

SERVQUAL: review, critique, research agenda

Francis Buttle

Manchester Business School, Manchester, UK

8

Received October 1994
Revised April 1995

SERVQUAL: a primer

SERVQUAL provides a technology for measuring and managing service quality (SQ). Since 1985, when the technology was first published, its innovators Parasuraman, Zeithaml and Berry, have further developed, promulgated and promoted the technology through a series of publications (Parasuraman *et al.*, 1985; 1986; 1988; 1990; 1991a; 1991b; 1993; 1994; Zeithaml *et al.*, 1990; 1991; 1992; 1993).

The ABI/Inform database "Global edition", (September 1994) reports that service quality has been a keyword in some 1,447 articles published in the period January 1992 to April 1994. By contrast SERVQUAL has been a keyword in just 41 publications. These publications incorporate both theoretical discussions and applications of SERVQUAL in a variety of industrial, commercial and not-for-profit settings. Published studies include tyre retailing (Carman, 1990) dental services (Carman, 1990), hotels (Saleh and Ryan, 1992) travel and tourism (Fick and Ritchie, 1991), car servicing (Bouman and van der Wiele, 1992), business schools (Rigotti and Pitt, 1992), higher education (Ford *et al.*, 1993; McElwee and Redman, 1993), hospitality (Johns, 1993), business-to-business channel partners (Kong and Mayo, 1993), accounting firms (Freeman and Dart, 1993), architectural services (Baker and Lamb, 1993), recreational services (Taylor *et al.*, 1993), hospitals (Babakus and Mangold, 1992; Mangold and Babakus, 1991; Reidenbach and Sandifer-Smallwood, 1990; Soliman, 1992; Vandamme and Leunis, 1993; Walbridge and Delene, 1993), airline catering (Babakus *et al.*, 1993a), banking (Kwon and Lee, 1994; Wong and Perry, 1991) apparel retailing (Gagliano and Hathcote, 1994) and local government (Scott and Shieff, 1993). There have also been many unpublished SERVQUAL studies. In the last two years alone, the author has been associated with a number of sectoral and corporate SERVQUAL studies: computer services, construction, mental health services, hospitality, recreational services, ophthalmological services, and retail services. In addition, a number of organizations, such as the Midland and Abbey National banks have adopted it.

Service quality (SQ) has become an important research topic because of its apparent relationship to costs (Crosby, 1979), profitability (Buzzell and Gale, 1987; Rust and Zahorik, 1993; Zahorik and Rust, 1992), customer satisfaction (Bolton and Drew, 1991; Boulding *et al.*, 1993), customer retention (Reichheld and Sasser, 1990), and positive word of mouth. SQ is widely regarded as a driver of corporate marketing and financial performance.

SERVQUAL is founded on the view that the customer's assessment of SQ is paramount. This assessment is conceptualized as a gap between what the customer expects by way of SQ from a class of service providers (say, all opticians), and their evaluations of the performance of a particular service provider (say a single Specsavers store). SQ is presented as a multidimensional construct. In their original formulation Parasuraman *et al.* (1985) identified ten components of SQ:

- (1) reliability;
- (2) responsiveness;
- (3) competence;
- (4) access;
- (5) courtesy;
- (6) communication;
- (7) credibility;
- (8) security;
- (9) understanding/knowing the customer;
- (10) tangibles.

(See Appendix for definitions and examples.) In their 1988 work these components were collapsed into five dimensions: reliability, assurance, tangibles, empathy, responsiveness, as defined in Table I. Reliability, tangibles and responsiveness remained distinct, but the remaining seven components collapsed into two aggregate dimensions, assurance and empathy[1]. Parasuraman *et al.* developed a 22-item instrument with which to measure customers' expectations and perceptions (E and P) of the five RATER dimensions. Four or five numbered items are used to measure each dimension. The instrument is administered twice in different forms, first to measure expectations and second to measure perceptions.

SERVQUAL:
review, critique,
research agenda

Dimensions	Definition	Items in scale
Reliability	The ability to perform the promised service dependably and accurately	4
Assurance	The knowledge and courtesy of employees and their ability to convey trust and confidence	5
Tangibles	The appearance of physical facilities, equipment, personnel and communication materials	4
Empathy	The provision of caring, individualized attention to customers	5
Responsiveness	The willingness to help customers and to provide prompt service	4

Table I.
SERVQUAL dimensions

In 1991, Parasuraman *et al.* published a follow-up study which refined their previous work (1991b). Wording of all expectations items changed. The 1988 version had attempted to capture respondents' normative expectations. For example, one 1988 expectations item read: "Companies offering ----- services should keep their records accurately". The revised wording focused on what customers would expect from "excellent service companies". The sample item was revised thus: "Excellent companies offering ----- services will insist on error-free records". Detailed wording of many perceptions items also changed. Two new items, one each for tangibles and assurance, were substituted for two original items. The tangibles item referred to the appearance of communication materials. The assurance item referred to the knowledge of employees. Both references had been omitted in the 1988 version.

