Available online at www.sciencedirect.com
ScienceDirect
Journal of Retailing and Consumer Services 14 (2007) 60–72
JOURNAL OF RETAILING AND CONSUMER SERVICES
www.elsevier.com/locate/jretconser

Measuring perceived service quality in urgent transport service

Laura Martínez Caro*, Jose Antonio Martínez García

Department of Management and Marketing, Polytechnic University of Cartagena, Cartagena, Spain

Abstract

In marketing literature, the study of service quality has focused on customer evaluation. The knowledge of consumer perceptions with regard to this complex construct constitutes powerful information for measuring company performance and for managing strategic projects. This fact is a critical factor in industries, such as the urgent transport service sector, in an environment in which a growth in the number of quality certifications has been prominent in recent years. The development of a valid tool for measuring perceived quality in this specific service has been proposed testing a hierarchical and multidimensional model, where service quality is a higher-order construct underlying four primary dimensions, which are defined by nine subdimensions. This conceptualization has been conceived after a qualitative research and the literature revision. The authors test this multilevel structure through a series of confirmatory factor analysis based on the partial disaggregation technique, and support the results with the cross-validation study. The implications and limitations of this research are discussed.

Keywords: Service quality; Scale development; Hierarchical conceptualization

1. Introduction

Since the eighties, service quality has been one of the most important issues in marketing literature and is considered as a vital element in management strategies in order to succeed and/or to outlive in competitive environments (e.g. Parasuraman et al., 1985, 1990). The pressure of competition in most industries has forced managers to look for ways to enhance their competitive position. Many have decided to improve service quality in order to differentiate their services from those of their competitors.

The growing importance of service quality is independent of the type of industry. Specifically, the concern for quality in the urgent transport sector is evident. In fact, the proportion of ISO certifications has increased seriously. Maybe this is due to the economic importance of the urgent transport industry in Spain. This industry has grown by 8% in recent years, and unlike other countries, the business is dominated by national companies. Therefore, service quality is turned into an essential competitive element.

The focus of this paper is to present an alternative approach to measure service quality in the urgent transport service. Many business organizations have felt the critical need to use a tool for evaluating service quality in order to appropriately assess and improve their service performance. After the review of the literature, we have not found any study that measures quality perception in this type of service. Therefore, it is necessary to develop a reliable and valid instrument to determine which aspects of a particular service define its quality. The proposed instrument incorporates performance-based measures on the basis of scales developed by Dabholkar et al. (1996) and Brady and Cronin (2001).

2. Conceptualization and operationalization of service quality

Service quality has been studied for a long time in the field of business management. However, the service literature suggests that there is no consensus on how to conceptualize or operationalize perceived service quality (Cronin and Taylor, 1992; Rust and Oliver, 1994). Two different perspectives have been adopted regarding this issue.

*Corresponding author. Tel.: +34 968 32 59 39; fax: +34 968 32 70 08.
E-mail addresses: laura.martinez@upct.es (L. Martínez Caro), martinez.jose@tiscali.es (J.A. Martínez García).

0969-6989/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jretconser.2006.04.001

The first perspective suggests that perceived service quality is based on the disconfirmation paradigm. Service quality is a comparison between consumers' expectations and their perceptions of the service they actually received (Grönroos, 1984; Parasuraman et al., 1985, 1990).

According to this approach, Grönroos (1984) developed *The Nordic Model*. Grönroos identified two dimensions to study service quality. The first dimension was *technical quality* that refers to the outcome of the service performance; the second dimension was *functional quality* that refers to the subjective perception of how the service is delivered. Later, Rust and Oliver (1994) added a third dimension: *physical environment* to Grönroos' (1984) model and proposed the *Tri-Component Model of Service Quality*.

Moreover, based on the disconfirmation paradigm Parasuraman et al. (1988) developed the SERVQUAL scale, where service quality is viewed as the result obtained from carrying out a comparison between expectations and perceptions of performance. Parasuraman et al. (1988) argued that, regardless of the type of service, consumers evaluate service quality using similar criteria, which can be grouped into five dimensions: tangibles, reliability, responsiveness, assurance, and empathy. These five dimensions were derived from 10 overlapping dimensions, which were regarded as essential to service quality by Parasuraman et al.'s (1985) exploratory research.

Despite SERVQUAL having been applied across a wide range of service contexts, it has been criticized on methodological and psychometric grounds by many marketing researchers (Carman, 1990; Babakus and Boller, 1992; Cronin and Taylor, 1992; Teas, 1993; Buttle, 1996).

Later, in further empirical research, Parasuraman et al. (1994) revamped SERVQUAL's structure to embody not only the discordance between perceived service and desired service (labeled as a measure of service superiority, or MSS), but also the discrepancy between perceived service and adequate service (labeled as a measure of service adequacy, or MSA).

Nevertheless, and despite the modifications of the SERVQUAL scale, other models emerged in order to question the conceptualization and measurement of service quality proposed by Parasuraman et al. (1988). Thus, Teas (1993) developed the Evaluated Performance (EP) model. EP model measures the gap between perceived performance and the ideal amount of a feature rather than the customer's expectations. Teas (1993) considered that this model could overcome some of the problems associated with the perceptions minus expectations gap conceptualization of service quality.

On the other hand, the second alternative perspective suggests that service quality should be measured considering only consumer perceptions rather than expectations minus perceptions (Carman 1990; Cronin and Taylor, 1992; McDougall and Levesque, 1994; Brady and Cronin, 2001). McDougall and Levesque (1994) consider that including expectation scores on a service quality instrument may be inefficient and unnecessary. This is due to the fact that people tend to indicate consistently high expectation ratings and their perception scores rarely exceed their expectations (Babakus and Boller, 1992). This reason has given rise to the development of an alternative scale of SERVQUAL, such as SERVPERF (Cronin and Taylor, 1992), the Retail Service Quality Scale (Dabholkar et al., 1996) or the Hierarchical and Multidimensional Model (Brady and Cronin, 2001).

According to Cronin and Taylor (1992) the performance-based measure was a better means of measuring the service quality construct. They developed SERVPERF model using the same 22 performance items from the Parasuraman et al. (1988) research. This measure explained more of the variance in an overall measure of service quality than did SERVQUAL. They also indicated that a psychometrically superior assessment of service quality (in terms of construct validity and operational efficacy) could be obtained through the SERVQUAL performance items alone.

Another alternative is the proposal by Dabholkar et al. (1996). These authors argued that SERVQUAL has not been successfully adapted to and validated for the retail store environment, suggesting that the dimensionality of service quality in a retail setting may not be similar to that of service quality in pure service industries. They proposed a hierarchical model of retail service quality, the Retail Service Quality Scale (RSQS). This scale is regarded as suitable for use in retail businesses which offer a mixture of service and goods, such as department or specialty stores (Dabholkar et al., 1996). The RSQS is a multilevel model, where retail service quality is viewed as a higher-order factor that is defined by two additional levels of attributes (dimensions level and subdimension level). The instrument includes five primary dimensions: physical aspects, reliability, personal interaction, problem solving and policy; and six subdimensions: appearance, convenience, promises, doing it right, inspiring confidence and courteous. Dabholkar et al. (1996) used only performance-based measures and found that their scale possessed strong validity and reliability, and adequately captured customers' perceptions of retail service quality.

