A PSYCHOMETRIC ASSESSMENT OF E-S-QUAL: A SCALE TO MEASURE ELECTRONIC SERVICE QUALITY

Christo Boshoff
Department of Business Management, University of Stellenbosch
Stellenbosch, SOUTH AFRICA
cboshoff@sun.ac.za

ABSTRACT

As the Internet retailing market grows at an inceasing rate, those who use the Internet as a retailing channel have realised that service delivery and service quality are as important for the Internet environment as for the bricks-and-mortar environment. The absence of a valid and reliable instrument to measure service quality in this new environment has, however, bedeviled the endeavors of both scholars and practitioners to effectively measure and thus manage service-quality strategies. The first scale developed that effectively captured the nature of electronic service quality from the perspective of online shopping through a retail Website was the E-S-QUAL scale.

The objective of this study was to subject the E-S-QUAL scale to a psychometric assessment. The results revealed that the scale is effective in capturing the essence of electronic service quality, but that both scholars and practitioners must assess the underlying factor structure of their data before drawing any conclusions from their study. For this particular data set, a six-factor configuration proved to be superior to the four-factor configuration originally proposed. Several ways in which the data can be analyzed are demonstrated, and suggestions for additional scale development and refinement are offered.

Key words: Internet, Internet retailing, Electronic commerce, Electronic service quality, LISREL

1. Introduction

The importance of quality for any firm operating in a competitive environment is well documented [Garvin 1988]. Thanks to the "quality movement" that started around the early 1930's, quality is no longer a peripheral issue "outsourced" to engineers, the "techies," or the quality inspectors. Instead, quality has become part of the mainstream of business thinking once managers at all levels have realised that they need to think of consumers' quality needs as much as they need to think about finance, logistics, and profitability.

Although somewhat belatedly, the importance of excellence in service delivery has been equally well documented [Zeithaml 2000; Zeithaml 1988]. Whether as part of a market offering that is a "pure" service or part of a market offering that is "tangible-dominated," the importance of the quality of a service can hardly be over-estimated. In any event, the anecdotal and empirical evidence is compelling that desirable outcomes such as repeat purchases, customer loyalty, and eventually profitability will simply not occur unless firms have succeeded in satisfying their customers' service quality needs first [Rust, Zahorik and Keiningham 1995; Bates, Bates and Johnston 2002].

Researchers in the domain of electronic commerce, Internet marketing, and Internet retailing have realised that, as the Internet retailing market matures, service delivery and service quality are as important for the Internet environment as for the bricks-and-mortar environment**.** Several researchers have therefore investigated the impact of electronic service quality in, for instance, the financial services industry [Floh and Treiblmaier2006] and travel and retailing industry [Lennon and Harris 2002].

2. The need to measure customers' perceptions of electronic service quality

Although a generic instrument to measure service quality (called SERVQUAL) has been in existence for some time [Parasuraman, Zeithaml and Berry 1988], many have argued that electronic channels are simply so different from the "bricks-and-mortar" environment that we are accustomed to that a complete rethink of electronic service quality was called for. Voss [2000], for instance, pointed out that the Assurance and Empathy dimensions, as operationalized and measured by SERVQUAL, refer to human interaction between service provider and customer which simply may not take place during a Web-based service encounter.

The absence of a valid and reliable scale to measure electronic service quality "forced" early researchers to make use of some fairly unsatisfactory alternatives to measure electronic service quality, such as using selected "generalizable" items from the SERVQUAL scale to measure electronic service quality [Montoya-Weiss, Voss and Grewal 2003]. These unsatisfactory practices must compromise some of the empirical results that emanated from those early studies.

In response to the growing recognition that service quality in an electronic channel environment is different (and thus the way it ought to be measured), several researchers have attempted fill this gap. Examples include Loiacono, Watson and Goodhue's [2002], WebQual™ instrument, Yoo and Donthu's [2001] "SITEQUAL," and Wolfinbarger and Gilly's [2003] "eTailQ." However, many of these attempts were flawed [Finn and Kayande 2002] as they were either too narrowly focused or failed to address service delivery from the perspective of service delivered to customers through a Website [Zeithaml, Parasuraman and Malholtra 2002].

The first scale developed that effectively captured the nature of electronic service quality *from the perspective of online shopping through a Website* was the E-S-QUAL scale developed by Parasuraman, Zeithaml and Malholtra [2005]. The E-S-QUAL scale measures four dimensions of electronic service quality, namely Efficiency, System Availability, Fulfillment, and Privacy.

The objective of the present study was to subject the E-S-QUAL scale proposed by Parasuraman et al. [2005] to a psychometric assessment in order to assess its construct validity.

The discussion of E-S-QUAL's empirical evaluation is preceded by a discussion of the Internet as a marketing tool, the importance of service quality as a differentiating variable in Internet marketing, and a review of earlier attempts to conceptualize and measure the quality of service delivery in an electronic environment.

3. The internet as a marketing tool

It took some time before marketers fully appreciated the potential impact of the Internet on marketing practices. However, the realization eventually dawned that if this new technology is to be used as a channel of distribution, consumer needs and customer satisfaction will be as important as always [Wang, Tang and Tang 2001].

As a result, the focus of researchers, Internet marketing practitioners, and site developers shifted, over time, from the appearance of the Website and the impressive features that can be incorporated into a Website (a production orientation), to information content, to functionality (a product orientation), and belatedly to attempts to understand consumer needs and their interaction with the Internet (a customer orientation).

Some of the earlier attempts to study consumers' interaction with the Internet focused on what constitutes a compelling online experience [Novak, Hoffman and Yung 2000]. Using the so-called "flow construct," Novak et al. [2000], for instance, found that Websites that moderately "challenge" consumers and offer them some excitement will raise their level of involvement with a Website.