Analysis of SERVQUAL data can take several forms: item-by-item analysis (e.g. $P1 - E1$, $P2 - E2$); dimension-by-dimension analysis (e.g. $(P1 + P2 + P3 + P4/4) - (E1 + E2 + E3 + E4/4)$, where $P1$ to $P4$, and $E1$ to $E4$, represent the four perception and expectation statements relating to a single dimension); and computation of the single measure of service quality ($(P1 + P2 + P3 \dots + P22/22) - (E1 + E2 + E3 + \dots + E22/22)$), the so-called SERVQUAL gap.

Without question, SERVQUAL has been widely applied and is highly valued. Any critique of SERVQUAL, therefore, must be seen within this broader context of strong endorsement. What follows is a discussion of several criticisms which have been levelled at SERVQUAL elsewhere or have been experienced in the application of the technology by this author.

Criticisms of SERVQUAL

Notwithstanding its growing popularity and widespread application, SERVQUAL has been subjected to a number of theoretical and operational criticisms which are detailed below:

(1) Theoretical:

- Paradigmatic objections: SERVQUAL is based on a disconfirmation paradigm rather than an attitudinal paradigm; and SERVQUAL fails to draw on established economic, statistical and psychological theory.
- Gaps model: there is little evidence that customers assess service quality in terms of $P - E$ gaps.
- Process orientation: SERVQUAL focuses on the process of service delivery, not the outcomes of the service encounter.
- Dimensionality: SERVQUAL's five dimensions are not universals; the number of dimensions comprising SQ is contextualized; items do not always load on to the factors which one would a priori expect; and there is a high degree of intercorrelation between the five RATER dimensions.

(2) Operational:

- Expectations: the term expectation is polysemic; consumers use standards other than expectations to evaluate SQ; and SERVQUAL fails to measure absolute SQ expectations.
- Item composition: four or five items can not capture the variability within each SQ dimension.
- Moments of truth (MOT): customers' assessments of SQ may vary from MOT to MOT.
- Polarity: the reversed polarity of items in the scale causes respondent error.
- Scale points: the seven-point Likert scale is flawed.
- Two administrations: two administrations of the instrument causes boredom and confusion.
- Variance extracted: the over SERVQUAL score accounts for a disappointing proportion of item variances.

SERVQUAL:
review, critique,
research agenda

11

Each of the criticisms will be examined below.

Theoretical

Paradigmatic objections. Two major criticisms have been raised. First, SERVQUAL has been inappropriately based on an expectations-disconfirmation model rather than an attitudinal model of SQ. Second, it does not build on extant knowledge in economics, statistics and psychology.

SERVQUAL is based on the disconfirmation model widely adopted in the customer satisfaction literature. In this literature, customer satisfaction (CSat) is operationalized in terms of the relationship between expectations (E) and outcomes (O). If O matches E, customer satisfaction is predicted. If O exceeds E, then customer delight may be produced. If E exceeds O, then customer dissatisfaction is indicated.

According to Cronin and Taylor (1992; 1994) SERVQUAL is paradigmatically flawed because of its ill-judged adoption of this disconfirmation model. "Perceived quality", they claim, "is best conceptualized as an attitude". They criticize Parasuraman *et al.* for their hesitancy to define perceived SQ in attitudinal terms, even though Parasuraman *et al.* (1988) had earlier claimed that SQ was "similar in many ways to an attitude". Cronin and Taylor observe:

Researchers have attempted to differentiate service quality from consumer satisfaction, even while using the disconfirmation format to measure perceptions of service quality... this approach is not consistent with the differentiation expressed between these constructs in the satisfaction and attitude literatures.

Iacobucci *et al.*'s (1994) review of the debate surrounding the conceptual and operational differences between SQ and CSat concludes that the constructs "have not been consistently defined and differentiated from each other in the literature". She suggests that the two constructs may be connected in a number of ways. First, they may be both different operationalizations of the same construct, "evaluation". Second, they may be orthogonally related, i.e. they may

be entirely different constructs. Third, they may be conceptual cousins. Their family connections may be dependent on a number of other considerations, including for example, the duration of the evaluation. Parasuraman *et al.* (1985) have described satisfaction as more situation- or encounter-specific, and quality as more holistic, developed over a longer period of time, although they offer no empirical evidence to support this contention. SQ and CSat may also be related by time order. The predominant belief is that SQ is the logical predecessor to CSat, but this remains unproven. Cronin and Taylor's critique draws support from Oliver's (1980) research which suggests that SQ and CSat are distinct constructs but are related in that satisfaction mediates the effect of prior-period perceptions of SQ and causes revised SQ perceptions to be formed. SQ and CSat may also be differentiated by virtue of their content. Whereas SQ may be thought of as high in cognitive content, CSat may be more heavily loaded with affect (Oliver, 1993).

