

Assessing e-service quality: the current state of E-S-QUAL

Luc Honore Petnji Yaya (*)

University of Girona, Departament d'Organitzacio',
Gestio' Empresarial i Desenvolupament de Producte,
Av. Lluís Santalo', s/n,
17071 Girona, Spain
Tel: +34 67 118 08 85
E-mail: arinopetnji@hotmail.com

Frederic Marimon

Universitat Internacional de Catalunya
Faculty of Economics and Social Sciences
Immaculada 22
08017 Barcelona, Spain
Tel: +34 93 254 18 00
E-mail: fmarimon@cir.uic.es

Marti Casadesus Fa

University of Girona, Departament d'Organitzacio',
Gestio' Empresarial i Desenvolupament de Producte,
Av. Lluís Santalo', s/n,
17071 Girona, Spain
Tel: +34 972 41 8286
E-mail: marti.casadesus@udg.edu

(*) corresponding author

Abstract

Purpose: this study seeks to holistically undertake a comprehensive review of the current state of the E-S-QUAL scale including methodology used, suggestions, and limitations associated with the adoption of the scale.

Design/methodology/approach: The data collection process was done through an exhaustive search of the largest well-known databases and search engines such as ScienceDirect, Emerald Insight, EBSCOhost, ABI/INFORM and Google Scholar

Results: The dimensional structure of E-S-QUAL appears to be very unstable, even within a given sector. However, the general results revealed that the scale is effective in capturing the core e-service quality since it has been to a certain extent successfully replicated and applied in 11 countries and a variety of e-service settings. The dimensions of Efficiency, System Availability and Privacy appear consistently in the various models regardless of the type of e-service. In contrast, the dimension of Fulfillment seems not to be generic but specific to particular e-service contexts such as web sites selling physical goods

Management implication: Providing the scale dimensional structure appears to be very unstable, both scholars and practitioners must assess the underlying factor structure of their data before drawing any conclusions from their study. Managers should be careful in applying the Fulfillment dimension in contexts that have few elements in common with industry-specific which the sites did not promised about order delivery and item availability are fulfilled.

Keywords: E-S-QUAL, Electronic service quality, Scale development, dimensionality

INTRODUCTION

In recent years, the internet retailing market grows at an exponential rate and online customers are becoming more and more clued-up. Given that online customers are increasingly eager to share their experiences through social media. Subsequently, the smartest online companies have started reckoning how to tie together the potential of the Web by delivering superior service quality. Given that those who succeed in doing so could lock in many profitable relationships at the expense of slow-footed rivals (Reichheld and Scheffer, 2000). As the novelty of using electronic services wears off, consumers become less willing to tolerate poor service quality (Parasuraman et al., 2005). Therefore in the context of electronic services, quality may well be the most important determinant of long-term success (Zeithaml et al., 2002). It is widely accepted that to deliver superior electronic service quality (e-SQ), managers of companies with web presences must first understand how consumers perceive and evaluate e-SQ. Consequently, an increasing number of research studies have focused on understanding online service quality and many different scales have been developed. These include WebQual (Barnes and Vidgen 2002), WebQualTM (Loiacono et al 2000), SITEQUAL (Yoo and Donthu 2001), e-SERVQUAL (Zeithaml et al 2002), and eTailQ (Wolfenbarger and Gilly 2003), eTransQual (Bauer et al 2006) and PeSQ (Cristobal et al 2007). Overall these studies provided useful insights into criteria that are relevant for evaluating electronic Service Quality (e-SQ).

Furthermore, although there is general agreement that service quality is crucial in the online environment, significant confusion in the literature seems to exist with regard to the number and nature of the dimensions influencing consumer perceptions of e-SQ (Ladhari 2010). Especially in view of the fact that many of the prior service quality measures do not take into account the entire process including the pre-purchase and after-purchase stages (Parasuraman et al., 2005).

Nonetheless, one study (Parasuraman et al., 2005) may be considered to constitute the most thoroughly conducted research on the topic, producing a service quality scale (E-S-QUAL) and a recovery scale (E-RecS-QUAL) that appear to capture the general domain of e-SQ with considerable accuracy and to have approached the issue of the shopping or buying interaction in its totality (Boshoff 2007; Yen and Lu 2008; Akinci et al 2010; Marimon et al 2010). Still, there are many different types of website, and the various different phases in the process of developing the E-S-QUAL scales focused exclusively on websites selling physical products, in contrast to purely service-oriented sites such as those offering financial or information services. Therefore, the question that then arises is how stable is the E-S-QUAL scales across different industries and different cultures? This study seeks to undertake holistically a comprehensive review of the current state of the E-S-QUAL scale in term of the methodology used, new suggestions to be made, and the limitations associated with its adoption.

LITERATURE REVIEW:

E-S-QUAL development

Unlike the traditional service quality literature, the number of studies on e-SQ is still at its preliminary phase both from theoretical and empirical perspectives (Akinci et

al 2010). According to Zeithaml et al. (2000), e-SQ is the extent to which a web site facilitates efficient and effective shopping, purchasing, and delivery of products or services. Numerous researchers have sought to find the global services attributes that contribute the most significantly to relevant e-SQ assessment. Yet, many of the proposed scales to evaluate websites do not provide a comprehensive evaluation of the service quality of the website. Given that these previous studies neither define the exact domain of their quality construct nor provide a clear-cut definition of electronic services (Praeg and Spath, 2010). Secondly, most of the scales that have been developed do not provide a comprehensive evaluation of the website and have not been properly examined in term of their psychometric properties or of possible improvements that may be required (Ladhari, 2010). Thirdly, most of these studies often conceptualize electronic service quality as being identical to web interface design quality (Zeithaml et al., 2002; Bauer et al., 2006).