As the use of the Internet as a marketing tool growed, however, Internet marketers soon realized that a flashy Website, even if fully functional, will not be the differentiating variable some marketers may have hoped for, and that service quality will probably be, as is the case in the goods sector, the only durable means of differentiation [Zeithaml et al. 2002]. The question that remained unanswered was: how should we measure electronic service quality?

4. Measuring service quality in the internet environment

Although academic researchers' early attempts to evaluate or measure aspects related to Websites (such as SITEQUAL) were based on consumer perceptions, the researchers generally studied convenience samples in laboratory-like settings, which did not capture the realism of actual online buying. Many of those studies were also aimed more at providing feedback to Website developers and designers [Loiacono, Watson and Goodhue 2000] than at understanding actual buying decision-making. In addition, these early studies were based on conceptualizations that were too narrow and did not capture all aspects of a final consumer's buying process, or too broad in the sense that service was only a limited portion of what was being evaluated (in the case of Wolfbinbarger and Gilly's 2003 study). In other words, the early focus was not on service delivery to final consumers who actually shopped online (see Table 1 for a summary).

Due to the differing approaches used by the early electronic marketing researchers, and particularly due to the confusion caused by the sometimes poor conceptualization of what was being studied, earlier attempts to conceptualize electronic service quality did not enjoy universal acceptance (see Table 1 for a summary of the most important criticisms). The problem was that too many of these attempts were too narrow in focus, as they were limited to actual electronic transactions only. Instead, Zeithaml et al. [2000] argued, the focus ought to be on all cues and encounters that occur before, during, and after the transaction.

Table 1: Limitations of earlier attempts to measure electronic service quality

INSTRUMENT	CRITICISM	SOURCES
WebQual™ Loiacono et al. [2002]	1) … geared toward helping Website developers better design Websites …. 2) Surveyed students, not actual purchasers 3) Did not include fulfillment as a dimension of SQ	Zeithaml et al. [2002] Parasuraman et al. [2005]
e-TailQ Wolfinbarger and Gilly [2003]	1) Did not capture the entire buying process 2) Dimensionality questioned	Zeithaml et al. [2002] Parasuraman et al. [2005]
Szmanski and Hise [2000]	… did not include aspects of customer service of fulfillment, and dealt with satisfaction rather than service quality.	Zeithaml et al. [2002]
SITEQUAL Yoo and Donthu [2001]	1) Used convenience samples 2) Did not capture all aspects of the purchasing process	Parasuraman et al. [2005]

Against this background, Zeithaml et al. [2000] defined Website service quality as the extent to which a Website facilitates efficient and effective shopping, purchasing, and delivery of products and services.

Based on this conceptualization and founded on the extensive empirical research they had conducted over the past 20 years, Parasuraman et al. [2005] embarked on an extensive scale development process to measure the quality of service delivery of Websites developed for the purpose of online shopping by final consumers.

The outcome of this process was a 22-item scale that measures four dimensions of electronic service quality that were named *Efficiency* (8 items), *System availability* (4 items), *Fulfillment* (7 items), and *Privacy* (3 items). This scale, named E-S-QUAL, was used to collect the data in this study. The objective of the this study was to subject the E-S-QUAL scale proposed by Parasuraman et al. [2005] to a psychometric assessment in order to assess its construct validity.

5. Methodology

5.1 Measuring instrument

The E-S-QUAL instrument originally published by Parasuraman et al. [2005] consisted of two parts, the one measuring service quality and the other so-called "service recovery." The latter part was not used in this study. Another difference from the original study was that a 6-point Likert-type scale was used (as opposed to the original 5-point scale), in which a 6 implied strong agreement and a 1 meant strong disagreement with a statement.

5.2 Sample

The E-S-QUAL instrument was sent electronically to the customer base of an Internet marketing firm selling mainly books, DVDs, CDs, and other gifts online. All participants had opted to receive questionnaires when they bought their first product on the Website and had shopped on the site at least once during the previous three months. Of the 6 906 customers surveyed, 1 409 responded, representing a response rate of 20.4%.

The nature of the respondent group is described in Table 2. Table 2 reveals a slight gender bias in the sense that about two-thirds of those who participated were females, and the majority of respondents (58.4%) were in the 25-39 age bracket. Table 2 also shows that 96.2% of the respondents had completed their secondary school education, and almost half of them (44.6%) had obtained some tertiary education qualification.

These demographic characteristics compare favorably with the Parasuraman et al. [2005] study, in which the female bias was slightly stronger (above 70%), but more than half were younger than 40 years of age, and around 40% had graduated.

5.3 Assessment of non-response bias

To assess the possibility of non-response bias, a time-trends approach, proposed by Armstrong and Overton [1977], was used. They suggest that those who respond late are similar (in respect of characteristics such as demographics) to non-respondents. In other words, if the early respondents (termed quartile 1) are the same as the late respondents (quartile 4), the realized sample will not differ from the non-respondents, and non-response bias should be minimal.

A comparison between the first quartile and the last quartile of respondents, using all available demographic-type information, was then conducted, using a Chi-square difference test. The results revealed that there were no significant differences between quartile 1 and quartile 4 at the 1% level of significance in terms of gender and level of education. However, in terms of age, there were some significant ($p < 0.01$) differences. Those who responded early (quartile 1) were more likely to be younger than those who responded later (quartile 4). This analysis suggested that non-response bias was minimal.

Assessing response bias will never be perfect. Although the Armstrong and Overton [1977] approach was no different, it did provide some evidence of non-response bias, and as their study showed, it was also superior to some other methods.

According to company executives, the composition of the respondents closely resembled their actual customer base. The composition was also in line with the general demographic characteristics of Internet users globally. Statistics by international firms such the Gartner Group point out that Internet users are generally in higher-income groups and relatively well-educated, with a slight bias towards male users.

