ELSEVIER

Developing a multidimensional and hierarchical service quality model for the travel agency industry

Laura Martínez Caro*, Jose Antonio Martínez García

Department of Management and Marketing, Universidad Politécnica de Cartagena, Facultad de Ciencias de la Empresa,
Paseo Alfonso XIII, 50, 30203, Cartagena, Spain

Abstract

This paper develops a comprehensive model measuring service quality in the travel industry based on an extensive literature review and qualitative and empirical research. Following Brady and Cronin's [2001. Some new thoughts on conceptualizing perceived service quality: A hierarchial approach. *Journal of Marketing* 65(July), 34–49] conceptualization, we propose a multidimensional and hierarchical model of service quality. In this model, service quality is a third-order construct which is composed of several dimensions and subdimensions. Our findings show that customers' perceptions of service quality of travel agencies consist of three primary dimensions: personal interaction, physical environment and outcome, which are defined by seven subdimensions: conduct, expertise, problem solving, equipment, ambient conditions, waiting time and valence. Our model has important implications for the measurement of service quality in this specific industry as well as for the development of valid measures of quality performance in the context of services.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: Service quality; Travel agencies industry; Multidimensional and hierarchical structure

1. Introduction

Travel agencies in Spain are currently facing a rapidly changing environment where substantial transformations are bound to take place. As a consequence, the development of new strategies is essential in order to obtain long-term competitive advantage (Millán & Esteban, 2004). Proving high quality service is increasingly recognized as a critical factor in the success of firms in the travel and tourism industry (Fick & Ritchie, 1991). Travel agencies need to assess their service quality in order to identify the level of service quality in the travel business and the key drivers for service quality improvements. The majority of studies that measure travel agencies' service quality use the SERVQUAL approach (Johns, Avci, & Karatepe, 2004; Lam & Zhang, 1999; Ryan & Cliff, 1997). However, this approach has been substantially criticized, especially with

regard to its expectations component and its dimensionality (Buttle, 1996). Moreover, SERVQUAL or SERV-PERF, as very general instruments, are inadequate to assist travel agencies in making satisfactory service related decisions as the dimensionality of service quality is dependent on the type of service offered (Babakus & Boller, 1992).

Instead of using an existing approach based on SERVQUAL, we develop an instrument tailored to specific travel agencies' service needs. Such an approach is in accord with Brady and Cronin's (2001) multi-level proposal. They suggest that customers form their service quality perceptions on the basis of performance evaluations on multiple levels. Customers ultimately combine these evaluations, resulting in an overall service quality perception.

The aim of this study is to develop a comprehensive model to measure service quality in the Spanish travel industry based on Brady and Cronin's (2001) conceptualization. The contribution of this paper is twofold. Firstly, we conceptualize and measure travel agencies' service quality by using a multidimensional and hierarchical

---

*Corresponding author. Tel.: +34 968 32 59 39; fax: +34 968 32 70 81.
E-mail addresses: laura.martinez@upct.es (L. Martínez Caro), josean.martinez@upct.es (J.A. Martínez García).

approach. This approach helps to overcome some of the weaknesses of traditional SERVQUAL or SERVPERF models and thus provides a more accurate tool for assessing service quality in the travel sector. In addition, we identify the key manifestation of service quality from the customers' point of view. Secondly, we discuss some practical implications for using this type of model for measuring perceived service quality in applied research.

The next section provides a review of the relevant literature. Thereafter, we introduce the research methodology and results of our empirical study. Finally, we discuss managerial implications and provide directions for future research.

2. Literature review

2.1. Service quality

The perception of service quality has been extensively studied during the past three decades. Owing to the intangible, heterogeneous and inseparable nature of services, service quality can be defined as "the customer's assessment of the overall excellence or superiority of the service" (Zeithaml, 1988, p. 3) or as "the consumer's overall impression of the relative inferiority/superiority of the organization and its services" (Bitner & Hubbert, 1994, p. 77). Many models have been developed to measure customer perceptions of service quality. The first attempt to measure service quality was based on Grönroos' (1984) service quality paradigm. He distinguished between technical quality, which refers to the outcome of the service performance, and functional quality, which relates to the subjective perception of how the service is delivered. Rust and Oliver (1994) expanded Grönroos' model by adding a new dimension: service environment.

Parasuraman, Zeithaml, and Berry (1985) carried out the most famous and influential studies on service quality relating to the development of the SERVQUAL instrument. In their initial study, they found that ten determinants characterize customers' perceptions of service quality. However, as a result of a later study, they reduced the ten dimensions of service quality to five, i.e. tangibles, reliability, responsiveness, assurance and empathy (Parasuraman, Zeithaml, & Berry, 1988). The SERVQUAL model was devised to assess the difference between quality expectations and perceived service across these five dimensions.

In the two decades since the publication of Parasuraman et al.'s (1985, 1988) research, many marketing researchers have criticized their methodology as well as the psychometric setting (Buttle, 1996; Carman, 1990; Karatepe, Yavas, & Babakus, 2005; Ko & Pastore, 2005). These criticisms led to the development of alternative models to measure customer perceptions of service quality. The Evaluated Performance Model (Teas, 1993) measures the gap between perceived performance and the ideal amount of a feature rather than the customers' expectations. Teas

(1993) considered that this model could overcome some of the problems associated with the gap conceptualization of service quality (Grönroos, 1984; Parasuraman et al., 1985, 1988). Cronin and Taylor (1992) developed the SERVPERF model, which is a performance-only based measure of service quality. This measure explained more of the variance in an overall measure of service quality than SERVQUAL. Moreover, it obtained psychometrically superior assessment of service quality in terms of construct validity and operational efficacy through its performance items.

Dabholkar, Thorpe, and Rentz (1996) argued that SERVQUAL has not been successfully adapted to and validated for the retail store environment. They suggested that the dimensionality of service quality in a retail setting may be dissimilar to that of service quality in pure service industries. Using both quantitative and qualitative research methods, they proposed RSQS (Retail Service Quality Scale), a multilevel model in which retail service quality is viewed as a higher-order factor that is defined by two additional levels of attributes. The instrument includes five primary dimensions such as physical aspects, reliability, personal interaction, problem solving and policy as well as six subdimensions, i.e. appearance, convenience, promises, doing it right, inspiring confidence and being courteous/helpful.

Most recently, Brady and Cronin (2001) suggested a hierarchical and multidimensional model. These authors combined the traditional approach to service quality (i.e., the Tri-component model of service quality by Rust and Oliver (1994)) with the multi-level conceptualization of service quality (i.e., Dabholkar et al. (1996)). They described a third-order factor model, in which service quality is explained by three primary dimensions, i.e. interaction quality, physical environment quality and outcome quality. Each of these dimensions consists of three corresponding subdimensions, such as attitude, behaviour and experience (interaction quality); ambient conditions, design and social factors (physical environment quality); waiting time, tangibles and valence (outcome quality). According to these authors, the proposed model improves our understanding of three basic issues: (1) what defines service quality perceptions, (2) how service quality perceptions are formed, and (3) how important it is where the service experience takes place.

2.2. Service quality in the travel industry

A number of empirical studies have focused on service quality in the travel agency industry. Most studies have measured service quality by replicating or adapting the SERVQUAL model (Fick & Ritchie, 1991; Karatepe & Avei, 2001; Lewis, 1987; Saleh & Ryan, 1991). Lam and Zhang (1999) used and adapted the SERVQUAL instrument to study the service quality of travel agencies in Hong Kong. They used descriptive research to identify and test 26 attributes relevant to the travel industry. Five factors

were derived using factor analysis such as reliability, responsiveness and assurance, empathy, resources and corporate image, and tangibility. These factors were taken as service quality dimensions. According to regression analysis, the most important service quality dimension in predicting customers' overall satisfaction was reliability, followed by responsiveness and assurance. These two dimensions were especially crucial for excellent-service travel agencies. Resources and corporate image was the least influential dimension in predicting customer satisfaction. This result was fundamentally different from LeBlanc's (1992) study in Canada, in which corporate image was the most significant driver in explaining customers overall satisfaction with services provided by travel agencies.

Similarly, Johns et al. (2004) employed SERVQUAL to investigate travel agencies' service quality in Northern Cyprus. This scale was administered to 337 users of travel agencies who had used the service at some time during the previous three months. Their findings showed gap scores were poorer predictors of satisfaction than performance scores. Furthermore, factor analysis showed an underlying uni-dimensionality which is consistent with other authors' findings (e.g., Babakus & Mangold, 1992; Cronin & Taylor, 1992; Karatepe & Avei, 2001). The authors admitted that although the uni-dimensionality findings are of empirical and theoretical significance, it undermines the practical utility of scale, since in principle it becomes impossible to allocate service excellence or failure to specific causative areas.