More recently, in an attempt to address all the concerns mentioned above, Zeithaml et al (2000) systematically investigated, analyzed and identified a number of website features at the perceptual attribute level and categorized them in term of an e-SERVQUAL scale. On the basis of a comprehensive review and synthesis of the extant literature on e-SQ, e-SERVQUAL was painstakingly examined by Zeithaml et al (2002). The results produced a more comprehensive conception of e-service quality in terms of seven dimensions that evaluates the whole service encounter, including both the transaction and the post-transaction process. Based on this explorative study, Parasuraman et al., (2005) have recently refined and validated the instrument. The seven dimensions proposed by Zeithaml et al (2002) were divided into two different scales. The first scale called “E-SQUAL” addresses core service quality aspects and consists of 22 items grouped into four quality factors (efficiency, fulfilment, system availability and privacy).

The second scale, entitled E-RecS-QUAL, was designed to be appropriate when customers had non-routine encounters to measure the effectiveness of handling problems and return, compensation for problems cropped up and availability for assistance. The E-RecS-QUAL scale is composed of 11 items and grouped in three quality dimensions (responsiveness, compensation and contact). Both scales were subjected to exploratory and confirmatory factor analyses and provided an important step forward in the conceptualization of e-service quality, as they address and resolve many of the concerns about previous scales (Connolly et al., 2010). However, the question remains about whether the scale is an appropriate tool to capture the entire field of e-SQ and whether it is universally applicable. Given that the scale development focused only on websites that sold physical products.

RESEARCH METHOD AND DATA COLLECTION OF THIS STUDY

The data collection process was carried out through an exhaustive search of the leading well-known databases and search engines such as ScienceDirect, Emerald Insight, EBSCOhost, ABI/INFORM and Google Scholar. The searching keywords comprised: E-SQUAL; website or web site or online or electronic service quality evaluation or assessment or measurement, etc. Acknowledging that the direct search may lead to some limited results, we also searched the references of the articles that were retrieved. This process yielded a total of 21 published articles since the appearance of E-SQUAL six

years ago. Noteworthy that some articles could contain the analysis of e-SQ in more than one sector or country in such case we considered each sector examined or each sample per country to be the sample unit, which gave a total of 25 observations. Table 1 summarized and categorized the review of the paper and detailed discussions of these subjects are presented below.

Table 1 Selected studies, summary of methodology use and sample characteristics

Author(s)/Year	Country	Domain of measure	Sample	Items loading ^(a)	Internal consistency ^(b)	Final number of dimensions (number of items)
Petnji et al., 2011	Spain	E-banking service quality	428 consumer of e-banking	0.68 to 0.92	0.88 to 0.90	3 dimensions = Efficiency (7), System availability(7), Privacy (3)
Chang 2011	Taiwan	Electronic service quality in a pure service	123 experienced users of Yahoo auction and Free e-services	N/A	N/A	7 dimensions = Efficiency (1), System availability (1), Fulfillment (1) Privacy (1), Responsiveness (1), Compensation (1) and Contact (1)
Rafiq et al., 2011	UK	Internet retail service quality	491 e-grocery shoppers	0.64 to 0.89	0.83 to 0.96	3dimensions = Efficiency (8), System availability(4), Fulfillment(6)
Meng and Mummalaneni, 2010	USA and China	Service quality in online shopping	147 Chinese and 228 African American students consumers of online shopping	0.51 to 0.88 for both AA and Chinese	0.73 to 0.94 for AA and from 0.69 to 0.80 for Chinese	7 dimensions both AA and China = Efficiency(8), Fulfillment (7), System availability (4), Privacy(3), responsiveness (5), Compensation (3) and contact(3)
Akinci et al., 2010	Turkey	Electronic service quality in a pure service	2017 consumer of e-banking	0.66 to 0.88 after erasing 12 items ranges from 0.64 to 0.85	0.84 to 0.92 after erasing 12 items 0.83 to 0.87	4 dimensions = Efficiency (2), System availability(2), Fulfillment (2), Privacy(2)
Marimon et al., 2010	Spain	Service quality in online shopping	131 Actual online purchases	0.47 to 0.90	0.75 to 0.88	4 dimensions = Efficiency (3), System availability(4), Fulfillment (5), Privacy(3)
Fuentes-Blasco et al., 2010	Spain	E-service quality in retailing	191 consumers of online travel agencies and online book store	0.63 to 0.94	0.80 to 0.91	5 dimensions =System Efficiency (5), System availability (3), Fulfillment (4) Security (3), Responsiveness/Contact (3)
Connolly et al., 2010	Ireland	Service quality in tax filing and collection system	6661 users of online taxation	0.50 to 0.90	0.79 to 0.90	5 dimensions = Efficiency (8), Ease of Completion (3), System availability (4), Privacy (2) and contact (3)
Yang et al 2010	Taiwan	e-Service quality in online shopping	234 experience online shoppers	0.55 to 0.91	0.84 to 0.91	5 dimensions = Efficiency (8), System availability(4), Fulfillment (7), Privacy(3), Enjoyment (5)
Chiou et al., 2009	Taiwan	E-service quality for buyer and seller	221 Online auction buyers	0.46 to 0.97 Auction website and 0.62 to 0.96 Seller website	0.79 to 0.97	5 dimensions for Online Auction = Efficiency (10), System availability (4), Privacy (3), Contact (3) compensation (3); 4 dimensions for seller Auction = Fulfillment (7),