Cronin and Taylor suggest that the adequacy-importance model of attitude measurement should be adopted for SQ research. Iacobucci *et al.* (1994) add the observation that "in some general psychological sense, it is not clear what short-term evaluations of quality and satisfaction are if not attitudes". In turn, Parasuraman *et al.* (1994) have vigorously defended their position, claiming that critics seem "to discount prior conceptual work in the SQ literature", and suggest that Cronin and Taylor's work "does not justify their claim" that the disconfirmation paradigm is flawed.

In other work, Cronin and Taylor (1994) comment that:

Recent conceptual advances suggest that the disconfirmation-based SERVQUAL scale is measuring neither service quality nor consumer satisfaction. Rather, the SERVQUAL scale appears at best an operationalization of only one of the many forms of expectancy-disconfirmation.

A different concern has been raised by Andersson (1992). He objects to SERVQUAL's failure to draw on previous social science research, particularly economic theory, statistics, and psychological theory. Parasuraman *et al.*'s work is highly inductive in that it moves from historically situated observation to general theory. Andersson (1992) claims that Parasuraman *et al.* "abandon the principle of scientific continuity and deduction". Among specific criticisms are the following:

First, Parasuraman *et al.*'s management technology takes no account of the costs of improving service quality. It is naïve in assuming that the marginal revenue of SQ improvement always exceeds the marginal cost. (Aubrey and Zimble (1983), Crosby (1979), Juran (1951) and Masser (1957) have addressed the issue of the costs/benefits of quality improvement in service settings.)

Second, Parasuraman *et al.* collect SQ data using ordinal scale methods (Likert scales) yet perform analyses with methods suited to interval-level data (factor analysis).

Third, Parasuraman *et al.* are at the "absolute end of the street regarding possibilities to use statistical methods". Ordinal scales do not allow for

investigations of common product-moment correlations. Interdependencies among the dimensions of quality are difficult to describe. SERVQUAL studies cannot answer questions such as: Are there elasticities among the quality dimensions? Is the customer value of improvements a linear or non-linear function?

Fourth, Parasuraman *et al.* fail to draw on the large literature on the psychology of perception.

Gaps model. A related set of criticisms refer to the value and meaning of gaps identified in the disconfirmation model.

Babakus and Boller (1992) found the use of a “gap” approach to SQ measurement “intuitively appealing” but suspected that the “difference scores do not provide any additional information beyond that already contained in the perceptions component of the SERVQUAL scale”. They found that the dominant contributor to the gap score was the perceptions score because of a generalized response tendency to rate expectations high.

Churchill and Surprenant (1982), in their work on CSat, also ponder whether gap measurements contribute anything new or of value given that the gap is a direct function of E and P. It has also been noted that:

while conceptually, difference scores might be sensible, they are problematic in that they are *notoriously unreliable*, even when the measures from which the difference scores are derived are themselves highly reliable (Iacobucci *et al.*, 1994).

Also, in the context of CSat, Oliver (1980) has pondered whether it might be preferable to consider the P – E scores as raw differences or as ratios. No work has been reported using a ratio approach to measure SQ.

Iacobucci *et al.* (1994) take a different tack on the incorporation of E-measures. They suggest that expectations might not exist or be formed clearly enough to serve as a standard for evaluation of a service experience. Expectations may be formed simultaneously with service consumption. Kahneman and Miller (1986) have also proposed that consumers may form “experience-based norms” after service experiences, rather than expectations before.

A further issue raised by Babakus and Inhofe (1991) is that expectations may attract a social desirability response bias. Respondents may feel motivated to adhere to an “I-have-high-expectations” social norm. Indeed, Parasuraman *et al.* report that in their testing of the 1988 version the majority of expectations scores were above six on the seven-point scale. The overall mean expectation was 6.22 (Parasuraman *et al.*, 1991b).

Teas (1993a; 1993b; 1994) has pondered the meaning of identified gaps. For example, there are six ways of producing P – E gaps of -1 (P = 1, E = 2; P = 2, E = 3; P = 3, E = 4; P = 4, E = 5; P = 5, E = 6; P = 6, E = 7). Do these tied gaps mean equal perceived SQ? He also notes that SERVQUAL research thus far has not established that all service providers within a consideration or choice set, e.g. all car-hire firms do, in fact, share the same expectations ratings across all items and dimensions.

A further criticism is that SERVQUAL fails to capture the dynamics of changing expectations. Consumers learn from experiences. The inference in much of Parasuraman *et al.*'s work is that expectations rise over time. An E-score of seven in 1986 may not necessarily mean the same as an E-score in 1996. Expectations may also fall over time (e.g. in the health service setting). Grönroos (1993) recognizes this weakness in our understanding of SQ, and has called for a new phase of service quality research to focus on the dynamics of service quality evaluation. Wotruba and Tyagi (1991) agree that more work is needed on how expectations are formed and changed over time.