Moreover, LeBlanc (1992) carried out an exploratory study of customers' perceptions of service quality in travel agencies in Canada. He administered a questionnaire to 227 customers with 50 variables, which related to different aspects of the travel agencies' service offerings. These variables were identified after a literature review and personal interviews of directors of travel agencies. In order to measure service quality, the data were directly collected in terms of the perception-expectation difference. The factor analysis resulted in nine dimensions of service quality, namely: physical evidence, competence, corporate image, timeliness, courtesy, competitiveness, responsiveness, confidentiality and accessibility. These factors are similar to those identified by Parasuraman et al. (1985, 1988), except for the corporate image dimension. Regression analysis showed that service quality is derived principally from corporate image.

Bigné, Martínez, Miquel, and Belloch (1996) adapted generic SERVQUAL scale using the Delphi method. Six travel agency managers, four managers related to travel agencies and four academics were part of an expert panel. The resulting scale was composed of the five SERVQUAL dimensions. The main problem with the technique used by Bigné et al. (1996) is that they started from the 22 items on the SERVQUAL scale. This aspect impedes the consideration of other specific relevant dimensions of travel agencies service. Likewise, Setó (2003) applied Bigné's et al. (1996)

scale to travel agencies in Spain. They obtained a classic five-factor structure including tangibles, responsiveness, assurance, reliability and empathy, using only performance measures.

Also, Ryan and Cliff (1997) employed the SERVQUAL questionnaire to study travel agencies in New Zealand. Although the results support the reliability of the SERVQUAL scale, they showed that only three factors (reassurance, reliability, and tangibles) could be replicated out of the five in the original SERVQUAL research (Parasuraman et al., 1988). Furthermore, Ryan and Cliff (1997) suggest that the difference may be due to data analysis problems and the application of the SERVQUAL scale to a service with the potential for a strong emotive outcome.

In another study of the perception of travel agencies' service quality, Albrecht (1992) suggested that customers evaluate service quality of their travel agencies using eight factors, i.e. error-free mechanics, "can do" attitude, information support, proactive safeguarding, personalized treatment, agent continuity, price confidence and recovery. The first five elements were similar to SERVQUAL. Although agent continuity is not explicitly accounted for in SERVQUAL, it does relate to the empathy and reliability dimensions. Price confidence and recovery are further factors different from the SERVQUAL dimensions.

To sum up, in the travel agencies literature, service quality has been conceptualized in different ways. The literature agrees that perceptions of travel agencies' service quality are of a multi-dimensional nature. However, the number and content of these dimensions remain open to debate. Many marketing researchers admit that the use of the generic models such as SERVQUAL or SERVPERF to measure service quality across industries is not feasible. Likewise, there are several studies that failed to replicate these models. This may be due to the theoretical and operational problems of SERVQUAL (Babakus & Boller, 1992; Cronin & Taylor, 1992; Teas, 1993) or the generic dimensions of SERVPERF (Buttle, 1996). According to Babakus and Boller (1992), there is a need to develop measures of service quality that are industry-specific, since the use of generic models forces researchers to drastically alter the items these models intend to apply in a specific context (Babakus & Boller, 1992; Babakus & Mangold, 1992; Carman, 1990; McAlexander, Dennis, & Koenig, 1994). Moreover, there is a need to develop service quality measures that are country/culture specific, since the measures that are developed in one culture may not be applicable in a different cultural setting (Karatepe et al., 2005; Mattila, 1999). According to Mattila (1999), cultural factors are said to have greater influence on people's evaluation of services than on their evaluations of physical goods due to involvement of customer contact and interaction with employees while a service is delivered. Also, Ueltschy and Krampf (2001) considered that differences in culture affect measure of quality in a service sector. However, the studies developed in Spanish travel

agencies context have the weakness that none of them used qualitative research that allowed identifying service quality dimensions. These dimensions could be critical for consumer evaluations, and therefore they could overcome the definition of service quality provided by SERVQUAL or SERVPERF models.

Thus, in light of the problems associated with these models, the aim of this study is to develop a scale using a multidimensional and hierarchical model which takes the specific characteristics of the travel industry into account.

3. Scale development

The current literature emphasizes the hierarchical and multi-dimensional nature of perceived service quality (Brady & Cronin, 2001; Kim & Jin, 2002; Ko & Pastore, 2004). We adopted this approach to show the factors which determine the travel agencies' customer perceptions of service quality, since it provides a better explanation of the complexity of human perceptions than the earlier conceptualizations offered in the literature (Brady & Cronin, 2001).

This proposal has been applied by other investigators like Lui (2005), Ko and Pastore (2005), Jones (2005) and Martínez and Martínez (2007), although from different perspectives and with different degrees of precision.

Lui (2005) claimed that the concept of perceived service quality is multidimensional and multi-level. This author proposed the same dimensions and subdimensions of Brady and Cronin's (2001) quality model in several areas: fast food restaurants, gasoline stations, medical clinics, photo shops, mobile phone repair shops and 24 h grocery stores. However, he did not test the conceptualization with a confirmatory method.

For their part, Ko and Pastore (2005) adapted Brady and Cronin's (2001) conceptualization to the recreation sport industry. The result of this study showed how service quality can be measured by four primary dimensions defined by several corresponding subdimensions: (a) programme quality—range of programme, operating time, and information, (b) interaction quality—client–employee interaction and inter-client interaction, (c) outcome quality—physical change, valence, and sociability, and (d) environment quality—ambient condition, design, and equipment.

Finally, Jones (2005) integrated communication into the Brady and Cronin's three dimensions of overall service quality to investigate the critical role of this factor in service industries. Communications quality was divided into two sub-dimensions consisting of interpersonal and impersonal communications quality. The results showed that communications quality was a significant driver of overall service quality in three of the four industry samples (banking, fine dining, physician care, and sports). Nevertheless, in our view, Jones (2005) failed to capture the essence of Brady and Cronin's (2001, p. 46) final assertion: "Yet we acknowledge that it may be impossible to develop a model that is equally applicable across all service

industries (…). We therefore stress that modifications of this conceptualization to account for industry- specific factors are critical''.

We believe that researchers have to focus their efforts on building industry-specific models instead of trying to replicate any universal conception of service quality models, because it seems clear that critical factors of service quality are not the same across all service industries. This is the perspective adopted by Martínez and Martínez (2007) in their recent adaptation of Brady and Cronin's (2001) model to the urgent transport service. These authors tested a hierarchical and multidimensional model where service quality is composed of four dimensions: (a) personal interaction, (b) design, (c) physical environment and (d) outcome; and nine subdimensions: (a) conduct, expertise, problem solving, (b) range of service, operating time, (c) tangibles, information, and (d) punctuality and valence.

Therefore, we developed a specific hierarchical and multi-dimensional model of service quality for the travel agency industry. To accomplish this aim, we followed the procedure for scale development recommended by Churchill (1979), using both qualitative and quantitative research.

3.1. Qualitative research

According to Chumpitaz and Swaen (2002) the number and the nature of service quality dimensions is directly related to the service under investigation. To accomplish this goal, qualitative research was carried out to identify the factors which determine the service quality perceptions of travel agency customers.

For the qualitative research, we used in-depth interviews of travel agency employees and customers. In order to obtain information from the travel agents' point of view, five managers were interviewed. The managers were interviewed about several service quality issues: (1) what they perceived to be service quality from the consumer's perspective; (2) what steps they took to control or improve service quality; (3) what problems they faced in delivering high quality services (Parasuraman et al., 1985). With regard to the customers' side, ten customers who had used the services of one or more travel agencies during the previous twelve months were interviewed (Ryan & Cliff, 1997). 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), prices were eliminated from the decision set because they are not part of a 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. Bitner & Hubbert, 1994; Zeithaml, 1988). Findings from this qualitative study showed the need to adapt the structure
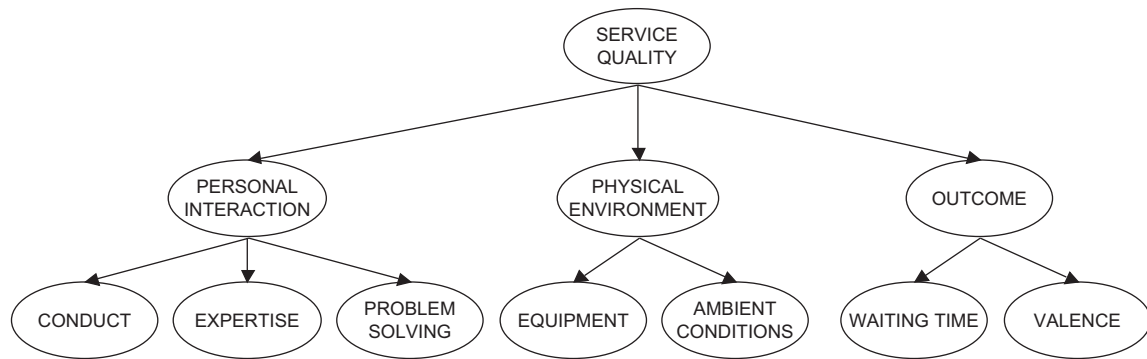
Fig. 1. Travel agencies service quality model.

proposed by Brady and Cronin (2001) in order to identify essential characteristics of this service. Qualitative study showed that customers do not make any distinction between employees' attitude and behaviour. Social factors were hardly taken into account by travel agencies' customers. However, the qualitative study revealed that both customers and managers considered that employees' interactions with customers were highly important in determining travel agencies' perceived service quality.