						Responsiveness (5) Compensation (3) Contact (3)
Herington and Weaven, 2009	Australia	Online banking service	200 consumers of online banking	0.77 to 0.95	0.80 to 0.96	4 dimensions = Personal needs (3); Site organization (4); User Friendliness (4) Efficiency of web site (3)
Lin et al., 2009	Taiwan	Electronic service quality of the HR service agency	309 online job seekers	N/A	0.77 to 0.93 for performance and from 0.80 to 0.9260 for Importance	7 dimensions = dimensions: Efficiency(27), Fulfillment (7), System availability (4), Privacy(4), responsiveness (4), Compensation (2) and contact(4)
Sun et al., 2009	China	Online banking service	276 consumer of online banking	N/A	0.72 to 0.86	5 dimensions = Efficiency (8), System availability(4), Fulfillment (7), Privacy(3), Trust/Assurance (6)
Yen and Lu, 2008	Taiwan	E-service quality for buyer and seller	619 Online auction buyers	0.68 to 0.90	0.87 to 0.94	3 dimensions for Online Auction = Efficiency (4), System availability (4), Privacy (4); 3 dimensions for seller Auction = Contact (4) Fulfillment (4), Responsiveness (4)
Sahadev and Purani, 2008	India	Electronic service quality in online job portals	350 online job seekers	0.68 to 3.92	0.71 to 0.80	4 dimensions = Efficiency (8), System availability(4), Fulfillment(5), Privacy(3)
Boshoff, 2007	South Africa	Online retailing service quality	1409 retailing consumers	0.62 to 0.95	0.76 to 0.93	6 dimensions = Efficiency (6), Delivery (3), Privacy (3), Speed (3), System Availability (3), Reliability(3)
Mekovec et al., 2007	Croatia	E-service quality in retailing	28 raters	N/A	N/A	Only the dimension of Efficiency was investigated
Wu and Ding 2007	USA	Electronic service quality in electronic retailing	276 e-tailers	0.84 to 0.93	0.93 to 0.97	4 dimensions = Efficiency (8), System availability(4), Fulfillment (7), Privacy(3)
Yang & Tsai, 2007	Taiwan	Electronic service quality in online shopping	278 e-shopper	0.66 to 0.89	0.863 to 0.937	4 dimensions = Efficiency (8), System availability(4), Fulfillment(7), Privacy(3)
Kim et al., 2006	USA	Apparel website quality	(No survey) instead collection of 111 apparel websites	N/A	N/A	9 dimensions = Efficiency (15), Fulfillment (6), System availability(2), Privacy(8), Responsiveness(2), Contact(2), Personalization (18), information (7), and graphic style (21)

Parasuraman et., 2005	USA	E-service quality in e-retailing	205 and 653 online store shopper	0.71 to 0.88 for Amazon; 0,68 to 0.91 for Walmart	0.85to 0.94 for Amazon and from 0,83 to 0,94 for Walmart	4 dimensions for both = Efficiency (8), System availability(4), Fulfillment (7), Privacy(3)
-----------------------	-----	----------------------------------	----------------------------------	---------------------------------------------------	----------------------------------------------------------	---------------------------------------------------------------------------------------------

(a) These represent the standardized loading estimates from CFA;

(b) Internal Consistency or Construct Reliability coefficients are represented by Croanbach's Alpha or Jorskog's p or Composite Reliabilit

RESULTS

Widespread use, methodology and sample characteristics

In order to assess robustness or identify limitations that may exist with respect to scale generalizability, the E-S-QUAL was tested in 11 countries and in different languages such as: English, Turkish, Chinese, Croatian, Taiwanese, Hindi, Spanish, Catalan, etc. (see Table 1). Moreover, the scale has been used to measure e-service quality in a variety of domains and types of service industries including: Sites offering music, books, DVDs, department stores, electronics, computers (Parasuraman et al 2005; Mekovec et al 2007; Wu and Ding 2007; Boshoff 2007; Marimon et al 2010). Whereas, other studies focus on specific sectors such as: online banking (Akinici et al., 2010; Herington and Weaven 2009; Petnji et al 2011), online Job Portals (Sahadev and Purani 2008), Online HR service (Lin et al 2009), online taxation filing service (Connolly et al 2010) online auction (Chiou et al 2009; Yen and Lu 2008), online travel agencies and online book store (Fuentes-Blasco et al 2010) and in pure online service (Chang 2011).

Given that E-S-QUAL is an existing theoretically supported scale; it is not a surprise that most of the studies listed in table 1 adopted a quantitative methodology. Furthermore, an online survey administration was most frequently used as the data collection method. However, Herington and Weaven (2009) used an off-line data collection, whereas Sahadev and Purani (2008) and Lin et al., (2009) used a combination of both online and off-line data collections in their studies. In general, several of the studies reviewed used limited sample sizes. However, the number of observations per scale item ranged from 5.95 (Marimon et al., 2010) to 333.05 (Connolly et al., 2010), thus, indicating that the sample size in each study at best exceeded the conventional requirement that around five observations per scale item are needed for conducting factor analyses (Hair et al., 1998). Additionally, the majority of studies used convenience sampling of the actual user in the chosen field. However, Mekovec et al (2007) used respondents from 2nd year university, whereas, Sahadev and Purani (2008) used a convenient sample of executives and final year students on three MBA courses. Similarly, Meng and Mummalaneni (2010) used Chinese and African American students.

Scales used

During the initial stage of their scale development process, Parasuraman et al (2005) used focus groups to understand respondents' reactions to alternative ways of phrasing scale items and anchors (Likert-type versus low or high performance anchors). On the basis of insights from focus groups, Parasuraman et al. (2005) adopted a 5-point Likert-scale (1 = strongly disagree, 5 = strongly agree) format for collecting responses. The studies reviewed showed that 66% of the studies adopted the same 5-point Likert-type scale format. However, Boshoff (2007) used a 6-point Likert-type. Moreover, in contrast to the majority, Yang & Tsai (2007), Yang et al (2010), Rafiq et al (2011) and Chang (2011) adopted a 7-point Likert scale. One of the justifications put forward is that the seven point scale extends the range and variability of responses and potentially increases the

reliability of the results. Obviously, some scholars advocate a 5-point scale where respondents can have a neutral middle point, whereas others prefer the respondents to be forced to select a negative or positive position with a 6-point scale. Nevertheless, Dawes (2008) observed that there are inconclusive results on the use of an even-point or neutral point. Still, the authors found that the 5- and 7-point scales produced exactly the same mean score as each other, once they were rescaled. Whereas, the 10-point format tended to produce slightly lower relative means than either the 5- or 7- point scales.

This review also observed that Kim et al (2006) avoided likert-type scales instead they extended the E-S-QUAL by using a coding guide. The majority of online service attributes were coded as “unavailable” “available in the text only” or “available both in in the text and with a picture”. In the same vein Mekovec et al (2007) advocated that the use of a likert-scale is a subjective evaluation of service quality attributes, given that; even the experts or trained evaluators may themselves differ and be inconsistent in their judgments when likert-scale are used for the assessment of various quality attributes. Thus, they redefine E-S-QUAL items in a way that replaces its likert type scales with check-lists that include specific observable elements for appraisal and a scoring procedure. In general, simulation studies and empirical studies have generally concurred that reliability and validity are improved by using 5- to 7-point scales rather than C-OAR-SE ones (those with fewer scale points). It is suffice to say that more finely graded scales do not further improve reliability and validity (Dawes, 2008).

