1-1-1999

Testing the SERVQUAL Scale in the Business-to-Business Sector: The Case of Ocean Freight Shipping Service

Srinivas Durvasula
Marquette University, srinivas.durvasula@marquette.edu

Steven Lysonski
Marquette University, steven.lysonski@marquette.edu

Subhash Mehta
University of Southern Queensland

Testing the SERVQUAL Scale in the Business-to-Business Sector: The Case of Ocean Freight   Shipping Service

Srinivas Durvasula

Senior Fellow of Marketing, National University of Singapore, Singapore and Associate Professor of Marketing, College of Business Administration, Marquette University
Milwaukee, WI

Steven Lysonski

Associate Professor of Marketing, College of Business Administration, Marquette University
Milwaukee, WI

Subhash C. Mehta

Associate Professor of Marketing, Faculty of Business Administration, National University of Singapore
Singapore

Abstract:

A key question is whether the instruments developed for consumer services can accurately gauge the service quality perceptions of organisational customers. Reports psychometric testing of the SERVQUAL as a measure of service quality in ocean freight services. Based on a survey of a cross-sectional sample of 114 business organisations in Singapore, which regularly utilise ocean freight services for their export needs, this study found that the psychometric properties of the SERVQUAL scale are at variance with those found in consumer services settings. Further, the SERVQUAL perceptions scores were found to be a better predictor than the SERVQUAL gap scores. In

sum, the service quality measures developed for consumer services can only be applied with caution in business-to-business marketing. Implications and future directions for research are discussed.

Introduction

Developing productive and profitable relationships with customers is a key goal of firms not only in the consumer sector but also in the business-to-business sector. Indeed, noteworthy attention has been given to understanding the dimensions associated with establishing and maintaining long-term relationships with customers. The emerging area of "relationship marketing" underscores the keen interest in explicating the dimensions associated with such relationships. In fact, some marketing scholars (Webster, 1992; Achrol, 1997) maintain that building relationships has become the major focus of an increasing number of businesses. The level of quality in the way a firm delivers its service to industrial customers is central to this relationship.

In recent years, substantial research has examined the level of quality (and its dimensions) in the performance of a service. Research on UK service-based firms has found empirical support for the belief that a service company's excellence in the way it delivers its service offering is related to its business performance (Caruana *et al.*, 1995). It is likely that certain dimensions in the service encounter have a greater impact on this relationship than others. Bitner (1995) and Berry (1995) argue that the paradigm of relationships provides a crucial way to understand services marketing. Berry (1995) refers to "level three relationship marketing" as a specific dimension in service marketing that focuses on the value-added benefits the service marketer provides to customers. Offering a high quality service is one visible way by which a company can distinguish itself from its competitors in building a close relationship with customers and attaining a competitive advantage. Service quality may not be easy to copy since certain characteristics in an organization's culture are often required to develop and sustain it. Hise and Gabel (1995) found that customer service could serve as a strategic weapon even in international logistics systems, while Lovelock and Yip (1996) argue

that services management is taking on greater international significance in the development of global strategies.

Building strong customer-focused relationships requires understanding the needs of specific customers and the firm's success in meeting these needs. As such, a means to measure the perceptions of customers' experiences in the services encounter becomes critical (Parasuraman *et al*., 1991). A customer's evaluation of service quality and the concomitant satisfaction is thought to be connected to repurchase, loyalty, and ultimately a firm's profitability (Iacobucci *et al.*, 1994). Towards this end, researchers have devised various instruments to measure the level of marketing orientation such as marketing audits. In particular, in the service realm, considerable research has focused on measuring service quality in the consumer sector.

Among the several quality measures that have been proposed, the SERVQUAL scale as developed and subsequently modified by Parasuraman *et al.* (1991) has received the most recognition. The SERVQUAL scale has been a driving force in the research activity measuring service quality. The SERVQUAL scale looks at five dimensions inherent in the service relationship such as reliability, and responsiveness, among others. This approach has been applied to a host of consumer service organizations such as banks, insurance agencies, real estate brokerage firms, hospitals, dental clinics and telephone companies. The application of this scale to the consumer sector contrasts sharply with the relative absence of studies using this scale in the business-to-business sector. Cooper and Jackson (1988) suggest that "the characteristics of consumer services would seem to transfer readily to the industrial services sector without modification".

Only four studies have looked at this scale in the industrial arena compared with more than 15 studies in the consumer realm (Asubonteng *et al*., 1996). The results in using the SERVQUAL scale have been mixed in the industrial setting. For example, an application of the SERVQUAL scale in the mainframe software industry by Pitt *et al*. (1996) found that the instrument's reliability and validity scored well with only discriminant validity being problematic. Moreover, Young and Varble (1997) used the SERVQUAL scale successfully to examine a purchasing department. Similarly, Farley *et al.* (1990) used the SERVQUAL approach effectively to measure the service quality perceived by multinational customers of a manufacturer with

multinational manufacturing locations. In contrast, the study by Bresinger and Lambert (1990) using SERVQUAL on motor carriers' transportation services found the instrument to have inferior predictive validity. Van Dyke *et al.* (1997) provide an excellent discussion concerning the serious limitations of the SERVQUAL scale. The paucity of research in the industrial sector and the conflicting findings prohibits us from knowing if this scale is applicable for industrial uses.

The goal of this research is to further our understanding of the applicability and robustness of the SERVQUAL scale to business-to-business services. If firms wish to develop viable relationships, it is imperative that they understand the perceptions of their services in a business-to-business context. To achieve our objective we collected data from ocean freight shipping companies in Singapore to assess the psychometric properties of the SERVQUAL scale. Although the findings of this study focus on one industrial area, the results are applicable to other business-to-business services. The paper is organized as follows. The first part discusses the SERVQUAL scale, its background and dimensions. The methodology in terms of sampling, data collection and the measures is then discussed. The results section reports the psychometric analyses using tests for dimensionality, reliability, convergent, discriminant and nomological validity. The study also examines the correlation between customer satisfaction with the dimensions of the SERVQUAL scale. The paper concludes with discussion of the applicability of the SERVQUAL scale to the business-to-business sector, implications and directions for future research.

About the SERVQUAL scale

When applying the SERVQUAL scale as proposed by Parasuraman *et al.* (1988, 1991), a set of 22 items are used to measure perceptions of the actual service provided by a company and a similar set of 22 items are used to measure the level of service expected (i.e. expectations) from an excellent service provider. An example of a perceptions item is "XYZ shipping line has modern-looking equipment," responded to on a 1 = strongly agree to 7 = strongly disagree scale. The corresponding expectations item would read as "Excellent shipping lines have modern looking equipment," also measured on a 1 = not at all essential to 7 = absolutely essential scale. The 22-item gap scores are then obtained by subtracting the perceptions score of each item from the corresponding expectations

item score. It is the 22-item gap score that is known as the SERVQUAL scale.