Implicit in SERVQUAL is the assumption that positive and negative disconfirmation are symmetrically valent. However, from the customer's perspective, failure to meet expectations often seems a more significant outcome than success in meeting or exceeding expectations (Hardie *et al.*, 1992). Customers will often criticize poor service performance and not praise exceptional performance.

Recently, Cronin and Taylor (1992) have tested a performance-based measure of SQ, dubbed SERVPERF, in four industries (banking, pest control, dry cleaning and fast food). They found that this measure explained more of the variance in an overall measure of SQ than did SERVQUAL. SERVPERF is composed of the 22 perception items in the SERVQUAL scale, and therefore excludes any consideration of expectations. In a later defence of their argument for a perceptions-only measure of SQ, Cronin and Taylor (1994) acknowledge that it is possible for researchers to *infer* consumers' disconfirmation through arithmetic means (the P – E gap) but that "consumer perceptions, not calculations, govern behavior". Finally, a team of researchers, including Zeithaml herself (Boulding *et al.*, 1993), has recently rejected the value of an expectations-based, or gap-based model in finding that service quality was only influenced by perceptions.

Process orientation. SERVQUAL has been criticized for focusing on the process of service delivery rather than outcomes of the service encounter.

Grönroos (1982) identified three components of SQ: technical, functional and reputational quality. Technical quality is concerned with the outcome of the service encounter, e.g. have the dry cleaners got rid of the stain? Functional quality is concerned with the process of service delivery, e.g. were the dry cleaner's counter staff courteous? Reputational quality is a reflection of the corporate image of the service organization.

Whereas technical quality focuses on *what*, functional quality focuses on *how* and involves consideration of issues such as the behaviour of customer contact staff, and the speed of service.

Critics have argued that outcome quality is missing from Parasuraman *et al.*'s formulation of SQ (Cronin and Taylor, 1992; Mangold and Babakus, 1991; Richard and Allaway, 1993).

Richard and Allaway (1993) tested an augmented SERVQUAL model which they claim incorporates both process and outcome components, and comment that "the challenge is to determine which process and outcome quality

attributes of SQ have the greatest impact on choice”[2]. Their research into Domino Pizza’s process and outcome quality employed the 22 Parasuraman *et al.* (1988) items, modified to suit context, and the following six outcome items:

- (1) Domino’s has delicious home-delivery pizza.
- (2) Domino’s has nutritious home-delivery pizza.
- (3) Domino’s home-delivery pizza has flavourful sauce.
- (4) Domino’s provides a generous amount of toppings for its home-delivery pizza.
- (5) Domino’s home-delivery pizza is made with superior ingredients.
- (6) Domino’s prepared its home-delivery pizza crust exactly the way I like it.

These researchers found that the process-only items borrowed and adapted from SERVQUAL accounted for only 45 per cent of the variance in customer choice; the full inventory, inclusive of the six outcome items, accounted for 71.5 per cent of variance in choice. The difference between the two is significant at the 0.001 level. They conclude that process-and-outcome is a better predictor of consumer choice than process, or outcome, alone.

In defence of SERVQUAL, Higgins *et al.* (1991) have argued that outcome quality is already contained within these dimensions: reliability, competence and security.

Dimensionality. Critics have raised a number of significant and related questions about the dimensionality of the SERVQUAL scale. The most serious are concerned with the number of dimensions, and their stability from context to context.

There seems to be general agreement that SQ is a second-order construct, that is, it is factorially complex, being composed of several first-order variables[3]. SERVQUAL is composed of the five RATER factors. There are however, several alternative conceptualizations of SQ. As already noted, Grönroos (1984) identified three components – technical, functional and reputational quality; Lehtinen and Lehtinen (1982) also identify three components – interactive, physical and corporate quality; Hedvall and Paltschik (1989) identify two dimensions – willingness and ability to serve, and physical and psychological access; Leblanc and Nguyen (1988) list five components – corporate image, internal organization, physical support of the service producing system, staff/customer interaction, and the level of customer satisfaction.

Parasuraman *et al.* (1988) have claimed that SERVQUAL:

provides a basic skeleton through its expectations/perceptions format encompassing statements for each of the five service quality dimensions. The skeleton, when necessary, can be adapted or supplemented to fit the characteristics or specific research needs of a particular organization.

In their 1988 paper, Parasuraman *et al.* also claimed that “the final 22-item scale and its five dimensions have *sound and stable psychometric properties*”. In the

1991b revision, Parasuraman *et al.* found evidence of “*consistent factor structure ... across five independent samples*” (emphases added). In other words, they make claims that the five dimensions are generic across service contexts. Indeed, in 1991, Parasuraman *et al.* claimed that “SERVQUAL’s dimensions and items represent core evaluation criteria that transcend specific companies and industries” (1991b)[4].