Data analysis procedure for assessing factor structure (CFA, EFA)

Research on e-service quality has only recently started gaining momentum, and the main research question that all relevant studies try to address pertains to the factorial structure of the construct and measurement issues (Wolfenbarger & Gilly 2003). As a general rule, the dimensionality of the scale is assessed using exploratory factor analysis (EFA) and/or confirmatory factor analysis (CFA). EFA is used to uncover the underlying structure of a relatively large set of variables. The researcher's a priori assumption is that any indicator may be associated with any factor and factor loadings are used to discern the factor structure of the data. Several of the studies reviewed used EFA, including Parasuraman et (2005); Boshoff (2007); Yang & Tsai (2007); Akinci et al (2010); Herington and Weaven (2009); Marimon et al (2010); Fuentes-Blasco et al (2010); Connolly et al (2010); and Petnji et al (2011). Moreover, assuming that E-S-QUAL is an existing theoretical supported scale, studies such as Chiou et al (2009); Wu and Ding (2007); Sahadev and Purani (2008); Meng and Mummalaneni (2010); Yen and Lu (2008); Sun et al (2009); Yang et al (2010) and Rafiq et al (2011) only used CFA to see if the indicator variables load as predicted on the expected number of factors.

Even though EFA is often used before conducting CFA, some authors still criticized its use. Gerbing and Anderson (1988), for example, 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. In contrast, CFA allows researchers to compare several model specifications and to examine the invariance of a specific parameter in the factor solution. Hence given the limitations of EFA, researchers should use a combination of EFA and CFA (Ladhari 2010). This review also observed that some of the studies scales reported are problematic, since they did not applied either EFA or CFA to define the number of factors. For example Lin et al (2009) averred they

were provided 20 extra items from the HR service agency, whereas Chang (2011) used E-S-QUAL as a mixed initiative model for quality based e-service pricing. Both studies did not applied psychometric test to reject/confirm the items added in their representing factors.

Scale reliability

Generally, Scale internal consistency is often assessed by Cronbach's alpha or Composite reliability. Cronbach's alpha is widely believed to indirectly indicate the degree to which a set of items measures a single unidimensional latent construct. The interpretation of the composite reliability is similar to that of Cronbach's alpha, except that it also takes into account the actual factor loadings, rather than assuming that each item is equally weighted in the composite load determination (Sun et al., 2009). The studies reviewed in table 1 shows the preponderance of coefficient alpha, or that the Jorskog's value exhibited exceeded the conventional minimum of 0.7 (Nunally and Bernstein 1994), except in the case of the study conducted by Meng and Mummalaneni (2010), who used $\alpha = 0,69$. However, they exceeded the minimum satisfactory value of 0.6 (Malhotra 2004), thus demonstrating a high internal consistency and hence the reliability of each dimension presented. In general the findings validate the cohesiveness of the E-S-QUAL scale.

Scales convergent validity

It is the extent to which measures of constructs that theoretically should be interrelated to each other are in reality perceived to be related to each other on the same construct. The studies reviewed show that wide varieties of labels are used to describe the convergent validity of the measure. For example, Yang & Tsai (2007) ; Akinci et al (2010) ; Sun et al (2009) ; Marimon et al (2010) and Petnji et al (2011) adopted Fornell and Larker (1981) method whereby all indicator factor loadings should be significant and exceed 0.5, and AVE by each construct should exceed the variance due to measurement error for that construct (i.e. should exceed 0.5). Whereas Parasuraman et al (2005); Wu and Ding (2007); Rafiq et al (2011), and Boshoff (2007) adopted Hair et al (1998), by assessing factor loadings on individual items are all high and significant along with high coefficient alpha values. Other studies such as Chiou et al (2009); Fuentes-Blasco et al (2010), and Meng and Mummalaneni (2010) implemented Anderson and Gerbing (1988) by determining whether each indicator's estimated pattern coefficient on its proposed underlying construct is significant.

The factor loading pattern of E-S-QUAL in studies reviewed revealed some weakness in convergent validity, given that some items in these studies had a higher loading on dimensions that were different from those suggested for items by Parasuraman et al (2005). For example, Akinci et al. (2010) reported that only 8 out of 22 items (i.e. two for each dimension) loaded as expected. In addition, the item EFF5 "it loads its pages fast" was moved from the dimension of efficiency to system availability in the studies conducted by Petnji et al (2011); Marimon et al (2010) and Fuentes-Blasco et al (2010). This is not a surprise, since the intensified competition in the industry has forced most Internet providers to develop a broadband or fiber-optic version that offers more reliable, uncapped Internet connection speeds. Likewise, the results obtained by Rafiq et al. (2011) show that the dimensions of Efficiency and System Availability were too closely

correlated at the measurement level and therefore need more refinement to improve their discriminant validity. Therefore, they discarded FUL 7 as it loaded equally with FUL 1. Similarly, Boshoff (2007) study shows the items EFF5, EFF7 and SYS 2 were grouped under the dimension of Speed, and that the fulfillment dimensions split into two as follows: the items of FUL1 to 3 and FUL4 to 6 were congregated under the dimensions of Delivery and Reliability. Consequently, the author concluded that the 22 items were not as clearly illustrated by the 4 dimensions configuration of electronic service quality proposed by Parasuraman et al (2005) as by the six-factor configuration.

Scale Discriminant validity

This is the extent to which measures of constructs that theoretically should not be interrelated to each other are in fact observed not to be related to each other. As in the case of convergent validity, the studies reviewed show that different methods were used to describe the discriminant validity of the measurement process. For example, Yen and Lu (2008); Marimon et al (2010) and Petnji et al (2011) adopted the Fornell and Larcker (1981) method by examining whether Inter-factor correlations are less than the square root of the average variance (AVE). Chiou et al (2009) took the Smith and Barclay (1997) method that examine whether the confidence interval around the correlation between any two latent constructs does not include 1. Whereas, Boshoff (2007) followed Loiacono et al (2000) by comparing the fit of two correlated factors with the fit of a single factor model for each pair of dimensions, discriminant validity being established when the fit of two factors is better than the fit of one factor.

