E-Service Quality: Development of a Hierarchical Model

Abstract

Using survey data from 358 online customers, the study finds that the e-service quality construct conforms to the structure of a third-order factor model that links online service quality perceptions to distinct and actionable dimensions, including (1) website design, (2) fulfilment, (3) customer service, and (4) security/privacy. Each dimension is found to consist of several attributes that define the basis of e-service quality perceptions. A comprehensive specification of the construct, which includes attributes not covered in existing scales, is developed. The study contrasts a formative model consisting of 4 dimensions and 16 attributes against a reflective conceptualization. The results of this comparison indicate that studies using an incorrectly specified model overestimate the importance of certain e-service quality attributes. Global fit criteria are also found to support the detection of measurement misspecification. Meta-analytic data from 31,264 online customers are used to show that the developed measurement predicts customer behavior better than widely used scales, such as WebQual and E-S-Qual. The results show that the new measurement enables managers to assess e-service quality more accurately and predict customer behavior more reliably.

Keywords: Online retailing; online service quality; scale development; meta-analysis

Online retailers such as Amazon constantly work on improving their business models to provide their customers with a superior online shopping experience. For instance, Amazon recently started the AmazonFresh service, which allows customers to order groceries online for home delivery (AmazonFresh 2016). Besides experimenting with new assortments, Amazon CEO Jeff Bezos also revealed that the company has plans to use remotely controlled delivery drones as part of its Amazon Prime Air service in order to deliver individual packages to customers' doorsteps within just 30 minutes of ordering (Amazon Prime Air 2016). As these examples illustrate, it seems to be important for online retailers to constantly improve their offered service quality in order to satisfy customer needs. To deliver superior service quality, managers need to understand how customers perceive and evaluate online customer service.

Not surprisingly, the conceptualization and measurement of e-service quality has received considerable attention in the retailing literature (Blut, Chowdhry, Mittal, and Brock 2015). E-service quality refers to the extent to which an online store facilitates efficient and effective shopping, purchasing, and delivery (Parasuraman, Zeithaml, and Malhotra 2005). The construct is defined broadly to encompass all phases of a customer's interactions with an online store. Wolfinbarger and Gilly (2003) define e-service quality as the beginning to the end of the transaction, including information search, privacy policies, website navigation, the ordering process, customer service interactions, delivery, return policies, and satisfaction with the ordered product. Thus, available measures cover numerous attributes of the online store, such as the attractiveness of the online assortment, the convenience of the ordering process, and the quality of the return policies, among many other attributes (Table 1 provides an overview).

Despite significant progress with respect to construct measurement, a recent meta-study identified several shortcomings of existing scales (Blut et al. 2015). The authors give

specific recommendations regarding how to develop a new measurement based on the meta-analytic findings. The authors explain that (1) existing measures are incomplete since they do not capture all online store attributes that constitute the construct; (2) the relationships among online store attributes, dimensions, and overall quality assessment are not reflected in widely used measurements; and (3) the items of numerous measures have low predictive ability.

First, most prior e-service quality measurements are incomplete since they do not capture all attributes that have been found to affect the customer shopping experience. It has been suggested that a comprehensive measure should cover the 16 attributes and 4 dimensions identified in a recent meta-study (Blut et al. 2015). The two most prominent measurements, E-S-Qual and eTailQ, fall short with respect to the different criteria that customers use to assess online stores (Parasuraman et al. 2005; Wolfinbarger and Gilly 2003). This is problematic since customers may be dissatisfied with the online shopping experience and may switch to a different online store, while managers do not fully understand the reasons for customer churn. Thus, this study develops an improved e-service quality measure that comprehensively captures online store attributes.

Second, existing measurements largely employ reflective indicators to conceptualize e-service quality even though a formative conceptualization may be more suitable (Collier and Bienstock 2006). Arguing that service quality is "the sum total of a number of specific activities that make up the overall performance of a particular industry's service", Rossiter (2002, p. 314) suggests a formative construct conceptualization for the service quality construct. In an online context, Parasuraman et al. (2005, p. 220) also underline that it may be "more appropriate to treat the first-order dimensions [of the e-service quality construct] as formative indicators of the second-order latent construct." Jarvis, MacKenzie, and Podsakoff (2003) explain that the understanding of the appropriate conceptualization is essential since misspecification of the direction of causality may lead to inaccurate conclusions about the

structural relationships between constructs and may therefore lead to biased results. Thus, this study proposes a hierarchical formative measurement approach that may be more appropriate for measuring e-service quality. The study compares a correctly specified formative model with a misspecified reflective model in order to illustrate the different conclusions that researchers draw depending on the model specification. It also examines whether global fit criteria are able to detect measurement misspecification.

Third, the existing measurements differ in their ability to predict customer behavior. The recent meta-study reveals that the most widely employed scales, E-S-Qual and eTailQ, ranked only fifth and eighth with respect to their predictive ability. The authors of the study recommend assessing whether "new measures [can] be developed by using [the meta-analytic] insights and combining items from existing measures" (Blut et al. 201, p. 698). Thus, the present study combines items from best-performing scales in the online retailing literature in order to develop an improved measurement. This study also compares whether the new scale shows stronger correlations with customer outcomes, such as customer satisfaction with the online store, repurchase intention, and word-of-mouth intention, than existing measurements.

Since the existing measurement instruments in the literature have been developed over a decade ago, in a very early phase of e-commerce research, and since a definitive measurement has not been reported to date, the current study develops and tests a comprehensive conceptualization of e-service quality. Using survey data from 358 online customers, the present study develops and tests a hierarchical formative conceptualization of the construct using the construct validation procedures proposed by MacKenzie, Podsakoff, and Podsakoff (2011). Meta-analytic data from 31,624 online customers are used to contrast the predictive ability of the new measurement with established scales. In the following sections, the theoretical basis of the construct's conceptualization is briefly presented, and

existing scales are discussed. After presenting the empirical results of the scale development and comparison, implications for management practice and further research are discussed.

LITERATURE AND CONCEPTUAL BACKGROUND

Means-Ends-Chain Theory as a Theoretical Basis

Means-ends-chain theory represents the theoretical basis for the hierarchical conceptualization of e-service quality (Gardial et al. 1994). This theory assumes that consumers retain product information in memory at multiple levels of abstraction (Parasuraman et al. 2005). It suggests that consumers are able to evaluate their experiences (with an online store) in terms of very specific and concrete occurrences (at the attribute level), which may then be synthesized and related to higher-order dimensions, the latter being relatively more abstract (Johnson 1984). For instance, customers may evaluate their consumption experiences in terms of very specific online store attributes, such as the "width of the online assortment", which may then be related to higher-order performance dimensions, such as "website design". According to this theory, each specific attribute is associated with a higher-order dimension, and each dimension is associated with an overall higher-order summary construct, such as overall e-service quality. Overall e-service quality is then associated with key outcomes such as customer satisfaction, customer repurchase intentions, and word-of-mouth behavior (Parasuraman et al. 2005).

According to this theory, it is essential for the understanding of e-service quality to encompass (a) the different attributes included in the conceptualization of the construct, (b) the number of higher-order dimensions that are related to the attributes, and (c) the type of relationship between the attributes and dimensions, which can be either reflective (direction of causality is from dimension to attribute) or formative (direction of causality is from attributes to dimensions). Means-ends-chain theory suggests a formative relationship between

attributes and dimensions, as the customers assess the attributes first before they form their more abstract judgments about the different dimensions.

Existing Measurements, Common Conceptualizations, and Predictive Ability

The meta-study suggests that e-service quality measures should cover 16 attributes, which can be assigned to 4 dimensions. (1) The first dimension is website design, which refers to all elements of the consumer's experience related to the website (except for customer service), including navigation, information search, order processing, shipment tracking, product availability, product and price offerings, personalization, and system availability. (2) The second dimension is fulfilment, which includes activities that ensure that customers receive what they thought they ordered based on the display and description provided on the website and/or delivery of the right product at the right price (i.e., billed correctly) in good condition within the time frame promised. (3) The third dimension is customer service, which refers to helpful, responsive service that responds to customer inquiries and addresses returns/complaints quickly during or after the sale. (4) The fourth dimension is security/privacy, which is defined as the security of credit card payments and privacy of shared information during or after the sale. Table 1, which is a condensed version of a table reported by Blut et al. (2015), provides an overview over the 16 attributes and shows how the most prominent measures in the literature capture them.

_____

Insert Table 1 here

_____

As Table 1 indicates, several similarities and differences among the available scales can be observed. First, existing measurements do not fully cover all 16 attributes. The E-S-Qual measurement, which was developed by Parasuraman et al. (2005), is the most often

employed scale in the online retailing literature. This measurement covers 13 of the 16 attributes and assigns them to seven dimensions. The meta-study compared Parasuraman et al.'s (2005) seven-dimensional conceptualization of the construct with a four-dimensional conceptualization and finds that the latter conceptualization better fits the data (Blut et al. 2015). Wolfinbarger and Gilly's (2003) eTailQ includes 13 of the 16 attributes and uses the same four dimensions to synthesize the attributes as the present study. The WebQual measurement developed by Loiacono, Watson, and Godhue's (2002) covers only 9 of the 16 attributes and proposes four initial dimensions. Thus, this measurement is the least comprehensive approach.

Second, the three measurements have in common that they do not consider the formative hierarchical structure between attributes and dimensions, as suggested by means-end-chain theory. According to this theory, attributes are assessed individually by customers and then synthesized and related to higher-order dimensions. Hence, attributes and dimensions represent a hierarchy in which they build on each other. Dabholkar, Thorpe, and Rentz (1996) identify and test a hierarchical conceptualization of general service quality perceptions in a retail setting and propose three levels: (1) customers' overall perceptions of service quality, (2) primary dimensions, and (3) subdimensions. They argue that such a structure more completely accounts for the complexity of human perceptions. MacKenzie et al. (2011) explain that a formative hierarchical measurement approach has the general advantage that it allows the decomposition of the total effect of each subdimension on an outcome variable into the proportion of the effect that is due to the superordinate construct and the proportion that is due to subdimension specific factors. Parasuraman et al. (2005) suggest employing a formative approach when developing the E-S-Qual measurement, but since this would require additional indicators that were not available in their study, the authors use a reflective conceptualization instead. While general service quality research has

suggested that the service quality construct may be most appropriately conceptualized as a formative construct (Rossiter 2002; Dabholkar, Shepherd, and Thorpe 2000; Parasuraman et al. 2005), the literature on e-service quality mainly uses reflective measurements.