To code the qualitative data thus obtained, similar to prior studies (Brady & Cronin, 2001; Karatepe et al., 2005; Richins, 1997), a content analytic approach was employed. In order to ensure the validity of this research, a semi-structured questionnaire was used in the interviews. This was formulated from the emerged framework of references (travel agencies service quality literature and Brady and Cronin's (2001) study) in order to answer the research problem and the research questions. Furthermore, in order to ensure the reliability of this research, a tape recorder was also used during the interviews. This method provided us with the opportunity to double check the answers and convinced us that we did not miss any important information.

Carman (1990) suggests context customization may involve adding or dropping relevant dimensions and appropriately adjusting the attributes considered to make up those dimensions. Therefore, after listening to the tapes, three independent coders prepared field notes based on Brady and Cronin's (2001) study and on the rest of the literature about travel agencies service quality. Taking all that information into account, we proposed three distinct dimensions and seven subdimensions as our service quality variables.

3.2. Proposed factor structure

After combining the findings from our qualitative research with the service quality literature revision, we propose the following model: a hierarchical and multi-dimensional model in which quality is a higher order factor that is defined by three primary dimensions and seven subdimensions. The primary dimensions are personal interaction, physical environment and outcome, which

are defined by seven subdimensions: conduct, expertise, problem solving, equipment, ambient conditions, waiting time and valence (see Fig. 1).

The first dimension is "personal interaction". Several researchers have indicated the importance of this factor in the delivery of services and have identified it as having the most significant effect on service quality perceptions (Bigné et al, 1996; Grönroos, 1982; LeBlanc, 1992). Both the service quality literature and our qualitative research suggest that there are three subdimensions: *conduct*, *expertise*, and *problem solving*. The first two subdimensions coincide with Capelleras and Veciana's (2002) proposal. According to Grönroos (1990), attitudes, behaviour and skills of employees are factors that largely influence service quality perceptions. Based on our qualitative study and literature review, we identified a third subdimension, problem solving. The results of this qualitative study reveal that consumers are indeed quite sensitive to how service providers address problems and complaints.[1] Although Dabholkar et al. (1996) considered this factor a separate dimension of "personal interaction", others authors such as Kim and Jin (2002) suggested that personal interaction and problem solving were not separate dimensions of service quality. This indicates that customers do not make any distinction between the store's ability to make them feel well treated and reassured (personal interaction) and the store's ability to handle potential problems (problem solving).

The second dimension is "physical environment". Authors such as Brady and Cronin (2001), Bitner (1990, 1992) and Spangenberg, Crowley, and Henderson (1996) consider that this factor can have a significant influence on perceptions of the overall quality of service encounter. It has a broader meaning than the SERVQUAL (Parasuraman et al., 1988) tangibles dimension. According to the findings from our qualitative research and the literature, two subdimensions can be identified to explain "physical

---

[1]Participants in the interviews commented that this aspect was important to them. ("I want to know that when I have a problem there will be someone who can help me"; "I like travel agents that are able to handle my complaints when the hotel or place booked is not what we agreed".)

environment": *equipment* and *ambient conditions*. *Equipment* includes the computers and other implements used to achieve effective service delivery (Ko & Pastore, 2004). Thus, some customers pointed out that they valued the technology used to book a travel, flight or hotel. Theory suggests that customers use any tangible evidence of the service outcome as a proxy for judging performance (Booms & Bitner, 1981; Zeithaml, Berry, & Parasuraman, 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 travel agencies' service, where the final result is not carried out in a physical environment. Nevertheless, and given the importance that users assign to this topic, it seems more appropriate to include this subdimension within the "physical environment" dimension, as stated by Dabholkar et al. (1996). The second subdimension is *ambient conditions*. This factor is proposed by Bitner (1992), and it relates to the non visual aspects associated with the service-producing system and includes factors such as temperature, scent and music. LeBlanc (1992) defined this factor as appearance, decor and ambience, and furniture layout. In addition, the conclusion from the qualitative study indicated that consumers value the *ambient conditions* more than the experts who had been interviewed. In the travel agencies' context, several authors identified "physical environment" as a determinant of perception service quality (e.g. Bigné et al., 1996; Johns et al., 2004; Lam & Zhang, 1999; LeBlanc, 1992; Ryan & Cliff, 1997).

The third dimension of travel agencies' service quality is "outcome". Many marketing researchers agree that the outcome of the service encounter significantly affects customer perceptions of service quality (Carman, 2000; Grönroos, 1984, 1990; McDougall & Levesque, 1994; Rust & Oliver 1994). This factor is labelled "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 Brady and Cronin's (2001) findings and our qualitative research to identify the subdimensions of *waiting time* and *valence*. *Waiting time* has been considerably supported in the literature. Parasuraman et al. (1985) showed that consumers identify service punctuality as an integral part of their overall evaluation. In the travel agency context, LeBlanc (1992) identified timeliness as a factor of service quality. This factor is associated with the ability of the service firm to manage demand in an effective manner. The qualitative research showed that the waiting time before being served affects consumer service evaluation. The last subdimension is *valence*. In line with Brady and Cronin (2001), valence captures attributes that control whether customers believe that a service outcome is good or bad, regardless of any other experienced aspects. In a travel agency setting, a traveller can measure various

psychological benefits such as relaxation, rest, and stress reduction, etc. Findings from Brady and Cronin (2001), Ko and Pastore (2005) and Martínez and Martínez (2007) 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

The generation of a list of items was developed adapting the items of existing generic scales (e.g., Brady & Cronin, 2001; Dabholkar et al., 1996; Ko & Pastore, 2005; Parasuraman et al., 1988) and specific travel agencies' scales (Bigné et al., 1996; LeBlanc, 1992; Ryan & Cliff, 1997). On the basis of our literature review, we generated an initial pool of 62 items using a performance-only measurement. These items were indicators of each theoretical subdimension. A 5-point Likert scale was applied to measure the different items anchored, from strongly disagree to strongly agree.

The purification of the scale was carried out in two steps: the first step consisted of an assessment of content and face validity through a panel of experts and a field test (Ko & Pastore, 2005). The panel members were five executives from travel agencies and three academics. As result of this panel, 31 items were dropped; in the second step, we developed a questionnaire containing the 31 remaining items. This questionnaire was pilot-tested with 50 respondents (customers of travel agencies). The aim was to study the correlation structure of the items of each subdimension. To accomplish this end, Cronbach's alpha together with item-to-total correlation and exploratory factor analysis for each one of seven subdimensions were achieved (Parasuraman et al., 1988; Steenkamp & Van Trijp, 1991). These patterns of correlations are relevant information for selecting the final model, as we will show later.

4. Data collection

The sample was collected from users of travel agencies in a specific geographic area (Murcia, a south-eastern region of Spain), during the first three months of 2006. The procedure to obtain the data was a personal interview with the individuals and a self-administered questionnaire. We trained business and marketing students for this purpose. Respondents had to fulfil the condition of having used the service during the previous 12 months (Ryan & Cliff, 1997). An initial convenience sample of 230 respondents was obtained, although 28 questionnaires related to online travel agencies were dropped. According to service quality literature, to measure online service quality it is necessary to develop a specific scale for these types of services (Grönroos, Heinonen, Isoniemi, & Lindholm, 2000; Liljander, Van Riel, & Pura, 2002; Zeithaml, Parasuraman, & Malhotra, 2002). The final sample was of 202 subjects and the data was representative of the heterogeneity of the individuals selected for the research design. Ninety-one

percent of the respondents were between the ages of 18 and 45. The majority of the respondents (56 percent) were female. Fifty-six percent of the respondents had university degrees. It appeared that respondents with a "leisure" purpose for their trip (85%) dominated the sample, while "business" accounted for only 15%. About 63% of the respondents used travel agencies' services once a year; 20% used such services twice or three times a year; and 17% used such services four times or more a year. Twenty-two different travel agencies were evaluated.