Table 2: Demograhpic chracteristics of the respondent group

GENDER	**NUMBER**	**%**
Males	484	34.4
Females	925	65.6
Total	**1409**	**100.0**
AGE CATEGORY		
Under 18 years	12	0.9
Between 18 and 24 years	194	13.8
Between 25 and 29 years	324	23.0
Between 30 and 34 years	300	21.3
Between 35 and 39 years	199	14.1
Between 40 and 44 years	140	9.9
Between 45 and 49 years	99	7.0
Between 50 and 55 years	109	7.7
Older than 56 years	28	2.0
No response	4	0.3
Total	**1409**	**100.0**
HIGHEST LEVEL OF EDUCATION COMPLETED		
Primary school	53	3.7
Secondary school	728	51.7
Other tertiary	255	18.1
University degree	373	26.5
Total	**1409**	**100.0**

5.4 Data analysis

The first phase of the data analysis procedure was to conduct an exploratory factor analysis to assess whether the data contain the four factors or dimensions of electronic service quality as suggested by Parasuraman et al. [2005]. The next phase was to assess the reliability of the E-S-QUAL scale using Cronbach's alpha, followed by a confirmatory factor analysis to assess the construct validity of the scale. Finally, as recommended by Steenkamp and van Trijp [1991], the nomological validity of the E-S-QUAL scale was assessed by empirically assessing its relationships with other constructs in a nomological net.

When the exploratory factor analyses were conducted, the estimation method used was Maximum Likelihood, and Direct Quartimin was specified as the rotation method because all the factors were positively correlated with each other [see Appendix A].

6. The empirical results

6.1 Exploratory factor analysis results

Initially, the number of factors to be extracted for the first exploratory factor analysis was not specified. Several factor solutions were considered. The initial indications, based on the scree plot, the Eigen values, and the interpretability of the factors that emerged, were that the data might contain five or six factors rather than the expected four factors or dimensions reported by Parasuraman [2005: 220].

The six-, five-, and four-factor solutions are reported in Tables 3, 4, and 5 respectively.

Table 3: Rotated factor loadings:six-factor solution(1)

	Factor 1 **Efficiency**	Factor 2 **Delivery**	Factor 3 **Privacy**	Factor 4 **Speed**	Factor 5 **System availability**	Factor 6 **Reliability**
EFF4	**0.848**	0.027	-0.004	-0.050	-0.070	0.050
EFF8	**0.750**	0.049	-0.004	0.096	-0.033	0.057
EFF1	**0.740**	0.019	0.030	-0.019	0.038	-0.025
EFF2	**0.665**	-0.063	-0.028	0.094	0.091	0.031
EFF6	**0.609**	0.033	0.096	0.161	0.021	-0.015
EFF3	**0.389**	0.103	0.110	-0.034	0.263	-0.011
FUL3	0.021	**0.927**	0.019	0.022	0.016	0.006
FUL2	0.006	**0.904**	0.025	0.033	0.017	-0.019
FUL1	0.032	**0.788**	-0.014	0.026	-0.032	0.102
PRI1	0.053	0.015	**0.766**	0.020	0.056	-0.013
PRI2	0.003	-0.044	**0.907**	0.018	-0.056	0.062
PRI3	-0.028	0.038	**0.886**	-0.009	-0.001	-0.030
EFF5	0.006	0.068	-0.022	**0.799**	0.002	-0.020
EFF7	0.084	-0.014	0.047	**0.773**	-0.044	0.040
SYS2	-0.003	0.011	0.034	**0.725**	0.119	0.003
SYS4	0.099	0.044	0.031	-0.031	**0.729**	-0.008
SYS3	-0.080	0.004	0.014	0.179	**0.644**	0.072
SYS1	0.193	-0.083	0.092	0.109	**0.356**	0.157
FUL6	0.077	-0.061	0.063	0.059	-0.002	**0.807**
FUL5	0.037	0.232	-0.063	-0.022	0.031	**0.590**
FUL4	-0.052	0.161	0.167	-0.020	0.115	**0.367**
Eigen value	2.871	2.419	2.270	1.862	1.197	1.190
α-value	0.889	0.938	0.899	0.858	0.768	0.764

1) Loadings greater than 0.35 were considered significant

Table 3 shows that the six-factor solution is a fairly clear one (after the removal of item FUL7), with all loadings greater than 0.35 and no cross loadings. All the factors that are retained have Eigen values exceeding one, have at least three items loading on them, and are interpretable. All the items expected to load on the *Efficiency* factor (Factor 1) loaded as expected, except for items EFF5 and EFF7. These two items refer to speed: "The Web pages load fast" (EFF5) and "This site enables me to get on quickly" (EFF7), and loaded on a separate factor (Factor 4). The third item (SYS 2) that loaded on this factor (Factor 4) also referred to speed: "This site launches and runs right away." Accordingly, Factor 4 was labeled *Website speed*. The items expected to measure the *Fulfillment* dimension proposed by Parasuraman et al. [2005] split into two, however. The items that loaded on Factor 2 refer to *Delivery,* and those that loaded on Factor 6 refer to *Reliability* (mainly honesty/truthfulness).

The two factors that remained intact from the Parasuraman et al. [2005] configuration were *Privacy* (all the items loaded as expected, on Factor 3) and *System Availability* (all the items loaded as expected, on Factor 5, except for SYS2).

Item FUL7 was the only item that had to be deleted because it consistently loaded to a significant extent on two factors, namely the *Delivery* factor (Factor 2) and the *Reliability* factor (Factor 6). If one carefully considers the wording of the item: "It makes **accurate promises** about **delivery** of products," one can understand why respondents were unable to fully appreciate the distinction between the two factors.

Table 3 also shows that the items used to measure each of the six factors (dimensions) can be regarded as reliable, as the Cronbach alpha value for all of them exceeds the 0.7 normally regarded as the cut-off value for internal consistency [Nunnally and Bernstein 1994].

The six-factor solution explains 65.6% of the variation in the data.

The next step was to consider a five-factor solution.