Based on factor analysis of several applications of the scale, Parasuraman *et al.* (1991) identified that the SERVQUAL scale has five perceptual dimensions, namely:

1) tangibles,

2) reliability,

3) responsiveness,

4) assurance, and

5) empathy.

To quantify each of these dimensions, questions were developed. For example, items related to tangibles are described as "XYZ shipping line's physical facilities are visually appealing", "XYZ's employees are neat-appearing". Reliability is measured by items such as "XYZ shipping line performs the service right the first time", "XYZ shipping line insists on error-free records." Examples of responsiveness measures are "Employees of XYZ shipping line give you prompt service," "Employees of XYZ shipping line are always willing to help you." Assurance is measured by items such as "You feel safe in your transactions with XYZ shipping line," "Employees of XYZ are consistently courteous with you." Finally, examples of items categorized under empathy are "XYZ shipping line gives you individual attention," "XYZ shipping line has your best interests at heart." Items measuring expectations can be similarly constructed, but worded such as "Excellent shipping lines will have your best interests at heart," etc.

Several researchers examined the properties of the SERVQUAL scale using data from the consumer sector. While some of them questioned the five-factor dimensionality of the SERVQUAL scale, others provided it support (see Parasuraman *et al.,* 1991). The usefulness of the gap scores to represent service quality was also questioned by some researchers on conceptual and empirical grounds. For example, Brown *et al.* (1993) showed that gap scores in general demonstrate poor reliability. The construct validity of gap scores is also suspect because gap scores would have a theoretically high correlation with their component scores (e.g. perceptions and expectations). As a result, gap scores are not likely to be distinct from

their component scores. Further, Brown *et al.* (1993) suggested that gap scores suffer from "variance restriction" problems that would prevent their usage in certain types of statistical analyses. Empirically, Cronin and Taylor (1992) and Babakus and Boller (1992) showed that the perceptual components of SERVQUAL outperformed SERVQUAL gap scores in predicting overall service quality, suggesting that service quality is better measured by perceptions than by gap scores. However, some of these criticisms were refuted by Parasuraman *et al.* (1993), who argued that gap scores indeed provide useful information about service quality. The debate concerning the efficacy of the SERVQUAL scale continues. Since the bulk of extant studies examining the statistical reliability of the SERVQUAL scale were conducted in the consumer sector, judgments about the scale must be made in this context. In this study, the SERVQUAL scale will be examined for its psychometric properties using a sample of industrial customers.

Methodology

Sample and questionnaire administration

The data were collected from shipping managers of various organizations in Singapore who regularly use the services of ocean freight companies for their exporting requirements. A sample of 114 shipping managers participated in the study. Their names were randomly drawn from a list of all shippers, obtained from a large shipping company. While some of the shippers on the list represented customers of that shipping company, others were customers of competing shipping lines. Skilled interviewers who had experience in conducting managerial interviews personally administered the survey. A profile of the surveyed organizations that the shipping managers represented is as follows: 41.2 percent of the organizations dealt with trading only, 17.5 percent dealt with manufacturing activities, and 26.3 percent had both trading and manufacturing activities. About 50 percent of the organizations were domestic, 18.4 percent foreign owned, and 31.6 percent joint ventures with a foreign partner. The annual turnover of the organizations varied from less than US$10 million (for 49.1 percent of the companies) to more than US$25 million (for about 26 percent). The annual export freight expenses for these organizations ranged from less than US$25,000 (for 25.7 percent of the companies) to more than US$100,000 (for 35.4 percent).

Measures

Each respondent was first asked to indicate his/her company's most preferred shipping line on that company's busiest route for exports. (The reason for focusing on the most preferred shipping lines is that the respective organizations that participated in the study used services of these shipping lines on a regular basis. As a result, subjects are familiar with the most preferred shipping lines and are in a position to better evaluate their service quality. By obtaining service quality perceptions of the most preferred shipping lines, this method allows for establishing benchmarks for service quality in the shipping industry. Given the focus of this investigation on the psychometric properties of SERVQUAL scale as opposed to assessing mean service quality ratings, it is not critical to obtain data from a range of shipping lines that varied from most preferred to least preferred. However, by limiting the evaluations to a specific type of shipping line such as the most preferred shipping line, our method provides a common reference point for data collection from various organizations.) The SERVQUAL scale was then administered to the respondents, and their perceptions (of their most preferred shipping lines' service) and expectations (of service from excellent shipping lines) were obtained using the 22-item scales (as explained in the "About the SERVQUAL scale" section). SERVQUAL gap scores were then obtained by subtracting SERVQUAL perception scores from expectation scores.

Other measures were also employed for SERVQUAL validation purposes. For example, overall evaluation of the shipping line's service was measured on a scale, 1 = extremely poor, 7 = excellent. Next, shipping lines often divide their organizations into a number of specialized departments, with each department having an interface with the customers. As such, customers' perceptions of the shipping lines' service quality is shaped by their experiences with the interfacing departments. In this study, four such interfacing departments:

1) booking services,

2) documentation,

3) operations, and

4) claims

and three supporting activities:

1) marketing/sales department,

2) telephone services, and

3) service rendered during personal visits

were identified. Respondents were asked to rate these interfacing departments and activities on multi-item seven-point rating scales. Composite indices of these scales all exhibited high coefficient alpha reliabilities exceeding 0.9.

Results

Analysis overview

The major objective of this study is to assess the psychometric properties of SERVQUAL when applied to ocean freight services. Given this objective and consistent with previous studies in this area (see Babakus and Boller, 1992), the data were analyzed using covariance structure analysis via Lisrel VII (Joreskog and Sorbom, 1989). Analyses were performed on both SERVQUAL perception scores and gap scores. The results were used to answer the following questions:

1) Does the SERVQUAL scale exhibit the same five-dimensional factor structure (representing the dimensions of tangibles, reliability, responsiveness, assurance, and empathy) in a business- to-business context as found in consumer marketing?

2) Is the SERVQUAL measure internally consistent? Do the items of each SERVQUAL dimension provide a reliable measure of that dimension?

3) Do the five dimensions of the SERVQUAL measure exhibit convergent validity?

4) Are the five SERVQUAL dimensions indeed distinct from each other and do they exhibit discriminant validity?

5) Are the SERVQUAL dimensions related to other types of service quality measures? Hence, does the SERVQUAL measure have nomological validity?