Number of dimensions. When the SERVQUAL instrument has been employed in modified form, up to nine distinct dimensions of SQ have been revealed, the number varying according to the service sector under investigation. One study has even produced a single-factor solution.

Nine factors accounted for 71 per cent of SQ variance in Carman’s (1990) hospital research: admission service, tangible accommodations, tangible food, tangible privacy, nursing care, explanation of treatment, access and courtesy afforded visitors, discharge planning, and patient accounting (billing)[5].

Five factors were distinguished in Saleh and Ryan’s (1992) work in the hotel industry – conviviality, tangibles, reassurance, avoid sarcasm, and empathy. The first of these, conviviality, accounted for 62.8 per cent of the overall variance; the second factor, tangibles, accounted for a further 6.9 per cent; the five factors together accounted for 78.6 per cent. This is strongly suggestive of a two-factor solution in the hospitality industry. The researchers had “initially assumed that the factor analysis would confirm the [SERVQUAL] dimensions but this failed to be the case”.

Four factors were extracted in Gagliano and Hathcote’s (1994) investigation of SQ in the retail clothing sector – personal attention, reliability, tangibles and convenience. Two of these have no correspondence in SERVQUAL. They conclude “the [original SERVQUAL scale] does not perform as well as expected” in apparel speciality retailing.

Three factors were identified in Bouman and van der Wiele’s (1992) research into car servicing – customer kindness, tangibles and faith[6]. The authors “were not able to find the same dimensions for judging service quality as did Berry *et al.*”.

One factor was recognized in Babakus *et al.*’s (1993b) survey of 635 utility company customers. Analysis “essentially produced a single-factor model” of SQ which accounted for 66.3 per cent of the variance. The authors advance several possible explanations for this unidimensional result including the nature of the service, (which they describe as a low-involvement service with an ongoing consumption experience), non-response bias and the use of a single expectations/perceptions gap scale. These researchers concluded: “With the exception of findings reported by Parasuraman and his colleagues, empirical evidence does not support a five-dimensional concept of service quality”.

In summary, Babakus and Boller (1992) commented that “the domain of service quality may be factorially complex in some industries and very simple and unidimensional in others”. In effect, they claim that the number of SQ dimensions is dependent on the particular service being offered.

In their revised version, Parasuraman *et al.* (1991b) suggest two reasons for these anomalies. First, they may be the product of differences in data collection and analysis procedures. A “more plausible explanation” is that “differences among empirically derived factors across replications may be primarily due to across-dimension similarities and/or within dimension differences in customers’ evaluations of a specific company involved in each setting”.

Spreng and Singh (1993) have commented on the lack of discrimination between several of the dimensions. In their research, the correlation between Assurance and Responsiveness constructs was 0.97, indicating that they were not separable constructs. They also found a high correlation between the combined Assurance-Responsiveness construct and the Empathy construct (0.87). Parasuraman *et al.* (1991b) had earlier found that Assurance and Responsiveness items loaded on a single factor, and in their 1988 work had found average intercorrelations among the five dimensions of 0.23 to 0.35.

In testing their revised version (Parasuraman *et al.*, 1991b), Parasuraman and colleagues found that the four items under Tangibles broke into two distinct dimensions, one pertaining to equipment and physical facilities, the other to employees and communication materials. They also found that Responsiveness and Assurance dimensions showed considerable overlap, and loaded on the same factor. They suggested that this was a product of imposing a five-factor constraint on the analyses. Indeed, the additional degrees of freedom allowed by a subsequent six-factor solution generated distinct Assurance and Responsiveness factors.

Parasuraman *et al.* (1991a) have now accepted that the “five SERVQUAL dimensions are interrelated as evidenced by the need for oblique rotations of factor solutions...to obtain the most interpretable factor patterns. One fruitful area for future research”, they conclude, “is to explore the nature and causes of these interrelationships”.

It therefore does appear that both contextual circumstances and analytical processes have some bearing on the number of dimensions of SQ.

Contextual stability. Carman (1990) tested the generic qualities of the SERVQUAL instrument in three service settings – a tyre retailer, a business school placement centre and a dental school patient clinic. Following Parasuraman *et al.*'s suggestion, he modified and augmented the items in the original ten-factor SERVQUAL scale to suit the three contexts. His factor analysis identified between five and seven underlying dimensions.

According to Carman, customers are at least partly context-specific in the dimensions they employ to evaluate SQ. In all three cases, Tangibles, Reliability and Security were present[7]. Responsiveness, a major component in the RATER scale, was relatively weak in the dental clinic context. Carman also commented: “Parasuraman, Zeithaml and Berry combined their original Understanding and Access dimensions into Empathy... our results did not find this to be an appropriate combination”. In particular he found that if a dimension is very important to customers they are likely to be decomposed into a number of sub-dimensions. This happened for the placement centre where

Responsiveness, Personal attention, Access and Convenience were all identified as separate factors. According to Carman, this indicates that researchers should work with the original ten dimensions, rather than adopt the revised five-factor Parasuraman *et al.* (1988) model.