Table 4: Rotated factor loadings:five-factor solution(1)

	Factor 1 **Fulfillment**	Factor 2 **Efficiency**	Factor 3 **Privacy**	Factor 4 **Speed**	Factor 5 **System availability**
FUL3	**0.968**	-0.025	-0.011	0.076	-0.051
FUL2	**0.935**	-0.046	-0.009	0.089	-0.053
FUL1	**0.884**	0.011	-0.030	0.048	-0.061
FUL5	**0.567**	0.128	0.003	-0.100	0.150
FUL6	**0.385**	0.214	0.171	-0.056	0.170
FUL4	**0.367**	0.001	0.199	-0.072	0.204
EFF4	0.040	**0.869**	-0.003	-0.044	-0.082
EFF8	0.069	**0.774**	-0.001	0.088	-0.030
EFF1	-0.011	**0.738**	0.020	0.000	0.020
EFF2	-0.064	**0.682**	-0.028	0.091	0.102
EFF6	0.013	**0.617**	0.094	0.163	0.014
EFF3	0.083	**0.377**	0.098	0.000	0.236
PRI2	-0.028	0.008	**0.934**	0.010	-0.045
PRI3	0.006	-0.046	**0.892**	0.010	-0.021
PRI1	-0.008	0.041	**0.778**	0.043	0.030
EFF5	0.071	0.031	-0.001	**0.758**	0.024
EFF7	0.015	0.126	0.075	**0.705**	0.007
SYS2	0.017	0.025	0.051	**0.681**	0.155
SYS3	0.026	-0.078	0.006	0.178	**0.694**
SYS4	0.020	0.083	0.029	0.039	**0.660**
SYS1	-0.017	0.223	0.102	0.080	**0.423**
Eigen value	3.223	3.037	2.380	1.655	1.301
α-value	0.894	0.889	0.899	0.858	0.768

1) Loadings greater than 0.35 were considered significant

Table 4 shows that when five factors are extracted, a factor structure that is starting to resemble Parasuraman et al.'s [2005] four-factor configuration emerges. The only exception is that, after the removal of item FUL7 after again cross-loading, the factor labeled *Website speed* in Table 4 again split out of the *Efficiency* factor (measured by items EFF5, EFF7 and SYS2). It thus appears that *Efficiency*, as measured by Parasuraman et al. [2005], is not a uni-dimensional construct.

It must be noted that in the five-factor solution the items (FUL 4, FUL5 and FUL6) that were regarded as measures of *Reliability* in the six-factor solution (Table 3) now load on the *Fulfillmen*t factor.

The five-factor solution explains 63.4% of the variation in the data.

The next question is whether, if four factors are extracted, whether all 22 items of the E-S-QUAL instrument load as suggested by Parasuraman et al. [2005].

Table 5 shows that all the items (including item FUL7) load to a significant extent on the factors as expected, with the exception of items EFF5 and EFF7. Both these items refer to speed: "The Web pages load fast" (EFF5) and "This site enables me to get on quickly" (EFF7). Although they were expected to measure *Efficiency*, our respondents interpreted them as additional measures of the *System Availability* dimension.

Table 5: Rotated factor loadings:four-factor solution(1)

	Factor 1 **Fulfillment**	Factor 2 **Efficiency**	Factor 3 **System Availability**	Factor 4 **Privacy**
FUL3	**0.970**	-0.042	0.031	-0.050
FUL2	**0.942**	-0.066	0.045	-0.051
FUL1	**0.920**	-0.013	-0.008	-0.074
FUL7	**0.791**	0.038	0.027	0.012

FUL5	**0.619**	0.121	-0.044	0.039
FUL6	**0.436**	0.212	0.019	0.206
FUL4	**0.402**	-0.003	0.032	0.250
EFF4	0.044	**0.884**	-0.120	-0.025
EFF8	0.073	**0.783**	0.054	-0.028
EFF1	-0.009	**0.753**	-0.007	0.021
EFF2	-0.061	**0.694**	0.138	-0.016
EFF6	0.007	**0.625**	0.168	0.070
EFF3	0.095	**0.384**	0.125	0.156
SYS2	-0.006	-0.010	**0.844**	-0.010
EFF5	0.038	0.011	**0.816**	-0.100
EFF7	-0.013	0.111	**0.748**	-0.022
SYS3	0.061	-0.036	**0.530**	0.161
SYS4	0.058	0.113	**0.380**	0.193
SYS1	0.008	0.229	**0.322**	0.194
PRI2	-0.011	0.028	-0.012	**0.894**
PRI3	0.014	-0.031	0.005	**0.870**
PRI1	0.006	0.057	0.060	**0.767**
Eigen value	4.065	3.133	2.559	2.404
α-value	0.915	0.889	0.859	0.899

1) Loadings greater than 0.35 were considered significant
The four-factor solution explains 61.1% of the variation in the data.

6.2 Confirmatory factor analysis results

The next step in the data analysis phase of the study was to subject the three measurement models (Tables 3, 4 and 5) to a confirmatory factor analysis.

The initial analysis suggested that the assumption of normality does not hold for this data set (the relative kurtosis was 1.8). As a result, the Robust Maximum Likelihood estimation method was used for the confirmatory factor analysis, as recommended by Jöreskog and Sörbom [2003].

For comparative purposes, the original E-S-QUAL instrument (all 22 items measuring four factors) as reported by Parasuraman et al. [2005] was also subjected to a confirmatory factor analysis. The results are reported in Table 6.

Table 6 Comparative fit indices for four meeasurement models

CFA MODEL	χ^2	df	RMSEA	GFI	AGFI	ECVI
Original Parasuraman et al. model	2675.70	203	0.072	0.83	0.79	2.26
Four-factor model – Table 5	2239.35	203	0.067	0.85	0.82	1.96
Five-factor model – Table 4	1575.33	179	0.056	0.90	0.87	1.29
Six-factor model – Table 3	996.34	174	0.040	0.94	0.91	0.81

The fit statistics reported in Table 6 clearly show that the six-factor model fits the data better than any of the other configurations.