The decision regarding whether a construct should be conceptualized as formative or reflective should be based on several criteria (MacKenzie, Podsakoff, and Jarvis 2005). It is argued that the construct indicators are formative if (a) they are viewed as defining characteristics of the construct, (b) changes in the indicators are expected to explain changes in the construct, (c) the indicators do not have the same or similar content, and (d) the elimination of an indicator may change the conceptual domain of the construct. With respect to e-service quality, it is observed that different attributes define what constitutes a good online store (e.g., excellent product selection and 15 other unique attributes); changes in an attribute such as website personalization improve website design assessment; attributes are not necessarily highly correlated because they do not have the same content; and the exclusion of a single attribute would eliminate an essential aspect of the construct's conceptual domain (e.g., the price attribute is not covered by the remaining attributes). Thus, a formative measurement of the e-service quality construct seems to be justified.

Third, the measurements differ with respect to their predictive ability. Among all measurements in the online retailing literature, the WebQual measurement performed best in the meta-study, as it had highest correlations with outcome variables such customer satisfaction and repurchase intention (Blut et al. 2015). The more popular E-S-Qual and eTailQ measures ranked only fifth and eighth with respect to their predictive ability. Although the WebQual items work very well, the measurement has a narrow focus, as numerous attribute are not covered; thus, Blut et al. (2015) suggest combining items from this measurement with best performing items from eTailQ and E-S-Qual measurements. While the items from Wolfinbarger and Gilly (2003) do not perform very well with respect to

customer service and security, Parasuraman et al.'s (2005) items show higher correlations on these two dimensions.

CONCEPTUALIZING E-SERVICE QUALITY: A HIERARCHICAL APPROACH

In line with Blut et al. (2015), the present study proposes that e-service quality is best conceptualized as a hierarchical construct with 16 attributes within 4 dimension, where the 4 dimensions affect overall service quality perceptions (figure 1). Although they are the exception, some studies on the service quality construct have previously used a similar conceptualization. For instance, Brady and Cronin (2001) examine service quality in four offline contexts, including fast food, photograph development, amusement parks, and dry cleaning, and explain that their conceptualization of the construct includes three subdimensions. They explain that customers are assumed to aggregate their evaluations of the subdimensions to form their perceptions of an organization's performance on each of the three primary dimensions. Finally, those perceptions then lead to an overall service quality perception. While Fassnacht and Koese (2006) also propose a hierarchical conceptualization in an online context when examining the quality of electronic services (QES), the study employed reflective rather than formative relationships between attributes, dimensions, and overall quality, which are contrary to the decision-making criteria proposed by MacKenzie (2005), means-ends-chain theory, and suggestions in the recent service quality literature.

_____

Insert Figure 1 about here

_____

HYPOTHESES DEVELOPMENT

Website design

Website design is the first dimension of e-service quality, and it has received most attention in the literature. Early conceptualizations had a very narrow focus and mainly examined attributes related to the website interface (Wolfinbarger and Gilly 2003). Research has indicated that many of the early service quality measures do not account for the entire purchase process (Holloway and Beatty 2008). Bauer, Falk, and Hammerschmidt (2006) propose that the online shopping process comprises four phases, including an information phase, an agreement phase, a fulfillment phase, and an after-sales phase. They argue that attributes used by online shoppers differ with respect to the customer needs in the different phases. Attributes related to website design are of major importance during early stages of the shopping process.

Customers in this stage have a major interest in information related to the products, and the easier information can be accessed, the more favorable the website will be assessed (Holloway and Beatty 2008). Customers in this phase are interested in the quality of the provided information, the aesthetics of the website when forming a purchase decision, the smoothness of the purchasing process, the convenience of browsing the website, the product selection, the merchandise availability, the information about pricing, the personalization of information, and the permanent availability of the website. According to means-ends-chain theory, customers use these attributes and assess their experiences in terms of these concrete occurrences. Furthermore, they synthesize their experiences and relate them to the more abstract assessment of the online store's ability to ensure an effective website design. Customers further aggregate their experience to an overall assessment of the online store's performance. Hence,

H1:   Perceptions of website design directly contribute to overall e-service quality perceptions.

H2a-i: Perceptions of website design are directly influenced by a website's (a) information quality, (b) website aesthetics, (c) purchase process, (d) website

convenience, (e) product selection, (f) merchandise availability, (g) price offerings, (h) website personalization, and (i) system availability.

Fulfilment

The second dimension is fulfilment of orders, which refers to the online store' ability to ensure that customers receive what they thought they ordered. This stage is more relevant in the later shopping process since it primarily relates to postsales issues (Bauer et al. 2006). In this phase, customers have already placed their order, and they want to receive what they expect to receive. Hence, attributes related to accuracy of order fulfillment, order timing, and the condition of delivered products are of utmost importance. Customers expect the online store to ensure that the display and description provided on the website are accurate and that the right products are delivered at the right price (i.e., billed correctly), in a good condition, and within a timely manner. Using the critical incident technique, Holloway and Beatty (2008) find that customers use these attributes to assess online store performance, and they report that in about 64 percent of the examined depth interviews, customers mentioned problems with fulfilment, which caused dissatisfaction. One can therefore assume that delivery timeliness, order accuracy, and delivery condition are important attributes that affect perceptions of an online store's ability to ensure effective order fulfillment. It is also assumed that order fulfilment represents a major determinant of overall quality assessments. Thus,

H3: Perceptions of fulfilment directly contribute to overall e-service quality perceptions.

H4a-c: Perceptions of fulfilment are directly influenced by an online retailer's (a) delivery timeliness, (b) order accuracy, and (c) delivery condition.

Customer Service

The third dimension, customer service, relates to online customer support—prior to, during, or after the online order has been placed (Bauer et al. 2006). It includes attributes

related to the online store's general service level and return handling/policies. While online stores do not offer personnel contact, they can still provide an 800 number for customer service or the option to contact service employees via email. Holloway and Beatty's (2008) study indicates that reliable service support, combined with fair, well-communicated return/exchange policies, is an important attribute for driving customer satisfaction. Hence, it is assumed that service level and return handling/policies contribute to perceptions of superior customer service for the online retailer. Moreover, customer service is proposed to contribute to overall quality assessments when customers form their overall judgment of the online store. Therefore,

> H5: Perceptions of customer service directly contribute to overall e-service quality perceptions.

> H6a-b: Perceptions of customer service are directly influenced by an online retailer's (a) service level and (b) return handling/policies.

Security/privacy

Finally, the fourth dimension is the security/privacy of an online store, which refers to customers' concerns regarding potential security/privacy lapses. Customers frequently avoid online stores when they have the feeling that their credit card payment is not secure and that their information during or after the sale is not private. Customers may have the concern that their private information, such as their name or address, may be given to every credit card company (Holloway and Beatty 2008). Online customers expect online stores to protect them against fraud, theft, and "junk" emails after an online purchase. Against this background, it is assumed that a website's security and privacy are important online store attributes, and that the security/privacy dimension affects consumers' overall assessment of an online store. Hence,

> H7: Perceptions of security/privacy directly contribute to overall e-service quality perceptions.

H8a-b: Perceptions of security/privacy are directly influenced by an online store's (a) security and (b) privacy.

METHOD

This research tests the proposed measurement in five steps and uses several tests recommended by MacKenzie et al. (2011) to develop and assess the new measurement. First, after discussing the data collection of this research, the study presents the employed measurements, discusses difficulties of model identification associated with formative constructs, and assesses measurement reliability and validity of the construct.

Second, the study uses structural equation modeling to assess whether the attributes are related to the dimensions and whether the dimensions, in turn, are related to the overall quality assessment as proposed in the hypotheses.

Third, the study assesses the extent of the model misspecification when choosing the wrong measurement model. It compares the results of the correctly specified formative model with a misspecified reflective model in order to illustrate the severity of the model misspecification and the differential conclusions that are drawn by researchers. It also tests whether the fit criteria detect measurement misspecification.

Fourth, the study examines the nomological network in which the e-service quality construct is embedded in order to specify the nature of the relationships between the focal construct and the outcomes variables discussed in the literature (e.g., customer satisfaction). It also uses the nomological network to assess the validity of the multidimensional structure.

Finally, secondary data from a recent meta-study are used to test the developed measurement against extant e-service quality measures. Using meta-analytic data, the study compares the scale's predictive validity with that of WebQual measurment, E-S-Qual, and eTailQ.

Data Collection, Construct Measurement, and Measurement Reliability and Validity

Data collection. The data for this study were collected with the help of a market research firm, which distributed the survey to a random sample of Internet users in the U.S. Participants were screened to ensure that they remember their last online shopping experience. They were included in the survey only if they had made a least one purchase from an online retailer within the last six months, and they were instructed to respond with respect to the last online store that they used within this time (Parasuraman et al. 2005). The participants were directed to a website containing the questionnaire, which they then self-administered. The sample consists of 358 U.S. customers with an average age of 34.28 years (SD=10.91) and an annual income of $35,985 (SD=14,554). The gender composition of the sample is almost even (51.8% female; 48.2% male).

Construct measurement. The first-order reflective constructs in the proposed model are measured by combining items from WebQual, E-S-Qual, and eTailQ based on their displayed performance in the meta-study (Blut et al. 2015). Website design items are mainly adapted from WebQual (Loiacono et al. 2002), while fulfilment and customer service items are taken from E-S-Qual (Parasuraman et al. 2005) and eTailQ (Wolfinbarger and Gilly 2003). For merchandise availability, price offerings, and delivery condition, none of the measures provided a sufficient number of items. For these attributes, items are derived from in-depth interviews and definitions provided by Holloway and Beatty (2008).