6) Does the SERVQUAL measure predict overall customer satisfaction with the service provider? Which measure of service quality, perceptions or the gap scores, has the better predictive ability?, and

7) Is the perceptions component of the SERVQUAL measure indeed distinct from the gap scores?

The following results provide answers to these questions.

Dimensionality

The dimensionality of the SERVQUAL scale was assessed by comparing the fit of the hypothesized model for that scale with a number of other competing models. As applied to consumer services, the SERVQUAL measure developed by Parasuraman *et al.* (1991) represents five dimensions, namely:

- tangibles,
- reliability,
- responsiveness,
- assurance, and
- empathy.

Hence, we considered the five-factor SERVQUAL measure as the initial hypothesized model to be tested. Several competing models of SERVQUAL measure were also examined. For example, in a subsequent study, Parasuraman *et al*. (1991) performed factor analysis and found that items representing responsiveness and assurance loaded on the same factor. Based on these findings, an alternative model is the four-factor model where responsiveness and assurance dimensions are combined to form one factor. An extension of this model is the three-factor model, where in addition to the separate tangibles and reliability factors, responsiveness, assurance, and empathy were combined to form the third factor.

Babakus and Boller (1992) identified only two dimensions in the SERVQUAL scale. Exploratory factor analysis of the shipping data also identified two dimensions, one representing tangibles and the other representing intangibles (where reliability, responsiveness, assurance, and empathy are combined into one factor). Hence the two-factor

model was also considered as an alternative model representing the dimensionality of the SERVQUAL measure. Finally, the one-factor model, where all five SERVQUAL dimensions were combined into one factor was also considered. The same competing models were considered for both SERVQUAL perception scores and gap scores.

Table I features several fit indices used for evaluating the various competing models of the SERVQUAL measure. Examining the SERVQUAL perception scores, first, the $\chi 2$ fit of the hypothesized (five-factor) model (449.44, 199 df.) was compared to the $\chi 2$ fit of the four-factor model (451.43, 203 df.). The fit of the five-factor model is not significantly different from that of the four-factor model (c2 diff. = 1.99, df. = 4, p > 0.05), implying that the parsimonious four-factor model is a better representation of SERVQUAL dimensionality than the five-factor model. Next, the four-factor model was compared to the fit of the three-factor model ($\chi 2$ = 456.91, df. = 206). The difference in $\chi 2$ fit between these two models (5.48, 3 df.) is also not significantly different (p > 0.05), implying that the relatively more parsimonious three-factor model is preferred. The three-factor model also provides a significantly better fit when compared to the two-factor model ($\chi 2$ diff. = 23.53, df. = 2, p < 0.05) and the one-factor model ($\chi 2$ diff. = 34.77, df. = 3, p < 0.05). Hence, the $\chi 2$ fit statistics support the three-factor model, representing the dimensions of tangibles, reliability, and the combination of responsiveness, assurance, and empathy.

[Table I]

The use of $\chi 2$ fit statistics for comparing the fit of alternative models was criticized by some researchers (see Marsh, 1994) because the $\chi 2$ value is sensitive to sample size and for larger samples, even a small degree of lack of fit becomes statistically significant. This would lead to the rejection of a more parsimonious model (e.g. two-factor or one-factor models) in favour of a less parsimonious or a more elaborate model (e.g. three-factor model). Hence, to overcome this problem, a variety of other fit indices were examined as seen in Table I. Specifically, the fit indices considered in this study were the root-mean-square residual (RMR), goodness-of-fit index (GFI), adjusted goodness-of-fit index (AGFI), normed fit index (NFI), the $\chi 2$/df ratio,

comparative fit index (CFI), and the Tucker-Lewis Index (TLI). Most of these indices are available as part of the Lisrel output. Bollen and Long (1993) compiled articles from researchers who evaluated most of the above fit indices. When examining the above fit indices, GFI, AGFI, CFI, NFI, and TLI values of "1" provide a perfect fit of the underlying model to the data. For RMR, a residual of close to zero, and for $\chi2/df$ ratio, a value of three or less is preferred. When using these fit indices, it can be seen that for SERVQUAL perception scores, all of the fit indices are relatively stable when going from the five-factor model to the one-factor model, implying that perhaps the most parsimonious one-factor model provides the best representation of the data. This would imply that the SERVQUAL measure is unidimensional.

When repeating this analysis for gap scores, the results of $\chi2$ fit support the four-factor model, whereas other fit indices point to a one-factor model. In sum, the dimensionality tests indicate that, while the hypothesized five-factor model as conceptualized by Parasuraman *et al*. (1991) provides acceptable level of fit in terms of various fit indices, the more parsimonious models offer equally good fit and that they should be preferred because of their parsimony or simplicity.

Reliability

The composite reliability estimates (Fornell and Larcker, 1981), which are analogous to coefficient alpha estimates, were computed next. Table II provides the reliability estimates for both SERVQUAL perception scores and gap scores. For any SERVQUAL dimension, a reliability estimate of 0.7 or better is considered acceptable (Nunnally and Bernstein, 1994). For both perception scores and gap scores, the reliability estimates of the five-factor model are generally acceptable (with the exception of the tangibles dimension whose reliability estimates are relatively modest). When responsiveness and assurance dimensions are combined into one factor (to represent the four-factor model), the reliability estimates of this factor are still acceptable. Similarly, as the various dimensions are combined one after the other, the reliability estimates are all found to be acceptable and increasing, ultimately leading to the one-factor model whose reliability estimates are the highest. This is not unexpected, as the reliability levels generally increase when additional items are added to a factor. In sum, the reliability analysis provides some support for the five-factor

model (given the relatively modest reliability score for the tangibles dimension), but stronger support for more parsimonious models.

[Table II]

Convergent validity

Convergent validity of the SERVQUAL perception and gap scores was examined next. This was done by computing the "average variance extracted" by each SERVQUAL dimension from its underlying indicators. An average variance extracted of at least 0.50 (i.e. 50 percent) provides support for convergent validity (see Gerbing and Anderson, 1988; Fornell and Larcker, 1981). From examining the average variance extracted scores in Table III, it is clear that with the exception of the tangibles dimension, these estimates are above 0.5 for the remaining dimensions of the SERVQUAL perception scores. As for gap scores, only the reliability and responsiveness dimensions have average variance extracted scores in excess of 0.5. In sum, the convergent validity test provides moderate support for the five-factor model of SERVQUAL perception scores, but only a weak support for the five-factor model of SERVQUAL gap scores.