6.3 Nested model comparisons

To further investigate the difference between the Parasuraman et al. [2005] four-factor configuration of electronic service quality and our proposed six-factor model, a Chi-square difference test for nested models was conducted. To be able to use the nested model technique suggested by Satorra [2000], item FUL7 had to be included in our model as well. As this item (FUL7) loaded on both the *Delivery* factor and the *Reliability* factor when the data were factor-analyzed, we considered this item as a measure of both in two separate models. Thus in Table 7, item FUL7 is regarded as a measure of *Delivery* in Model A, and a measure of *Reliability* in Model B. The null hypothesis is based on the original Parasuraman et al. [2005] four-factor model.

The null hypothesis is:

H0: Electronic service quality is measured by four underlying dimensions

Table 7 Results of the satorra-bentler chi-square difference test:four-factor model vs six-factor model A[1]

MODEL	χ^2	df	TRD/df	Significance
Original Parasuraman et al. model	3088.39	203		
Six-factor model – A	1370.28	194	553.65/9	$p < 0.000$

1. Chi-square value is the Normal Theory Weighted Least Square value
   TRD = Satorra-Bentler scaled Chi-square difference test statistic value
   In Model A, item FUL7 is a measure of the latent variable RELIABILITY, and in Model B, a measure of the latent variable DELIVERY

The results reported in Tables 7 and 8 clearly show that the null hypothesis that electronic service quality is measured by four underlying dimensions must be rejected, irrespective of whether item FUL 7 is regarded as a measure of the latent variable *Reliability* (Model A) or as a measure of the latent variable *Delivery* (Model B). In other words, for this data set, the six-factor configuration of electronic service quality is superior to the four-factor configuration.

Table 8 Results of the satorra-bentler chi-square difference test:four-factor model vs six-factor model B[1]

MODEL	χ^2	df	TRD/df	Significance
Original Parasuraman et al. model	3088.39	203		
Six-factor model - B	1382.61	194	457.81/9	$p < 0.000$

1. Chi-square value is the Normal Theory Weighted Least Square value
   TRD = Satorra-Bentler scaled Chi-square difference test statistic value
   In Model A, item FUL7 is a measure of the latent variable RELIABILITY and in Model B, a measure of the latent variable DELIVERY

To assess the construct validity of the alternative six-factor configuration of electronic service quality as measured by E-S-QUAL, the next step was to assess its convergent validity.

6.4 Convergent validity

One way to assess the convergent validity of an instrument is to assess factor loadings on individual items [Parasuraman et al. 2005: 220]. The loadings of the six-factor E-S-QUAL model for the exploratory factor analysis (Table 3) and the confirmatory factor analysis loadings (Table 9) confirm the convergent validity of the six-factor model.

Table 9 Factor loadings of the six-factor CFA

	Efficiency	**Delivery**	**Privacy**	**Speed**	**System availability**	**Reliability**
EFF1	0.750					
EFF2	0.751					
EFF3	0.628					
EFF4	0.781					
EFF6	0.803					
EFF8	0.833					
FUL1		0.866				
FUL2		0.919				
FUL3		0.957				
PRI1			0.841			
PRI2			0.901			
PRI3			0.855			
EFF5				0.796		
EFF7				0.826		
SYS2				0.829		
SYS1					0.700	
SYS3					0.740	
SYS4					0.739	
FUL4						0.633
FUL5						0.731
FUL6						0.821

6.5 Predictive validity

"Predictive validity" is the extent to which an individual's future level of some variable can be predicted by his or her performance on a current measure of the same or a different variable. For instance: Will a measure of an attitude predict future purchases? [Tull and Hawkins 1993: 318]. To assess the predictive validity of the six-factor model of electronic service quality, each one of the six factors was correlated with a measure of future purchases on the Website ("I will continue to buy on the XXX Website in the future"). Table 10 shows that all the correlations were in the predicted direction, and all the correlations were statistically significant ($p < 0.001$). These results lend support to the predictive validity of the six-factor model of electronic service quality.

Table 10 Correlation analysis results:predictive validity(1)

	Loyal	**Efficiency**	**Fulfillment**	**Privacy**	**Speed**	**System**	**Reliability**
Loyal	1.000						
Efficiency	0.484	1.000					
Fulfillment	0.481	0.433	1.000				
Privacy	0.459	0.524	0.354	1.000			
Speed	0.395	0.625	0.355	0.440	1.000		
System	0.454	0.597	0.364	0.527	0.603	1.000	
Reliability	0.560	0.507	0.679	0.474	0.381	0.477	1.000

1) All correlation coefficients significant at $p < 0.001$

6.6 Nomological validity

The "nomological validity" of a construct is assessed by investigating the relationships of the construct with other constructs in a nomological net. The relationships in the nomological net are based on a theoretical (causal) model for the constructs involved. Although this is often assessed by means of a correlation or regression analysis, these techniques do not allow for formal testing of the nomological net (theory), and they do not incorporate measurement errors for the latent constructs of the nomological net [Steenkamp and Trijp 1991]. On the other hand, Structural Equation Modeling with Latent Variables does allow for measurement error and a formal test of the nomological net. Consequently, Structural Equation Modeling is a powerful statistical tool to assess the nomological validity of a construct.

In this study, the nomological net that was tested was based on empirical evidence such as the PIMS studies of the 1990's, scholarly research [Rust and Zahorik 1993; Sirohi, Mclaughlin and Wittink 1998; Cronin, Brady and Hult 2000], and on the anecdotal evidence provided by theorists [Heskett, Jones, Loveman, Sasser, and Schlesinger 1994; Oliver 1997; Bateson and Hoffman 1999: 290], and depicted in Figure I. The bulk of evidence from all these sources suggests that effective service delivery will influence value perceptions that will enhance customers' long-term loyalty intentions.