For the second- and third-order formative constructs, the literature argues that formative measurement models have to fulfill several conditions in order to be identified. First, in isolation, formative models are underidentified and cannot be estimated (Bollen, 1989; Bollen and Davis, 1994). According to the 2+ emitted paths rule, formative latent variables need to emit at least two paths to other (reflective) constructs or indicators. Thus, Jarvis et al. (2003) and MacKenzie et al. (2005) suggest including at least two additional

reflective items to the set of formative indicators. In the present study, the construct's four dimensions and overall quality are measured with additional reflective indicators, which were adapted from Zeithaml, Berry, and Parasuraman (1996) and Wolfinbarger and Gilly (2003). Second, formative constructs need to be scaled for model identification (scaling rule); thus, this study standardizes the formatively measured constructs by fixing their variance to unity, as advised by Edwards (2001). Third, MacKenzie et al. (2005) explain that model identification requires that the number of nonredundant elements in the covariance-matrix of the observed variables is greater than (or equal to) the number of unknown parameter in the model (t-rule).

For all reflective items, 5-point Likert scales are used, with the anchors 1 (=strongly agree) to 5 (=strongly disagree). Table 2 summarizes the measurement in this study.

_____

Insert Table 2 about here

_____

Measurement reliability and validity. MacKenzie et al. (2011) suggest examining measurement reliability for reflective constructs in the model by using confirmatory factor analysis and calculating Cronbach's alpha coefficients and composite reliabilities. It can be noted that most of the coefficient alphas and composite reliabilities of the examined attributes are larger than .7 (Panel A, Table 3), a threshold generally proposed in the literature (Nunnally 1978). The employed measurement originally included 48 items to capture these attributes and test their impact. Since two items were excluded because of their low factor loadings, the final measurement consists of 46 items. All 120 attribute pairs in the model are tested for discriminant validity using Fornell and Larcker's (1981) stringent criteria. Almost all attribute pairs met these criteria. In the cases of an exception, the χ^2 test for discriminant validity is employed, which supports the discriminant validity of these construct pairs

(Anderson and Gerbing 1988; Garver and Mentzer 1999).[1] Panel B in Table 3 shows that coefficient alphas and composite reliabilities are also sufficiently large for reflective measurements of four dimensions and that discriminant validity is given according to the Fornell and Larcker (1981) criteria. Therefore, the reliability and validity of the reflective constructs in the model are within acceptable boundaries.

_____

Insert Table 3 about here

_____

Furthermore, MacKenzie et al. (2011) suggest assessing the validity of the set of attributes by using Edwards' (2001) adequacy coefficient (R^2_a), which is calculated by summing the squared correlations between the construct and its attributes and dividing by the number of attributes. It is argued that R^2_a values greater than .50 mean that a majority of the variance in the lower-order attributes is shared by the higher-order latent construct. This study finds satisfying coefficients for all dimensions: R^2_a (website design) = .50; R^2_a (fulfilment) = .63; R^2_a (customer service) = .66; R^2_a (security) = .87. The literature argues that reliability assessment in an internal consistency sense is not meaningful for formative indicators since the correlations between formative indicators may be positive, negative or zero (Bollen 1984; MacKenzie et al. 2011).

Subdimension Validity and Hypotheses Testing

To assess subdimension validity, MacKenzie et al. (2011) suggest examining whether each subdimension is significantly related to the second-order latent construct. Similar to Brady and Cronin (2001), this research follows a step-wise testing approach that examines

---

[1] The construct pairs that met the discriminant test of Anderson and Gerbing (1988) were security and privacy (χ^2_{diff}=130.8), system availability and website convenience (χ^2_{diff}=84.1), merchandise availability and product selection (χ^2_{diff}=96.7), purchase process and product selection (χ^2_{diff}=124.4), purchase process and website convenience (χ^2_{diff}=135.9), purchase process and system availability (χ^2_{diff}=90.1), product selection and system availability (χ^2_{diff}=102.9), and website convenience and product selection (χ^2_{diff}=169.8). For more information about this test, see Parasuraman et al. (2005), who also employ this test when developing the E-S-Qual scale.

the dimensions first (Model 1, Figure 2) before testing the attributes (Model 2) and the full

model (Model 3). As indicated in Figure 2, each formative construct in the three models is

measured with additional reflective indicators. All models in this study are estimated by using

the Mplus software package.

_____

Insert Figure 2 about here

_____

In the first step, the second-order factor model is assessed (Model 1).[2] In this model,

the impact of the four dimensions website design, fulfilment, customer service, and

security/privacy on overall e-service quality is tested. The model fit is summarized in Table

4, and the results suggests that the model fits the data well ($\chi^2_{df:80}$=75.263, CFI=1.000,

TFI=1.000, RMSEA=.00, SRMR=.017). Each dimension is found to affect overall e-service

quality.

In the second step, the sixteen attributes are assessed (Model 2, Figure 2). It is argued

that these attributes are the main determinants of website design, fulfilment, customer

service, and security/privacy. The fit indices in Table 4 support the use of most attributes

($\chi^2_{df: 1453}$=2216.031, CFI=.945, TFI=.937, RMSEA=.038, SRMR=.070), and the results

indicate that 13 of the 16 attributes are significant. Only purchase process, website

convenience, and product selection are not related to website design.

The third step combines the two previous steps and tests the hierarchical model

simultaneously (Model 3, Figure 2). Again, the fit indices in Table 4 support this model ($\chi^2_{df:}$

$_{1629}$=2496.651, CFI=.941, TFI=.933, RMSEA=.039, SRMR=.069). This model largely

---

[2] The study uses several techniques to examine the potential for common method variance. First, CFA approach to Harmon's one-factor test is used (McFarlin and Sweeney 1992). The fit is considerably worse for the unidimensional model than for the measurement model. Second, the study uses the marker variable technique (Lindell and Whitney 2001). Similar to Griffith and Lusch (2007), respondents' age is employed as the marker variable. Since the marker variable is not significantly related to any of the variables in the model, common method variance is not a serious problem.

confirms the proposed measurement, with the exception of purchase process, website

convenience, and product selection (Model 3, Table 4).

_____

Insert Table 4 about here

_____

Table 4 shows the results of the formative conceptualization (Model 3). As the

formative model indicates, the four dimensions website design (β=.29, p<.05), fulfilment

(β=.32, p<.05), customer service (β=.09, p<.05), and security/privacy (β=.29, p<.05) are

significantly related to overall service quality. These findings are in line with the proposed

relationships in hypotheses H1, H3, H5, and H7. In addition, thirteen of the sixteen attributes

are found to affect the dimensions. Only purchase process (β=-.02, p>.05), website

convenience (β=.13, p>.05), and product selection (β=.02, p>.05) are nonsignificant, while

the other attributes have significant relationships with the four dimensions. The importance of

the different attributes vary and range from β=.11 to β=.65 (each p<.05). Therefore, the

proposed hypotheses H4a-c, H6a-b, and H8a-b are supported, and hypotheses H2a-e are

partially supported by the findings. Regarding nonsignificant attributes in formative

measurement models, the literature suggests assessing the extent of multicollinearity in the

data set, which may be an issue with formative measurement models (MacKenzie et al.

2011).[3] The study tested multicollinearity using variance inflation factors (VIFs). Since all

VIF values are below the discussed threshold of 4 (website design attributes: 3.244,

---

[3] The study also assesses whether the attributes are nonsignificant because of face validity issues. The employed items are assigned to attributes and dimensions based on previous research. For instance, the study conducted by Holloway and Beatty (2008) uses independent coders in several steps to assign attributes to dimensions with high agreement rates among coders. Moreover, Wolfinbager and Gilly (2003) employ focus groups and a sorting exercise to assign items to the four dimensions used in this study. Similarly, Blut et al. (2015) suggests assigning attributes to the four dimensions. The present research conducts an additional sorting task with five PhD students, and the high agreement rate of 97 percent also indicates that there is face validity.

fulfilment: 2.393, customer service: 1.397, and security/privacy: 2.459), the extent of multi-collinearity can be concluded to be nonproblematic.[4]

Assessment of Model Misspecification

Measurement misspecification may bias the estimates in measurement models and thereby undermine the validity of statistical conclusion. To assess the severity of any misspecification, the correctly specified formative model (model 3 in Figure 2 and Table 4) is compared with two misspecified reflective models (models 4-5 in Figure 2 and Table 4). While the correctly specified model (model 3) assumes that the relationships between attributes, dimensions, and overall quality are formative, the misspecified models (models 4-5) assume reflective relationships. Model 5 differs from model 4 with regard to the additional reflective indicators that are used in model 3 for model identification. The additional items are added to model 5 to ensure the comparability of model tests similar to MacKenzie et al.'s (2005) Monte Carlo study on measurement misspecification.

The results of the comparison generate two insights regarding model misspecification. First, a comparison of fit indices of different specifications shows that all fit criteria indicate that the formative model 3 shows a better fit than the reflective models 4 and 5. While the literature argues that SRMR in particular is the most sensitive goodness-of-fit index for detecting misspecified relationships between latent constructs, the CFI and RMSEA are discussed to be the most sensitive goodness-of-fit indices for detecting measurement model misspecification (Hu and Bentler 1998). This research also finds that the chi-square (per degree of freedom) statistic and TLI support the detection of misspecification.

Second, comparing the results of the formative conceptualization (model 3, Table 4) with the reflective conceptualization (model 5, Table 4), several differences can be observed. While each attribute shows a significant factor loading between $\lambda=.55$ and $\lambda=.97$ (each p<.05)

---

[4] Because the literature indicates that instances where an entire subdimension can be dropped from a formative model without eliminating an essential aspect of the construct domain are very rare, these attributes are kept in the measurement model, as suggested by Diamantopoulos, Riefler, and Roth (2008).

in the reflective model, one can observe that three attributes in the formative model are not significantly related to higher-order dimensions. These results indicate that studies using an incorrectly specified model overestimate the importance of some e-service quality attributes. The finding is in line with previous research on model misspecification. Jarvis et al. (2003), Law and Wong (1999), and MacKenzie et al. (2005) examine the impact of model misspecification and show an overestimation of structural parameters when the latent variable is misspecified. More specifically, they find that in some cases, the (incorrect) reflective specification yields a significant parameter estimate, while the parameter estimate is not significant in the (correct) formative specification. This finding underlines the relevance of the correct specification of the e-service quality construct.

_____

Insert Table 5 about here

_____

Testing the Construct in the Nomological Network

MacKenzie et al. (2011) further discuss the importance of the nomological network for the development and testing of formative constructs. The authors explain that it is also important to (1) specify the nature of the lawful relationships between the focal construct (i.e., e-service quality) and other constructs and (2) use the nomological network to assess the validity of the multidimensional structure.