5. Procedure for model testing

We conducted a test of the hierarchical third-order structure following the model testing methodology proposed by Hayduk (1996) and Hayduk and Glaser (2000) in the context of structural equation modelling. This view is based on the "one step procedure",[2] the relevancy of causal effects for determining construct validity, and researcher commitment to the meaning of latent variables through the correspondence between the latent and the best indicator of the concept (the "gold standard indicator"). Therefore, we followed these steps (Hayduk, 1996): (1) choose the indicator we believe is the single best available indicator of each subdimension; (2) fix the causal coefficient (lambda) from latent to this item at 1.0; (3) fix the error variance (theta) of this indicator at a specific value; and (4) enter free lambda and theta variances for the second and third best indicator(s) of the concept.

We selected the "gold standard indicator" for each concept from the items battery resulting from the qualitative depuration process (see Appendix A). Our image of each latent subdimension determines the proximity between that abstract concept and the real world manifestation. Thus, the definition of the theoretical subdimension determines the choice of the best indicator. The next step was fixing the error variance of those indicators. The extensive literature review and the qualitative research justify that our commitment to the fictional meaning of the latents is strong. We thought that all the best indicators shared between the 85% and 90% of the latent variance; consequently the error variance had to be fixed to between 10% and 15% of the indicator variance. We finally decided to choose the more conservative estimation of 15%.

Since we have different hierarchical levels, the reasoning of Hayduk (1996) regarding the commitment of the latents (and their relationship with the indicators) can be extended to the higher order latents. Therefore, if the indicators are manifestations of the latents in the first order level, the first order latents are manifestations of the second order latents in the second order level, and successively for higher order

levels (see Law, Wong, & Mobley, 1998). This means that the definition of the higher order latent variables determines the proximity between those variables, the lower order latents and the indicators.[3] Following the methodology of Hayduk (1996), some gamma and beta paths have to be fixed together with the error variances of their respective endogenous latents.

We can test the strength of our conceptualization adding two or more indicators for each subdimension. Adding indicators enhances the demand of proportionality constraints and provides scale-dependent statements of the quality of these indicators. This has to be done carefully and several factors need to be considered. We have to add what are thought to be the second-best and the third-best indicators in separate estimations (Hayduk, 1996), but the increase of number of parameters to be estimated could be inadequate considering the sample size (Jackson, 2003), and the chi-square test of model fit could be biased.

The idea of fixing the error variances of gold standard indicators has been called "revolutionary" by some researchers (e.g. Bentler, 2000, p. 84). Certainly, the postulates of Hayduk (1996) and Hayduk and Glaser (2000) have received many criticisms (e.g. Bentler, 2000; Mulaik & Millsap, 2000). However, our view is that these criticisms do not diminish the real meaning of the Hayduk (1996) thesis: freeing the error variance of all observable indicators, a new meaning is assigned to a similarly labelled concept each time a new data matrix is used and a new indicator error variance is estimated. The result is a weak theory.

Therefore, we propose the following sequence for testing our third-order model. First, we test our more restricted model ($M_{1,1}$), only with the best indicator per subdimension, and fixing the error variances of the indicators and the error variances of the latents that define the higher order dimensions. If the model fits, we have the chance to test our conceptualization, adding the second best indicator per subdimension ($M_{1,2}$). Second, if $M_{1,1}$ does not fit, we can test a less restricted model without fixing the error variances of the correspondent latents ($M_{2,1}$). This is a model with a more fictional nature of the meaning of the higher order latents. If $M_{2,1}$ fits, we can test the strength of the theoretical model again, adding the second best observable indicator ($M_{2,2}$).[4] Third, if $M_{2,1}$ does not fit, our theoretical perspective regarding the meaning of the latents is questionable. However, we can conduct another test on $M_{2,2}$, freeing all the indicators error variances ($M_{3,2}$). If this latter model fits, we are not sure what those

---

[2]The one step procedure tests the causal model without considering previous analysis based on confirmatory factor analysis (two step procedure) or exploratory and confirmatory factor analysis (four step procedure). See Mulaik and Millsap (2000) for an extensive debate.

[3]Hayduk (22/10/2006). SEMNET (a web discussion group on structural equation modeling. http://bama.ua.edu/archives/semnet.html).

[4]A third indicator could be added if $M_{2,2}$ fits (we can notate it as $M_{2,3}$) for a more robust test of the model. This is equally applicable to $M_{1,2}$ (in this case it would be $M_{1,3}$). However, as Hayduk (1996) points out, it is rare to find three well-behaved indicators of a concept. Furthermore, adding a third indicator to our model dramatically decreases the ratio sample size/parameters to estimate, causing the mentioned non-desirable consequences for the chi-square test (Jackson, 2003).

latents truly represent (Bentler, 2000), but it is unquestionable that the causal structure proposed is supported by the data, and hence we have an interpretable model (although perhaps with an undesirable degree of abstraction). Finally, if $M_{3,2}$ does not fit, we have to study residuals and provide some explanation of the failing model, and then reconsider the theory.

6. Selection of the "gold standard" latents

We thought that the dimension "outcome" is the best definition of the service quality concept. Although as has been mentioned before, several studies have shown that "personal interaction" plays an important role for the evaluation of service quality, there are other researchers that have found "outcome" as a key manifestation of perceived quality (Brady & Cronin, 2001; Ko & Pastore, 2005; Martínez & Martínez, 2007). As "outcome" refers to results obtained by the consumer once service is delivered, its meaning is very close to the global evaluation of service quality implemented in other studies (e.g. Brady, Cronin, & Brand, 2002; Cronin, Brady, & Hult, 2000). Similarly, the "valence" subdimension is the one chosen as most closely defining the "outcome" dimension. It captures attributes that control whether customers believe that service outcome is good or bad, as an overall evaluation also. Authors such as Grace and O'Cass (2005) or Patterson and Spreng (1997) have defined a similar concept in the context of overall customer satisfaction evaluation.

Regarding the "personal interaction" dimension, we believed that "conduct" is its best manifestation. Grönroos (1990) pointed out that attitudes, behaviour and skills of employees are very important for quality evaluations. However, we believed that skills were more difficult for customers to evaluate than the "face to face" behaviour and attitudes of the front-line employees. Attitude and behaviours have been considered to be vital elements in service encounters (Bitner, 1990; Bowen & Schneider, 1985).

Finally, we selected the "equipment" subdimension as the key indicator of the "physical environment" concept. This tangible evidence is more closely related with the meaning of the concept than "ambient conditions". However, the non-visual aspects could play a major role in other service contexts, such as retail stores, and therefore they could be a very good definition of the "physical environment" dimension in that situation.

7. Results

We began with the test of $M_{1,1}$ using the LISREL 8.50 (Jöreskog & Sörbom, 2001) maximum likelihood method. As we have explained earlier, we assigned 15% error variance (θ) to all the gold standard indicators. Since the causal effects lose their strength from higher order factors to the observable indicators (Edwards & Bagozzi, 2000; Hayduk, Ratner, Johnson, & Bottorff, 1995), we had to be equally restrictive fixing the error variances of the latents (ψ) as we had done for indicators. Therefore, the diagonal elements $\psi_{3,3}$, $\psi_{4,4}$, $\psi_{7,7}$, $\psi_{10,10}$ of the Psi matrix were fixed to represent 15% of the variance of their respective latents.[5] However, the model did not fit properly: $SB\chi^2$: 30.35 (15); $p = 0.011$.

The next stage was the test of a less restricted model ($M_{2,1}$) freeing all diagonal elements of the Psi matrix. This model yielded an excellent fit: $SB\chi^2$: 6.54 (11); $p = 0.840$. Nevertheless, this conceptualization lacks the dictated meaning of the higher order latents. Hence, a more robust test of the theoretical model is needed, adding a second indicator per subdimension. The selection of the second best indicator was achieved according to the subjective view of the second best indicator and the patterns of correlations showed in the pretest analysis.[6] However, in order to avoid the second indicators being mere repetitive items that could not help to add meaning to the latent, we made sure that they were not extremely correlated, following the recommendations of Epstein (1983). Thus, 90% confidence intervals were built to the inter-item correlation of each subdimension. To accomplish this, we, used R2 software (Steiger & Fouladi, 1992). The results of this correlation analysis showed that the upper limit of the confidence interval of each pairwise correlation was not above 0.85. Therefore, second best indicators were not extremely correlated with gold standard indicators. We included these items in model $M_{2,2}$, and the test yielded a poor fit: $SB\chi^2$: 121.02 (74); $p < .001$.