Although the Dabholkar et al.'s (1996) study contributed to a greater understanding of service quality in certain retail stores, it was criticized, because it failed to investigate the relationship between customer perceptions of the quality of the products a retailer carries and customer perceptions of the service quality provided by the retailer (Finn and Kayandé, 1997).

Finally, Brady and Cronin (2001) combined the Rust and Oliver's model and Dabholkar's et al. (1996) hierarchical approach to develop a hierarchical and multidimensional model of perceived service quality.

According to Brady and Cronin (2001) customers form their service quality perceptions on the basis of an evaluation of performance at multiple levels and ultimately combine these evaluations to arrive at an overall service quality perception. They described a third-order factor

model, where quality service is explained in terms of three primary dimensions: interaction quality, physical environment quality and outcome quality. Each of these dimensions consists of the three corresponding subdimensions: attitude, behavior, and experience (interaction quality); ambient conditions, design, and social factors (physical environment quality); waiting time, tangibles and valence (outcome quality). This approach is believed to better explain the complexity of human perceptions than the conceptualizations currently offered in the literature (Dabholkar et al., 1996; Brady, 1997).

In short, the model proposed by Brady and Cronin (2001) allows us to know what defines service quality perceptions, how service quality perceptions are formed, and how important where the service experience takes place is.

3. Methodology

The literature review has not identified any study that measures the perception of the quality in urgent transport service and, therefore, there was no previous validated scale that we could use. It was therefore deemed valuable to develop a measurement instrument, in accordance with the procedure for scale development recommended by Churchill (1979).

3.1. Qualitative research

According to Chumpitaz and Swaen (2002) the number and the nature of service quality dimensions is in direct relation to the service under analysis. To accomplish this end, qualitative research was carried out to identify the factors which determine the service quality perceptions of the urgent transport consumers.

The qualitative research was conducted using in depth interviews of employees and consumers. In order to obtain information from the point of view of the companies, five managers from the industry were interviewed. With regards to demand, ten consumers with a high degree of experience in urgent transport service consumption were interviewed.

Following Brady and Cronin (2001) the respondents were encouraged to list all factors that influenced their perception according to their experience.

As pointed out by Brady and Cronin (2001) and Dabholkar et al. (1996) the price is eliminated from the decision set because price is not part of generally accepted understanding of service quality in the literature (Dabholkar et al., 1996). The literature clearly suggests that price is a determinant of service value (e.g. Zeithaml, 1988; Bitner and Hubbert, 1994).

3.2. Proposed factor structure

Carman (1990) suggests context customization may involve adding or dropping relevant dimensions and appropriately adjusting the attributes considered to make up those dimensions. Therefore, findings from qualitative research, together with the review of the quality literature have been conducted to propose the following model: a hierarchical and multidimensional model where quality is a higher-order factor that is defined by four primary dimensions and ten subdimensions (see Fig. 1).

The first dimension identified is "personal interaction". Several researchers have indicated the importance of this factor in the service delivery and it is regarded as having the most significant effect on service quality perceptions (Grönroos, 1982; Bitner et al., 1994). Both the service literature and our qualitative research suggest that "personal interaction" has four subdimensions: *attitude*, *behavior*, *expertise*, and *problem solving*. The first three subdimensions coincide with the proposal of Brady and Cronin (2001). According to Czepiel et al. (1985) and Grönroos (1990), attitudes, behavior, and skills of employees are factors that define the quality of service delivery.

Based on the qualitative study, we identified a fourth subdimension: *problem solving*. Several researchers have indicated the importance of this factor in service delivery. Dabholkar et al. (1996) and Terblanche and Boshoff (2001) suggest that this factor can be included in the "personal interaction" dimension. Furthermore, Westbrook (1981) found that consumers are quite sensitive to how service

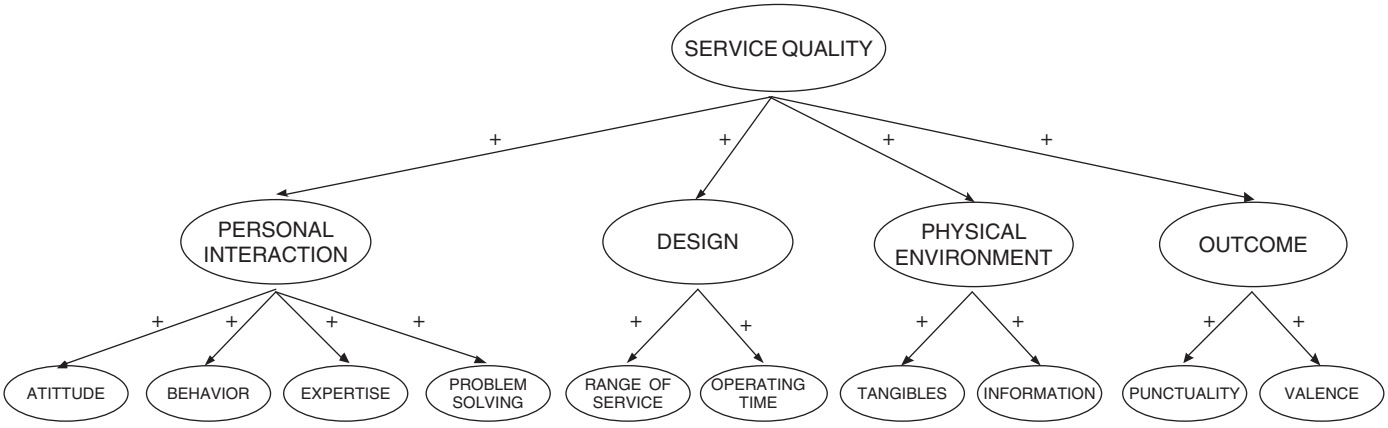

Fig. 1. Proposed scale for measuring perceived service quality.

providers attend to problems and complaints. Due to the peculiarities of the urgent transport service, where the contact is mainly carried out by telephone (90% of the contacts, according to the expert interviewed), this subdimension includes both physical contact and telephone.

The second dimension of service quality is called "design". Findings from our qualitative study reveal that is necessary to include this dimension. Following the research by Ko and Pastore (2005) this dimension is added to the model proposed by Brady and Cronin (2001). This dimension includes all aspects related to the configuration of the service and is defined by the next two subdimensions: *range of service* and *operating time*. The *range of service* has been considered in numerous studies (Kim and Kim, 1995; Howat et al., 1996; Chelladurai and Chang, 2000). It refers to the variety and attractiveness of services offered to consumers, whereas *operating time* evaluates whether the service provider has convenient hours (Howat et al., 1996; Ko and Pastore, 2005).

The third dimension we propose is "physical environment". Authors such as Bitner (1990, 1992), Spangenberg et al. (1996) and Brady and Cronin (2001) reflect on the importance of this factor in customer service evaluations.

As we mentioned earlier, urgent transport service has certain essential characteristics that are distinct from other services, but maybe the most outstanding one is that consumers do not need to visit the physical establishment to make a shipment. Consequently, aspects such as design or the store layout are not evaluated.

Therefore, in agreement with both the findings from our qualitative research and the study of Dabholkar et al. (1996), two subdimensions are identified to explain "physical environment": *tangibles* and *information*. Theory suggests that customers use any *tangible evidence* of the service outcome as a proxy for judging performance (Booms and Bitner, 1981; Zeithaml et al., 1985). According to Parasuraman et al. (1985) tangible evidence is a factor that service customers consider when forming quality perceptions.