First, with respect to examination of potential consequences of the focal construct, Cronbach and Meehl (1955, p. 290) explain that "to 'make clear what something is' means to set forth the laws in which it occurs. We shall refer to the interlocking system of laws which constitute a theory as a nomological network." In the context of e-service quality, the nomological validity of the construct is regularly tested by examining its impact on customers' satisfaction, intention to repurchase from the online store, and intention to

recommend the online store to family and friends (Parasuraman et al. 2005). Thus, the present study uses these outcomes of excellent e-service quality to test nomological validity.

Customer satisfaction in this study is measured with a scale provided by Fornell (1992). Repurchase intentions and word of mouth are measured with items provided by Zeithaml et al. (1996). Items for all constructs are assessed on a 5-point Likert scale (5=strongly agree; 1=strongly disagree). Participants of the e-service quality survey also responded with regard to these constructs. All measurements have proven to be reliable since the coefficient alphas are larger than .7 (customer satisfaction=.78; repurchase intentions=.83; word of mouth=.94). The employed scales are displayed in Table 2. The results of the structural equation model with latent outcome variables shown in Table 6 (model A: indirect effects model) indicate that all four e-service quality dimensions are associated with overall quality: website design (β=.33, p<.05), fulfilment (β=.29, p<.05), customer service (β=.11, p<.05), and security (β=.29, p<.05). With respect to the customer outcomes, overall e-service quality is found to be related to customer satisfaction (β=.95, p<.05), repurchase intention (β=.80, p<.05), and word of mouth (β=.79, p<.05). These findings suggest that nomological validity is given in the present study. The overall fit of the indirect effects model is also good ($\chi^2_{df:2199}$=3419.006, CFI=.930, TFI=.923, RMSEA=.040, SRMR=.069).

Second, the formative measurement literature indicates that the adequacy of the multidimensional structure of the focal construct can be assessed with the help of the nomological network (Edwards 2001). MacKenzie et al. (2011) propose assessing the adequacy of the hypothesized multidimensional structure by testing whether the subdimensions of the multidimensional (formative) focal construct have significant direct effects on a consequence construct—over and above the direct effect of the focal construct. Adding direct relationships from the four dimensions to the three outcome variables (model B: direct and indirect effects model, Table 6), the study finds that website design positively

affects customer satisfaction (β=.14, p<.05), repurchase intention (β=.19, p<.05), and word of mouth (β=.19, p<.05). Customer service is also related to customer satisfaction (β=.08, p<.05) and repurchase intention (β=.07, p<.05), and the effects are in the expected direction. This study also tests the significance of these direct paths with a chi-square difference test of the model with these paths (direct and indirect effects model, Table 6) and without these paths (indirect effects model). The results indicate that the subdimensions have significant direct effects on the consequence constructs over and above the indirect effects ($\Delta \chi 2$ (*df*=12): 21.105, p<.05).[5] MacKenzie et al. (2011) also note that the measurement model is supported when the effect of the focal construct on the outcome construct is substantially larger than the direct effects of the subdimensions on this construct. In the present study, overall quality has a much stronger effect on outcomes than the four dimensions indicating the importance of the overall quality of the construct. Overall quality fully mediates the relationship between dimensions and outcomes for fulfillment and security, and it partially mediates the relationships for website design and customer service.[6]

_____

Insert Table 6 about here

_____

Comparing the Performance of the Measurement

To assess the performance of the new e-service quality measurement with that of existing measures, meta-analytic data from Blut et al. (2015) are used. The dataset consists of 573 correlations between e-service quality and outcome variables from 31,264 individuals

---

[5] A similar chi-square difference test is conducted for direct relationships between attributes and outcome variables, and it was found that the attributes also have significant direct effects on the consequence constructs over and above the indirect effects ($\Delta \chi 2$ (*df*=60): 108.51, p<.05). The conducted chi-square difference tests use correction factors, as suggested by Satorra and Bentler (2010).

[6] MacKenzie et al. (2011) indicate that the literature has different views on the proposed test. While the authors recommend using this test, they explain that Franke et al. (2008) argue that the formative measurement model specification implies that the latent construct should completely mediate the effect of its indicators on all other outcome variables. Since MacKenzie et al. (2011) and Franke et al. (2008) expect the focal construct to have much stronger effects than the subdimensions, the findings of this research support using a hierarchical model.

from 89 independent samples reported in 69 articles. The articles have been published within the last 15 years.[7] They examined e-service quality in the context of online stores and were included in the meta-analysis when reporting correlations between e-service quality construct and customer satisfaction, repurchase intentions, and word-of-mouth behavior. The studies were coded by two coders according to the definitions of the 16 e-service quality attributes in Table 1. When samples used multiple measures of a construct, the correlations between two constructs were averaged and reported as a single study (Hunter and Schmidt 2004).

A random-effects approach was employed to calculate the average effect sizes using the SPSS macros from Lipsey and Wilson (2001). The meta-analytic data were analyzed by estimating the averaged bivariate relationships between e-service quality attributes and outcomes. Following Hunter and Schmidt (2004), effect sizes were corrected for potential biases, including dichotomization and range restriction. They were also corrected for measurement error by dividing them by the product of the square root of the respective reliabilities of the two constructs (Hunter and Schmidt 2004). The artifact-corrected effect sizes were transformed into Fisher's z coefficients, and they were weighted by the estimated inverse of their variance (N – 3), before they were converted back to correlation coefficients (Kirca, Jayachandran, and Bearden 2005). Table 7 reports the weighted artifact-corrected effect by the measurements used in prior studies. The meta-study reports these correlations to assess the predictive validity of different e-service quality measures (Web Appendix A reports the comprehensive correlation table from the meta-study).

_____

Insert Table 7 about here

_____

_____

[7] The complete list of included articles can be found online: http://dx.doi.org/10.1016/j.jretai.2015.05.004.

Similar to the meta-study, the present study uses the correlations between the 16 e-service quality attributes and the three outcome variables to assess the scale's predictive validity. It compares the correlations that have been corrected according to the same procedures in the meta-study with the correlations of other measurements. Correlation coefficients have been chosen for this comparison since they allow the aggregation of findings across studies and also comparisons between different studies. Hunter and Schmidt (2004) argue that other effect sizes such as path coefficients should not be used as surrogates for correlation coefficients in meta-analyses.[8] The comparability of findings with the meta-study is ensured in several ways.

First, the items of the 16 attributes are assigned to the four dimensions, and the average correlations for website design, fulfilment, customer service, and security/privacy are calculated. This procedure is similar to the meta-study and enables a comparison of the calculated correlations with correlations reported in the meta-study (Table 7).

Second, the correlations are corrected for artifacts. More specifically, the correlations are corrected for measurement error in the dependent variable and measurement error in the independent variable (Hunter and Schmidt 2004). Specifically, they are divided by the product of the square root of the respective reliabilities of the two constructs (Hunter and Schmidt 2004). Artifact corrections other than these are not necessary since dichotomization and range restriction of artifacts do not apply to the present data.

Third, the correlations reported in the meta-study are based on a random-effects meta-analytic model, which is more conservative than a fixed-effects model. Since the present study is the first to test the new measurement by using a sample of 358 customers, a fixed-effects model had to be calculated. To ensure that the conclusions still hold, the data from

---

[8] As an additional test, the study compares the standardized total indirect effect from this research with the standardized total indirect effect reported by the meta-study. This comparison shows that in 7 of 8 cases, the indirect effect is higher in the present study that in the meta-study. Only for the relationship between customer service and satisfaction is the indirect effect slightly higher in the meta-study than in the present study.

Blut et al. (2015) are reexamined, and fixed effects correlations are calculated. The drawn conclusions are found to be the same independent of the chosen model.[9]

The results of the measurement comparison in Table 7 indicate that the newly developed scale shows consistently higher correlations ($p<.05$) than Loiacono et al.'s (2002) WebQual measurement, Parasuraman et al.'s (2005) E-S-Qual, and Wolfinbarger and Gilly's (2003) eTailQ. The only correlation that is not significantly different is the customer service-repurchase intentions relationship when comparing the new measurement with WebQual. Thus, it is concluded that the present measurement is better suited to predict important customer outcomes than existing measures.

GENERAL DISCUSSION

The literature on e-service quality provides numerous measures that conceptualize the construct differently. The present study builds on these measures and develops a new measurement that addresses several shortcomings. Following MacKenzie et al.'s (2011) construct validation procedures, the study develops a hierarchical conceptualization and measurement of the construct that is more comprehensive than existing scales. The study responds to a recent call to develop a new measurement covering all 16 attributes that define a superior service quality of the online store and that improves the predictive ability of the measurement by combining best performing items from existing scales.

Contribution to the E-Service Quality Literature

The developed hierarchical model and measurement contribute to the discipline in several ways. First, this study tests the conceptualization suggested by Blut et al. (2015) and gives an answer to the question regarding which attributes of an online store matter to customers. The study shows that numerous attributes related to website design, fulfilment,

---

[9] For instance, the fixed-effects correlations for Webqual are $r_c=.78$ (versus .71 in the random-effects model), $r_c=.78$ (versus .71), $r_c=.73$ (versus .67), $r_c=.47$ (versus .47), $r_c=.28$ (versus .28), and $r_c=.68$ (versus .63).

customer service, and security/privacy have to be considered. Excellent website design is determined by superior information quality, website aesthetics, website convenience, merchandise availability, website personalization, and system availability. With respect to customer fulfilment, a superior online store ensures timeliness of delivery, accuracy of orders, and good condition of delivered items. The online store should also provide a good service level and communicate return handling/policies. Finally, online stores have to ensure data security and privacy.

Contrary to expectations, the attributes purchase process, website convenience, and product selection did not affect perceptions of website design. According to Blut et al. (2015), the attributes should be of particular relevance in countries with high uncertainty avoidance. Since the present study examined online shoppers in the U.S. and since the U.S. shows only a moderate level of uncertainty avoidance (Hofstede 2016), it is concluded that in other countries, the attributes are of greater relevance. Given that existing measures of e-service quality are limited with respect to the number of included attributes and their relationships with dimensions and outcomes, the study provides new insights into relevance of choice criteria used by online customers.