Discriminant validity

This was assessed in several ways. First, the various dimensions of the SERVQUAL scale are considered to be distinct from each other (and exhibit discriminant validity) if the correlations among those dimensions are small and if the confidence intervals around those correlations do not contain a value of "1". The correlations among the SERVQUAL dimensions for both perception and gap scores as well as the standard errors are provided in Table IV. These estimates were obtained using Lisrel VII. When examining the 95 percent confidence intervals around these correlations, only the tangibles dimension appeared to be distinct from the other four dimensions for both the perception and gap scores. This result would imply that apart from tangibles, the other four dimensions are not distinct from each other, and that they could be combined into one factor. Hence, there is support only for the two-factor model, where tangibles and intangibles

(representing reliability, responsiveness, assurance, and empathy) are the two dimensions.

Nunnally and Bernstein (1994) criticize the assumption that the maximum correlation between any two dimensions is "1", if multiple items are used to measure those dimensions. They showed that the correlations in that case would be less than one, and the maximum correlation possible between any two dimensions is a function of the reliability scores of the two dimensions. For example, the maximum correlation between tangibles and responsiveness would be much less than one, as tangibles has a low reliability score. Given this possibility, we computed the maximum correlations among the various SERVQUAL dimensions. Confidence intervals around the correlations provided in Table IV were then compared to their respective maximum correlations. This analysis revealed that most of the confidence intervals indeed contained the respective maximum correlations, indicating that the SERVQUAL dimensions are not really distinct from each other. This result applied to both perception scores and gap scores, and indicates support for only a one-factor model.

[Table III]

A third way of assessing the discriminant validity between any two dimensions is to compare the average variance extracted by each dimension from its underlying indicators with the variance shared between those two dimensions. If the two dimensions are indeed distinct from each other, then the average variance extracted estimates would be greater than the shared variance estimates (Fornell and Larcker, 1981). The shared variance estimate between any two dimensions is obtained by squaring the correlation estimate between those two dimensions as provided in Table IV.

[Table IV]

For example, the shared variance between tangibles and reliability dimensions of SERVQUAL perception scores is 0.58 (i.e. square of 0.76). This shared variance estimate is not smaller

compared to the average variance extracted scores in Table III of either tangibles or reliability, implying that tangibles and reliability dimensions are not distinct from each other. Proceeding with this analysis, with the exception of a few shared variance estimates featuring tangibles dimension, most of the shared variance estimates are indeed greater than the corresponding average variance estimates, for both perception and gap scores. In sum, discriminant validity tests do not provide support for the five-factor model. Instead, it appears that in general, the SERVQUAL factors are not really distinct from each other.

Nomological validity

This was assessed by examining the correlation between SERVQUAL scores (i.e. perceptions and gap scores) and the composite ratings of the four interfacing departments (e.g. booking services, documentation) and the three supporting activities (e.g. telephone services). It is expected that customers' perceptions of service quality are influenced by their experiences when dealing with the interfacing departments and supporting activities. Hence the underlying correlations are expected to be relatively high and significant. Table V presents the correlation results. It is evident that the correlations of SERVQUAL dimensions and the composite SERVQUAL score with interfacing departments and support activities are moderately high and significant for SERVQUAL perceptions. For the SERVQUAL gap scores, only about 60 percent of the correlations are significant, but even they are relatively low in size. In sum, the network of relationships examined by inspecting the various correlations in Table V support nomological validity of only the SERVQUAL perception scores, but not of the SERVQUAL gap scores.

Predictive ability

The ability of SERVQUAL measure in predicting the overall evaluation of service (provided by the most preferred shipping line) was assessed next. For each of the SERVQUAL dimensions, their correlations were computed with the overall evaluation of service for both perceptions and gap scores. These correlations for SERVQUAL perceptions scores are 0.29 for tangibles, 0.50 for reliability, 0.54 for responsiveness, 0.51 for assurance, and 0.45 for empathy. Further, the correlation of the composite score of all 22 items of the SERVQUAL

perception measure with the overall customer satisfaction measure is 0.50. The corresponding correlations for SERVQUAL gap scores are 0.11 for tangibles, 0.23 for reliability, 0.23 for responsiveness, 0.21 for assurance, and 0.29 for empathy. Moreover, the composite 22-item SERVQUAL gap score has a correlation of 0.23 with the overall customer satisfaction measure. With one exception, all of the above correlations are statistically significant ($p < 0.05$). Examining the above correlations, it is clear that the tangibles dimension has a relatively lower correlation with the overall satisfaction measure. The other four SERVQUAL dimensions have significant correlations with the overall satisfaction measure for both perceptions and gap scores. Next, the composite SERVQUAL score also has a significant correlation with overall satisfaction, implying that SERVQUAL measure indeed serves as a useful predictor of the overall evaluation of service.

When comparing the SERVQUAL perceptions scores with the gap scores, however, it is clear that the perception scores outperform the gap scores in predicting overall evaluation of service, as evidenced by the correlation of 0.50 for composite perception score compared to 0.23 for composite gap score as shown in Table VI. To determine the relative predictive ability of perceptions and gap scores, the test for dependent correlations (Steiger, 1980) was used. It was found that the correlation between the composite SERVQUAL perception score and the overall evaluation of service was significantly higher than the corresponding correlation between the composite SERVQUAL gap score and the overall evaluation of service ($z = 2.62$, $p < 0.05$).

Other tests

The above results show support for using SERVQUAL perception scores instead of gap scores when assessing service quality. The gap scores are obtained by subtracting perceptions of actual service provided (i.e. perception scores) from the expected level of service expected by customers from the best service provider (expectation scores). Hence gap scores are a function of the perception scores. If the expected level of service from the best service provider is relatively high and constant across the respondents, then the correlation between perceptions and gap scores would be very high.

[Table V]

[Table VI]

In that case the perceptions score would serve as a surrogate for gap scores, and there is no need to examine the gap scores separately. To assess this possibility, we computed the correlation between the composite SERVQUAL perception score and the composite SERVQUAL gap score. This correlation was found to be -0.47 and the confidence interval around this correlation did not contain the maximum possible correlation between the two scores. This would imply that the SERVQUAL gap scores are indeed different from the SERVQUAL perception scores.

Discussion

A central premise in this paper is that successful relationship marketing in the service sector is facilitated by knowledge about customers' perceptions of the quality of the service. Though service quality measures have been developed and widely tested for consumer services, it is not known whether these measures possess sound psychometric properties when applied in the business-to- business context. Therefore, a key question arises as to whether the measures developed for consumer services can be applied to assess service quality perceptions of organizational customers. To address this issue our paper presents an application of SERVQUAL as a measure of service quality in ocean freight services.