Although initially this dimension is included by Brady and Cronin (2001) within the outcome dimension, it does not seem pertinent to insert this subdimension for explaining the outcome of the service, due to the peculiarities of urgent transport service where the final result is not carried out in a physical environment. Nevertheless, and given the importance that users concede to this, it seems more appropriate to include this subdimension within the "physical environment" dimension, as stated by Dabholkar et al. (1996).

In addition, the conclusion from the qualitative study indicated that consumers value the *tangible elements* more than the experts who had been interviewed. Some consumers pointed out that they valued both the transport technology and the uniform of the couriers as important.

The second subdimension identified in the qualitative research is *information*. This factor is proposed by Dabholkar et al. (1996), and concerns the ease of finding merchandise and moving within the store. As we pointed out earlier, in urgent transport service, consumers hardly visit the physical establishments, so the only way of easily and comfortably knowing about the variety of a company's products is via the phone or internet. The *information* subdimension is defined by Howat et al. (1996) as the feasibility of obtaining up-to-date information about services variety.

The last dimension of service quality is "outcome". There is agreement in the literature that the outcome of the service encounter significantly affects customer perceptions of service quality (Grönroos, 1984, 1990; McDougall and Levesque, 1994; Rust and Oliver, 1994; Carman 2000). This factor is labeled "technical quality" by Grönroos (1984, p. 37), who defines it as "what the consumer is left with when the production process is finished".

We used the findings from Brady and Cronin (2001) and our qualitative research to identify the subdimensions of this dimension: *punctuality* and *valence*.

Punctuality is the most cited factor by respondents in the qualitative research. This factor has considerable support in the literature. The findings of Parasuraman et al. (1985) showed that consumers identify service *punctuality* as an integral part of their overall evaluation. Houston et al. (1998) incorporated *punctuality* into their analysis of service quality encounters and found it to be an important predictor. Moreover, according to Taylor and Claxton (1994) and Brady and Cronin (2001) *punctuality* predicts a positive relationship: more favorable perceptions of *punctuality* are associated with enhanced outcome quality perceptions.

The last subdimension is *valence*. In line with Brady and Cronin (2001) *valence* captures attributes that control whether customers believe that service outcome is good or bad, regardless of their evaluation of any other aspect of the experience. For example, customers may have a positive perception of each service quality dimension, but the negative valence of the outcome can ultimately lead to an unfavorable service quality. Findings from Brady and Cronin (2001) are consistent with the results of our qualitative study, which reveals that *valence* is a key determinant of service outcome.

3.3. Generation of scale items and scale purification

A list of items was generated by adapting the items of existing generic scales (e.g. Parasuraman et al., 1988; Dabholkar et al., 1996; Brady and Cronin, 2001; Terblanche and Boshoff, 2001; Ko and Pastore, 2005). On the basis of the review of the service literature, we generated an initial pool of 68 items. According to Dabholkar et al. (1996) we used a 5-point Likert-type scale (1 = strongly disagree, 5 = strongly agree) for respondent's responses.

The purification of scale was carried out in two steps. The first step was an assessment of content and face validity through a panel of experts and a field test (Ko and

Pastore, 2005). The panel members were five executives from urgent transport companies. Inputs from the panel resulted in 26 items being removed.

After this revision, the second step was to develop a questionnaire with the 42 remaining items. This questionnaire was pilot-tested among 50 respondents (users of urgent transport service). According to Parasuraman et al. (1988) the purification of the instrument begins with the computation of Cronbach's alpha coefficient, item-to-total correlation and exploratory factor analysis for the ten subdimensions. Items with very low correlations and/or those correlations that produced a sharp drop in the plotted pattern were discarded. After that, we examined the dimensionality of the scale in order to prove that the factor patterns that emerged were independent of one another.

Based on the results of the pilot-test, some items were dropped and others were reworded to avoid confusion. The final instrument had a total of 38 items reflecting 10 subdimensions of transport urgent service quality. The complete questionnaire is depicted in the Appendix A.

3.4. Data collection

The sample was collected from a customer population using the particular service in the Región de Murcia which is in the south-east of Spain. An initial sample of 400 respondents was obtained, although 25 questionnaires were dropped due to non-valid responses. The final sample was of 375 subjects and the data was representative of the heterogeneity of the individuals selected for the research design. Thus, the sample was representative of the population with regard to demographic questions such as sex, age, occupation and incomes. Additionally, the final distributions of responses concerned with the consumption experience, showed balanced groups of low, medium and high levels of urgent transport service use.

Data was obtained through a personal interview with the individuals during which a self-administered questionnaire was used. Trained business and marketing students were entrusted to filter the individuals who fulfilled the demographic quotas and the condition of having used the service within the previous 12 months.

The final sample was split by random procedure (Steenkamp and Baumgartner, 1998), using one part to test the model and the other part to validate the solution obtained from the first portion (Anderson and Gerbing, 1988). A calibration sample of 200 was finally randomly selected in order to test the model with a critical sample size[1] (Hair et al., 1999).

[1]The initial research was designed to obtain two samples of 200 individuals in order to calibrate the data with the first half and validate the results with the second, as Anderson and Gerbing (1988) recommended. However, when the non-valid questionnaires were subtracted, we decided to maintain the first number of responses—200—to test the model, following the recommendations of Boomsma and Hoogland (2001) who propose this number of individuals as a sufficient sample size for acceptable bias of parameter estimates.

4. Results

4.1. Exploratory assessment of the measures

The items of the proposed model were evaluated with exploratory techniques for assessing the reliability and dimensionality of the measures. In the first stage, each subdimension was assessed using the item-to-total correlation, Cronbach's alpha and exploratory factor analysis. The decision to retain items was based on recommendations proposed by Nurosis (1993), Nunnally (1978) and Hair et al. (1999) with regard to statistical criteria,[2] and no-one item was eliminated. During the second stage, a maximum likelihood exploratory factor analysis with oblique rotation was computed for each dimension in order to test the factor structure of these constructs. The underlying factors for the "design", "physical environment" and "outcome" dimensions supported the proposed structure with factors loading exceeding 0.54 on its hypothesized factor and communality above 0.40. For the "interaction personal" dimension two items of the "behavior" subdimension were dropped because of their poor correlation with underlying factors and the low communality. The configuration matrix showed that the other two items of the subdimension *behavior* were highly correlated with *attitude,* thus the new proposed structure was composed by three manifest constructs representing "personal interaction", where the new concept labeled *conduct* was perceived by the customer as a composite of attitudes and behaviors of the employees (Capelleras and Veciana, 2002). A reliability and dimensionality analysis for the new subdimension was achieved and the results supported the significance of the measures with a lower item-to-total correlation and factor loading of 0.70 and 0.78, respectively, and a coefficient alpha of 0.90.

4.2. Test of the third-order proposed structure

A series of confirmatory factor analysis were estimated to confirm the proposed structure, using the covariance matrix as input via LISREL 8.50 (Jöreskog and Sörbom, 2001) maximum likelihood (ML) method. The two step approach procedure (Anderson and Gerbing, 1988) was developed to test the structure in a separate estimation of the measurement model prior to the simultaneous estimation of measurement and structural models. Thus, the first stage in this procedure was to confirm the quality of the measures assessing the psychometric properties by confirmatory factor analysis of the subdimensions. The descriptive statistics obtained for the items and other diagnostic statistics are presented in Table 1, and the construct correlation matrix is showed in Table 2.