These results indicate that our view about the meaning of the latents is questionable, or at least, it is not as close to the gold standard observable indicator as we thought at first. However, if a less restricted model without error variance constraints fits, the model is not causally misspecified, and we cannot reject the conceptualization unless some unexpected estimates appear. We tested $M_{3,2}$ with an excellent result: $SB\chi^2$: 70.91 (67); $p = 0.350$, and highly significant gamma, beta and lambda paths. The square multiple correlations for structural equations and observable indicators are depicted in Table 1, confirming the meaningfulness of the estimated parameters in the causal analysis.

7.1. Alternative models

The analysis of alternative models is a highly recommendable practice in model testing methodology (Kline, 2005). The comparison among plausible theoretical explanations of the same phenomenon enriches the strength of the original research model. We considered two alternative models: (1) A model with a general factor

---

[5]See the form of structural equations in Appendix B. The procedure for fixing the error variances of the indicators and the error variances for the latents is depicted in Appendix C.

[6]Items are presented in Appendix A.

Table 1
Square multiple correlations for structural equations and indicators of Model $M_{3,2,}$ and total effects on indicators

Personal interaction						Physical environment				Outcome			
0.88						0.46				1.09[a]			
Conduct		Expertise		Problem solving		Equipment		Ambient conditions		Waiting time		Valence	
0.71		0.87		0.88		1.18		0.76		0.78		0.78	
y1	y2	y3	y4	y5	y6	y7	y8	y9	y10	y11	y12	y13	y14
0.69	0.82	0.81	0.69	0.71	0.61	0.67	0.75	0.74	0.57	0.71	0.72	0.77	0.86
0.66[b]	0.71	0.78	0.72	0.74	0.69	0.60	0.64	0.51	0.44	0.77	0.78	0.81	0.85

[a]See Jöreskog (1999) for an explanation of how standardized loadings (and hence square multiple correlations) could be above 1.0. In this case, the error variance of the latent (ψ_{33}) is not significantly distinct from 0, what supports Jöreskog's explanation.
[b]Total effects of service quality on indicators.

Table 2
Summary of the model testing sequence

	θ	ψ	Parameter to estimate	Indicators per subdimensions (total)	Satorra–Bentler chi-square[a] (df)	*p* Value
$M_{1,1}$	15%[b]	15%[d]	13	1 (7)	30.35 (15)	0.011
$M_{2,1}$	15%	Free	17	1 (7)	6.54 (11)	0.840
$M_{2,2}$	15%[c]	Free	31	2 (14)	121.02 (74)	<0.001
$M_{3,2}$	Free	Free	38	2 (14)	70.91 (67)	0.350
Alternative models						
M_{A1}	Free	Free	28	2 (14)	408.16 (77)	<0.001
M_{A2}	Free	Free	35	2 (14)	147.74 (70)	<0.001

[a]Previous analysis of the data distribution with PRELIS shows non-extreme skewness and kurtosis of the different covariance matrices provided. The asymptotic covariance matrices were generated to obtain the scaled chi-square (Satorra & Bentler, 1988) and robust estimation of standard errors.
[b]$\theta_{1,1}, \theta_{2,2}, \theta_{3,3}, \theta_{4,4}, \theta_{5,5}, \theta_{6,6}, \theta_{7,7}$.
[c]$\theta_{2,2}, \theta_{4,4}, \theta_{5,5}, \theta_{8,8}, \theta_{9,9}, \theta_{12,12}, \theta_{14,14}$.
[d]$\psi_{3,3}, \psi_{4,4}, \psi_{7,7}, \psi_{10,10}$.

accounting for all items covariances.[7] The aim was to analyse if service quality perception could exert a halo effect on observable indicators, as seen in studies regarding the multidimensional structure of corporate image evaluations (Caruana, 1997; Dowling, 1988; Fombrun & Shanley, 1990). Thus, if customers are not capable of judging certain service attributes properly, they can infer these evaluations from the overall judgement based on a few well-known service cues. However, the analysis of this model (M_{A1}) revealed that it is extremely inconsistent with the data: $SB\chi^2$: 408.16 (77); $p < .001$. (2) A model with a general factor accounting for all sudimensions covariances. This model hypothesizes that subdimensions are not capable of reflecting the divergent meaning of their respective dimensions, and only one factor (instead of three) emerges in the second order level. Again, the analysis of the model (M_{A2}) showed that it is inconsistent with the data: $SB\chi^2$: 147.74 (70); $p < .001$.

[7]This type of model is known as Harman's single factor test (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003).

Finally, a summary of the model testing sequence is depicted in Table 2.

8. Discussion, limitations and further research

There is a growing acceptance among researchers that to achieve and maintain competitiveness and profitability, travel agencies should make efforts to continuously improve the level of service quality offered to their customers. However, a basic principle of quality management is that to improve quality it must first be measured. This study proposes a multilevel and multidimensional model of service quality in accordance with the hierarchy of perceptions suggested by Brady and Cronin (2001). Therefore, customers form their service quality perceptions on the basis of their evaluations of three primary dimensions: personal interaction, physical environment and outcome, which are composed of seven subdimensions. Qualitative research and the literature revision support this view, and indicate that personal interaction is formed by conduct, expertise, and problem solving subdimensions; physical environment consists of equipment and ambient

conditions; and finally outcome is formed by waiting time and valence subdimensions. Thus, the combination of all theses evaluations constitutes a customer's overall perception of the travel agencies' service quality. This study contributes towards filling a gap existing in service quality research by proposing an integrated model of service quality in the travel agency industry.

The results of the third order confirmatory factor analysis have shown that we can count on the validity of a battery of 14 items that partially support the original purpose. We say ''partially'' because we had to relax the original constraints of the model that were consistent with our view regarding the meaning of the latents. Recall that we have partly failed in the determination of the meaning of the latents and that they are not as close to the observable indicators as we thought. The study of the residuals pattern of failing models, like $M_{2,2}$, is somewhat embarrassing because there are several problematic residuals. This fact makes it more difficult to detect misspecification (Hayduk, 1996). However, the test of the less constrained model $M_{3,2}$, can reveal valuable information with regard to the pattern of causal relationships, assuming we sacrifice the precise meaning of latents. This is summarized in the following lines.

First, ''outcome'' is the key manifestation of service quality. This factor is unexpectedly close to the meaning of the higher order construct as it perfectly reflects the consumer evaluations of service quality. Surprisingly, the reviewed literature on service quality in travel agencies does not define this aspect in its measurement instruments. Findings from our qualitative research support the importance of evaluations about consumer gains from the service. If we could isolate this part of the model, respecifying the error variance restrictions,[8] we would obtain an adequate model fit. Although we recognize this is a poor methodological device, it serves as a heuristic signal of the proper meaning of latents representing this portion of the model. The relevance of ''outcome'' evaluations is related with consumer sensitivity to service failure and its impact on global evaluations of the service (Mittal, Ross, & Baldaser, 1998). This could be dangerous from the management point of view because, as Brady and Cronin (2001) claim, some factors that shape the valence of the outcome are outside the direct control of service management. Therefore, for example, consumers' not obtaining the required location at a desired destination because of an overbooking problem could generate a negative perception of the global service delivered by the travel agency. This negative perception would be independent of the perceived quality of the rest of attributes. Another example could be the next case: a consumer who has encountered bad weather conditions during vacations. This customer might

have a positive perception of each service quality dimension, but the negative valence of the outcome could ultimately lead to an unfavourable service experience (Brady & Cronin, 2001).

This is one of the major findings of our research, because this fact questions the implementation of service quality studies to fully analyse company performance. We encourage travel agencies to design questionnaires that also consider those things that are outside their control and could have influence on the customer overall evaluation of the service. This draws on the locus-of-causality literature in attribution theory (Oliver, 1993), where the locus dimension is concerned with the source of causality; that is, either the cause resides with the customer, in the situation, or with the company. Thus, we believe it is important to implement measures of company performance that inform companies not only about the third source but also about the other two. Because of the peculiarities of the travel agency industry, we think that this is a very interesting starting point for further research about perceived service quality.

The meaning of ''valence'' subdimension yields another important contribution from the managerial point of view; managers could use this subdimension alone to obtain a very good measure of global perceived service quality. Service quality has to be measured systematically by the companies (Parasuraman et al., 1988), and this is a key principle of quality management systems such as ISO certification. The length of questionnaires is associated with problems from the point of view of data validity (e.g. halo effect, learning effect, customer tiredness, etc.) and from the perspective of managerial costs (e.g. design, tabulation and analysis). Therefore, depending on the objectives, resources and budget, travel agencies could evaluate service quality using a more disaggregated tool (using the entire model), or a simpler device (using only the ''valence'' subdimension). The former, logically, provides a more complete picture of service excellence, but the latter, by simplifying data collection and analysis, is an attractive option for travel agencies that systematically evaluate service quality several times a year. Both methods, as we have discussed above, should be rounded out with information about things outside companies' control.