There is also an indication from one piece of cross-cultural research that the scale may not always travel well. Ford *et al.* (1993) computed alphas for a SERVQUAL application in the higher education contexts of New Zealand and the USA markets which the authors describe as “intuitively” similar. Table II displays the results.

Table II.
SERVQUAL alphas in
New Zealand and the
USA

Dimensions	Cronbach alpha	
	USA	New Zealand
Tangibles	0.7049	0.6833
Reliability	0.8883	0.8514
Responsiveness	0.8378	0.8063
Assurance	0.8229	0.7217
Empathy	0.8099	0.7734

These results challenge Zeithaml’s (1988) claim that consumers form higher level abstractions of SQ that are generalized across contexts.

Item loadings. In some studies (e.g. Carman, 1990), items have not loaded on the factors to which they were expected to belong. Two items from the Empathy battery of the Parasuraman *et al.* (1988) instrument loaded heavily on the Tangibles factor in a study of dental clinic SQ. In the tyre retail study, a Tangibles item loaded on to Security; in the placement centre a Reliability item loaded on to Tangibles. An item concerning the ease of making appointments loaded on to Reliability in the dental clinic context, but Security in the tyre store context. He also found that only two-thirds of the items loaded in the same way on the expectations battery as they did in the perceptions battery. Carman supplies other examples of the same phenomena, and suggests that the unexpected results indicate both a face validity and a construct validity problem. In other words, he warns against importing SERVQUAL into service setting contexts without modification and validity checks.

Among his specific recommendations is the following: “We recommend that items on Courtesy and Access be retained and that items on some dimensions such as Responsiveness and Access be expanded where it is believed that these dimensions are of particular importance”. He also reports specific Courtesy and Access items which performed well in terms of nomological and construct validity.

Carman (1990) further suggested that the factors, Personal attention, Access or Convenience should be retained and further contextualized research work be done to identify their significance and meaning.

Item intercorrelations. Convergent validity and discriminant validity are important considerations in the measurement of second-order constructs such as SERVQUAL. One would associate a high level of convergent validity with a high level of intercorrelations between the items selected to measure a single RATER factor. Discriminant validity is indicated if the factors and their component items are independent of each other (i.e. the items load heavily on one factor only)[8]. Following their modified replication of Parasuraman *et al.*'s work, Babakus and Boller (1992) conclude that rules for convergence and discrimination do not indicate the existence of the five RATER dimensions.

The best scales have a high level of intercorrelation between items comprising a dimension (convergent validity). In their development work in four sectors (banking, credit-card company, repair and maintenance company, and long-distance telecommunications company) Parasuraman *et al.* (1988) found inter-item reliability coefficients (alphas) varying from 0.52 to 0.84. Babakus and Boller (1992) report alphas which are broadly consistent with those of Parasuraman, varying from 0.67 to 0.83 (see Table III). In their 1991b version, Parasuraman *et al.* report alphas from 0.60 to 0.93, and observe that "every alpha value obtained for each dimension in the final study is higher than the corresponding values in the...original study". They attribute this improvement to their rewording of the 22 scale items.

Spreng and Singh (1993), and Brown *et al.* (1993) are highly critical of the questionable application of alphas to difference scores. They evaluate the reliability of SERVQUAL using a measure specifically designed for difference scores (Lord, 1963). Spreng and Singh conclude that "there is not a great deal of difference between the reliabilities correctly calculated and the more common [alpha] calculation", an observation with which Parasuraman *et al.* (1993) concurred when they wrote: "The collective conceptual and empirical evidence neither demonstrates clear superiority for the non-difference score format nor warrants abandoning the difference score format".

Operational

Expectations. Notwithstanding the more fundamental criticism that expectations play no significant role in the conceptualization of service quality, some critics have raised a number of other concerns about the operationalization of E in SERVQUAL.

In their 1988 work, Parasuraman *et al.* defined expectations as "desires or wants of consumers, i.e. what they feel a service provider *should* offer rather than *would* offer" (emphasis added). The expectations component was designed to measure "customers' normative expectations" (Parasuraman *et al.*, 1990), and is "similar to the ideal standard in the customer satisfaction/dissatisfaction literature" (Zeithaml *et al.*, 1991). Teas (1993a) found these explanations "somewhat vague" and has questioned respondents' interpretation of the expectations battery in the SERVQUAL instrument. He believes that respondents may be using any one of six interpretations (Teas, 1993b):