We decided to subject the six-factor E-SQUAL model to one final empirical assessment to assess its nomological validity, by testing the theoretical model depicted in Figure I (see Appendix B for the items used to measure these variables). This model is similar to Parasuraman et al.'s [2005] assessment of nomological validity.

As proposed by Steenkamp and Trijp [1991], LISREL for Windows [Jöreskog and Sörbom 2003] was used to fit the model depicted in Figure I to the data, to avoid the limitations associated with correlation and regression analyses.

The fit indices show that the six-factor model to assess nomological validity are: $\chi^2 = 1102.29$; df = 355; RMSEA = 0.0387; GFI = 0.913; AGFI = 0.894; NFI = 0.989; ECVI = 1.493. These indices (Table 11) suggest a close fit to the data for the six-factor model, and provide strong empirical support for the nomological validity of the six-factor model.

For comparative purposes, the four-factor model proposed by Parasuraman et al. [2005] was also modeled as antecedent of Value perceptions and Loyalty. The indices of this model are also shown in Table 11. Again, the six-factor model seems to offer a superior fit of the model to the data, which suggests better nomological validity.

Table 11 Comparative fit indices for the measurement models

CFA MODEL	χ^2	**df**	**RMSEA**	**GFI**	**AGFI**	**NFI**	**ECVI**
Six-factor model	1102.29	355	0.0387	0.913	0.894	0.989	1.493
Four-factor model	2338.93	394	0.0592	0.833	0.803	0.979	3.109

The estimates of the path coefficients are shown in Table 12. Although the relationships among the various latent constructs were not the purpose of the analysis, it must be noted that Reliability is the strongest predictor of Value perceptions.

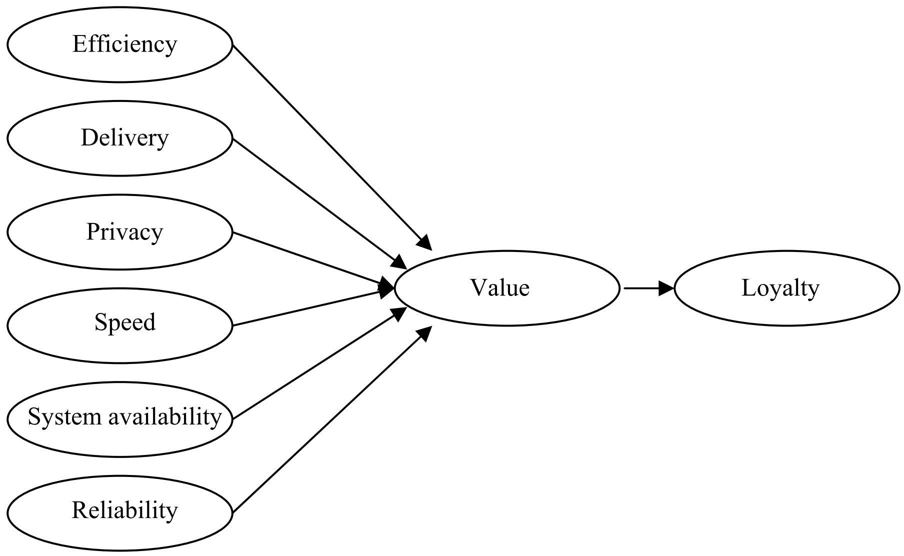

Figure 1 Nomological NET

Table 12 Nomological assessment: parameter estimates for six-factor model in Figure 1

PATH	**PARAMETER ESTIMATE [t-value]**
VALUE → LOYALTY	0.731 [17.12]***
EFFICIENCY → VALUE	0.228 [4.18]***
DELIVERY → VALUE	-0.043 [-0.748]
PRIVACY → VALUE	0.090 [2.36]*
SPEED → VALUE	0.107 [1.57]
RELIABILITY → VALUE	0.594*** [7.37]

*** = $p < 0.001$
** = $p < 0.01$
* = $p < 0.05$

7. Conclusion

There can be no doubt that a need exists to effectively measure the quality of service offered by Internet retailers. It is also true that the earlier efforts of the pioneers studying electronic service quality have, unfortunately, demonstrated some serious if not fatal limitations. To date, the most effective scale to measure the quality of service offered by Internet retailers is the E-S-QUAL scale developed by Parasuraman et al. [2005].

However, our analyses have shown fairly conclusively that the four-dimensional configuration of electronic service quality proposed Parasuraman et al. [2005] does not fit *this data set* as well as the six-factor configuration does. However, this conclusion may not be valid for other data sets. Clearly more research is needed before any final conclusions on the dimensionality of electronic service quality can be reached.

From a managerial perspective, we believe that the E-S-QUAL instrument is an excellent instrument to measure electronic service quality. We also believe, nevertheless, that when the scale is used for managerial purposes by future researchers, they must examine and evaluate the dimensionality that may have been captured by their data.

Academic researchers need to do more empirical and psychometric assessments of E-S-QUAL. In particular, the dimension of *Reliability* needs to be explored further. In our data set, *Reliability* has emerged as a separate dimension measured by three items:

The XXX site sends out the items ordered
The XXX site has in stock the items the company claims to have
The XXX site is truthful about its offerings.

These items suggest a narrower definition of reliability that borders on honesty/truthfulness compared to the broader definition initially proposed by Zeithaml, Parasuraman and Malholtra [2002: 364] to reflect "other meanings" such as "technical reliability" and "the proper functioning of the site."

Future researchers may have to consider a more appropriate operationalization of the *Reliability* dimension.

Reliability may have to be operationalized as a Website/electronic retailer that can be trusted (to send out the correct items) and is truthful (has in stock the items the company claims to have; accurate product presentations). It may be appropriate to consider the possibility that *Reliability* could have two sub-dimensions. The first could be the trust/truthful aspects described above, and the second could be the reliability of the site itself (the ability to connect to the site when one desires; access). The problem with the second sub-dimension of reliability would be the distinction between "reliability of the Website" on the one hand, and possible overlap with aspects related to *Efficiency, System availability* and *Fulfillment* on the other hand. In fact, Zeithaml et al. [2002: 364], in their earlier attempt to identify the dimensions of electronic service quality, failed to make a proper distinction between *Fulfillment* and *Reliability*.