[2]Item-to-total correlation above 0.30 (Nurosis, 1993); Cronbach's alpha above 0.70 (Nunnally, 1978); factors loading exceeding 0.40 considering this sample size (Hair et al., 1999).

Table 1
Measurement results of first-order confirmatory factor analysis with total and partial disaggregation

	Mean	Std. dev.	Standardized loading		Minimum t-value		Composite reliability		AVE	
			TD[a]	PD[b]	TD	PD	TD	PD	TD	PD
Conduct	3.33	0.80	0.71–0.85	0.92–0.94	11.20	18.03	0.90	0.93	0.61	0.87
Expertise	3.39	0.75	0.68–0.79	0.76–0.90	10.34	11.65	0.84	0.82	0.57	0.70
Problem solving	3.12	0.77	0.70–0.80	0.81–0.90	10.73	12.99	0.85	0.85	0.58	0.74
Range of service	3.75	0.83	0.73–0.86	0.96–0.97	11.35	19.48	0.84	0.96	0.64	0.93
Operating hours	3.40	0.86	0.66–0.90	0.86–0.91	9.97	14.87	0.84	0.88	0.58	0.79
Tangibles	3.12	0.80	0.62–0.83	0.75–0.86	9.06	11.06	0.78	0.80	0.55	0.66
Information	3.13	0.85	0.70–0.79	0.74–0.79	10.60	11.46	0.81	0.74	0.53	0.59
Punctuality	3.36	0.95	0.79–0.90	0.89–0.91	13.18	17.37	0.91	0.90	0.72	0.82
Valence	3.31	0.83	0.77–0.85	0.91–0.92	12.63	16.99	0.89	0.92	0.68	0.85

[a]Total disaggregation.
[b]Partial disaggregation.

Table 2
Construct correlation matrix (Φ) with total and partial disaggregation

	CON	EXP	PRS	RAN	OPH	TAN	INF	PUN	VAL
Conduct (CON)	1.00	0.75	0.68	0.57	0.58	0.52	0.72	0.70	0.74
Expertise (EXP)	0.76	1.00	0.74	0.62	0.54	0.62	0.68	0.69	0.79
Problem solving (PRS)	0.68	0.74	1.00	0.46	0.58	0.55	0.72	0.70	0.77
Range of service (RAN)	0.62	0.67	0.50	1.00	0.55	0.52	0.60	0.61	0.62
Operating hours (OPH)	0.59	0.53	0.57	0.61	1.00	0.54	0.71	0.61	0.70
Tangibles (TAN)	0.54	0.61	0.57	0.57	0.52	1.00	0.79	0.69	0.76
Information (INF)	0.66	0.64	0.71	0.55	0.64	0.81	1.00	0.85	0.86
Punctuality (PUN)	0.70	0.68	0.68	0.65	0.60	0.70	0.76	1.00	0.87
Valence (VAL)	0.77	0.80	0.77	0.67	0.70	0.77	0.79	0.87	1.00

Note: Correlations with total disaggregation are presented in the lower triangle of the matrix and correlations with partial disaggregation are given in the upper triangle.

The model fit was evaluated using the most stable and robust approximate fit indices following Gerbing and Anderson (1992): DELTA2 (Bollen, 1989), the comparative fit index (CFI) (Bentler, 1990) and the relative noncentrality index (RNI) (McDonald and Marsh, 1990). Additionally, other fit indices were used for evaluative purposes: the χ^2 statistic, the root-mean-square error of approximation (RMSEA) (Steiger, 1990), and the parsimonious fit index (PNFI) (James et al., 1982). Although χ^2 is recommended with moderate samples (e.g. 100–200; Tabachnick and Fidell, 1996), others fit indices are needed to cover the well-known shortcomings of this index (e.g. Hair et al., 1999; Gefen et al., 2000). Hence, the properties of DELTA2, RMSEA, RNI-CFI, and PNFI with regard to: the slight bias by sample size (Bollen, 1989), the noncentrality distribution in the construction of the index (Curran et al., 2002), the association with specific population parameters (Medsker et al., 1994), and the appropriateness for comparing models with different degrees of freedom (Hair et al., 1999), respectively, justify their selection. The criteria for assessing the indices were established following the recommendations of: (1) Bagozzi and Heatherton (1994) (RNI > 0.90); (2) Kline (1998) (CFI > 0.90); (3) Hu and Bentler (1999) (RMSEA < 0.08); (4) Widaman and Thompson (2003) (DELTA2 > 0.95); (5) and Hair et al. (1999) (difference of PNFI between two models above 0.09). Additionally, the Satorra–Bentler scaled χ^2 (Satorra and Bentler, 1988) was computed to better approximate χ^2 under non-normality, in order to correct the standard errors affected by the non-normality data.[3]

The indices showed in Table 3 indicate that the comprehensive model fits the data well, and the examination of the modification indices does not suggest the need for any change or modification of the model. The internal consistency of each scale were also evaluated; standardized loadings of individual items were highly significant and the values were larger than the recommended threshold of 0.70 (Steenkamp and Van Trijp, 1991) except for three items. Composite reliability (Bagozzi and Yi, 1988) and average variance extracted (AVE) indices (Fornell and Larcker, 1981) presented values higher than the evaluation criteria of 0.6 (Bagozzi and Yi, 1988). The convergent validity of

[3]Previous analysis of the data distribution with PRELIS shows the excessive skewing and moderate kurtosis of the different covariance matrices provided (Kline, 1998). The asymptotic covariance matrices were generated to obtain the scaled χ^2 and robust estimation of standard errors.

Table 3
Summary results of confirmatory factor analysis for the proposed factor structure of urgent transport perceived service quality

	$(\chi^2)^a$	(S–B χ^2)[b]	df	RMSEA (90% IC)	CFI; RNI[c]	DELTA 2	PNFI
Calibration sample							
Test of the subdimensions (TD)	939.45	—[d]	558	0.055 (0.048, 0.062)	0.921	0.922	0.734
Test of the subdimensions	149.48	119.60	99	0.032 (0.000, 0.051)	0.983	0.984	0.617
Test of second-order factor	181.88	143.49	120	0.031 (0.000, 0.049)	0.980	0.980	0.740
Test of third-order factor	206.13	163.37	123	0.040 (0.021, 0.056)	0.973	0.973	0.752
Validation sample							
Test of the subdimensions (TD)	1155.38	—	558	0.076 (0.070, 0.083)	0.850	0.848	0.660
Test of the subdimensions	198.12	155.48	99	0.057 (0.039, 0.074)	0.954	0.943	0.591
Test of second-order factor	242.89	192.95	120	0.059 (0.043, 0.074)	0.943	0.940	0.701
Test of third-order factor	253.72	201.27	123	0.061 (0.045, 0.075)	0.938	0.938	0.714

[a]Minimun fit function χ^2.
[b]Scaled Satorra–Bentler χ^2.
[c]When RNI is between 0 and 1 (all models tested). CFI equals RNI.
[d]The scaled Satorra–Bentler χ^2 was not computed cause of the asymptotic covariance matrix was not generated by PRELIS due to the small sample size.

the measures was highly supported by the significant item *t*-value, all items loaded on their proposed hypothesized dimensions and the parameters estimates were 10–20 times as large as the standard errors (Anderson and Gerbing, 1988). Discriminant validity was assessed by calculating the shared variance between pairs of constructs (Φ^2) and verifying that it was lower than the average variances extracted for the individual constructs (Fornell and Larcker, 1981). The shared variances between pairs of all possible scale combinations indicated that the variances extracted were higher than the associated shared variances in all cases (see Tables 1 and 2) except in seven. In the interest of thorough discriminant validity an additional test was conducted, supporting this assumption since the confidence interval (± 2 SE) around the correlation estimate between any two latent indicators never includes 1.0 (Anderson and Gerbing, 1988).