Second, service quality evaluations are quite well defined by the judgements regarding how the service is delivered, and this approach agrees with the relevance of ''personal interaction'' in the service marketing literature (Grönroos, 1990). One reason explaining why our original constrained model does not fit could be the somewhat unexpected roles of the ''expertise'' and ''problem solving'' subdimensions. We thought that employee attitudes and behaviours were the best manifestation of the personal interaction dimension. However, the skills of the employees and their ability to solve problems are seen as important attributes of the service quality evaluation. Both subdimensions are highly correlated (0.81) with ''valence'' and ''waiting time'', which could be interpreted as follows; the customer associates the

---

[8]We could specify zero error variance in ''outcome'', and the same values of the original proposal allocated to the error variance of ''valence'' and to the error variances of gold standard indicators of ''valence'' and ''waiting time''.

ability of the employees to respond to requests and their competence at solving problems with the outcome of the service. If there are no service failures, then (1) there are no problems to resolve, or (2) the problems were solved successfully. Regrettably, we have not specifically asked about service failures in the questionnaire, and thus we cannot evaluate this important issue. This is one of the limitations of our study. If there were not any kinds of problems in the service encounters, then we could argue that customer evaluations of "outcome" could exert a halo effect on judgements of "problem solving". In this case, the customer might not have sufficient information to make a valid evaluation of this subdimension of service quality.

Third, results clearly indicate that the environment in which the service delivery occurs is not as good a definition of service quality as the other two dimensions. This is an expected finding, because of the high degree of intangibility of this kind of service. The study shows that functional aspects define quality better than aesthetic aspects. However, again we could have failed to correctly define the latents; in this case the "equipment" subdimension. Travel agencies try to make the service delivered more tangible, especially with flyers, handouts, etc. This fact could have not been properly interpreted by customers in the evaluation of the store equipment.

On the other hand, there are several methodological issues that deserve some discussion.

First of all, our test of the third-order structure assumes that the three dimensions are different forms manifested by perceived service quality. Likewise, these three dimensions are explained by their respective subdimensions, and finally the indicators are observable representations of each subdimension. This view agrees with the taxonomy of reflective multidimensional constructs by Law et al. (1998) and Edwards (2001), and it reproduces the proposal of Dabholkar et al. (1996). Nevertheless, this does not coincide with the Brady and Cronin's (2001) perspective on 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 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. This view regarding measurement agrees with the entity realism of latent variable theory and the ontological concept of validity (Borsboom, Mellenbergh, & van Heerden, 2003, 2004), as opposed to instrumentalism and the formative models. We conceptualize service quality similar to an attitude (Parasuraman et al., 1988), and we consider the implementation of measures about attitudes more consistent with the reflective view. In addition, studies that use their proposed dimensions of service quality for predicting an overall evaluation of quality (e.g. Brady & Cronin, 2001; Cronin & Taylor, 1992) or a satisfaction judgement (e.g. Cronin & Taylor, 1992; Lam & Zhang, 1999) are in disaccord with researchers who advocate for the difficulties of inferring causality in cross-sectional designs (Kaplan, Harik, & Hotchkiss, 2000; Kline, 2006). This problem is widely explained in the context of service quality literature by Martínez (2006).

The presence of equivalent models is another interesting issue. Our third order structure is equivalent to a model with only first and second order factors. However, if we consider service quality as a multidimensional construct, it does not exist separately from its dimensions (Edwards, 2001). Thus, both models are not theoretically plausible and do not have semantic equivalence (Markus, 2004). The challenge is therefore to consider dimensions of quality as separate from overall service quality, i.e. to distinguish the attribute-level performance from global judgements about the service as disparate entities, with divergent meaning, antecedents and consequences. Mittal et al. (1998) show how a customer can evaluate two attributes of a service in a disparate way. If two attributes are very low correlated, the reflective view is open to debate. We have not detected this issue in our study, and all latents are highly correlated. However, the question is still on the table: Could the customer's evaluations of personal interaction be, for example, low correlated with facilities layout or tangible elements in a specific service? If we say yes, we should discard hierarchical multidimensional reflective models as a universal representation of the service quality construct.

The utility of multidimensional constructs has been seriously questioned (see Edwards, 2001 or Hayduk et al., 1995), specifically due to the operational problems for causal relationships among the multidimensional construct and other divergent concepts. Nevertheless, the interest of this research is not in analysing causal networks between different latents. We are working in the area of measurement validity, and we invite discussion regarding whether validity requires a descriptive theory of meaning that implies nomological networks (Borsboom et al., 2004; Cronbach & Meehl, 1955).

We encourage academics to take all these considerations into account in the study of service quality in services. Specifically in the travel agency industry, our study shows that a lesser degree of abstraction in the definition of certain attributes of service quality would be desirable. Therefore, our model serves as a starting point for improving the measurement of service performance in this sector. We have made some progress on the diagnostic properties of a specially designed service quality model for travel agencies. In an industry where different quality management systems are adopted (such as ISO certification or Q for tourist Quality) this instrument could serve as a diagnostic tool that will allow managers to determine service areas that are weak and in need of attention. Although this study is based on a non-random sample of customers, results of our hypothesized model are interpretable if we assume causal homogeneity and a specific sub-population of Spanish superpopulation. We expect

other researchers to replicate our model in other contexts and test situations (Cohen, 1994; Hitchcock, 2002) in order to generalize results; especially with more balanced stratum samples regarding age. We also stress the systematic utilization of service quality questionnaires by travel agencies in order to obtain a dynamic picture of evaluations and consumer future intentions over time, with the aim of enriching the dynamic view of customer attitudes towards service (Johnson, Herrmann, & Huber, 2006; Mittal, Katrichis, & Kumar, 2001).

Finally, further research should consider two key aspects in the analysis of service quality evaluations: the role of customer expectations distributions and the placebo effect of price and brand information on quality perception. Rust, Inman, Jia, and Zahorik (1999) have shown that experience is a crucial factor for quality evaluations, and thus companies should pay more attention to the quality delivered to customers with a low degree of experience in the service. This type of customer is more sensitive to disconfirmation that highly experienced customers. Therefore, travel agencies should include this variable to compare service quality evaluations for different groups of customers with disparate degrees of experience. On the other hand, as Shiv, Carmon, and Aricly (2005) and Rao (2005) have indicated, price or brand information could exert an unconscious influence on expectations about quality, similar to a placebo effect. These variables also have to be considered in order to obtain a more detailed view of service performance; thus, perception of service quality mediocrity or excellence should be interpreted with caution because it could have been influenced by several aspects that are not a specific part of quality performance.

Appendix A. Dimensions, subdimensions and indicators[9]

Personal interaction	It refers to the customers' subjective perception of how the service is delivered during the service encounter. It is interaction between the service provider and customer.
Conduct	It refers to attitudes and behaviour of employees.
	• Employees seek the best for the customers • The attitude of employees demonstrates their willingness to help me
Expertise	It refers to degree of knowledge of employees about service.
	• The employees have knowledge enough about different services to respond to consumer's requests
	• The employees are competent
Problem solving	It measures the store personnel's ability to handle potential problems and complaints.
	• When a customer has a problem, this company shows a sincere interest in solving it • The employees are able to handle customer complaints efficiently
Physical environment	It refers to the built environment in which service delivery occurs as opposed to the natural or social environment.
Equipment	Includes the computers and other implements used to achieve effective service delivery.
	• The equipment at this store is in good condition • Employees have a neat and professional appearance
Ambient conditions	It relates to the non-visual aspects associated with the service-producing system and includes factors such as temperature, scent and music, as well as appearance, decor and ambience, and furniture layout.
	• The store's ambiance (clean, music, temperature…) is appropriate • The store is safe and comfortable
Outcome	It refers to the outcome of the service act and represents what the consumer gains from the service.
Waiting time	This factor is associated with the ability of the service firm to manage demand in an effective manner.
	• This company provides service to the consumers punctually • This company show its interest in accelerating the service
Valence	It 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.
	• When I leave this company, I always feel that I got what I wanted • I would evaluate the outcome of this service favourably

[9]First and second best indicators, respectively.