We believe that *Reliability* as a separate dimension is of particular importance, given the fact that it proved to be the strongest predictor of customer satisfaction and the second strongest predictor of loyalty in Wolfinbarger and Gilly's [2003] study. In the present study, *Reliability* was the strongest predictor of Value perceptions (Table 12).

We also believe that the possibility of operationalizing *Reliability* as trust/trustworthiness/honesty along the lines of the assurance/trust dimension originally proposed by Parasuraman [2005: 219] is the most appropriate option.

It is also possible that *Reliability* is a higher-order factor that should capture the reliability of the Website itself as well as the reliability of the delivery system.

Our results have also shown that if the reliability-related items are removed from the *Fulfillment* dimension, the remaining items all refer to delivery. If *Delivery* is indeed a separate electronic service quality dimension, its operationalization also needs to reconsidered.

It is clear that some work still needs to be done on the dimensionality of the construct "electronic service quality."

This conclusion is confirmed by the fact that some of the items expected to measure the *Efficiency* dimension which referred to the speed of the Website, consistently loaded on a separate factor. These items were: "The Web pages load fast" (EFF5); "This site enables me to get on quickly" (EFF7); and "This site launches and runs right away" (SYS 2). Accordingly, this factor was labeled *Website* s*peed* in this study. This dimension has not received much attention in the literature, as it has consistently been seen as part of other dimensions such as *Responsiveness* or *Ease of navigation* [Parasuraman et al. 2005: 219] rather than as a separate dimension. If follow-up studies do confirm *Website speed* as an electronic service quality dimension, there is an obvious need to operationalize this dimension as well.

In summary: This study has shown that the E-S-QUAL instrument is a valid and reliable instrument to measure service quality in an electronic shopping environment, provided its dimensionality is properly analyzed. We also believe that the results of this study suggest that more scholarly research is needed to re-assess the dimensionality of the electronic service quality construct.

REFERENCES

Armstrong, J.S. and T.S. Overton, "Estimating non-response bias in mail surveys," *Journal of Marketing Research*, 14, 396-402, 1977.

Bates, K., H. Bates, and R. Johnston, "Linking Service to Profit: The Business Case for Service Excellence," Proceedings of the QUIS8 Conference: Quality in Service: Crossing Boundaries, (pp. 86-95), Victoria, Canada, 2002.

Bateson, J.E.G. and K.D. Hoffman, *Managing Services Marketing*, Dryden Press, 1999.

Cronin, J.J., M.K. Brady, and G.T. Hult, "Assessing the Effects of Quality, Value and Satisfaction on Consumer Behavioral Intentions in Service Environments", *Journal of Retailing*, Vol. 76, No. 2: 193-218, 2000.

Finn, A. and U. Kayande, "Reassessing the Construct Validity of Website Quality", Proceedings of the QUIS8 Conference: Quality in Service: Crossing Boundaries, Victoria, Canada, 141-143, 2002.

Floh, A. and H. Treiblmaier, "What Keeps the e-banking Customer Loyal? A Multigroup Analysis of the Moderating Role of Consumer Characteristics on e-loyalty in the Financial Service Industry," *Journal of Electronic Commerce Research,* Vol. 7, No. 2: 97-110, 2006.

Garvin, D.A., *Managing Quality*, New York: The Free Press, 1988.

Heskett, J.L., T.O. Jones, G.W. Loveman, W.E Sasser Jr., and L.A. Schlesinger, "Putting the Service-Profit Chain to Work", *Harvard Business Review*, Vol. 72, March-April: 164-174, 1994.

Jöreskog, K.G. and D. Sörbom, "*LISREL 8.54 for Windows*", Lincolnwood, IL: Scientific Software International, Inc., 2003.

Lennon, R. and J. Harris, "Customer Service on the Web: A Cross-Industry Investigation," *Journal of Targeting, Measurement and Analysis for Marketing*, Vol. 10, No. 4: 325-338, 2002.

Loiacono, E., R.T. Watson, and D. Goodhue, "*WebQual^{TM}: A Web Site Quality Instrument*", Paper read at the American Marketing Association's Winter Marketing Conference, Austin, Texas, 2002.

Montoya-Weiss, M., G.B. Voss, and D. Grewal, "Determinants of Online Channel Use and Overall Satisfaction with a Relational, Multichannel Service Provider", *Journal of the Academy of Marketing Science*, Vol. 31, No. 4: 448-458, 2003.

Novak, T. P., D.L. Hoffman, and Y. Yung, "Measuring the Customer Experience in Online Environments: A Structural Modeling Approach", *Marketing Science*, Vol. 19, No. 1, 22-42, 2000.

Nunnally, J.C. and I.H. Bernstein, *Psychometric Theory*, New York: McGraw-Hill, 1994.

Oliver, R.L., *Satisfaction*: *A Behavioral Perspective on the Consumer*, New York : McGraw-Hill International, 1997.

Parasuraman, A., V.A. Zeithaml, and L.L. Berry, "SERVQUAL: A Multiple-Item Scale for Measuring Service Quality," *Journal of Retailing*, Vol. 64, No. 1, 12-40, 1988.

Parasuraman, A., V.A. Zeithaml, and A. Malholtra, "E-S-QUAL: A Multiple-Item Scale for Assessing Electronic Service Quality," *Journal of Service Research*, Vol. 7 No. 3: 213-235, 2005.

Rust, R.T. and A.J. Zahorik, "Customer Satisfaction, Customer Retention, and Market Share", *Journal of Retailing*, Vol. 69, Summer: 193-215, 1993.