4.3. Nested models sequence with partial disaggregation

Once the adequate psychometric properties of the scales were assessed, the next step was to evaluate the proposed structure with the development of a series of tests for the hierarchical model, using the partial disaggregation procedure (see Bagozzi and Heatherton, 1994; Dabholkar et al., 1996). Although total disaggregation allows the most explicit test of construct quality (Baumgartner and Homburg, 1996), the partial disaggregation method is achieved to decrease the number of parameter estimates of nearly 5:1 respect to sample size, as recommended by Bagozzi and Edwards (1998), to reduce the likelihood of computational problems and to obtain smaller standard errors (Perugini and Conner, 2000). The rationale for this method is to avoid the difficulties associated with categorical item level-data and to achieve a higher level of reliability for each of the scores on which the confirmatory factor analysis were based than would be realized from responses on each of the individuals items. Previous literature provides additional support for this approach (e.g. Bagozzi and Heatherton, 1994; Gibbons and Hocevar, 1998; Paik and Michael, 1999). Operationally, partial disaggregation is accomplished by randomly aggregating items that relate to a given construct (subdimension) where there are two composite measures of each concept. The basis for random combination of items is that all items or indicators related to a latent variable should correspond in the same way to that latent variable; thus any combination of these items should yield the same model fit (Dabholkar et al., 1996) (see Appendix A). The psychometric results of the previous total disaggregation analysis support the use of this procedure.

The results of first-order confirmatory factor analysis with partial disaggregation yielded an excellent fit (Table 3). As for the total disaggregation model, the same analysis of internal consistence and construct validity were achieved and the results supported the adequacy of the measures (Table 1). Only three cases of the test of discriminant validity with the Fornell and Larcker procedure were not confirmed, although they are supported in terms of the confidence interval test (Anderson and Gerbing, 1988). The high correlations between the subdimensions suggested the presence of second-order factors to explain this common variance. A second confirmatory factor analysis was computed using the same indicators for the subdimensions, establishing relationships between the subdimensions (first order) and their respective theoretical dimension (second order). The assessment of the fit indices supported the adequacy of second-order structure (CFI, RNI and DELTA2: 0.980) (Table 3), and all paths were statistically significant. In order to provide greater confidence in this model a χ^2 difference test was undertaken. Anderson and Gerbing (1988) recommend this procedure to compare nested models; a nonsignificant χ^2 difference test would lead to the acceptance of more parsimonious. Due to the departures from multivariate normality, the scaled χ^2 was used to compare models, computing the difference of

scaled χ^2 statistics proposed by Satorra and Bentler (1999). The results of the test showed a nonsignificant change (difference of scaled χ^2: 23.98; df: 21; p: 0.29), thus the second-order structure was a better reproduction of the observed data, and supported the theoretical structure of relationships among dimensions and subdimensions. Nevertheless, the conceptual framework considers quality dimensions as manifestations of a higher-order construct: the perceived service quality. Examining the correlations among the dimensions in the second-order model (from 0.84 to 0.94), and considering the strong theoretical evidences of a third-order structure that captures the common variances shared by all dimensions, a higher-order construct should be proposed to conform the hierarchical model of three-level service quality. As Table 3 shows, the model yielded an excellent fit (CFI, RNI and DELTA2: 0.973). All the paths depicted in the research model were supported, and the t-values of the paths were positive and significant ($p \leqslant 0.001$). The standardized solution of the proposed relation between the concepts has been reflected in Fig. 2. A new scaled χ^2 difference test was undertaken for the third- and the second-order model. The results of the test indicated significant differences between the models (difference of scaled χ^2: 31.90; df: 3; $p < 0.001$). Although the χ^2 difference test is the most common method of examining the difference between nested models, other fit indices were also considered in order to evaluate overall model fit (Woehr et al., 2003). Further to the indices mentioned above, the parsimony gain of the third-order model was significant (difference of PNFI above 0.09) and the RMSEA index presented a low value (0.040). The high correlation between dimensions, the strong theoretical framework and the excellent fit of the model, support the third-order structure of perceived service quality in urgent transport service. Given acceptable convergent and discriminant validities, the test of the structural model then constitutes a confirmatory assessment of nomological validity of the perceived service quality construct (Anderson and Gerbing, 1988).

5. Cross-validation study

The result reported for the scale using the calibration data might be upwardly biased because they are based on the sample on which the instrument was developed (Baumgartner and Steenkamp, 1996). A second study with the second part of the initial sample (175 individuals) was achieved to validate the findings. The development of the analysis was accomplished following the same stages of the first study. The model fit obtained was acceptable to each analysis (Table 3). Furthermore, for both samples, all factor loadings were significant and positive, and the composite reliabilities and the average variances extracted were above the levels recommended. Thus, the validation study also supports the entire hierarchical factor structure illustrated in Fig. 2.

6. Discussion

The perceived service quality concept is a complex construct that has created a large debate in the scientific literature regarding its conceptualization and measurement. The need to develop specific measurement tools for different services (Carman, 1990; Chumpitaz and Swaen, 2002), contradicts the proposal of universality of the dimensions provided by the SERVQUAL perspective (Parasuraman et al., 1988). The qualitative study of this research supports this view; customers make their judgments of service quality on the basis of a series of factors that are specific to the evaluated service. The concept of hierarchical structures of perceptions (Dabholkar et al., 1996; Brady and Cronin, 2001) has been adopted for developing the measurement instrument, and has been confirmed with the qualitative study and the empirical results. Hence, customers base their evaluation of the primary dimensions on the assessment of the corresponding subfactors. The combination of all these constitutes a customer's overall perception of service quality (Brady and Cronin, 2001). Thus, perceived service quality in urgent

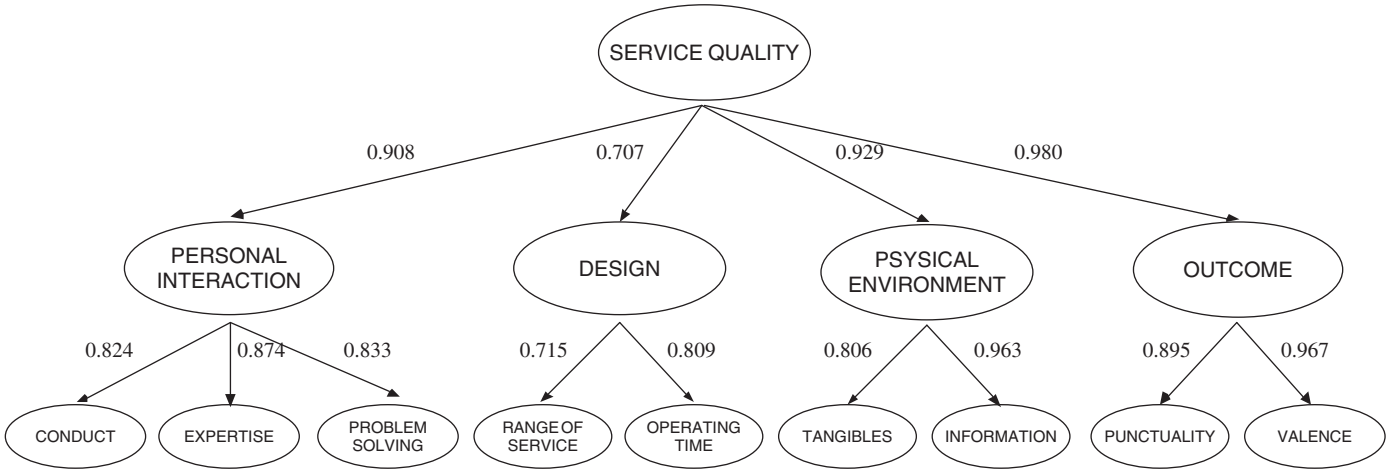

Fig. 2. Final hierarchical structure of service quality in urgent transport industry (standardized loadings).