Appendix B. The form of structural equations model
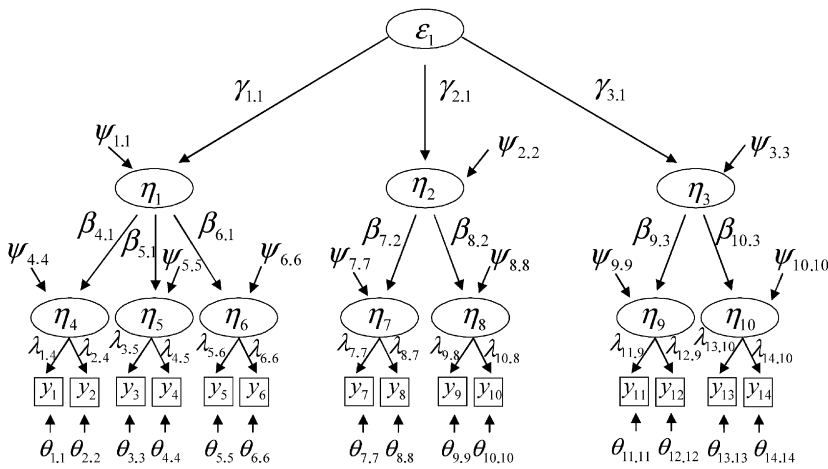
Appendix C

C.1. Derivation of the starting values for theta parameters ($\theta_{i,i}$)

Let y_i, with $i = 1 \ldots 14$ the observable indicators, η_k with $k = 1 \ldots 10$ the endogenous variables, $\lambda_{i,k}$ the causal coefficients from latents to indicators, and $\theta_{i,i}$ the indicator error variances.

For the gold standard indicator (y_{14}) of the "valence" subdimension (η_{10}), the variance of the indicator is:

$$Var(y_{14}) = (\lambda_{14,10})^2 Var(\eta_{10}) + \theta_{14,14}. \tag{1}$$

As ($\lambda_{14,10}$) is fixed to 1.0, the endogenous variance will be close to the variance of the indicator by an amount determined by $\theta_{14,14}$. Therefore, $\theta_{14,14}$ could be fixed to a percentage of the observable indicator variance. In this case: $\theta_{14,14} = 15\% \, Var(y_{14})$.

This reasoning is extensible to the remaining gold standard indicators.

C.2. Derivation of the starting values for Psi parameters ($\psi_{k,k}$)

Let $\psi_{k,k}$ the error variances of the endogenous latent variables, ε_j with $j = 1$ the exogenous variable, $\gamma_{k,j}$ the causal coefficients from exogenous to endogenous variables, and $\beta_{k,k}$ the causal coefficients among endogenous variables.

For the "valence" subdimension (η_{10}), the variance of this latent is:

$$Var(\eta_{10}) = (\beta_{10,3})^2 Var(\eta_3) + \psi_{10,10}. \tag{2}$$

As $\beta_{10,3}$ is fixed to 1.0, and considering Eq. (1), we can write the next expression:

$$Var(y_{14}) - \theta_{14,14} = Var(\eta_3) + \psi_{10,10}. \tag{3}$$

Being $\theta_{14,14}$ fixed, $\psi_{10,10}$ can be fixed to a percentage of the left-hand side of Eq. (3). In this case: $\psi_{10,10} = 15\%(Var(y_{14}) - \theta_{14,14})$.

The derivation of $\psi_{7,7}$ and $\psi_{4,4}$ is done by a similar way. Finally, for deriving $\psi_{3,3}$, we have to count with the next equation:

$$Var(\eta_3) = (\gamma_{3,1})^2 Var(\varepsilon_1) + \psi_{3,3}. \tag{4}$$

As $\gamma_{3,1}$ is fixed to 1.0, and considering Eq. (3), we can write the following expression:

$$Var(y_{14}) - \theta_{14,14} - \psi_{10,10} = Var(\varepsilon_1) + \psi_{3,3}. \tag{5}$$

Being $\theta_{14,14}$ and $\psi_{10,10}$ fixed, $\psi_{3,3}$ can be fixed to a percentage of the left-hand side of Eq. (4). In this case: $\psi_{3,3} = 15\%(Var(y_{14}) - \theta_{14,14} - \psi_{10,10})$.

References

Albrecht, K. (1992). *The only thing that matters*. New York: Harper Business.

Babakus, E., & Boller, G. W. (1992). An empirical assessment of the SERVQUAL scale. *Journal of Business Research, 24*, 235–268.

Babakus, E., & Mangold, D. W. G. (1992). Adapting the SERVQUAL scale to hospital services: and empirical investigating. *Health Services Research, 26*(6), 767–786.

Bentler, P. M. (2000). Rites, wrongs, and gold in model testing. *Structural Equation Modeling, 7*, 82–91.

Bigné, E., Martínez, C., Miquel, M. J., & Belloch, A. (1996). La Calidad de Servicio en las Agencias de Viaje. Una Adaptación de la Escala SERVQUAL. *Revista Europea de Dirección y Economía de la empresa, 5*(2), 7–18.

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. J., & Hubbert, A. R. (1994). Encounter satisfaction versus overall satisfaction versus quality. In R. Rust, & R. Oliver (Eds.), *Service quality: New directions in theory and practice* (pp. 72–94). London: Sage Publications, Inc.

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

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review*, *110*(2), 203–219.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, *111*(4), 1061–1071.

Bowen, D. E., & Schneider, B. (1985). Boundary-spanning-role employees and the service encounter: Some guidelines for management and research. In J. A. Czepiel, M. R. Solomon, & C. F. Surprenant (Eds.), *The service encounter: Managing employee/customer interaction in service business* (pp. 127–147). Lexington, MA: D.C. Heath and Company.

Brady, M. K., & Cronin, J. J., Jr. (2001). Some new thoughts on conceptualizing perceived service quality: A hierarchical approach. *Journal of Marketing*, *65*(July), 34–49.

Brady, M. K., Cronin, J. J., & Brand, R. R. (2002). Performance-only measurement of service quality: A replication and extension. *Journal of Business Research*, *55*, 17–31.

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

Capelleras, J. L., & Veciana, J. M. (2002). Service quality in university education: An empirical assessment. *In Thirty first European marketing academy conference*. Braga (Portugal).

Carman, J. (1990). Consumer perceptions of service quality: An assessment of the SERVQUAL dimensions. *Journal of Retailing*, *66*, 33–55.

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

Caruana, A. (1997). Corporate reputation: Concept and measurement. *Journal of Product and Brand Management*, *6*(2), 109–118.

Chumpitaz, R., & Swaen, V. (2002). Service quality and brand loyalty relationships: investigating the mediating effect of costumer satisfaction. *In Thirty first European Marketing Academy Conference*, Braga (Portugal).

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

Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*, *49*, 1304–1312.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological test. *Psychological Bulletin*, *52*, 218–302.

Cronin, J. J., & Taylor, S. (1992). Measuring service quality: A rexamination and extension. *Journal of Marketing*, *56*, 55–68.

Cronin, J. J., Brady, M. K., & Hult, G. T. M. (2000). Assessing the effects of quality, value, and customer satisfaction on consumer behavioral intentions in service environments. *Journal of Retailing*, *76*(2), 193–218.

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.

Dowling, G. R. (1988). Measuring corporate images: A review of alternative approaches. *Journal of Business Research*, *17*, 27–34.

Edwards, J. R. (2001). Multidimensional constructs in organizational behavior research: An integrative analytical framework. *Organizational Research Methodology*, *4*, 144–192.

Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods*, *5*, 155–174.

Epstein, S. (1983). Aggregation and beyond: Some basic issues on the prediction of behaviour. *Journal of Personality*, *51*, 360–392.

Fick, R. P., & Ritchie, R. B. (1991). Measuring service quality in the travel and tourism industry. *Journal of Travel Research*, *30*(2), 2–9.

Fombrun, C. J., & Shanley, M. (1990). What's in a name? Reputation building and corporate strategy. *Academy of Management Journal*, *33*(2), 233–258.

Fullerton, G. (2005). The service quality–loyalty relationship in retail services: Does commitment matter? *Journal of Retailing and Consumer Services*, *12*(2), 99–111.

Grace, D., & O'Cass, A. (2005). Service branding: Consumer verdicts on service brands. *Journal of Retailing and Consumer Services*, *12*(2), 125–139.

Grönroos, C. (1982). *Strategic management and marketing in the service sector*. Swedish School of Economics and Business Administration.

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 the moment of truth in service competition*. United States: Lexington Book.

Grönroos, C., Heinonen, F., Isoniemi, K., & Lindholm, M. (2000). The Netoffer model: A case example from the virtual marketspace. *Management Decision*, *38*(4), 243–252.

Hayduk, L. A. (1996). *LISREL issues, debates and strategies*. Baltimore, MD: Johns Hopkins University Press.

Hayduk, L. A., & Glaser, D. N. (2000). Jiving the four-step, waltzing around factor analysis, and other serious fun. *Structural Equation Modeling*, *7*, 1–35.

Hayduk, L., Ratner, P. A., Johnson, J. L., & Bottorff, J. L. (1995). Attitudes, ideology and the factor model. *Political Psychology*, *16*(3), 479–507.

Hitchcock, C. (2002). Probabilistic causation. In E. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2002 ed). Stanford: Stanford University Press. Available at ⟨http://plato.stanford.edu/archives/fall2002/entries/causation-probabilistic/⟩.