Furthermore, Rafiq et al (2011) followed Hair et al (1998) method by making sure that each AVE exceeded its respective shared variance (squared correlations) between the factors. Moreover, the studies of Parasuraman et al. (2005); Sahadev and Purani (2008); and discriminant validity was evidenced following Gerbing and Anderson (1988) by constraining each of the correlations (one at a time) to unity in the measurement model (leaving all other parameters free), and repeated the CFA. In every case, the discriminant validity is confirmed when the constrained model in CFA produced an increase in the chi-square statistic (χ^2) compared with non-constrained model. In other studies discriminant validity is confirmed by examining all the differences in χ^2 values were significant and the confidence interval around the correlation estimate between any two construct did not include 1 (Yang & Tsai 2007). Moreover, authors such as Akinci et al (2010) adopted combine methods. They first compared the fit of two correlated factors with the fit of a single factor (did not work). After trimming the scale by deleting 12 problematic items, they deep-rooted discriminant validity by (i) constraining the correlation between two factors to one and comparing with the factor model and (ii) Square root of AVE greater than the correlation presented by each construct with other construct. Finally, certain authors (Meng and Mummalaneni (2010); and Connolly et al., 2010) did not report discriminant validity in their study.

Fitness of the structural model used

Since the commonly used goodness-of-fit indices such as χ^2 , Goodness-of-Fit Index (GFI) and Adjusted Goodness-of-Fit Index (AGFI) are significantly subjective by variations in sample size and non-normality of the variables. Hair et al., (1998) recommend that a model reporting the relative chi-square χ^2/df and the more robust

measures such as Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), and the Root Mean Square Error of Approximation (RMSEA) will often provide sufficient unique information to evaluate a model. Before discussing the different results, it is worth mentioning that during the E-S-QUAL scale development, the overall goodness-of-fit statistic result for the CFA was well above conventional cutoff value, with the possible exception of RMSEA, which was slightly above the cutoff value of .06 suggested by Hu and Bentler (1999). Even so, it was well within the criteria suggested by Browne and Cudeck (1993) (lower than .08) for inferring an acceptable fit.

The results of the various different fit indices and the cutoff values from the studies reviewed are as follows : χ^2/df range from 1.24 to 6.29 (≤ 3); CFI range from .90 to 1.00 (≥ 0.90) ; NFI range from .90 to .98 (≥ 0.90); NNFI range from .89 to .94 (≥ 0.90); GFI range from .82 to .99 (≥ 0.90); AGFI from .91 to .98 (≥ 0.90); BB-NFI range from .90 to .93 (≥ 0.90); RFI/IFI range from .90 to .97 (≥ 0.90); TLI/TFI range from .91 to .99 (≥ 0.90) and RMSEA range from .04 to .11 ($\leq .06$) (See note1). Taken as a whole, the statistics result of the majority of the studies suggested an adequate model fit for the dimension structure. However, some of the fit indices appear to be doubtful and suspect. For example in the studies of Parasuraman et al (2005); Boshoff (2007) and Chiou et al (2009) the reported value of χ^2/df was somewhat above the acceptable threshold value (≤ 5) recommended by Bagozzi and Yi (1988). Whereas, Wu and Ding 2007; Yang et al., (2010) reported GFI slightly lower than the conventionally cutoff value (≥ 0.90) recommended by Hu and Bentler (1999). Moreover, the studies conducted by Yang et al., (2010) and Wu and Ding (2007) show a RMSEA value equal to .09. Similarly, unlike the acceptable results of the original scale, Parasuraman et al (2005) reported RMSEA values (.09 for Amazon and .11 for Walmart) somewhat above the acceptable threshold value ($\leq .08$) recommended by Browne and Cudeck (1993) during their scale reassessment.

This review shows all the studies have the minimum acceptable sample sizes (see table 1) necessary for achieving specified levels of statistical power when testing structural models with varying degrees of freedom. Regarding the fairly high RMSEA values, Parasuraman et al (2005) argued that it is worth noting that the interpretation of any fit index in isolation could be problematic because trade-offs between Type I and Type II errors call for the interpretation of combinations of indices in various model contexts. Hence, the high statistical power and exemplary values of different fit indices presented in most of the cases mitigate the fairly high root mean square error of the approximation (RMSEA) values (Hu and Bentler, 1999).

Nomological validity

The testing of the relationship strength between e-SQ and the predictive dimensions is intended to demonstrate further validation of the instrumentation. If the constructs perform as predicted by theory (based on traditional service or emerging evidence on the field), then it can infer that the measurement of the constructs is nomologically valid. To

¹ CFI = Comparative Fit Index; NFI = Normed Fit Index; NNFI = non-normalized fit index; GFI = goodness-of-fit index; AGFI = adjusted goodness-of-fit index; BB-NFI = Bentley-Bonnet non-normed fit index; RFI = Relative Fit Index; TLI = Tucker-Lewis Index; RMSEA = root mean square error of approximation;

test the nomological validity, some authors' modeled e-SQ as a first order exogenous construct that influences the higher order constructs of predictive dimensions in the structural model (Akinci et al 2010; Chiou et al 2009; Marimon et al 2010). While others preferably treat e-SQ as a second-order latent construct in the structural model with the dimensions of e-SQ serving as first order constructs, which in turn are presented by the scale items (Wu and Ding 2007; Yang & Tsai 2007). Except Kim et al (2006) and Mekovec et al (2007) who did not examined the predictive validity, the studies reviewed results indicated that regardless of the method used to test the nomological validity, most of the relationship between the dimensions of E-S-QUAL and the predictive dimensions were confirmed. Henceforth, were signifying further authentication of the psychometric properties of the E-S-QUAL scale. Nonetheless, there is no consensus on the nature of the relationship strength between e-SQ dimensions and the dependent variable used in assessing nomological validity across studies. These observations are discussed in greater detail below.