Based on a cross-sectional sample of 114 business organizations in Singapore, which regularly utilize ocean freight services for their export needs, the study asked the shipping managers in these organizations to identify their most preferred shipping lines for their heaviest used export routes and provide their evaluation of services offered by their preferred suppliers. Specifically, the questionnaire included SERVQUAL measures of expectations and evaluations on various service quality dimensions, evaluation measures of various interfacing departments in the shipping line, and overall evaluation of the most preferred shipping line. The gap scores in service quality were then computed by subtracting the actual evaluation scores from the service expectation scores. The service quality measures,

perceptions and gap scores were then examined using covariance structure analysis.

The SERVQUAL scale was hypothesized to be five-dimensional, representing tangibles, reliability, responsiveness, assurance, and empathy. Results indicate that, while the fit of the hypothesized five-dimensional factor structure is reasonable, the measures may be better represented by a more parsimonious (i.e. three-dimensional) factor structure where responsiveness, assurance, and empathy are combined into a single dimension. The composite reliability indices are acceptable for both the five- and three- dimension models. Convergent validity tests provide moderate support for the various dimensions excluding tangibles. Several tests were performed to assess discriminant validity. Some of the tests provide moderate support for the tangibles and reliability dimensions; others indicate that none of the five dimensions has discriminant validity. Nomological validity of the five dimensions is moderately supported.

Overall, the perception scores provide a better measure of service quality than the gap scores, based on composite reliability, fit indices, and correlation indices. Further, tests for dependent correlations indicate that the perception scores provide a better ability to predict overall evaluation of service than the gap scores. However, the gap scores were found to be distinct from the perception scores, implying that the gap scores do have a useful role to play in identifying the areas of weakness for an organization wherever the gaps in service quality are high, and the relative strengths wherever the gap scores are small. In sum, the service quality measures developed for consumer services can only be applied with caution in business-to-business marketing.

Implications and future research directions

The findings of this study raise a concern about the ready extension of the SERVQUAL scale to the industrial setting. Because of differences between consumer and industrial characteristics, it is likely that the instrument needs to be specifically tailored to the industry under investigation. Moreover, the dimensions of SERVQUAL may not be universally applicable across different types of industrial services (Bienstock *et al*., 1997). Such advice is compatible with the findings of several studies that show that service quality measures exhibit a factor

structure that varies across industries (see Babakus and Boller, 1992; Brown *et al.,* 1992; Cronin and Taylor, 1992).

This study reinforces the concerns by others who have identified serious psychometric limitations with the gap or difference scores. One way to resolve this problem may be to eliminate the gap scores and use non-difference measures instead (Peter *et al.,* 1993). It is likely that service quality is influenced directly by perceptions of performance. For example, Babakus and Boller (1992) and Babakus and Mangold (1992) find that performance-based measures of service quality are superior to gap measures. In fact, Cronin and Taylor (1992, 1994) claim that performance scores alone (using their SERVPERF scale) are superior to SERVQUAL's gap scores. Both Carman (1990) and Babakus and Boller (1992) recommend revising the SERVQUAL scale by combining the expectations and perceptions into a single question. Parenthetically, the originators of the scale have provided strong evidence that managers can obtain a truer assessment of service quality by comparing perceptions against expectations (i.e. computing a difference score) than by interpreting perceptions only (Parasuraman *et al.,* 1993, 1994a, 1994b). These authors also argue that the SERVQUAL scale offers richer diagnostic properties than other scales. Clearly more research is needed to resolve this debate.

Other researchers have developed different models to measure perceptions of quality. For example, Dabholkar (1996) and Dabholkar *et al.* (1996) propose hierarchical factor structure and attribute models to capture dimensions related to retail customers' perceptions of service quality. Boulding *et al.* (1993) offer a Bayesian-like model where service perceptions are used to predict intended behavior. Unfortunately, these models are aimed at the consumer setting. Thus, their extension to the industrial arena is somewhat problematic until more research is undertaken. More research is also needed on models that integrate service quality, customer satisfaction, and service value and service provider- customer relationships.

Our research also found that SERVQUAL's five dimensions could be reduced to a smaller number. In fact this finding is in line with research results by Mels *et al.* (1997) who found a two-factor structure (called intrinsic and extrinsic factors) emerging from using the SERVQUAL scale. As a result, they advocate that the SERVQUAL scale be used with the "utmost care". Hence, other research is needed to determine if the SERVQUAL scale can be reduced to a more

parsimonious structure. Interestingly, the Nordic School argued long ago that service quality consists of only two or three dimensions (see Grönroos, 1978; 1984).

Although the instrument may not be directly transferable to the industrial sector, it does offer direction in terms of identifying the important attributes associated with a specific service. Johns and Tyas (1997) find evidence that there is a high level of randomness among scaled perceptions using the SERVQUAL approach. They suggest that a much more effective approach would be to use a critical incident analysis where customers are interviewed during the provision of the service. The list of incidents can then be analyzed to identify those that are truly critical in forming perceptions of the service. This technique they argue is more holistic (or Gestalt) and less reductionist in the way it identifies important service attributes. Their research suggests a data collection technique based on open-ended interviews or free elicitation questionnaires that can provide a more realistic insight into how customers' expectations affect their perceptions of service quality. Similarly, Kong and Mayo (1993) have used open-ended questions in their examination of service quality in the industrial sector with great success while Bienstock *et al.* (1997) used interviews to develop a scale to measure service quality in physical distribution.

Other theorists argue that script theory may be a more powerful way to assess expectations of service experiences (Hubbert *et al.* 1995). These researchers found that their data support the notion that scripts operationalize expectations and that customer-provided scripts can reveal expectations that are not being currently met. Clearly, more research is needed on these techniques as a replacement for the SERVQUAL approach in the industrial sector.

Another area of research deals with expectation management of service quality. Berry and Parasuraman (1991), for example, advise firms to manage expectations by not overpromising. Research is needed to understand what happens when a firm, which is providing excellent service, promises too much in the way of service quality. Does this affect their overall perceptions negatively and if so how can the firm counteract the negative effect of these negative perceptions?

Given the growing importance of industrial services in the domestic and international arena, understanding the perceptions of service is crucial. Though the findings of the study may apply to various business-to-business services including shipping services in

other parts of Asia-Pacific or elsewhere, future research is needed to evaluate service quality measures when applied to other types of business-to-business services. If service firms are to establish loyal relationships with their business-to-business clients, they need to not only conceptualize the meaning of good service but also quantify their customers' reactions to their service offerings. However, by examining SERVQUAL's application to ocean freight services, this study offers insight into the application of the scale to business-to-business marketing.