Second, many of the above attributes are often measured with only single items in existing measures, and they are often merged with items of other attributes. The results of the present study indicate that all 16 attributes are distinct from each other and that they cannot be simply merged or even deleted from a study (just to achieve a high alpha coefficient), without changing the meaning of the e-service quality construct (Diamantopoulos et al. 2008). Against this background, a hierarchical model of e-service quality is proposed, which can be described as a third-order factor model that links service quality perceptions to distinct and actionable dimensions and attributes. The results of the study suggest that the relationship between attributes, dimensions, and overall assessment is of formative rather than of

reflective nature. The attributes are assumed to affect dimensions, and the dimensions, to affect overall assessments. The decision-making criteria proposed by MacKenzie et al. (2005), means-ends-chain theory, and extant service quality literature support the use of a formative model. Also, the global fit criteria in the SEM indicate the superiority of this model. Since few studies on e-service quality have considered this hierarchy so far, future studies are encouraged to make more use of hierarchical measurement models.

Third, given that many studies still employ reflective indicators, it may be that previous studies report biased results that do not accurately explain and describe service quality (Collier and Bienstock 2006). The model comparison in the present study indicates that an incorrectly specified reflective model overestimates the importance of certain online store attributes. It is found that some estimates (purchase process, website convenience, product selection) are not significant in the (correct) formative specification, while they are significant in the (incorrect) reflective specification. Thus, researchers and managers may overestimate the relevance of certain online store attributes depending on the chosen model specification. The study also indicates that researchers can rely on global fit criteria to detect misspecification.

Fourth, the developed measurement comprised items from existing measures based on their displayed performance to develop a new measurement. The reliability and validity of this measure is assessed, and its performance is contrasted with existing measures. Thus, this study offers answers to the questions raised by Blut et al. (2015) in their research agenda: Can new measures be developed by using these insights and combining items from existing measures? Do new measures outperform WebQual, E-S-Qual, or eTailQ? The results of this research indicate that the new measure shows high reliability and validity and that it better predicts customer behavior than established measurements. Hence, researchers and managers are encouraged to use this measurement in future studies.

Managerial Implications

The findings assist retail managers in several ways. First, the findings of the study help managers to better understand (1) what defines perceptions of e-service quality, (2) how e-service quality perceptions are formed, and (3) how important certain attributes and dimensions are for satisfying and retaining online customers. Based on these insights, managers can make strategic investments to improve the service quality of their online store and become better in important customer outcomes variables. Because of the comprehensive measurement of the construct, managers ensure that they do not ignore important online store aspects that customers consider to be important when choosing among online stores.

Second, the developed conceptualization of e-service quality can be used to categorize customers with respect to their perception of the 16 attributes and identify attractive segments of customers that are most likely to respond positively to the online store's service quality. These insights help firms to concentrate their scarce marketing budget on promising market segments with the greatest likelihood of purchasing from a specific online store. While online retailers such as Amazon target mass markets, specialized online retailers with assortments targeting niche market may find the scale particularly helpful for identifying target segments.

Third, managers can use the developed measurement to permanently monitor performance improvement initiatives of the online store. The measurement can be part of the annual customer satisfaction survey to monitor customers' changing preferences and acceptance of performance improvement initiatives. The developed measurement is applicable to not only a retailer's own online store but also competitors' online stores, as the measurement can be used for benchmarking the online store's performance.

Limitations

The study has several limitations that should be addressed in future research. First, data for this research were collected from only one country, the United States. Since country

culture was found to affect the relevance of e-service quality construct (Blut et al. 2015), future research should test the scale invariance of this measurement before applying it to other countries. In this way, whether the measurement works equally well in different countries and cultures can be clarified.

Second, the participants for this study were sampled based on prior experience with the online store. The employed approach to examine only participants who purchased at least once from an online retailer within the last six months ensures comparability of the used sample with the study from Parasuraman et al. (2005), who used the same approach. Nonetheless, the screening procedure leads to a higher level of familiarity with the online store being rated. The literature argues that the less familiar and more ambiguous the object being rated, the greater the tendency for common method bias (Podsakoff et al. 2003), which could affect the factor structure. Future research should replicate this study's findings with samples that differ regarding the extent of online store familiarity.

Third, the conceptual model and measurement were developed to assess the service quality of online stores in an online retailing context. Blut et al. (2015) suggest developing industry-specific measurements since some items and attributes are not applicable across industry settings. Thus, future research is encouraged to develop similar measures for other online contexts, such as online banking. These measures should also consider using the proposed hierarchical conceptualization of the construct.

Fourth, the nomological validity of the construct was tested by examining its impact on three important customer outcomes. The findings suggest that overall e-service quality represents a full mediator for two dimensions and a partial mediator for two other dimensions. Future research should examine which further outcomes are particularly affected by customer service and security/privacy. While website design and fulfilment have strongest relationships with customer satisfaction , repurchase intention, and positive word of mouth,

one can assume that particularly excellent customer service motivates customers to pay more

for the received the service. Thus, future research could examine customers' willingness to

pay in an online context since online customers are considered to be more price sensitive than

their offline counterparts.

REFERENCES

AmazonFresh (2016), Introducing AmazonFresh, (accessed May 1, 2016), [available at https://fresh.amazon.com/about]

Amazon Prime Air (2016), Introduction to Amazon Prime Air, (accessed May 1, 2016), [available at http://www.amazon.com/b?node=8037720011]

Anderson, James C. and David W. Gerbing (1988), "Structural Equation Modeling in Practice: A Review and Recommended Two-Step Approach," *Psychological Bulletin*, 1003 (3), 411-423.

Bauer, Hans H., Tomas Falk, and Maik Hammerschmidt (2006), "eTransQual: A Transaction Process-Based Approach for Capturing Service Quality in Online Shopping," *Journal of Business Research*, 59 (7), 866-75.

Blut, Markus, Nivriti Chowdhry, Vikas Mittal, and Christian Brock (2015), "E-Service Quality: A Meta-Analytic Review," *Journal of Retailing*, 91(4), 679-700.

Bollen, Kenneth A. (1984), "Multiple Indicators: Internal Consistency or No Necessary Relationship?," *Quality and Quantity*, 18(4), 377-385.

Bollen, Kenneth A. (1989). *Structural Equations with Latent Variables*, New York: John Wiley.

Bollen, Kenneth A. and Walter R. Davis (1994), "Causal Indicator Models: Identification, Estimation, and Testing," Paper presented at the *American Sociological Association Convention*, Miami.

Brady, Michael K., and J. Joseph Cronin Jr. (2001), "Some New Thoughts on Conceptualizing Perceived Service Quality: A Hierarchical Approach," *Journal of Marketing*, 65 (3), 34-49.

Collier, Joel E. and Carol C. Bienstock (2006), "Measuring Service Quality in E-Retailing," *Journal of Service Research*, 8 (3), 260-275.

Cronbach, Lee J. and Paul E. Meehl (1955), "Construct Validity in Psychological Tests," *Psychological Bulletin*, 52(4), 281-302.

Dabholkar, Pratibha A., C. David Shepherd, and Dayle I. Thorpe (2000), "A Comprehensive Framework for Service Quality: An Investigation of Critical Conceptual and Measurement Issues through a Longitudinal Study," *Journal of Retailing*, 76(2), 139-173.

Dabholkar, Pratibha A., Dayle I. Thorpe, and Joseph O. Rentz (1996), "A Measure of Service Quality for Retail Stores: Scale Development and Validation," *Journal of the Academy of Marketing Science*, 24(1), 3-16

Diamantopoulos, Adamantios, Petra Riefler, and Katharina P. Roth (2008), "Advancing Formative Measurement Models," *Journal of Business Research*, 61(12), 1203-1218.

Edwards, Jeffrey R. (2001), "Multidimensional Constructs in Organizational Behavior Research: An Integrative Analytical Framework," *Organizational Research Methods* 4(2), 144-192.

Fassnacht, Martin and Ibrahim Koese (2006), "Quality of Electronic Services: Conceptualizing and Testing a Hierarchical Model," *Journal of Service Research*, 9(1), 19-37.

Fornell, Claes (1992), "A National Customer Satisfaction Barometer: The Swedish Experience," *Journal of Marketing*, 56 (1), 6–21.

Fornell, Claes and David F. Larcker (1981), "Evaluating Structural Equation Models with Unobservable Variables and Measurement Error," *Journal of Marketing Research,* 18 (1), 39-50.

Franke, George R., Kristopher J. Preacher, and Edward E. Rigdon (2008), "Proportional Structural Effects of Formative Indicators," *Journal of Business Research*, 61(12), 1229-1237.

Gardial, Sarah Fisher, D. Scott Clemons, Robert B. Woodruff, David W. Schumann, and Mary Jane Burns (1994), "Comparing Consumers' Recall of Prepurchase and Postpurchase Product Evaluation Experiences," *Journal of Consumer Research*, 20 (4), 548-560.

Garver, Michael S. and John T. Mentzer (1999), "Logistics Research Methods: Employing Structural Equation Modeling to Test for Construct Validity," *Journal of Business Logistics*, 20 (1), 33-57.

Griffith, David A. and Robert F. Lusch (2007), "Getting Marketers to Invest in Firm-Specific Capital," *Journal of Marketing*, 71(1), 129-145.

Hofstede, Geert H. (2016), "Cultural Values per Country" (accessed May 1, 2016), [available at http://www.geert-hofstede.com].

Holloway, Betsy Bugg, and Sharon E. Beatty (2008), "Satisfiers and Dissatisfiers in the Online Environment: A Critical Incident Assessment," *Journal of Service Research*, 10 (4), 347-64.

Hu, Li-tze and Peter M. Bentler (1999), "Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria versus New Alternatives," *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1-55.

Hunter, John E. and Frank L. Schmidt (2004), *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*, Thousand Oaks, CA: Sage Publications.

Jarvis, Cheryl Burke, Scott B. MacKenzie, and Philip M. Podsakoff (2003), "A Critical Review of Construct Indicators and Measurement Model Misspecification in Marketing and Consumer Research," *Journal of Consumer Research*, 30 (2), 199-218.

Johnson, Michael D. (1984), "Consumer Choice Strategies for Comparing Noncomparable Alternatives," *Journal of Consumer Research,* 11 (3), 741-753.