Rust, R.T., A.J. Zahorik, and T.L. Keiningham, "Return on Quality: Making Service Quality Financially Accountable", *Journal of Marketing*, Vol. 59, April: 58-70, 1995.

Satorra, A., Scaled and Adjusted Restricted Tests in Multi-sample Analysis of Moment Structures. In: Heijmans, D.D.H., Pollock, D.S.G. and A. Satorra, (eds.), *Innovations in Multivariate Statistical Analysis*. A Festschrift for Heinz Neudecker (pp. 233-247). London: Kluwer Academic Publishers, 2000.

Sirohi, N., E.W. McLaughlin, and D.R. Wittink, "A Model of Consumer Perceptions and Store Loyalty Intentions for a Supermarket Retailer", *Journal of Retailing*, Vol. 74, No. 2: 223-245, 1998.

Steenkamp, J-B.E.M. and H.C.M. Trijp, "The use of LISREL in validating marketing constructs", *International Journal of Research in Marketing*, Vol. 8: 283-299, 1991.

Szymanski, D.M. and R.T. Hise, "e-Satisfaction: An Initial Examination", *Journal of Retailing*, Vol. 76, No. 3: 309-322, 2000.

Tull, D.S. and D.I. Hawkins, *Marketing Research: Measurement and Method,* 6th edition, New York: Macmillan, 1993.

Voss, C., "Rethinking Paradigms of Service – The Impact of the Web", Conference Proceedings of the 7th International Conference of the Operations Management Association, Ghent, Belgium, 2000.

Wang, Y., T. Tang and J.E. Tang, "An Instrument for Measuring Customer Satisfaction Toward Websites that Market Digital Products and Services," Journal of Electronic Commerce Research, Vol. 2, No. 3: 89-102, 2001.

Wolfbinbarger, M. and M.C. Gilly, "e-TailQ: Dimensionalizing, Measuring, and Predicting Retail Quality", *Journal of Retailing*, Vol. 79, No. 3: 183-98, 2003.

Yoo, B. and N. Donthu, "Developing a Scale to Measure the Perceived Quality of an Internet Shopping Site (SITEQUAL)", *Quarterly Journal of Electronic Commerce*, Vol 2, No. 1: 31-45, 2001.

Zeithaml, V.A. "Consumer Perceptions of Price, Quality, and Value: A means-end Model and Synthesis of Evidence", *Journal of Marketing*, Vol. 52, July: 2-22, 1988.

Zeithaml, V.A., "Service Quality, Profitability, and the Economic Worth of Customers: What We Know and What We Need to Learn", *Journal of the Academy of Marketing Science*, Vol. 28, No.1: 67-85, 2000.

Zeithaml, V.A., A. Parasuraman, and A. Malhotra, "Service Quality Delivery Through Websites: A Critical Review of Extant Knowledge", *Journal of the Academy of Marketing Science*, Vol. 30, No. 4: 362 – 375, 2002.

APPENDIX A

FACTOR CORRELATIONS FOR ROTATED FACTORS: SIX-FACTOR SOLUTION

	Efficiency	Delivery	Privacy	Speed	System availability	Reliability
Efficiency	1.000					
Delivery	0.368	1.000				
Privacy	0.504	0.339	1.000			
Speed	0.619	0.306	0.443	1.000		
System avail.	0.504	0.301	0.525	0.592	1.000	
Reliability	0.472	0.674	0.484	0.346	0.415	1.000

FACTOR CORRELATIONS FOR ROTATED FACTORS: FIVE-FACTOR SOLUTION

	Fufilment	Efficiency	Privacy	Speed	System availability
Fufilment	1.000				
Efficiency	0.443	1.000			
Privacy	0.419	0.527	1.000		
Speed	0.284	0.572	0.389	1.000	
System availability	0.403	0.531	0.560	0.528	1.000

FACTOR CORRELATIONS FOR ROTATED FACTORS: FOUR-FACTOR SOLUTION

	Fufilment	Efficiency	Privacy	System availability
Fufilment	1.000			
Efficiency	0.475	1.000		
Privacy	0.418	0.664	1.000	
System availability	0.431	0.516	0.508	1.000

APPENDIX B

PERCEIVED VALUE

XXX's delivery cost is reasonable
The products on the XXX Website are reasonably priced
The products on the XXX Website are of good quality
XXX offer products that represent good value
The products offered by XXX function the way they are supposed to

LOYALTY

I am a loyal XXX client
I recommend XXX to others when I get the opportunity
I often praise XXX when talking to others

APPENDIX C
REFINED E-S-QUAL INSTRUMENT

(Previous codes in brackets)

EFFICIENCY

EFF 1	The XXX website makes it easy for me to find what I need / what I am looking for
EFF 2	XXX makes it easy to move around on the website
EFF 3	XXX enables me to complete a transaction quickly
EFF 4	The information on the XXX website is well organized
EFF 6	XXX is easy to use
EFF 8	XXX is well organized

DELIVERY

DEL 1	XXX delivers orders when promised / within the specified time (FUL 1)
DEL 2	XXX delivers ordered items within a reasonable time frame (FUL 2)
DEL 3	I receive my XXX orders in good time (FUL 3)

PRIVACY

PRI 1	XXX protects information about my web-shopping behaviour.
PRI 2	XXX does not share my personal information with others.
PRI 3	XXX protects my credit card information.

SPEED

SPEED 1	The XXX website pages load fast (EFF 5)
SPEED 2	I am able to access the XXX website quickly (EFF 7)
SPEED 3	XXX launches and runs right away (SYS 2)

SYSTEM AVAILABILITY

SYS 1	XXX is always available for business (SYS1)
SYS 2	XXX does not crash (SYS 3)
SYS 3	Pages on XXX do not freeze after I entered my order information (SYS 4)

RELIABILITY

REL 1	XXX sends out the correct items I ordered (FUL 4)
REL 2	XXX has in stock the items that it claims to have (FUL 5)
REL 3	XXX is truthful about its offerings (FUL 6)