References

Achrol, R.S. (1997), "Changes in the theory of interorganizational relations in marketing: toward a network paradigm", *Journal of the Academy of Marketing Science*, Vol. 25 No. 1, pp. 56-71.

Asubonteng, P., McCleary, K. and Swanson, J. (1996), "SERVQUAL revisited: a critical review of service quality", *The Journal of Services Marketing*, Vol. 10 No. 6, pp. 62-81.

Babakus, E. and Boller, G. (1992), "An empirical assessment of the SERVQUAL scale", *Journal of Business Research*, Vol. 24, pp. 253-68.

Babakus, E. and Mangold, W.G. (1992), "Adapting the SERVQUAL scale to hospital services: an empirical investigation", *Health Service Research*, Vol. 26 No. 6, pp. 267-80.

Berry, L.L. (1995), "Relationship marketing of services – growing interest, emerging perspectives", *Journal of the Academy of Marketing Science*, Vol. 23 No. 4, pp. 236-45.

Berry, L.L. and Parasuraman, A. (1991), *Marketing Services: Competing through Quality*, The Free Press, New York, NY.

Bienstock, C.C., Mentzer, J. and Bird, M.M. (1997), "Measuring physical distribution service quality", *Journal of the Academy of Marketing Science*, Vol. 25 No. 1, pp. 31-44.

Bitner, M.J. (1995), "Building service relationships: it's all about promises", *Journal of Academy of Marketing Science*, Vol. 23 No. 4, pp. 246-51.

Bollen, K.A. and Long, J.S. (1993), *Testing Structural Equation Models*, Sage Publications, Newbury Park, CA.

Boulding, W., Kalra, A., Staelin, R. and Zeithaml, V. (1993), "A dynamic process model of service quality: from expectations to behavioral intentions", *Journal of Marketing Research*, Vol. 30 No. 1, pp. 7-27.

Brensinger, R. and Lambert, D. (1990), "Can the SERVQUAL scale be generalized to business to business?", *Proceedings of Knowledge Development in Marketing*, AMA Summer Educators' Conference.

Brown, T.J., Churchill, G.A. Jr and Peter, J.P. (1993), "Improving the measurement of service quality", *Journal of Retailing*, Vol. 69 No. 1, pp. 127- 39.

Carman, J.M. (1990), "Consumer perceptions of service quality: an assessment of the SERVQUAL dimensions", *Journal of Retailing*, Vol. 66 No. 1, pp. 33-55.

Caruana, A., Pitt, L. and Morris, M. (1995), "Are there excellent service firms, and do they perform well?", *Service Industries Journal*, Vol. 15 No. 3, pp. 243-56.

Cooper, P.D. and Jackson, R.W. (1988), "Applying a service marketing orientation to the industrial services sector", *Journal of Services Marketing*, Vol. 2, Fall, pp. 66-70.

Cronin, J. and Taylor, S.A. (1992), "Measuring service quality: a re-examination and extension", *Journal of Marketing*, Vol. 56, July, pp. 55-68.

Cronin, J. and Taylor, S.A. (1994), "SERVPERF versus SERVQUAL: reconciling performance-based and perceptions minus-expectations measurements of service quality", *Journal of Marketing*, Vol. 58 No. 1, pp. 125-31.

Dabholkar, P. (1996), "Consumer evaluations of new technology-based self-service options: an investigation of alternative models of service quality", *International Journal of Research in Marketing*, Vol. 13 No. 1, pp. 29-51.

Dabholkar, P.A., Thorpe, I.D. and Rentz, J.O. (1996), "A measure of service quality for retail stores: scale development and validation", *Journal of the Academy of Marketing Sciences*, Vol. 24 No. 1, pp. 3-16.

Farley, J.M., Daniels, C.P. and Pearl, D.H. (1990), "Service quality in a multinational environment", *1990 ASQC Quality Congress Transactions*, San Francisco, CA.

Fornell, C. and Larcker, D. (1981), "Evaluating structural equation models with unobservable variables and measurement error", *Journal of Marketing Research*, Vol. 18, February, pp. 39-50.

Gerbing, D.W. and Anderson, J.C. (1988), "An updated paradigm for scale development incorporating unidimensionality and its assessment", *Journal of Marketing Research*, Vol. 25, May, pp. 186-92.

Grönroos, C. (1978), "A service-oriented approach to marketing of services", *European Journal of Marketing*, Vol. 12 No. 8, pp. 588-601.

Grönroos, C. (1984), "A service quality model and its marketing implications", *European Journal of Marketing*, Vol. 18 No. 4, pp. 36-55.

Hise, R. and Gabel, T. (1995), "Customer service as a strategic weapon in international operations", *Journal of Global Marketing*, Vol. 8 Nos. 3/4, pp. 151 70.

Hubbert, A.R., Sehorn, A. and Brown, S. (1995), "Service expectations: the consumer versus the provider", *International Journal of Service Industry Management*, Vol. 6 No. 1, pp. 6-21.

Iacobucci, D., Grayson, K.A. and Ostrom, A.L. (1994), "The calculus of service quality and customer satisfaction: theoretical and empirical differentiation and integration", in Swartz, T.A., Bowen, D.A. and Brown, S.W. (Eds), *Advances in Services and Marketing and Management*, Vol. 3, JAI Press, Greenwich, CT and New York, NY, pp. 1-67.

Johns, N. and Tyas, P. (1997), "Customer perceptions of service operations: Gestalt, incident or mythology?", *Service Industries Journal*, Vol. 17 No. 3, pp. 474-88.

Joreskog, K.G. and Sorbom, D. (1989), *LISREL 7: A Guide to the Program and Applications*, SPSS Inc., Chicago, IL.

Kong, R. and Mayo, M. (1993), "Measuring service quality in the business-to business context", *Journal of Business & Industrial Marketing*, Vol. 8 No. 2, pp. 5-15.

Lovelock, C. and Yip, G. (1996), "Developing global strategies for service businesses", *California Management Review*, Winter, Vol. 38 No. 2, pp. 64- 86.

Marsh, H.W. (1994), "Confirmatory factor analysis models of factorial invariance: a multifaceted approach", *Journal of Structural Equations Modeling*, Vol. 1, pp. 5-34.

Mels, G., Boshoff, C. and Nel, D. (1997), "The dimensions of service quality: the original European perspective revisited", *Service Industries Journal*, Vol. 17 No. 1, pp. 173-89.

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

Parasuraman, A., Berry, L.L. and Zeithaml, V. (1991), "Refinement and reassessment of the SERVQUAL scale", *Journal of Retailing*, Vol. 67 No. 4, pp. 420-50.