Kirca, Ahmet H., Satish Jayachandran, and William O. Bearden (2005), "Market Orientation: A Meta-Analytic Review and Assessment of its Antecedents and Impact on Performance," *Journal of Marketing*, 69(2), 24-41.

Law, Kenneth S. and Chi-Sum Wong (1999), "Multidimensional Constructs in Structural Equation Analysis: An Illustration Using the Job Perception and Job Satisfaction Constructs," *Journal of Management*, 25(2), 143-160.

Lindell, Michael K. and David J. Whitney (2001), "Accounting for Common Method Variance in Cross-Sectional Research Designs," *Journal of Applied Psychology*, 86(1), 114-121.

Lipsey, Mark W. and David B. Wilson (2001). *Practical Meta-Analysis*. Thousand Oaks, CA: Sage.

Loiacono, Eleanor T., Richard T. Watson, and Dale L. Goodhue (2002), "WEBQUAL: A Measure of Website Quality," *Marketing Theory and Applications,* 13 (3), 432-8.

MacKenzie, Scott B., Philip M. Podsakoff, and Cheryl Burke Jarvis (2005), "The Problem of Measurement Model Misspecification in Behavioral and Organizational Research and Some Recommended Solutions," *Journal of Applied Psychology*, 90(4), 710-730.

MacKenzie, Scott B., Philip M. Podsakoff, and Nathan P. Podsakoff (2011), "Construct Measurement and Validation Procedures in MIS and Behavioral Research: Integrating New and Existing Techniques," *MIS Quarterly*, 35(2), 293-334.

Nunnally, Jum C. (1978), *Psychometric Theory*, New York, NY: McGraw-Hill.

Parasuraman, A., Valarie A. Zeithaml, and Arvind Malhotra (2005), "E-S-QUAL A Multiple-Item Scale for Assessing Electronic Service Quality," *Journal of Service Research,* 7 (3), 213-33.

Podsakoff, Philip M., Scott B. MacKenzie, Jeong-Yeon Lee, and Nathan P. Podsakoff (2003), "Common Method Bias in Behavioral Research: A Critical Review of the Literature and Recommended Remedies," *Journal of Applied Psychology*, 88, 879-903.

Rossiter, John R. (2002), "The C-OAR-SE Procedure for Scale Development in Marketing," *International Journal of Research in Marketing*, 19 (4), 305-335.

Satorra, Albert and Peter M. Bentler (2010), "Ensuring Positiveness of the Scaled Difference Chi-Square Test Statistic," *Psychometrika*, 75(2), 243-248.

Wolfinbarger, Mary, and Mary C. Gilly (2003), "eTailQ: Dimensionalizing, Measuring and Predicting Etail Quality," *Journal of Retailing* 79 (3), 183-98.

Zeithaml, Valarie A., Leonard L. Berry, and A. Parasuraman (1996), "The Behavioral Consequences of Service Quality," *Journal of Marketing*, 60 (2), 31-46.

TABLE 1
DEFINITIONS OF E-SERVICE QUALITY ATTRIBUTES

Attributes	Definition	Loiacono et al. (2002): WebQual	Wolfinbarger and Gilly (2003): eTailQ	Parasuraman et al. (2005): E-S-Qual	Hierarchical model
Website design					
- Information quality	The information provided is accurate, updated, and appropriate. (Loiacono et al. 2002)	X	X	X	X
- Website aesthetics	The aesthetics of website. (Loiacono et al. 2002)	X	X	X	X
- Purchase process	Allowing all or most necessary transactions to be completed online (e.g., purchasing over the Web site). (Loiacono et al. 2002)	X	X	X	X
- Website convenience	Website is easy to read and understand. (Loiacono et al. 2002)	X	X	X	X
- Product Selection	Good selection/diversity of product offerings on website. (Holloway and Beatty 2008)		X		X
- Merchandise availability	Items listed on Website were not out of stock. (Holloway and Beatty 2008)			X	X
- Price offerings	Perception of price of offerings. (Holloway and Beatty 2008)				X
- Website personalization	Communications can be tailored to meet the user's needs. (Loiacono et al. 2002)	X	X		X
- System availability	The correct technical functioning of the site. (Parasuraman et al. 2005)	X		X	X
Fulfillment					
- Timeliness of delivery	Ordered products were delivered on time. (Holloway and Beatty 2008)		X	X	X
- Order accuracy	Right products were delivered and website communicates accurately about products. (Holloway and Beatty 2008)		X	X	X
- Delivery condition	Products were delivered in good condition. (Holloway and Beatty 2008)		X	X	X
Customer service					
- Service level	The availability of assistance through telephone or online representatives. (Parasuraman et al. 2005)	X	X	X	X

- Return handling/policies	Effective handling of problems and returns through the site. (Parasuraman et al. 2005)		X	X	X
Security/privacy					
- Security	Security of credit card payments during or after the sale. (Holloway and Beatty 2008)	X	X	X	X
- Privacy	Privacy of shared information during or after the sale. (Holloway and Beatty 2008)	X	X	X	X
Number of attributes		9	13	13	16
Number of dimensions		4	4	7	4
Conceptualization		no hierarchy	no hierarchy	no hierarchy	hierarchy
Causality		reflective	reflective	reflective	formative
Predictive ability[a]		high	moderate	moderate	highest

a. Assessment of predictive ability is based on the recent meta-study from Blut et al. (2015) and the results of the present study.

TABLE 2
MEASUREMENT OF LATENT CONSTRUCTS

Attributes	Items	
Website design		
- Information quality	IQ1	The information on the website is pretty much what I need to carry out my tasks.
	IQ2	The website adequately meets my information needs.
	IQ3	The information on the website is effective.
- Website aesthetics	WA1	The website is visually pleasing.
	WA2	The website displays a visually pleasing design.
	WA3	The website is visually appealing.
- Purchase process	PP1	The website has no difficulties with making a payment online.
	PP2	The purchasing process was not difficult.
	PP3	It is easier to use the website to complete my business with the company than it is to use a telephone or fax or mail a representative.
- Website convenience	WC1	The display pages within the website are easy to read.
	WC2	The text on the website is easy to read.
	WC3	The website labels are easy to understand.
- Product Selection	PS1	All my business with the company can be completed via the website.
	PS2	This website has a good selection.
	PS3	The site has a wide variety of products that interest me.
- Merchandise availability	MA1	The website has the items the company claims to have in stock.
	MA2	All products on the website are available.
	MA3	Items listed at website were out of stock. (R)[a]
- Price offerings	PO1	The website offers discount or free shipping.
	PO2	The website has low prices.
	PO3	The website has lower prices than offline stores.
- Website personalization	WP1	The website allows me to interact with it to receive tailored information.
	WP2	The website has interactive features, which help me accomplish my task.
	WP3	I can interact with the website in order to get information tailored to my specific needs.
- System availability	SA1	When I use the website, there is very little waiting time between my actions and the website's response.
	SA2	The website loads quickly.
	SA3	The website takes a long time to load. (R)[a]
Fulfillment		
- Timeliness of delivery	TD1	The product is delivered by the time promised by the company.
	TD2	This website makes items available for delivery within a suitable time frame.
	TD3	It quickly delivers what I order.
- Order accuracy	OA1	You get what you ordered from this website.
	OA2	The website sends out the items ordered.
	OA3	The website is truthful about its offerings.
- Delivery condition	DC1	The product was damaged during delivery. (R)
	DC2	The ordered products arrived in a good condition.
	DC3	The products arrived with a major damage. (R)
Customer service		
- Service level	SL1	This site provides a telephone number to reach the company.
	SL2	This site has customer service representatives available online.
	SL3	It offers the ability to speak to a live person if there is a problem.
- Return handling/policies	RP1	It provides me with convenient options for returning items.
	RP2	This site handles product returns well.
	RP3	This site offers a meaningful guarantee.
Security/privacy		
- Security	SEC1	I feel safe in my transactions with the website.
	SEC2	The website has adequate security features.
	SEC3	This site protects information about my credit card.
- Privacy	PRI1	I trust the website to keep my personal information safe.
	PRI2	I trust the website administrators will not misuse my personal information.

Attributes	Items	
	PRI3	It protects information about my web-shopping behavior.

Reflective indicators for higher-order constructs

Attributes	Items	
Website design	WD1	Overall, my experience at the online retailer's website is excellent.
	WD2	Overall, the quality of the online retailer's website is excellent.
	WD3	I am generally very satisfied with the website.
Fulfillment	FUL1	Overall, the online retailer's order fulfillment is excellent.
	FUL2	Overall, the quality of the online retailer's order fulfillment is excellent.
	FUL3	I am generally very satisfied with the order reliability.
Customer service	CS1	Overall, the online retailer's customer service is excellent.
	CS2	Overall, the quality of the online retailer's return handling is excellent.
	CS3	I am generally very satisfied with the customer service.
Security/privacy	SEP1	Overall, the online retailer's handling of data security is excellent.
	SEP2	Overall, the quality of the online retailer's security is excellent.
	SEP3	I am generally very satisfied with handling of private information.
Overall quality	OQ1	Overall, my purchase experience with this online retailer is excellent.
	OQ2	The overall quality of the service provided by this online retailer is excellent.
	OQ3	My overall feelings toward this online retailer are very satisfied.

Outcomes of e-service quality

Attributes	Items	
Customer satisfaction	SAT1	I am satisfied with this online retailer.
	SAT2	The online retailer is getting close to the ideal online retailer.
	SAT3	The online retailer always meets my needs.
Repurchase intention	RI1	I intend to use this online retailer within the next few years.
	RI2	I consider this online retailer to be my first choice for future transactions.
	RI3	I consider doing more business with this online retailer in the coming months.
Word of mouth	WM1	I say positive things about this online retailer to other people.
	WM2	I recommend this online retailer to someone who seeks my advice.
	WM3	I encourage friends and others to do business with this online retailer.

a. Two of the initial 48 e-service quality items have been excluded because of low factor loadings. Thus, the final measurement consists of 46 items.