transport service has a third-order factor structure, with four specific dimensions and nine specific subdimensions, in a multi-level hierarchical structure where the factors in the same level are highly correlated. The correlation between factors was indicated by Parasuraman et al. (1988, 1991) as a key issue for futures studies about conceptualization of service quality. Our proposed scale presents high indices of correlation between concepts on each level. To the extent that the first-order factors include common variance, the second-order factors captures the shared variance across factors (Bagozzi and Heatherton, 1994), and the same occurs with the common variance of the second-order dimensions and the third-order factor. Nevertheless, as pointed out by Bagozzi and Heatherton (1994), discriminant validity among factors may be less defensible and the dimensions should be interpreted as subcomponents of a higher-order concept. Hence, the conclusions about discriminant validity may be ambiguous when drawing the line between high and moderate correlations. The empirical evidence supports the rest of the psychometrical properties of the scale, its internal consistency, convergent and nomological validity. The paths in the research model are all confirmed, which indicates that each factor is appropriately conceived as an aspect of service quality. Finally, the cross-validation study makes the findings more robust.

Our test of the third-order structure assumes that the four dimensions are different forms manifested by perceived service quality. Likewise, these four dimensions are explained by their respective subdimensions, and finally the indicators are observable representations of each subdimension. This view agrees with the taxonomy of multidimensional constructs under latent model by Law et al. (1998), and it reproduces the proposal of Dabholkar et al. (1996). Nevertheless, this does not coincide with the Brady and Cronin (2001) perspective of hierarchical structure, in spite of the fact that they base their structure on Dabholkar's model. Thus, Brady and Cronin (2001) add items to represent the primary dimensions and the overall service quality, as an implicit assumption of the dimensions as antecedents of service quality. This conceptualization has been adopted by later investigations (see Fullerton, 2005). According to our research proposal, the dimensions are not antecedents of service quality but rather expressions of the complexity of the construct. Service quality is a higher-order factor underlying the dimensions. Thereby changes in the perceived service quality produced by the variation in the perception of one dimension will affect the perception of the rest of dimensions due to the correlation among them.

The nested model sequence for testing the hierarchical structure differs from the methodology of Dabholkar et al. (1996) and Brady and Cronin (2001). Although Dabholkar et al. (1996), did not test the global third-order model, both tested the primary dimensions considering the global service quality as a higher-order factor, in an underlying assumption of unidimensionality of those dimensions. This fact constitutes a contradiction regarding their theoretical explanation of multidimensionality of the dimensions. If customers form their perceptions of service quality on the basis of the low level of abstraction (subdimensions), the sequence of analysis must be accomplished from the low to the high level of abstraction, using the nested sequence to compare multi-levels alternatives (e.g. Marsh and Hocevar, 1985).

7. Managerial implications

This hierarchical conceptualization can help managers to understand how customers assess service encounters. They can obtain the judgments of customers about each of the key aspects of the service, from the most disaggregated level (subdimensions) to the unique value of global service quality. Thus, it is a valuable strategic tool to know about the weakness and strength of their performance. In an industry where the ISO certification is a quality management system adopted by several companies, this instrument could serve as a diagnostic tool that will allow managers to determine service areas that are weak and in need of attention. In addition, the analysis of the paths determine the contribution of each concept to its respective higher-order factor, therefore managers can obtain the key dimensions of perceived service quality and, from an aggregate perspective, they can find the determinants of industry competitiveness.

The findings of this study show the outcome as a principal determinant of service quality. Customers give high importance to the service results; "…the shipment has to arrive untouched and punctual to the destiny". Therefore companies can establish priorities in the decisions related to service improvement, considering the global evaluation in each crucial factor, and adjusting their communication strategies to differentiate their service on the basis of main drivers of consumer quality perceptions.

Managers must consider the hierarchical structure of service quality and the high correlation between factors. Low levels of perception in one of these subdimensions highly influence the perception of the rest of factors.

As pointed out by Brady and Cronin (2001), high level of service quality is associated with several key organizational outcomes, including high market share (Buzzell and Gale, 1987), improved profitability relative to competitors (Kearns and Nadler, 1992), enhanced customer loyalty (Zeithaml et al., 1996), the realization of a competitive price premium (Zeithaml et al., 1996), and an increased probability of purchase (Zeithaml et al., 1996). Furthermore, service quality is positively related to customer satisfaction (Cronin and Taylor, 1992; Anderson and Sullivan, 1993) and corporate image (Grönroos, 1984), though the causal order of these relationships has produced controversy. Thereby, the study of service quality can provide companies with a powerful instrument for obtaining their strategic goals.

8. Limitations and further research

This study has several limitations. First, the data were gathered in a specific geographic area of Spain, and the results can be specific for this area. With the aim of enhancing the efficiency of this tool, companies should augment the research scope, in order to obtain a representative sample of their target. Second, the cross-validation results support the hypothesized model, but with lower approximate fit indices. In addition, the cross-validation sample was collected from the same population. In order to generalize the proposed model, further studies should replicate this model in other populations. Third, the relationships with other constructs have not been analyzed in this study. Future research could focus on the determination of the antecedents of service quality and how customers form their perceptions. In this sense, the antecedents proposed by Parasuraman et al. (1985), should be studied in depth. Finally, the need of structural equation modeling to obtain the paths could be an obstacle for managers for using this tool in the most powerful way.

Acknowledgments

The authors acknowledge the helpful comments of the editor and reviewers on previous drafts of this article.