Jackson, D. L. (2003). Revisiting sample size and number of parameter estimates: Some support for the N:q hypothesis. *Structural Equation Modeling*, *10*(1), 128–141.

Johns, N., Avci, T., & Karatepe, O. (2004). Measuring service quality of travel agents: Evidence from Northern Cyprus. *The Service Industries Journal*, *24*(3), 82–100.

Johnson, M. D., Herrmann, A., & Huber, F. (2006). The evolution of loyalty intentions. *Journal of Marketing*, *70*, 122–132.

Jones, E. (2005). *The importance of communication quality in services*. Doctoral dissertation. The Florida State University. USA.

Jöreskog, K. (1999). *How large can a standardized coefficient be*? Cited in Jöreskog, K. G., & Sörbom, D. (2001).

Jöreskog, K. G., & Sörbom, D. (2001). *LISREL 8.50*. Chicago: Scientific Software International.

Kaplan, D., Harik, P., & Hotchkiss, L. (2000). Cross-sectional estimation of dynamic structural equation models in disequilibrium. In R. Cudeck, S. H. C. du Toit, & D. Sorbom (Eds.), *Structural equation modeling: Present and future. A Festschrift in honor of Karl G. Joreskog* (pp. 315–339). Lincolnville: Scientific Software International.

Karatepe, O. M., & Avei, T. (2001). *Trait validity assessment of a service quality measure: Evidence from lodging services in Northern Cyprus*. Unpublished manuscript, Gazimagusa, Northern Cyprus. Eastern Mediterranean University, School of Tourism and Hospitality Management.

Karatepe, O. M., Yavas, U., & Babakus, E. (2005). Measuring service quality of banks: Scale development and validation. *Journal of Retailing and Consumer Services*, *12*, 373–383.

Kim, S., & Jin, B. (2002). Validating the retail service quality scale for us and Korean customer of discount stores: An exploratory study. *Journal of Services Marketing*, *16*(3), 223–237.

Kline, R. B. (2005). *Principles and practice of structural equation modelling* (second ed.). New York: The Guildford Press.

Kline, R. B. (2006). Reverse arrow dynamics. In G. R. Hancock, & R. O. Mueller (Eds.), *A second course in structural equation modelling* (pp. 43–68). Greenwich, CT: Information Age Publishing.

Ko, Y. J., & Pastore, D. L. (2004). Current issues and conceptualizations of service in the recreational sport industry. *Sport Marketing Quarterly*, *13*(3), 159–167.

Ko, Y. J., & Pastore, D. L. (2005). A hierarchical model of service quality in the recreational sport industry. *Sport Marketing Quarterly*, *14*(2), 84–97.

Lam, T., & Zhang, H. Q. (1999). Service quality of travel agents: The case of travel agents in Hon Kong. *Tourism Management*, 20, 341–349.

Law, K. S., Wong, C., & Mobley, W. H. (1998). Towards a taxonomy of multidimensional constructs. *Academy of Management Review*, 23(4), 741–755.

LeBlanc, G. (1992). Factor affecting customer evaluation of service quality in travel agencies: An investigation of customer perception. *Journal of Travel Research*, 30(4), 10–17.

Lewis, R. C. (1987). The measurement of gaps in the quality of hotels services. *International Journal of Hospitality Management*, 6, 38–88.

Liljander, V., Van Riel, A. C. R., & Pura, M. (2002). Customer satisfaction with e-services: The case of an on-line recruitment portal. In M. Bruhn, & B. Stauss (Eds.), *Jahrbuch Dienstleistungsmanagement* (pp. 407–432). Wiesbaden: Gabler, Electronic Services.

Lui, C. M. (2005). The multidimensional and hierarchical structure of perceived quality and customer satisfaction. *International Journal of Management*, 22(3), 426–435.

Markus, K. A. (2004). Statistical equivalence semantic equivalence, eliminative induction, and the Raykov Marcoulides proof of infinite equivalence. *Structural Equation Modeling*, 9(4), 503–522.

Martínez, J. A. (2006). *La relación entre la calidad percibida, satisfacción e imagen corporativa. Una propuesta de modelización causal con ecuaciones estructurales, en diseños de corte transversal.* Unpublished Dissertation. Universidad Politécnica de Cartagena, Spain.

Martínez, L., & Martínez, J. A. (2007). Measuring perceived service quality in urgent transport service. *Journal of Retailing and Consumer Services*, 14(1), 60–72.

Mattila, A. S. (1999). The role of culture in the service evaluation process. *Journal of Service Research*, 1(3), 250–261.

McAlexander, J. H. K., Dennis, O., & Koenig, H. F. (1994). Service quality measurement. *Marketing Health Services*, 14(3), 34–44.

McDougall, G., & Levesque, T. (1994). A revised view of service quality dimensions: An empirical investigation. *Journal of Professional Service Marketing*, 11(1), 189–209.

Millán, A., & Esteban, A. (2004). Development of a multiple-item scale for measuring customer satisfaction in travel agencies services. *Tourism Management*, 25, 533–546.

Mittal, V., Katrichis, J. M., & Kumar, P. (2001). Attribute performance and customer satisfaction over time: Evidence from two field studies. *Journal of Services Marketing*, 15(5), 343–356.

Mittal, V., Ross, W. T., Jr., & Baldaser, P. M. (1998). The asymmetric impact of negative and positive attribute-level performance on overall satisfaction and repurchase intentions. *Journal of Marketing*, 62(January), 33–47.

Mulaik, S. A., & Millsap, R. E. (2000). Doing the four-step right. *Structural Equation Modeling*, 7, 36–73.

Oliver, R. L. (1993). Cognitive, affective, and attribute bases of the satisfaction response. *Journal of Consumer Research*, 20, 418–430.

Parasuraman, A., Zeithaml, V., & Berry, L. (1985). A conceptual model of service quality and its implications for future research. *Journal of Marketing*, 49(Fall), 41–50.

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.

Patterson, P. G., & Spreng, R. A. (1997). Modelling the relationship between perceived value, satisfaction and repurchase intentions in a business-to-business services context: An empirical examination. *International Journal of Service Industry Management*, 8(5), 414–434.

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5), 879–903.

Rao, A. R. (2005). The quality of price as a quality cue. *Journal of Marketing Research*, 42(November), 401–405.

Richins, M. L. (1997). Measuring emotions in the consumption experience. *Journal of Consumer Research*, 24, 127–146.

Rust, R. T., Inman, J. J., Jia, J., & Zahorik, A. (1999). What you don't know about customer-perceived service quality: The role of customer expectations distributions. *Marketing Science*, 18(1), 77–92.

Rust, R. T., & Oliver, R. L. (1994). Service quality: Insights and managerial implications from the Frontier. In R. T. Rust, & R. L. Oliver (Eds.), *Service quality: New directions in theory and practice* (pp. 1–19). London: Sage Publications.

Ryan, C., & Cliff, A. (1997). Do travel agencies measure up to customer expectation? An empirical investigation of travel agencies' service quality As measured by SERVQUAL. *Journal of Travel and Tourism Marketing*, 62, 1–32.

Saleh, F., & Ryan, C. (1991). Analysing service quality in the hospitality industry using the SERVQUAL model. *The Service Industries Journal*, 11(3), 324–343.

Satorra, A., & Bentler, P. M. (1988). Scaling corrections for chi-square statistics in covariance structure analysis. In: American Statistical Association (Ed.), *In Proceedings of the business and economic statistics section* (pp. 308–313). Alexandira, VA: American Statistical Association.

Setó, D. (2003). La Influencia de la Calidad de Servicio, Imagen, la Satisfacción y la Confianza en la Fidelidad del Cliente. *Revista Española de Investigación de Marketing*, 7(1), 27–55.

Shiv, B., Carmon, Z., & Aricly, D. (2005). Placebo effects of marketing actions: Consumers may get what they pay for. *Journal of Marketing Research*, 42(November), 383–393.

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(April), 67–80.

Steenkamp, J. B. E. M., & 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., & Fouladi, R. T. (1992). R2: A computer program for interval estimation, power calculation, and hypothesis testing for the squared multiple correlation. *Behavior Research Methods, Instruments, and Computers*, 4, 581–582.

Teas, R. (1993). Expectations, performance evaluation, and consumer's perceptions of quality. *Journal of Marketing*, 57, 18–34.

Ueltschy, L. C., & Krampf, R. F. (2001). Cultural sensitivity to satisfaction and service quality measures. *Journal of Marketing Theory and Practice*, 9, 12–31.

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

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

Zeithaml, V. A., Parasuraman, A., & Malhotra, A. (2002). Service quality delivery through web sites: A critical review of extant knowledge. *Journal of the Academy of Marketing Science*, 30(4), 362–375.