Final painstaking dimensions and number of items structure

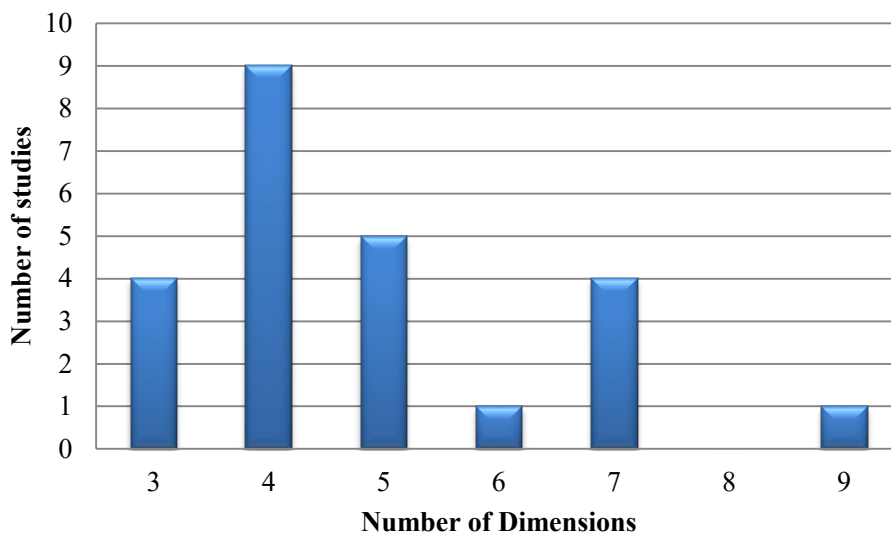
While the original study by Parasuraman et al (2005) proposed four universal dimensions which were supposed to measure e-service quality in any sector, the vast majority (more than 50%) of studies report a number of dimensions other than 4. Figure 1 illustrates the instability associated with the number of dimensions. Figure 1 also shows the use of E-S-QUAL in several sectors raises questions about the number of dimensions and their stability from one context to another. Excluding the study conducted by Mekovec et al (2007), who only examined the dimension of Efficiency, the remaining cases (96 percent) shows the number of dimensions varies between three (Yen and Lu 2008; Petnji et al 2011) and nine (Kim et al 2006). Hence, all of these studies confirmed that the construct of e-service quality is multidimensional. However, the result from the reviewed studies invalidates the invariance of the scale's structure. The dimensional structure is very unstable, even within a given sector. For example, while Akinci et al (2010) confirmed the four dimensions of E-S-QUAL in online banking in Turkey, Petnji et al (2011) found only three dimensions in Spain. Moreover, Wu and Ding (2007) deep-rooted the four dimensions of E-S-QUAL in online shopping in the USA, whereas Boshoff 2007 established 6 dimension in South Africa. Furthermore, Rafiq et al (2011) study in the UK originated only 3 dimensions. Likewise, in Taiwan, Yen and Lu (2008) study engender 3 dimensions for an auctioneers' website and 3 dimensions for an auction seller, while the Chiou et al (2009) study produced 5 and 4 respectively in the same country.

This result supports the work of Ladhari (2010), who intimated that e-SQ dimensions tend to be contingent on the service industry involve. Even in the same industry, the author found that the dimensions of e-SQ depend on the type of user service. According to the author information content is essential to portal web and Internet banking services and less important for companies such as Amazon.com that produce physical products. Nonetheless, among the various dimensions cited in the reviewed studies, the three dimensions "efficiency", "system availability" and "privacy" appear consistently (in decreasing order) in most of the studies. On the other hand, the dimension of fulfillment appears not to be generic but specific to particular e-service contexts. Studies related to pure service did not apply the dimension of fulfillment. In contrast,

Petnji et al (2011) included the dimension of fulfillment in their study, although it was subsequently discarded as it failed the psychometric test. In the same vein, the fulfillment dimension was divided into two (“speed” and “reliability”) during the EFA process in the Boshoff (2007) study.

Contrariwise, it can be accurately stated on the basis of this review that the fulfillment dimension is one of the prominent dimensions of the E-S-QUAL with regard to web sites that selling physical goods. Given that this dimension was taken as dyed-in-the-wool in most of the studies investigating web sites that sell physical goods (e.g. Wu and Ding 2007; Sahadev and Purani 2008; Yen and Lu 2008; Chiou et al 2009; Lin et al 2009; Marimon et al 2010; Fuentes-Blasco et al 2010; Chang 2011). Nonetheless, Rafiq et al (2011) study of online grocery services reported some covariance problems between the residuals of Fulfillment and Privacy dimensions and their regression weights also showed evidence of cross loading. Such misspecification means that the Fulfillment items could measure Privacy or vice versa. Hence, relying on Wolfinbarger and Gilly's (2003) study, Rafiq et al (2011) regarded Fulfillment as one of the core elements for online grocery services. Thus, they discarded the Privacy dimension instead for the reason that it was an augmented part of the service.

Figure 1 Final painstaking number of dimensions



Cultural context

Traditionally, several researchers questioned the applicability of the well-known SERVQUAL in certain cultural contexts. For example, it would seem that measurement of service quality in the brick-and-mortar banking industry is dependent on the cultural context (Ladari, 2009). Similarly, this review also raised the question of whether the applicability of E-S-QUAL in online banking depended on the cultural context. Sun et al., (2009) applied the E-S-QUAL scale to measure online banking service quality in the Chinese context and the results confirmed the four dimensions of E-S-QUAL. Likewise, Akinçi et al (2010) applied E-S-QUAL in the context of online banking in Turkey; however they were forced to drop 14 items during the CFA before confirming the 4 dimensions of E-S-QUAL. Petnji et al (2011) applied E-S-QUAL in the context of online

banking in Spain. Their findings showed that the dimensionality of E-S-QUAL yielded only three dimensions. These results support the controversy that the dimensionality of E-S-QUAL is unstable in e-banking between different cultural contexts.