TABLE 3
PANEL A: CORRELATIONS AMONG ATTRIBUTES

Attributes	IQ	WA	PP	WC	PS	MA	PO	WP	SA	TD	OA	DC	SL	RP	SEC	PRI
Information quality (IQ)	1.00															
Website aesthetics (WA)	.61	1.00														
Purchase process (PP)	.63	.39	1.00													
Website convenience (WC)	.69	.56	.82	1.00												
Product selection (PS)	.61	.46	.81	.77	1.00											
Merchandise availability (MA)	.50	.37	.67	.68	.75	1.00										
Price offerings (PO)	.51	.39	.61	.56	.60	.60	1.00									
Website personalization (WP)	.53	.46	.45	.51	.48	.61	.54	1.00								
System availability (SA)	.72	.59	.80	.84	.79	.69	.60	.59	1.00							
Timeliness of delivery (TD)	.55	.41	.53	.55	.52	.63	.54	.48	.62	1.00						
Order accuracy (OA)	.59	.46	.59	.59	.59	.58	.56	.44	.66	.83	1.00					
Delivery condition (DC)	.18	.05	.39	.29	.36	.13	.14	-.04	.30	.21	.35	1.00				
Service level (SL)	.33	.31	.22	.26	.28	.46	.35	.53	.37	.44	.37	-.14	1.00			
Return handling/policies (RP)	.42	.38	.44	.43	.39	.56	.49	.51	.53	.56	.58	.03	.62	1.00		
Security (SEC)	.64	.41	.59	.64	.62	.51	.54	.49	.66	.53	.60	.24	.38	.54	1.00	
Privacy (PRI)	.62	.43	.52	.53	.54	.47	.49	.49	.59	.49	.62	.19	.42	.52	.89	1.00
Alpha	.84	.94	.81	.91	.79	.65	.75	.89	.76	.89	.88	.89	.75	.87	.89	.87
CR	.85	.94	.82	.91	.82	.67	.77	.89	.78	.89	.88	.89	.76	.88	.90	.87
AVE	.65	.83	.61	.78	.61	.51	.53	.73	.63	.72	.71	.80	.51	.70	.74	.69

Notes. Alpha coefficients are presented in bold on the diagonal. The reported correlations are corrected correlations among reflective latent factors.

PANEL B: CORRELATIONS AMONG DIMENSIONS

Dimensions	WD	FUL	SER	SEP
Website design (WD)	1.00			
Fulfilment (FUL)	.57	1.00		
Customer service (SER)	.47	.37	1.00	

Security/privacy (SEP)	.59	.53	.45	1.00
Alpha	.93	.94	.93	.94
CR	.93	.94	.93	.94
AVE	.82	.85	.82	.84

Notes. Alpha coefficients are presented in bold on the diagonal. The reported correlations are corrected correlations among reflective latent factors.

TABLE 4
STRUCTURAL EQUATION RESULTS

Model (n=358)	χ^2	Degrees of Freedom	AIC	BIC	CFI	TLI	RMSEA	SRMR
Model 1: test of the dimensions	75.263	80	6744.351	6957.162	1.000	1.000	.000	.017
Model 2: test of the attributes	2216.031	1453	34311.389	35534.978	.945	.937	.038	.070
Model 3: test of the hierarchical model	2496.651	1629	35493.972	36743.755	941	.933	.039	.069
Model 4: test of the reflective model	1677.296	969	30592.504	31204.299	.914	.908	.045	.090
Model 5: test of the rival model	3229.065	1754	36159.224	36925.345	.899	.895	.049	.100

TABLE 5
COMPARING THE FORMATIVE WITH THE REFLECTIVE MODEL

Dimensions/attributes	Model 3: The formative model			Model 5: The rival model		
	Attrib.[a]	Dimen.[b]	Overall[b]	Attrib.[a]	Dimen.[a]	Overall[a]
Website design			.29*			.83*
- Information quality		.22*			.85*	
IQ1	.72*			.83*		
IQ2	.87*			.88*		
IQ3	.82*			.85*		
- Website aesthetics		.14*			.66*	
WA1	.90*			.90*		
WA2	.95*			.96*		
WA3	.89*			.89*		
- Purchase process		-.02			.81*	
PP1	.86*			.87*		
PP2	.80*			.81*		
PP3	.67*			.67*		
- Website convenience		.13			.86*	
WC1	.88*			.89*		
WC2	.92*			.92*		
WC3	.84*			.85*		
- Product selection		.02			.82*	
PS1	.60*			.61*		
PS2	.85*			.88*		
PS3	.87*			.86*		
- Merchandise availability		.18*			.79*	
MA1	.78*			.76*		
MA2	.64*			.67*		
- Price offerings		.15*			.69*	
PO1	.65*			.63*		
PO2	.81*			.84*		
PO3	.71*			.72*		
- Website personalization		.11*			.68*	
WP1	.85*			.85*		
WP2	.84*			.85*		
WP3	.87*			.88*		
- System availability		.18*			.90*	
SA1	.76*			.78*		
SA2	.83*			.83*		
Fulfillment			.32*			.72*
- Timeliness of delivery		.36*			.87*	
TD1	.79*			.84*		
TD2	.89*			.90*		
TD3	.86*			.88*		
- Order		.48*			.92*	
OA1	.84*			.84*		
OA2	.86*			.85*		
OA3	.83*			.82*		
- Delivery condition		.32*			.55*	
DC1	.92*			.94*		
DC2	.46*			.47*		
DC3	.86*			.84*		
Customer service			.09*			.54*

	A			B		
- Service level		.46*			.81*	
SL1	.64*			.63*		
SL2	.72*			.71*		
SL3	.77*			.78*		
- Return handling/policies		.53*			.82*	
RP1	.85*			.88*		
RP2	.86*			.88*		
RP3	.80*			.78*		
Security/privacy			.29*			.75*
- Security		.34*			.94*	
SEC1	.82*			.86*		
SEC2	.91*			.91*		
SEC3	.86*			.88*		
- Privacy		.65*			.97*	
PRI1	.88*			.88*		
PRI2	.86*			.86*		
PRI3	.77*			.77*		

Reflective indicators for higher-order constructs

	A	B
Website design		
WD1	.92*	.87*
WD2	.91*	.87*
WD3	.88*	.86*
Fulfillment		
FUL1	.93*	.92*
FUL2	.92*	.92*
FUL3	.91*	.91*
Customer service		
CS1	.90*	.89*
CS	.90*	.89*
CS3	.92*	.92*
Security/privacy		
SEP1	.93*	.91*
SEP2	.91*	.91*
SEP3	.91*	.90*
Overall quality		
OQ1	.85*	.79*
OQ2	.85*	.84*
OQ3	.89*	.87*

Goodness-of-fit statistics

	A	B
χ^2	2496.651	3229.065
df	1629	1754
AIC	35493.972	36159.224
BIC	36743.755	36925.345
CFI	.941	.899
TLI	.933	.895
RMSEA	.039	.049
SRMR	.069	.100

* $p < .05$ (one-tailed).

a. These are standardized loading estimates from CFA using the Mplus software package. B. These are standardized path estimates.

TABLE 6
RESULTS OF THE STRUCTURAL EQUATION MODEL

Relationship	Indirect Effects Model		Direct and Indirect Effects Model	
	Estimate	R^2	Estimate	R^2
Website design → overall quality	.33*	68.9%	.29*	68.5%
Fulfilment → overall quality	.29*		.32*	
Customer service→ overall quality	.11*		.09*	
Security→ overall quality	.29*		.29*	
Website design → customer satisfaction	—	91.0%	.14*	91.1%
Fulfilment → customer satisfaction	—		-.08	
Customer service→ customer satisfaction	—		.08*	
Security→ customer satisfaction	—		.03	
Overall quality→ customer satisfaction	.95*		.84*	
Website design → repurchase intention	—	63.7%	.19*	65.0%
Fulfilment → repurchase intention	—		-.07	
Customer service→ repurchase intention	—		.07*	
Security→ repurchase intention	—		.02	
Overall quality→ repurchase intention	.80*		.65*	
Website design → word of mouth	—	63.1%	.19*	64.4%
Fulfilment → word of mouth	—		-.05	
Customer service→ word of mouth	—		.03	
Security→ word of mouth	—		.07	
Overall quality→ word of mouth	.79*		.62*	

Goodness-of-fit statistics

$\chi2$	3419.006	3397.901
df	2199	2187
AIC	40307.378	40306.042
BIC	41684.848	41729.943
CFI	.930	.931
TLI	.923	.923
RMSEA	.040	.040
SRMR	.069	.067

Chi-square difference test

$\Delta \chi2$	21.105*
df	12
p	.05

* $p < .05$ (one-tailed).

TABLE 7

COMPARING THE DEVELOPED MEASUREMENT WITH RESULTS OF META-STUDY

Relationship	Hierarchical model	Loiacono et al. (2002): WebQual	Parasuraman et al. (2005): E-S-Qual	Wolfinbarger and Gilly (2003): eTailQ
Website design → Overall e-service quality	.81** PW	–	.26**	.42**
Fulfilment → Overall e-service quality	.75** PW	–	.41**	.47**
Customer service → Overall e-service quality	.59** W	–	–	.14**
Security → Overall e-service quality	.75** PW	–	.09*	.04
All dimensions → Overall e-service quality	.89** PW	–	.26**	.28*
Website design → Customer satisfaction	.83** LPW	.71**	.38**	.33**
Fulfilment → Customer satisfaction	.66** PW	–	.24**	.45**
Customer service → Customer satisfaction	.66** PW	–	.21	.14
Security → Customer satisfaction	.76** PW	–	.44	.16
All dimensions → Overall e-service quality	.90** LPW	.71**	.35**	.27**
Website design → Repurchase intentions	.76** LPW	.67**	.45**	.47**
Fulfilment → Repurchase intentions	.60** PW	–	.33**	.22**
Customer service → Repurchase intentions	.55** PW	.47**	.32**	.21**
Security → Repurchase intentions	.63** LPW	.28**	.30*	.19
All dimensions → Overall e-service quality	.78** LPW	.63**	.38**	.28**
Average	.73	.58	.32	.27
Rank	1.	2.	3.	4.

* p < .05 (two-tailed); ** p < .01 (two-tailed).

