to the assessment of e-commerce quality. *Journal of Electronic Commerce Research*, 3(3), 114-127.
- Boshoff, C. (2006). A psychometric assessment of E-S-QUAL: A scale to measure electronic service quality. *Journal of Electronic Commerce Research*, 8(1), 101-114.
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit In: Bollen KA, Long JS, eds. *Testing Structural Equation Models*. Beverly Hills, CA: Sage, 136-162.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological bulletin*, 56(2), 81.
- Child, D. (1990). *The essentials of factor analysis*, second edition. London: Cassel Educational Limited.
- Dolbec, P. Y., & Chebat, J. C. (2013). The impact of a flagship vs. a brand store on brand attitude, brand attachment and brand equity. *Journal of Retailing*, 89(4), 460-466.
- Field, A. (2009). *Discovering statistics using SPSS:(and sex and drugs and rock'n'roll)*. Sage.
- Garver, M.S., & Mentzer, J.T. (1999). Logistics research methods: Employing structural equation modeling to test for construct validity. *Journal of Business Logistics*, 20(1), 33-57.
- Hair J.F, Black W.C., Babin B.J., Anderson R.E. and Tatham R.L. (2006), *Multivariate Data Analysis* (6th ed.), Upper Saddle River, New Jersey: Prentice Hall.
- Hair, A. (1998). Tatham, and Black. *Análisis multivariante*.
- Hooper, D., Coughlan, J., & Mullen, M. (2008, September). Evaluating model fit: a synthesis of the structural equation modelling literature. In *7th European Conference on research methodology for business and management studies* (pp. 195-200).
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, 6(1), 1-55.

- Hubbard, R., Vetter, D.E., & Little, E.L. (1998). Replication in strategic management: Scientific testing for validity, generalizability, and usefulness. *Strategic Management Journal*, 9(3), 243–254
- Janda, S., Trocchia, P.J., & Gwinner, K.P. (2002). Consumer perception of Internet retail service quality. *International Journal of Service Industry Management*, 13(5), 412–431.
- Loiacono, E.T., Watson, R.T., & Goodhue, D.L. (2002). WebQual: A measure of web site quality. In K.R. Evans & L.K. Scheer (Eds.), *Marketing Theory and Applications*, 2002 Winter Educators' Conference, Proceedings pp. 432–438. Chicago: American Marketing Association
- MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological methods*, 1(2), 130.
- Marsh, H. W., Balla, J. R., & Hau, K. T. (1996). An evaluation of incremental fit indices: A clarification of mathematical and empirical properties. *Advanced structural equation modeling: Issues and techniques*, 315-353
- Miles, J. and Shevlin, M. (1998), "Effects of sample size, model specification and factor loadings on the GFI in confirmatory factor analysis," *Personality and Individual Differences*, 25, 85-90.
- Mulaik, S. A., James, L. R., Van Alstine, J., Bennett, N., Lind, S., & Stilwell, C. D. (1989). Evaluation of goodness-of-fit indices for structural equation models. *Psychological bulletin*, 105(3), 430.
- Nunnally, J.C. (1978). *Psychometric methods* (2nd ed.). New York: McGraw-Hill
- Parasuraman, A., Zeithaml, V. A., & Malhotra, A. (2005). ES-QUAL: A multiple-item scale for assessing electronic service quality. *Journal of service research*, 7(3), 213-233.
- Santos, J. (2003). E-service quality: A model of virtual service quality dimensions. *Managing Service Quality*, 13(3), 233–246.
- Saunders, M., Lewis, P., & Thornhill, A. (2003). *Research methods for business students*. Essex: Prentice Hall: Financial Times
- Shim, S., & Drake, M. F. (1990). Consumer intention to utilize electronic shopping. The Fishbein behavioral intention model. *Journal of direct marketing*, 4(3), 22-33.
- Steiger, J.H. (1990), "Structural model evaluation and modification," *Multivariate Behavioral Research*, 25, 214-12.
- Tabachnick, B. G. (2012). Fidell, I. S.(1996) *Using multivariate statistics*. Costello, A. B., & Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical assessment, research, and evaluation*, 10(1), 7.
- Vlachos, P. A., & Vrechopoulos, A. P. (2012). Consumer–retailer love and attachment: Antecedents and personality moderators. *Journal of Retailing and Consumer Services*, 19(2), 218-228.

- Wolfinbarger, M., & Gilly, M.C. (2003). eTailQ: Dimensionalizing, measuring and predicting etail quality. *Journal of Retailing*, 79(3), 183–198.
- Yang, Z. (2001). Consumer perceptions of service quality in Internet-based electronic commerce. *Proceedings of the 30th EMAC Conference* pp. 8–11. Bergen, Norway: The Norwegian School of Economics & Business Administration.
- Yoo, B., & Donthu, N. (2001). Developing a scale to measure the perceived quality of an Internet shopping site (SITEQUAL). *Quarterly Journal of Electronic Commerce*, 2(1), 31–45.
- Zeithaml, V.A., Parasuraman, A., & Malhotra, A. (2002). Service quality delivery through web sites: A critical review of extant knowledge. *Journal of Academy of Marketing Science*, 30(4), 362–375.