On the other hand, the E-S-QUAL scale appears to be relatively stable for websites selling physical goods in different cultural contexts, since the four dimensions of E-S-QUAL were confirmed in the USA (Parasuraman et al 2005; Wu and Ding 2007); in Taiwan (Yang & Tsai, 2007); in India (Sahadev and Purani, 2008); in Ireland (Connolly et al 2010), and in Spain (Marimon et al 2010). Moreover, Meng and Mummalaneni (2010) used the scale to test measurement equivalence on Chinese and African American consumers. They concluded that the instrument can be used to analyze web service quality perceptions in other culture and to compare web service quality perception between different cultures. However, they observed that contrary to their African American counterpart, Chinese consumers had significantly lower perceptions on all dimensions of e-SQ, with Efficiency, Privacy and System availability exhibited larger differences than Fulfillment. These findings are supported by Zhang et al., (2008) who reviewed several empirical studies and identified consistent results showing that service users from different countries and cultural backgrounds record different expectations, react differently to service encounters, and show dissimilar behavioral intentions.

CONCLUSIONS

Practitioners in general need a generic scale that provides the potential for cross-industry and cross-functional comparisons. In this new environment of electronic service, the absence of a valid and reliable instrument to measure e- service quality has hard-pressed early researchers to make use of some fairly inadequate alternatives, such as using subjective quality attributes or selected generalizable items from the SERVQUAL. Hence, confused the endeavors of both scholars and practitioners to effectively measure and consequently manage e-service quality strategies. Moreover, 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, particularly in cases where measures developed in one country are intended for use in other countries. As observed by Fariq et al (2011), replications not only help to determine the reliability and validity of newly developed measurement instruments but also help to define the scope and limits to their generalizability to other contexts. Drawing on the precedent, the measure of electronic service quality bids a challenge.

This review demonstrated that from among the various different instruments that have been proposed for assessing electronic service quality, it is the E-S-QUAL scale (Parasuraman et al., 2005) that has received the most recognition, since this scale has been to one extent or another successfully replicated and applied in 11 countries and a variety of e-service settings. Collectively, the findings of the studies reviewed reveal important differences in a number of final painstaking dimensions of E-S-QUAL and the number of items used in the scale across different industrial sectors. In addition, this study has raised concerns about some theoretical and empirical problems associated with the re-assessment of E-S-QUAL, such as the use of different scores, the scale's reliability, the convergent validity, the discriminant validity, the predictive/nomological validity and the applicability of the scale to different cultural contexts.

On the whole, this study has brought to light the results obtained by several authors who have duplicated the E-S-QUAL scale. The results revealed that the scale is effective in capturing the core of electronic service quality. Yet, both scholars and practitioners are called to first assess the structure of factors underlying their data before drawing any conclusions from their study. Given that the dimensional structure of E-S-QUAL appears to be very unstable even within a given sector. Nevertheless, the dimensions of efficiency, system availability and privacy appear consistently in the various models, indicating that there are some common factors of E-S-QUAL that are applied by customers in valuing e-SQ irrespective of the type of e-service delivered. On the other hand, the dimension of fulfillment appears not to be generic but specific to particular e-service contexts such as websites selling physical goods. Together, these results reinforce the support obtained in the scale development phase for the psychometric soundness of E-S-QUAL.

REFERENCES

- Akinci, S., Atilgan-Ina, E. and Aksoy, S. (2010) "Re-assessment of E-S-Qual and E-RecS-Qual in a pure service setting" *Journal of Business Research*, 63, pp. 232–240
- Anderson, J. C. and D. W. Gerbing (1988) "Structural Equation Modeling in Practice: A Review and Recommended Two-Step Approach", *Psychological Bulletin*, Vol.103, No.3, pp. 411-423.
- Barnes, S. J. And Vidgen, R.T. (2002) "An Integrative Approach to the Assessment of E-Commerce Quality", *Journal of Electronic Commerce Research* 3(3), 114 – 127.
- Bauer, H. H., Falk, T., Hammerschmidt, M., (2006). "eTransQual: A transaction process-based approach for capturing service quality in online shopping". *Journal of Business Research* 59, 866-875.
- Boshoff, Christo (2007), "A psychometric assessment of E-S-QUAL: a scale to measure electronic service quality", *Journal of Electronic Commerce Research*, Vol. 8, No. 1, pp.101-114.
- Carmel Herington, Scott Weaven, (2009) "E-retailing by banks: e-service quality and its importance to customer satisfaction", *European Journal of Marketing*, Vol. 43 Iss: 9/10, pp.1220 – 1231
- Chia-Hui Yen, Hsi-Peng Lu, (2008) "Effects of e-service quality on loyalty intention: an empirical study in online auction", *Managing Service Quality*, Vol. 18 Iss: 2, pp.127 – 146
- Connolly, R., Bannister, F. & Kearney, A. (2010). "Government website service quality: a study of the Irish revenue online service". *European Journal of Information Systems*, Volume: 19, Issue: 6 pp. 649-667.
- Cristobal, E., Flavian, C., & Guinaliu, M. (2007). Perceived e-service quality (PeSQ): Measurement validation and effects on consumer satisfaction and web site loyalty. *Managing Service Quality*, 17(3), 317 - 340.
- Dawes, John (2008). "Do Data Characteristics Change According to the number of scale points used? An experiment using 5-point, 7-point and 10-point scales". *International Journal of Market Research* 50 (1): 61–77.
- Fornell. C. & Larcker. D.F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*. 28(1). 39–50.
- Fuentes-Blasco, M.; Gil-Saura, I.; Berenguer-Contrí, G. & Moliner-Velázquez, B. (2010). "Measuring the antecedents of e-loyalty and the effect of switching costs on website", *The Service Industries Journal*, Vol. 30, N° 11, pp. 1837-1852.
- Gerbing, D.W., & Anderson, J.C. (1988). "An updated paradigm for scale development incorporating unidimensionality and its assessment". *Journal of Marketing Research*, 25(2), 186–192.