Notes. The reported correlations are reliability adjusted and weighted correlations to ensure comparability of effect sizes across studies. Please not that these correlations differ from raw correlations reported, for instance, in Table 3. Superscripts (L/P/W) indicate that z-test is significant at p < .05. The employed z-test takes the sample size into account.
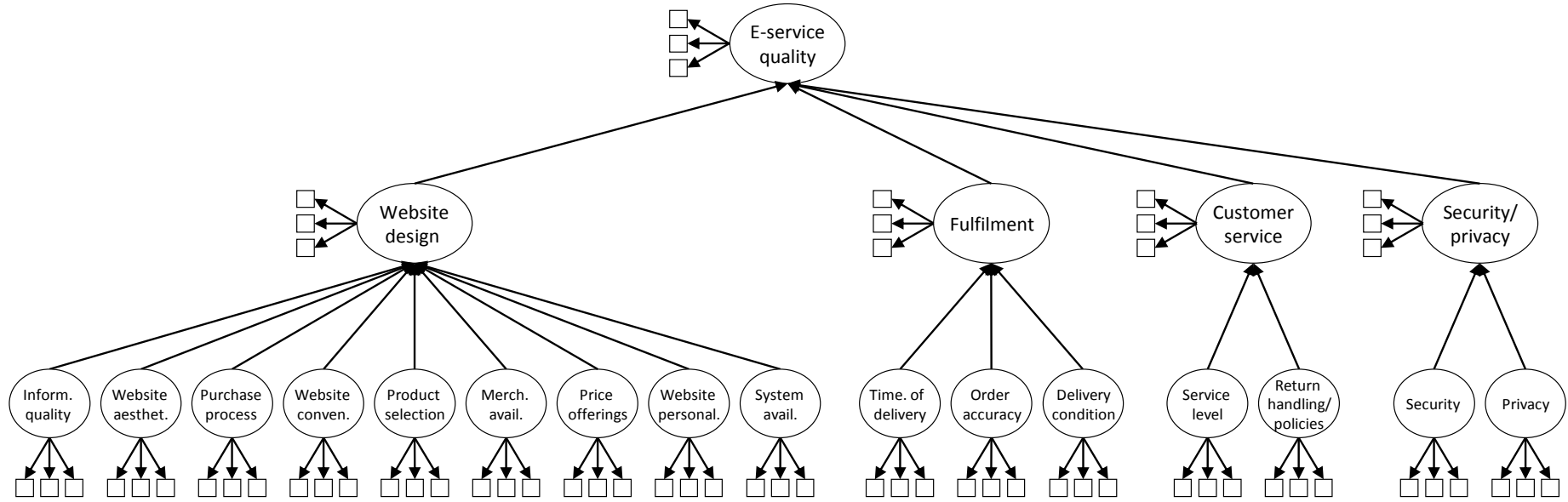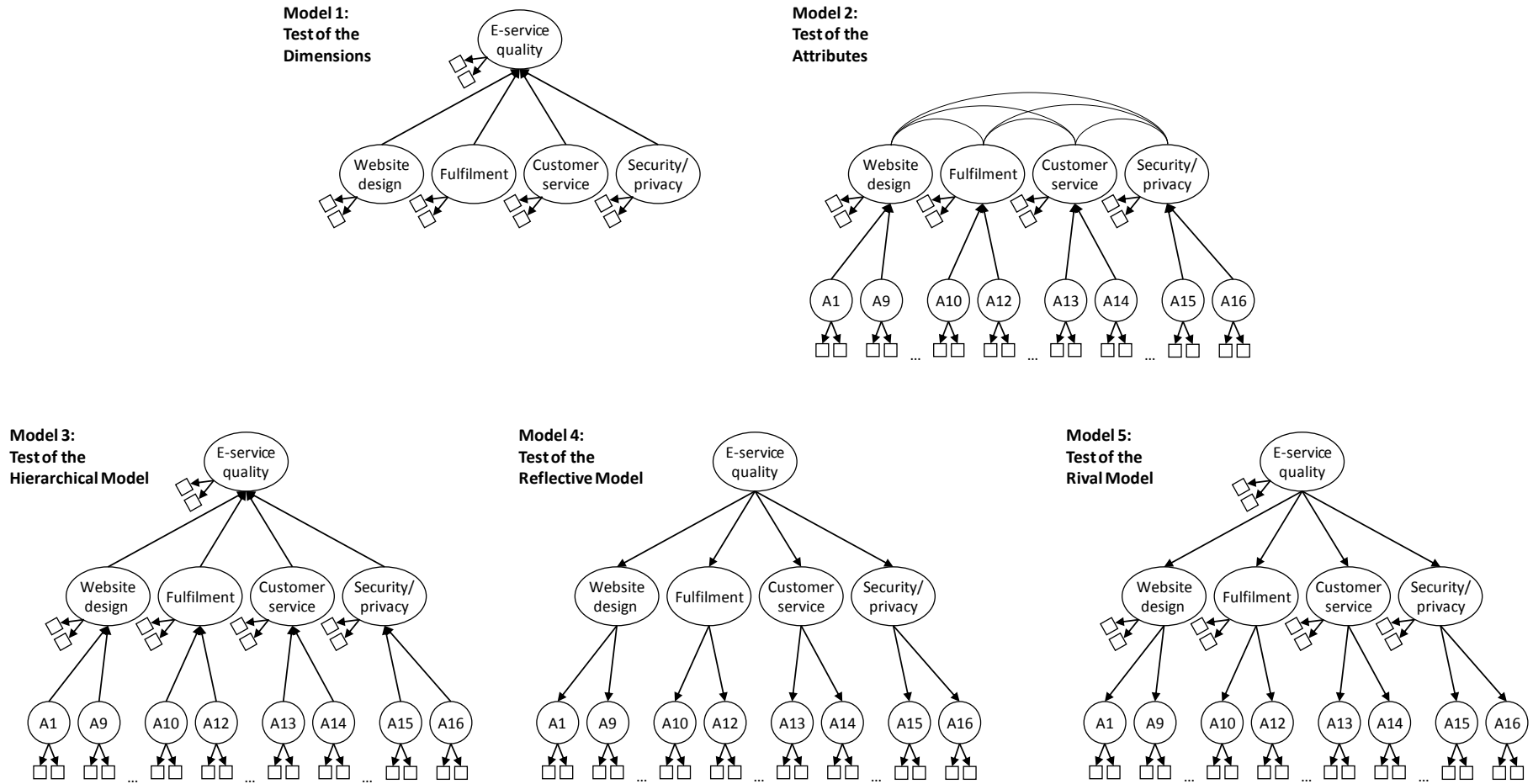
FIGURE 1
THE RESEARCH MODEL

FIGURE 2
STEP-WISE TESTING OF THE HIERARCHICAL MODEL

WEB APPENDIX A
DESCRIPTIVE STATISTICS AND CORRELATION OF E-SERVICE QUALITY WITH OUTCOME VARIABLES

Predictor Variable	Overall E-Service Quality				Customer Satisfaction				Repurchase Intention				Word-of-mouth			
	k	N	Aver. adj. r	Q	k	N	Aver. adj. r	Q	k	N	Aver. adj. r	Q	k	N	Aver. adj. r	Q
Website design	66	13,795	.489*	1,297*	133	15,616	.384*	1,967*	111	16,121	.461*	2,279*	17	2,828	.402*	325*
- Information quality	11	5,905	.593*	392*	28	7,534	.301*	286*	16	6,602	.418*	345*	5	986	.353*	68*
- Website organization	20	9,260	.522*	209*	22	9,390	.468*	283*	31	10,363	.399*	734*	4	1,717	.559*	28*
- Purchase process	3	904	.393*	18*	8	1,344	.293*	66*	11	2,232	.324*	141*	2	589	.376*	6
- Website convenience	9	4,942	.540*	214*	28	4,874	.453*	344*	14	3,637	.630*	206*	1	240	.536*	−
- Product Selection	2	753	.699*	−	7	2,805	.371*	235*	2	776	.485	−	2	450	.011	−
- Merchandise availability	1	271	.038	−	−	−	−	−	−	−	−	−	−	−	−	−
- Price offerings	1	472	.620*	−	12	1,198	.119	−	1	472	.647*	−	−	−	−	−
- Website personalization	6	2,380	.374*	22*	7	1,897	.412*	38*	7	2,240	.443*	46*	1	112	.490*	−
- System availability	7	2,586	.305*	86*	9	3,555	.446*	169*	13	4,290	.404*	250*	−	−	−	−
- Mixed measures	6	2,417	.402*	60*	12	3,641	.464*	202*	16	5,492	.566*	120*	2	589	.420*	−
Fulfillment	19	7,630	.529*	332*	26	7,876	.380*	252*	32	8,448	.297*	398*	5	2,138	.228*	38*
- Timeliness of delivery	3	411	.306*	4	2	280	.336*	−	2	431	.521*	−	−	−	−	−
- Order accuracy	−	−	−	−	3	459	.279*	5	−	−	−	−	−	−	−	−
- Delivery condition	1	1,258	.650*	−	−	−	−	−	−	−	−	−	−	−	−	−
- Mixed measures	15	7,459	.558*	305*	21	7,417	.399*	305*	30	8,257	.280*	371*	5	2,138	.228*	38*
Customer service	11	6,110	.414*	438*	43	7,090	.264*	649*	31	7,530	.371*	293*	6	1,553	.148	−
- Service level	11	6,110	.414*	438*	29	7,270	.327*	611*	28	7,211	.387*	279*	6	1,553	.148	−
- Return policies	−	−	−	−	14	1,608	.124*	7	3	1,073	.224*	2	−	−	−	−
- Mixed measures	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
Security	14	4,734	.342*	346*	32	8,863	.296*	450*	22	7,775	.316*	301*	5	1,927	.265*	93*
- Security	11	1,900	.357*	280*	14	4,638	.268*	262*	13	3,589	.307*	121*	4	892	.216*	42*
- Privacy	3	1,821	.291	61*	16	3,055	.290*	133*	8	3,151	.351*	154*	−	−	−	−
- Mixed measures	−	−	−	−	2	1,170	.358*	−	1	1,035	.145	−	1	1,035	.430*	−
All Predictors	110	13,795	.471*	2,950*	234	16,219	.351*	3,582*	196	16,504	.406*	3,583*	33	2,828	.312*	651*

Notes: k = number of correlations, N = total sample size across independent samples, Q = Q-statistic for homogeneity test. * indicates significant at *p* < .05. Operationally, Q-statistic was only calculated when there were a minimum of three correlations. A dash indicates that this condition was not met.