Appendix A. The survey items

Attitude
1. I can count on the employees being friendly
2. The attitude of employees demonstrates their willingness to help me
3. The employees search for what is in the best interests of the clients
4. The employees show their interest in accelerating the shipment

Behavior[a]
5. I can count on employees taking actions to address my needs
6. The behavior of the employees gives me trust
7. The employees are never too busy to respond to consumer's requests
8. This service has employees who give me individual attention

Expertise
9. Employees are able to answer my questions quickly
10. The employees understand that I rely on their knowledge to meet my needs
11. The employees understand consumers' specific needs
12. The employees are competent

Problem solving
13. When a customer has a problem, this company shows a sincere interest in solving it
14. The company is conscious of how important that the resolution of the complaints is for me
15. The company has a fair system for the handling of the complaints
16. The employees are able to handle customer complaints directly and immediately

Range of service
17. This company offers a wide range of shipment
18. The variety of shipment services offered by this company fit my needs
19. The types of shipment offered by this company are attractive to me

Operating time
20. The operating hours of this company are convenient
21. The shipment hours are convenient to my needs
22. This company has operating hours convenient to all its customers
23. The schedule of pick up fit my needs

Information
24. The company gives information on all the services that it offers
25. It is quick and easy to contact the company

26. The employees inform punctually and sincerely about all the conditions of service
27. This company informs appropriately about its promotions

Tangibles
28. The equipment at this store is visually appealing
29. The service has modern-looking electronic equipment
30. Employees have a neat and professional appearance

Punctuality
31. This company endeavors that the shipment arrive as soon as possible
32. The service provider understands that waiting time is important to me
33. This company provides service to the consumers punctually
34. This company provides its services at the time it promises to do so

Valence
35. When the service finishes, I usually feel that I have had a good experience
36. I believe this company tries to give me a good experience
37. I believe this company knows the type of experience its customers want
38. When I leave this company, I always feel that I got what I wanted

Note: The composite indicators with partial disaggregation are: I_1, P1 + P3 + P5; I_2, P2 + P4 + P6; I_3, P11 + P12; I_4, P9 + P10; I_5, P14 + P15; I_6, P13 + P16; I_7, P19; I_8, P17 + P18; I_9, P20 + P22; I_{10}, P21 + P23; I_{11}, P24 + P26; I_{12}, P25 + P27; I_{13}, P29; I_{14}, P28 + P30; I_{15}, P31 + P34; I_{16}, P32 + P33; I_{17}, P36 + P37; I_{18}, P35 + P38.

[a]The items 7 and 8 were dropped after the exploratory analysis, and a new subdimension labeled *conduct* were named as a composite of *attitude* and *behavior*.

References

Anderson, J.C., Gerbing, D.W., 1988. Structural equation modeling in practice: a review and recommended two-step approach. Psychological Bulletin 103 (3), 411–423.

Anderson, E., Sullivan, M.W., 1993. The antecedents and consequences of customer satisfaction for firms. Marketing Science 12 (2), 125–143.

Babakus, E., Boller, G.W., 1992. An empirical assessment of the Servqual scale. Journal of Business Research 24, 253–268.

Bagozzi, R.P., Edwards, J.R., 1998. A general approach for representing constructs in organizational research. Organizational Research Methods 1, 45–87.

Bagozzi, R.P., Heatherton, T.F., 1994. A general approach to representing multifaceted personality constructs: application to state self-esteem. Structural Equation Modeling 1 (1), 35–67.

Bagozzi, R.P., Yi, Y., 1988. On the evaluation of structural equation models. Journal of Academy of Marketing Science 16 (1), 74–94.

Baumgartner, H., Homburg, C., 1996. Applications of structural equation modeling in marketing and consumer research: a review. International Journal of Research in Marketing 13 (2), 139–161.

Baumgartner, H., Steenkamp, J.B., 1996. Exploratory consumer buying behaviour: conceptualization and measurement. International Journal of Research in Marketing 13, 121–137.

Bentler, P.M., 1990. Comparative fit indexes in structural models. Psychological Bulletin 107, 238–246.

Bitner, M.J., 1990. Evaluating service encounters: the effects of physical surrounding and employees responses. Journal of Marketing 54 (2), 69–81.

Bitner, M.J., 1992. Servicespaces: the impact of physical surroundings on customers and employee. Journal of Marketing 56, 57–71.

Bitner, M., Hubbert, A., 1994. Encounter satisfaction versus overall satisfaction versus quality. In: Rust, R., Oliver, R. (Eds.), Service Quality: New Directions in Theory and Practice. Sage Publications, Thousand Oaks, CA, pp. 72–94.

Bitner, M.J., Booms, B.H., Mohr, L.O., 1994. Critical service encounters: the employee's view. Journal of Marketing 58, 95–106.

Bollen, K.A., 1989. A new incremental fit index for general structural equation models. Sociological Methods and Research 17, 303–316.

Booms, B.H., Bitner, M.J., 1981. Marketing strategies and organization structures for service firms. In: Donnelly, J.H., George, W.R. (Eds.), Marketing of Services. American Marketing Association Chicago, pp. 47–52.

Boomsma, A., Hoogland, J.J., 2001. The robustness of Lisrel modelling revisited. In: Cudeck, R., Stephen Toit, S., Sörbom, D. (Eds.), Structural Equation Modelling: Present and Future. A Festschrift in Honor of Karl Jöreskog. Scientific Software International, Chicago, pp. 139–168.

Brady, M.K., 1997. Recoceptualizing Perceived Service Quality: Hierarchical Model. Unpublished Doctoral Dissertation. The Florida State University, Tallahassee.

Brady, M.K., Cronin, J.J., 2001. Some new thoughts on conceptualizing perceived service quality: a hierarchical approach. Journal of Marketing 5, 34–49.

Buttle, F., 1996. SERVQUAL: review, critique, research agenda. European Journal of Marketing 30 (1), 8–25.

Buzzell, R.D., Gale, B.T., 1987. The PIMS Principles. The Free Press, New York.

Capelleras, J.L., Veciana, J.M., 2002. Service quality in university education: an empirical assessment. In: 31st European Marketing Academy Conference, Braga, Portugal.

Carman, J., 1990. Consumer perceptions of service quality: an assessment of the SERVQUAL dimensions. Journal of Retailing 66, 33–35.

Carman, J., 2000. Patient perceptions of service quality: combining the dimensions. Journal of Services Marketing 49 (1), 57–65.

Chelladurai, P., Chang, K., 2000. Targets and standards of quality en sports services. Sport Management Review 3, 1–22.

Chumpitaz, R., Swaen, V., 2002. Service quality and brand loyalty relationships: investigating the mediating effect of customer satisfaction. In: 31st European Marketing Academy Conference, Braga, Portugal.

Churchill Jr., G.A., 1979. A paradigm for developing better measures of marketing constructs. Journal of Marketing Research 16 (1), 64–73.

Cronin, J.J., Taylor, S., 1992. Measuring service quality: a rexamination and extension. Journal of Marketing 56, 55–68.

Curran, P.J., Bollen, K.A., Paxton, P., Kirby, J., Chen, F., 2002. The noncentral chi-square distribution in misspecified structural equations