- Hair, J. F., Anderson, R. E., Tatham, R. L. and Black, W. C. (1998), *Multivariate data analysis*. 5th ed., Ed. Prentice Hall International, Inc., Upper Saddle River, NJ.
- Hu, L. and Bentler, P. (1999), "Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives", *Structural Equation Modeling*, Vol. 6, No. 1, pp. 1-55.
- Jyh-Shen Chiou, Lei-Yu Wu, Yi-Ping Sung, (2009) "Buyer satisfaction and loyalty intention in online auctions: Online auction web site versus online auction seller", *Journal of Service Management*, Vol. 20 Iss: 5, pp.521 – 543
- Kim, M., Kim, J-H., & Lennon, S.J. (2006). Online service attributes available on apparel retail web sites: An E-S-QUAL approach. *Managing Service Quality*, 16(1), 51–77.
- Kuang-Wen Wu, May-Ching Ding (2007). "Validating the American Customer Satisfaction Index Model in the Online Context: An Empirical Study of U.S. Consumer Electronics E-Tailers" *International Journal of Business and Information* Volume 2, Number 2, December 2007
- Ladhari, R. (2010): "Developing e-service quality scales: a literature review", *Journal of Retailing and Consumer Services*, Vol 17. pp. 464-477.
- Loiacono, E., Watson, R.T., & Dale, G. (2000). "WebQualTM: A Web site quality instrument". Working paper, *Worcester Polytechnic Institute*.
- Malhotra, N.K. (2004), *Marketing Research and Applied Orientation* (4th ed), Englewood Cliffs, NJ: Prentice Hall Inc.
- Marimon, F.; Vidgen, R.; Barnes, S.; Cristobal, E., 2010, "Purchasing behaviour in an online supermarket: the applicability of E-S-QUAL", *International Journal of Market Research*, 52, n° 1, 111-129
- Mekovec, R., Bubaš, G., Vrček, N., (2007). "A Method for Improvement of Objectivity of E-Service Quality Evaluation", *Journal of Information and Organizational Sciences*, Vol. 31, No.2, pp. 15-27
- Meng, J.G., Mummalaneni, V. (2010): "Measurement Equivalency of Web Service Quality Instruments: A Test on Chinese and African American Consumers" *Journal of International Consumer Marketing*, 1528-7068, Volume 22, Issue 3, Pages 259 – 269
- Mohammed Rafiq, Xiaoming Lu & Heather Fulford (2011): "Measuring Internet retail service quality using E-S-QUAL" *Journal of Marketing Management* DOI:10.1080/0267257X.2011.621441
- Nunnally, J. C. and Bernstein, I. H. (1994), *Psychometric Theory*, McGraw-Hill, New York.
- Parasuraman, A., Zeithaml, V. A., & Malhotra, A. (2005). "E-S-QUAL: A Multiple-Item Scale for Assessing Electronic Service Quality". *Journal of Service Research*, 7(3), 213-233.
- Petnji, Y.L.H., Marimon, F., & Casadesus, M. (2011). "Customer's loyalty and perception of ISO 9001 in online banking". *Industrial Management & Data System*, 111(8), 1194–1213.
- Praeg Claus-Peter; Spath. (2010), Dieter "Quality Management for IT Services: Perspectives on Business and Process Performance" Publisher: IGI Global,
- Reichheld, F.F., & Scheffer, P. (2000). E-loyalty: Your secret weapon on the web. *Harvard Business Review*, 87(4), 105–113.
- Riadh Ladhari, (2009) "A review of twenty years of SERVQUAL research", *International Journal of Quality and Service Sciences*, Vol. 1 Iss: 2, pp.172 – 198
- Ribbink, D., van Riel,, A. and Streukens, S. (2004), "Comfort your online customer: quality, trust and loyalty on the internet", *Managing Service Quality*, Vol. 14, No. 6, pp. 446-456.
- Sahadev, S. and K. Purani, 2008. Modeling the consequences of e-service quality. *Marketing Intelligence & Planning*, 26(6): 605-620.
- Shu-Ping Lin, Ya-Hui Chan & Ming-Chun Tsai (2009) "A transformation function corresponding to IPA and gap analysis", *Total Quality Management & Business Excellence*, 20:8, 829-846
- Sun, Q., C. Wang and H. Cao, 2009. "Applying E-S-QUAL Scale to Analysis the Factors Affecting Consumers to Use Internet Banking Services". *Proceeding of IITA International Conference on Services Science, Management and Engineering*, pp:242-245.
- Wei-Lun Chang (2011): "A mixed-initiative model for quality-based e-services pricing", *Total Quality Management & Business Excellence*, DOI:10.1080/14783363.2011.593853

- Wolfenbarger, Mary and Mary C. Gilly (2003), "eTailQ: Dimensionalizing, Measuring, and Predicting etail Quality," *Journal of Retailing*, 79 (3), 183-98.
- Yang H. E, Cheng W. J., Chan J. Y., Pan B. C., Chen C. S. (2010). "Applying an Extended E-S-Qual Scale to Assess the Effects of E-service Quality on Online Loyalty with Customer Satisfaction and Perceived Value as Mediators" 9th WSEAS Int. *Conference on Applied Computer and Applied Computational Science*. ISBN: 978-960-474-173-1
- Yang, H., & Tsai, F. (2007). General E-S-QUAL scales applied to websites satisfaction and loyalty model. *Communications of the IIMA*, 7(2), 115–126.
- Yoo, B., & Donthu, N. (2001). "Developing a scale to measure perceived quality of an internet shopping site (SITEQUAL)". *Quarterly Journal of Electronic Commerce*, 2(1), 31–46.
- Zeithaml, V.A., Parasuraman, A. and Malhotra, A. (2000), "e-Service quality: definition, dimensions and conceptual model", *working paper, Marketing Science Institute*, Cambridge, MA.
- Zeithaml, V.A., Parasuraman, A. and Malhotra, A. (2002), "Service quality delivery through web sites: a critical review of extant knowledge", *Journal of the Academy of Marketing Science*, Vol. 30 No. 4, pp. 362-75.
- Zhang, J., Beatty, S.E. and Walsh, G. (2008), "Review and future directions of cross-cultural consumer services research", *Journal of Business Research*, Vol. 61, pp. 211-24.

Acknowledgement:

This article was written as part of a research project titled "Comparative study between English and Spanish e-banking consumers (ref: TIN2011-13075-E)" financed by the Ministry of Economy and Competitiveness within the aid subprogram of complementary actions to research no orientated.