Parasuraman, A., Berry, L.L. and Zeithaml, V. (1993), "Research note: more on improving service quality measurement", *Journal of Retailing*, Vol. 69 No. 1, pp. 140-47.

Parasuraman, A., Zeithaml, V. and Berry, L. (1988), "SERVQUAL: a multiple item scale for measuring consumer perceptions of service quality", *Journal of Retailing*, Vol. 64 No. 1, pp. 12-40.

Parasuraman, A., Zeithaml, V. and Berry, L. (1994a), "Alternative scales for measuring service quality: a comparative assessment based on psychometric and diagnostic criteria", *Journal of Retailing*, Vol. 70 No. 3, pp. 201-29.

Parasuraman, A., Zeithaml, V. and Berry, L. (1994b), "Reassessment of expectations as a comparison in measuring service quality: implications for further research", *Journal of Marketing*, Vol. 58 No. 1, pp. 111-24.

Peter, J.P., Churchill, G. and Brown, T. (1993), "Caution in the use of difference scores in consumer research", *Journal of Consumer Research*, Vol. 19 No. 1, pp. 655-62.

Pitt, L., Morris, M. and Oosthuizen, P. (1996), "Expectations of service quality as an industrial market segmentation variable", *Service Industries Journal*, Vol. 16 No. 1, pp. 1-9.

Steiger, J.H. (1980), "Tests for comparing elements of a correlation matrix", *Psychological Bulletin*, Vol. 87 No. 2, pp. 245-51.

Van Dyke, T., Kappelman, L. and Prybutok, V. (1997), "Measuring information systems service quality: concerns on the use of SERVQUAL", *MIS Quarterly*, Vol. 21 No. 2, pp. 195-208.

Webster, F. (1992), "The changing role of marketing in the corporation", *Journal of Marketing*, Vol. 56, October, pp. 1-17.

Young, J. and Varble, D. (1997), "Purchasing's performance as seen by its internal customers: a study in a service organization", *International Journal of Purchasing & Materials Management*, Vol. 33 No. 3, pp. 36-41.

Executive summary and implications for managers and executives

This summary has been provided to allow managers and executives a rapid appreciation of the content of this article. Those with a particular interest in the topic covered may then read the article *in toto* to take advantage of the more comprehensive description of the research undertaken and its results to get the full benefit of the material present

SERVQUAL – not perfect but useful none the less

The development of reliable, easy-to-use measures of service quality represents a key aspect of service marketing research. We accept that service quality is important in influencing the choices of customers and we also know that service quality is something managers have control over. However – and despite its importance – we still lack a generally accepted way of assessing service quality.

The SERVQUAL scale is perhaps the closest we have got to creating a consistent, accepted service quality measure. But this measure still has its critics and, as more studies are done applying SERVQUAL to various settings, the more doubts are raised about its validity and value.

Durvasula et al. *apply the SERVQUAL measure to the business-to business sector. The authors note that service quality is as important to business buyers as it is to the ordinary consumer. At the same time as they study the application of SERVQUAL, we also see how more "parsimonious" measures with fewer dimensions compare.*

Before looking at the implications of Durvasula et al.*'s findings it is worth reminding ourselves of criticisms levelled at the SERVQUAL scale.*

SERVQUAL: good, bad or indifferent?

The SERVQUAL scale encompasses five dimensions of service quality:

(1) tangibles,

(2) reliability,

(3) responsiveness,

(4) assurance, and

(5) empathy.

Each of these dimensions represents one aspect of service quality and can be measured using a 22-point scale developed to apply SERVQUAL. The approach assesses gaps between expected and actual service.

The main concerns about the scale are:

- *The five dimensions are not separate – it is possible to obtain similar results using fewer dimensions (e.g. by combining responsiveness, assurance and empathy into one dimension).*

- *One dimension – reliability – is dominant, representing a good indicator of service quality on its own.*

- *The dimensions cannot be applied universally across business sectors, raising questions about the scale's reliability as a measure.*

- *The approach is less valid than an approach based on actual service performance alone.*

The creators of SERVQUAL (Parasuraman et al.) have responded to these criticisms with further work refining the scale and the re validation of the scale through empirical testing. The scale – and the measurement of service quality – remains under development. We cannot say that any measure provides a general test of overall service quality.

All this being said, we should not, however, dismiss SERVQUAL. While questions remain about its value in academic terms SERVQUAL remains an approach founded on sound logic. It seems wholly reasonable to approach assessments of quality by looking at the gap between what the customer wants and what we are giving in reality. The problem lies, I suspect, in the general nature of SERVQUAL rather than in the theory underlying the scale.

Business-to-business marketing – is it really different from consumer marketing?

I noted at the start that business-to-business marketers need to be concerned about service quality for the same reasons that consumer marketers are concerned. High quality represents a sustainable advantage to the business by encouraging customer loyalty, reducing the cost of correcting errors and improving sales to new customers. High quality also has a positive impact on brand and corporate image.

There are differences between selling to consumers and selling to businesses but these are mostly questions of scale and complexity – and a resulting difference in marketing mix – rather than profound differences in the ways in which marketing works. On this basis we can assume (for the time being) that issues of service quality will also remain very similar between consumers and businesses.

Given this similarity we can therefore suggest that an effective, general measure of service quality will be equally applicable to business-to-

business markets and to consumer markets. Durvasula et al.'s study raises some questions as to whether SERVQUAL is a generally applicable tool for measuring service quality. The findings suggest that "…the dimensions of SERVQUAL may not be universally applicable across different types of industrial service".

So what approach should we use to assess service quality? Given the problems with SERVQUAL (although we should not dismiss the scale out-of-hand) the search is on for an acceptable measure of service quality in business-to-business markets. Durvasula et al. suggest that "…perception scores provide a better measure of service quality than the gap scores…". This suggests that the SERVQUAL scale might still have value albeit in a different form from that intended by its creators.

As well as identifying some weaknesses in SERVQUAL, Durvasula et al. also indicate that all the five dimensions are not truly independent. We see that a smaller scale where responsiveness, assurance and empathy are combined could provide results as satisfactory as those from the full five dimensions.

For the manager this presents something of a dilemma – while we get a better overall assessment of our service quality from a more parsimonious measure than SERVQUAL, conversely, SERVQUAL provides a larger number of options for managerial action by having five dimensions. For the time being we need to combine more than one measure to provide information about our shortcomings in terms of service quality.

What managers want from measures is information to guide action plans rather than simply an indication of good or bad service quality. SERVQUAL, for all its weaknesses, does provide such a guide. Up to a point, this measure can help us in developing effective service marketing plans.