models: finite sample results from a Monte Carlo simulation. Multivariate Behavioral Research 37 (1), 1–36.
Czepiel, J.A., Solomon, M.R., Surprenant, C.F., 1985. The Service Encounter. Lexington Books, Lexington, MA.
Dabholkar, P.A., Thorpe, D.I., Rentz, J.O., 1996. A measure of service quality for retail stores: scale development and validation. Journal of the Academy of Marketing Science 24 (1), 3–16.
Finn, A., Kayandé, U., 1997. Consistency of the relationship between retailer product and service quality. In: Fourth Recent Advances in Retailing and Services Science Conference, Scottsdale, Arizona.
Fornell, C., Larcker, D.F., 1981. Evaluating structural equation models with unobservable variables and measurement error. Journal of Marketing Research 27, 39–50.
Fullerton, G., 2005. The service quality-loyalty relationship in retail services: does commitment matter? Journal of Retailing and Consumer Services 12, 99–111.
Gefen, D., Straub, D.H., Boudreau, M.C., 2000. Structural equation modelling and regression: guidelines for research practice. Communications of the Association for Information Systems 4 (7), 1–70.
Gerbing, D.W., Anderson, J.C., 1992. Monte Carlo evaluations of goodness of fit indices for structural equations models. Sociological Methods and Research 21, 132–160.
Gibbons, B.C., Hocevar, D., 1998. Levels of aggregation in higher level confirmatory factor analysis: application for academic self-concept. Structural Equation Modelling 5 (4), 377–390.
Grönroos, C., 1982. Strategic Management and Marketing in the Service Sector. Swedish School of Economies and Business Administration, Helsingfors.
Grönroos, C., 1984. A service quality model and its marketing implications. European Journal of Marketing 18 (4), 36–44.
Grönroos, C., 1990. Service Management and Marketing. Managing in the Moments of Truth in Service Competition. Lexington Books, Massachusetts.
Hair, J.F., Anderson, R.E., Tatham, R.L., Black, W.C., 1999. Análisis Multivariante. Prentice-Hall, Madrid.
Houston, M.B., Lance, A.B., Sutha, W., 1998. The relationship between waiting in a service queue and evaluations of service quality. Psychology and Marketing 15 (8), 735–753.
Howat, G., Absher, J., Crilley, G., Milne, I., 1996. Measuring customer service quality in sports and leisure centres. Managing Leisure 1, 77–89.
Hu, L.T., Bentler, P.M., 1999. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Structural Equation Modeling 6, 1–55.
James, L.R., Mulaik, S.A., Brett, J.M., 1982. Causal Analysis: Assumptions, Models, and Data. Sage Publications, Beverly Hills.
Jöreskog, K.G., Sörbom, D., 2001. LISREL 8.50. Scientific Software International, Chicago.
Kearns, D.T., Nadler, D.A., 1992. Prophets in the Dark: How Xerox Reinvented Itself and Beat Back the Japanese. Harper Collins Publishers, New York.
Kim, D., Kim, S.Y., 1995. QUESC: an instrument for assessing service quality in sports and leisure centres. Managing Leisure 1, 77–89.
Kline, R.B., 1998. Principles and Practice of Structural Equation Modelling. The Guilford Press, New York.
Ko, Y.J., Pastore, D.L., 2005. A hierarchical model of service quality for the recreational sport industry. Sport Marketing Quarterly 14 (2), 84–97.
Law, K.S., Wonk, C., Mobley, W.H., 1998. Towards a taxonomy of multidimensional constructs. Academy of Management Review 23 (4), 741–755.
Marsh, H.V., Hocevar, D., 1985. Application of confirmatory factor analysis to the study of self-concept: firs and higher order factor models and their invariance across groups. Psychological Bulletin 97 (3), 562–582.
McDonald, R.P., Marsh, H.W., 1990. Choosing a multivariate model: noncentrality and goodness of fit. Psychological Bulletin 107, 247–255.
McDougall, G.H., Levesque, T.J., 1994. A revised view of service quality dimensions: an empirical investigation. Journal of Professional Service Marketing 11 (1), 189–209.
Medsker, G., Williams, L.J., Holahan, P.J., 1994. A review of current practices for evaluating causal models in organizational behavior and human resources management research. Journal of Management 20 (2), 439–464.
Nunnally, J.C., 1978. Psychometric Theory. McGraw-Hill, New York.
Nurosis, M.J., 1993. SPSS. Statistical Data Analysis. SPSS Inc., Chicago, IL.
Paik, C., Michael, W.B., 1999. A construct validity investigation of scores on a Japanese version of an academic self-concept scale for secondary school students. Educational and Psychological Measurement 59 (1), 98–110.
Parasuraman, A., Zeithaml, V., Berry, L., 1985. A conceptual model of service quality and its implications for future research. Journal of Marketing 49, 35–48.
Parasuraman, A., Zeithaml, V., Berry, L., 1988. Servqual: a multiple-item scale for measuring consumer perceptions of service quality. Journal of Retailing 64 (1), 12–40.
Parasuraman, A., Zeithaml, V., Berry, L., 1990. An Empirical Examination of Relationships in an Extended Service Quality Model. Marketing Science Institute, Cambridge, MA.
Parasuraman, A., Zeithaml, V., Berry, L., 1991. Refinement and reassessment of the SERVQUAL scale. Journal of Retailing 67 (4), 420–450.
Parasuraman, A., Zeithaml, V., Berry, L., 1994. Alternative scales for measuring service quality: a comparative assessment based on psychometric and diagnostic criteria. Journal of Retailing 70 (3), 193–194.
Perugini, M., Conner, M., 2000. Predicting and understanding behavioral volitions: the interplay between goals and behaviours. European Journal of Social Psychology 20, 705–731.
Rust, R.T., Oliver, R.L., 1994. Service quality: insights and managerial implications from the frontier. In: Rust, R.T., Oliver, R.L. (Eds.), Service Quality: New Directions in Theory and Practice. Sage Publications, London, pp. 1–20.
Satorra, A., Bentler, P.M., 1988. Scaling corrections for chi-square statistics in covariance structure analysis. In: Proceedings of the Business and Economic Statistics Section. American Statistical Association, Alexandira, VA, pp. 308–313.
Satorra, A., Bentler, P.M., 1999. A scaled difference chi-square test statistic for moment structure analysis. Pychometrika 66, 507–514.
Spangenberg, E.R., Crowley, A.E., Henderson, P.W., 1996. Improving the store environment: do olfactory cues affect evaluations and behaviors? Journal of Marketing 60, 67–80.
Steenkamp, J.B., Baumgartner, H., 1998. Measurement invariance in cross-national consumer research. Journal of Consumer Research 25, 78–90.
Steenkamp, J.B., Van Trijp, H.C.M., 1991. The use of LISREL in validating marketing constructs. International Journal of Research in Marketing 8, 283–299.
Steiger, J.H., 1990. Structural model evaluation and modification: an interval estimation approach. Multivariate Behavioral Research 25, 173–180.
Tabachnick, B.G., Fidell, L.S., 1996. Using Multivariate Statistics. Harper Collins, New York.
Taylor, S., Claxton, J.D., 1994. Delays and the dynamics of service evaluations. Journal of the Academy of Marketing Science 22 (3), 254–264.
Teas, R., 1993. Expectations, performance evaluation, and consumer's perceptions of quality. Journal of Marketing 57, 18–34.
Terblanche, N.S., Boshoff, C., 2001. Measuring customer satisfaction with the controllable elements of the in-store shopping experience. South African Journal of Business Management 32 (4), 11–19.
Westbrook, R.A., 1981. Sources of consumer satisfaction with retail outlets. Journal of Retailing 57, 68–85.

Widaman, K.F., Thompson, J.S., 2003. On specifying the null model for incremental fit indices in structural equation modelling. Psycological Methods 8 (1), 16–37.

Woehr, D.J., Arciniega, L.M., Fowler, O., 2003. Measuring work ethic: an examination of the measurement equivalence of English and Spanish versions of the multidimensional work ethic profile. In: The European Applied Business Research Conference, Venice, Italy.

Zeithaml, V., 1988. Consumer perceptions of price, quality, and value: a means-end model and synthesis of evidence. Journal of Marketing 52, 2–22.

Zeithaml, V.A., Berry, L.L., Parasuraman, A., 1985. Problems and strategies in services marketing. Journal of Marketing 49, 33–46.

Zeithaml, V.A., Berry, L.L., Parasuraman, A., 1996. The behavioral consequences of service quality. Journal of Marketing 60, 31–46.