Factor	Item	Parasuraman <i>et al.</i> (1988)		Babakus and Boller (1992)	
		Coefficient alpha	Item-to-total correlations	Coefficient alpha	Item-to-total correlations
Tangibles	Q1	0.72	0.69	0.67	0.38
	Q2		0.68		0.59
	Q3		0.64		0.31
	Q4		0.51		0.54
Reliability	Q5	0.83	0.75	0.82	0.66
	Q6		0.63		0.58
	Q7		0.71		0.59
	Q8		0.75		0.75
	Q9		0.50		0.49
Responsiveness	Q10	0.82	0.51	0.68	0.44
	Q11		0.77		0.44
	Q12		0.66		0.45
	Q13		0.86		0.52
Assurance	Q14	0.81	0.38	0.83	0.64
	Q15		0.72		0.77
	Q16		0.80		0.65
	Q17		0.45		0.58
Empathy	Q18	0.86	0.78	0.71	0.46
	Q19		0.81		0.46
	Q20		0.59		0.48
	Q21		0.71		0.45
	Q22		0.68		0.47

Table III.
Reliability of
SERVQUAL

- (1) *Service attribute importance.* Customers may respond by rating the expectations statements according to the importance of each.
- (2) *Forecasted performance.* Customers may respond by using the scale to predict the performance they would expect.
- (3) *Ideal performance.* The optimal performance; what performance “can be”.
- (4) *Deserved performance.* The performance level customers, in the light of their investments, feel performance should be.
- (5) *Equitable performance.* The level of performance customers feel they ought to receive given a perceived set of costs.
- (6) *Minimum tolerable performance.* What performance “must be”.

Each of these interpretations is somewhat different, and Teas contends that a considerable percentage of the variance of the SERVQUAL expectations measure can be explained by the difference in respondents’ interpretations.

Accordingly, the expectations component of the model lacks discriminant validity. Parasuraman *et al.* (1991b; 1994) have responded to these criticisms by redefining expectations as the service customers would expect from “excellent service organizations”, rather than “normative” expectations of service providers, and by vigorously defending their inclusion in SQ research.

Iacobucci *et al.* (1994) want to drop the term “expectations” from the SQ vocabulary. They prefer the generic label “standard”, and believe that several standards may operate simultaneously; among them “ideals”, “my most desired combination of attributes”, the “industry standard” of a nominal average competitor, “deserved” SQ, and brand standards based on past experiences with the brand.

Some critics have questioned SERVQUAL’s failure to assess customer evaluations based on absolute standards of SQ. The instrument asks respondents to report their expectations of excellent service providers within a class (i.e. the measures are relative rather than absolute). It has been argued that SERVQUAL predicts that:

customers will evaluate a service favourably as long as their expectations are met or exceeded, regardless of whether their prior expectations were high or low, and regardless of whether the absolute goodness of the [service] performance is high or low. This unyielding prediction is illogical. We argue that “absolute” levels (e.g. the prior standards) certainly must enter into a customer’s evaluation (Iacobucci *et al.*, 1994).

Put another way, SERVQUAL assumes that an E-score of six for Joe’s Greasy Spoon Diner is equivalent to an E-score of six for Michel Roux’s Le Lapin French restaurant. In absolute terms, clearly they are not. Grönroos (1993) refers to a similar oddity, which he calls the bad-service paradox. A customer may have low expectations based on previous experience with the service provider; if those expectations are met there is no gap and SQ is deemed satisfactory.

Since Zeithaml *et al.* (1991) have themselves identified two comparison norms for SQ assessment (“desired service”, the level of service a customer believes can and should be delivered; “adequate service”, the level of service the customer considers acceptable) it seems unlikely that the debate about the meaning of expectations is over.

Item composition. Each factor in the 1988 and 1991 SERVQUAL scales is composed of four or five items. It has become clear that this is often inadequate to capture the variance within, or the context-specific meaning of, each dimension. Carman’s (1990) study of hospital services employed 40 items. Bouman and van der Wiele (1992) used 48 items in their car service research, Saleh and Ryan (1992) 33 items in their hospitality industry research, Fort (1993) 31 items in his analysis of software house service quality and Babakus and Mangold (1992) 15 items in their hospital research. Parasuraman *et al.* (1991b) acknowledge that context-specific items can be used to supplement SERVQUAL, but caution that “the new items should be similar in form to the existing SERVQUAL items”.

Moments of truth. Many services are delivered over several moments of truth or encounters between service staff and customer: hotel and hospital services for

example. Carman (1990) found evidence that customers evaluate SQ by reference to these multiple encounters. For example, in his hospital research he listed the three items below:

- (1) My discharge from the hospital was prompt.
- (2) Nurses responded promptly when I called.
- (3) My admission to the hospital was prompt.

These items did not load heavily on a single Responsiveness factor as might be expected; instead they loaded on factors which represented a particular hospital function, or moment of truth. Parasuraman *et al.*, in contrast, have declared the SQ is a more global construct, not directly connected to particular incidents.