(A précis of the article "Testing the SERVQUAL scale in the business-to business sector: the case of ocean freight shipping service". Supplied by Marketing Consultants for MCB University Press.)

Table I. Tests for dimensionality: confirmatory factor analysis results of chi-square and other fit indices for SERVQUAL factor models

| | Model | | | | |
	Five-factor	Four-factor	Three-factor	Two-factor	One-factor
χ^2					
Perceptions	449.44	451.43	456.91	480.44	491.68
Gap scores	359.84	367.76	385.57	378.33	419.97
Degrees of freedom					
Perceptions	199.00	203.00	206.00	208.00	209.00
Gap scores	199.00	203.00	208.00	206.00	209.00
Root-mean-square residual					
Perceptions	0.07	0.07	0.07	0.06	0.06
Gap scores	0.08	0.08	0.08	0.08	0.08
Goodness of fit index					
Perceptions	0.74	0.74	0.74	0.73	0.73
Gap scores	0.78	0.78	0.77	0.77	0.74
Adjusted goodness of fit index					
Perceptions	0.67	0.68	0.68	0.68	0.67
Gap scores	0.72	0.72	0.72	0.72	0.69
Normed fit index					
Perceptions	0.77	0.76	0.76	0.75	0.74
Gap scores	0.76	0.76	0.75	0.75	0.72
χ^2/df *ratio*					
Perceptions	2.26	2.22	2.22	2.31	2.35
Gap scores	1.81	1.81	1.84	1.85	2.01
Comparative fit index					
Perceptions	0.85	0.85	0.85	0.84	0.83
Gap scores	0.88	0.87	0.86	0.87	0.84
Tucker-Lewis Index					
Perceptions	0.83	0.83	0.83	0.82	0.81
Gap scores	0.86	0.86	0.85	0.85	0.82

Table II. Composite reliability indices for SERVQUAL factors

Service quality factors	Reliability indices for	
	Perceptions	Gap scores
Tangibles	0.69	0.69
Reliability	0.87	0.84
Responsiveness	0.88	0.85
Assurance	0.84	0.74
Empathy	0.85	0.79
Responsiveness and assurance	0.92	0.89
Responsiveness, assurance, empathy	0.94	0.92
Reliability, responsiveness, assurance, empathy	0.95	0.94
One-factor model	0.94	0.93

Table III. Assessment of convergent validity of SERVQUAL Scale: results of confirmatory factor analysis

Service quality factors	Average variance extracted score	
	Perceptions	Gap scores
Tangibles	0.30	0.39
Reliability	0.58	0.52
Responsiveness	0.67	0.58
Assurance	0.57	0.42
Empathy	0.53	0.45

Note: An average variance extracted score of 0.50 or above indicates that the construct in question has convergent validity

Table IV. Test for discriminant validity of SERVQUAL factors: correlations and (standard errors) among SERVQUAL factors

SERVQUAL factors	SERVQUAL factors			
	Reliability	Responsiveness	Assurance	Empathy
Tangibles				
Perceptions	0.76 (0.07)	0.69 (0.09)	0.77 (0.08)	0.69 (0.09)
Gap scores	0.49 (0.09)	0.52 (0.09)	0.58 (0.09)	0.61 (0.08)
Reliability				
Perceptions		0.92 (0.03)	0.93 (0.03)	0.87 (0.04)
Gap scores		0.93 (0.02)	0.96 (0.03)	0.90 (0.03)
Responsiveness				
Perceptions			0.99 (0.02)	0.97 (0.02)
Gap scores			0.96 (0.03)	0.97 (0.02)
Assurance				
Perceptions				0.98 (0.02)
Gap scores				0.97 (0.03)

Note: Confidence intervals around factor correlations indicate that in general SERVQUAL factors are not distinct from each other

Table V. Nomological validity of SERVQUAL: correlations of SERVQUAL factors with ratings of shipping lines

Ratings of most preferred shipping line on:

SERVQUAL factors	Marketing representatives	Telephone services	Personal visits	Booking services	Documentation	Operations	Claims
Tangibles							
Perceptions	0.35	0.43	0.43	0.37	0.36	0.45	0.48
Gap scores	0.13 ns	0.16*	0.11 ns	0.04 ns	0.06 ns	0.11 ns	0.09 ns
Reliability							
Perceptions	0.59	0.64	0.61	0.56	0.52	0.60	0.54
Gap scores	0.34	0.22	0.32	0.11 ns	0.08 ns	0.25	0.24 ns
Responsiveness							
Perceptions	0.59	0.54	0.60	0.54	0.52	0.59	0.49
Gap scores	0.34	0.26	0.31	0.14 ns	0.11 ns	0.37	0.22 ns
Assurance							
Perceptions	0.52	0.50	0.62	0.54	0.47	0.62	0.45
Gap scores	0.32	0.23	0.29	0.12 ns	0.12 ns	0.29	0.17 ns
Empathy							
Perceptions	0.52	0.46	0.54	0.49	0.45	0.47	0.43
Gap scores	0.34	0.28	0.40	0.19*	0.12 ns	0.24*	0.25*
Composite index							
Perceptions	0.58	0.55	0.64	0.57	0.52	0.61	0.53
Gap scores	0.33	0.26	0.34	0.14 ns	0.11 ns	0.30	0.22 ns

Note: All correlations are significant at a = 0.05 unless indicated otherwise as * (significant at a = 0.10) or ns (not significant)

Table VI. Predictive ability of SERVQUAL scale: correlations between SERVQUAL and overall satisfaction scores and t-test results for differences between dependent correlations

Correlation between	Correlation between	Z-score	$p <$
For most preferred shipping line:			
Evaluations – overall satisfaction	Expectations – overall satisfaction		
0.50	0.20	2.83	0.00
Evaluations – overall satisfaction	Gap scores – overall satisfaction		
0.50	0.23	2.62	0.01
Expectations – overall satisfaction	Gap scores – overall satisfaction		
0.20	0.23	−0.28	0.78
For second most preferred shipping line:			
Evaluations – overall satisfaction	Expectations – overall satisfaction		
0.36	0.15	1.86	0.06
Evaluations – overall satisfaction	Gap scores – overall satisfaction		
0.36	0.16	1.81	0.07
Expectations – overall satisfaction	Gap scores – overall satisfaction		
0.15	0.16	−0.09	0.93

Notes: Is the SERVQUAL scale for gap scores distinct from expectation score and evaluation score scales? Confidence intervals for correlations between: expectation score evaluation score (0.4), expectation score-gap scores (0.62), and evaluation score-gap scores (−0.47) do not contain one, suggesting discriminant validity for the three types of SERVQUAL scales