Polarity. Of the 22 items in the 1988 SERVQUAL scale, 13 statement pairs are positively worded, and nine pairs are negatively worded. The negatives are the full set of Responsiveness and Empathy statements. Parasuraman *et al.*'s goal was to reduce systematic response bias caused by yea-saying and nay-saying. This is accepted as good normative research practice (Churchill, 1979), yet has consequences for respondents who make more comprehension errors, and take more time to read items (Wason and Johnson-Laird, 1972).

In factor analysis of SERVQUAL data, Babakus and Boller (1992) found that all negatively-worded items loaded heavily on one factor while all positively-worded items loaded on another. They also found a significant difference between the average P, E and gap scores of positively and negatively-worded items. They conclude that the wording of the items produces a "method factor": "Item wording may be responsible for producing factors that are method artifacts rather than conceptually meaningful dimensions of service quality". Item wording creates data quality problems, and calls into question the dimensionality and validity of the instrument. Babakus and Mangold (1992), in their application of SERVQUAL to a hospital setting, therefore decided to employ only positively-worded statements. Parasuraman *et al.* (1991b) have responded to these criticisms by rewording all negatively-worded items positively.

Scale points. The use of seven-point Likert scales has been criticized on several grounds. Although none of these are specific to SERVQUAL applications, they bear repeating here. Lewis (1993) has criticized the scale for its lack of verbal labelling for points two to six. She believes this may cause respondents to overuse the extreme ends of the scale and suggests this could be avoided by labelling each point. Another issue is the respondents' interpretation of the meaning of the midpoint of the scale (e.g. is it a "don't know", "do not feel strongly in either direction" or a "do not understand the statement" response?) Lewis is also concerned about responses which suggest there is no gap when in fact a gap does exist. For instance a respondent may have expectations of 5.4 and perceptions of 4.6 (a gap of 0.8) but when completing SERVQUAL may rate each as 5, the nearest possible response in each case. This is an example of a Type II error.

Babakus and Mangold (1992) opted to use five-point Likert scales on the grounds that it would reduce the "frustration level" of patient respondents, increase response rate and response quality.

Two administrations. Respondents appear to be bored, and sometimes confused by the administration of E and P versions of SERVQUAL (Bouman and van der Wiele, 1992). Boredom and confusion imperil data quality.

Carman (1990) also comments on the timing of the two administrations. He is critical of Parasuraman *et al.* for asking respondents to complete the two questionnaires at a single sitting. In Parasuraman *et al.*'s 1988 work respondents were asked to report their expectations and perceptions, based on what they had experienced in the last three months. All self-reports were entirely *ex post*, a practice also criticized by Grönroos (1993). Carman also observed that it was impractical to expect customers to complete an expectations inventory prior to a service encounter and a perceptions inventory immediately afterwards. His solution was to collect data on the expectations-perceptions difference with a single question at a single administration, for example: "The visual appeal of XYZ's physical facilities is (much better, better, about the same, worse, much worse) than I expected". Lewis (1993) refers to work undertaken by Orledge who has also experimented with an alternative method of combining perceptions and expectations. He combined the two elements as in the following example:

Indicate on the scale using a "P" how well dressed the staff of company XYZ are. On the same scale indicate using an "E" how well dressed you expect the staff of companies in this industry to be.

smart ____: ____: E: ____: ____: P: ____ untidy

Bouman and van der Wiele (1992) also comment on the same problem. Babakus and Boller (1992), and Babakus *et al.* (1993b) solved the problem by employing a single seven-point scale to collect gap data. Recommended earlier by Carman (1990), the scale ranges from 7 = "greatly exceeds my expectations" to 1 = "greatly falls short of my expectations".

Clow and Vorhies (1993) argue:

When expectations and experience evaluations are measured simultaneously, respondents will indicate that their expectations are greater than they actually were before the service encounter.

They contend that expectations must be measured prior to receipt of services otherwise responses will be biased. Specifically, Clow and Vorhies found that:

Customers who had a negative experience with the service tend to overstate their expectations, creating a larger gap; customers who had a positive experience tend to understate their expectations, resulting in smaller gaps.

Variance extracted. Fornell and Larcker (1981) have suggested that "variance extracted" should be stringently employed as a measure of construct validity. Parasuraman *et al.* (1988) reported that the total amount of variance extracted by the five RATER factors in the bank, credit-card, repair and maintenance, and long-distance telephone samples was 56.0 per cent, 57.5 per cent, 61.6 per cent and 56.2 per cent respectively. Parasuraman *et al.* (1991a) report variance explained in a telephone company, insurance company 1, insurance company 2, bank 1 and bank 2 at 67.2 per cent, 68.3 per cent, 70.9 per cent, 71.6 per cent and 66.9 per cent, respectively. When the samples are combined, variance explained is 67.9 per cent.

Babakus and Boller's (1992) utility-sector replication reported 58.3 per cent. Carman's (1990) modified replication in the hospital sector, tyre store, business school placement centre and dental clinic reported 71 per cent, 61 per cent, 75 per cent and 71 per cent respectively. Saleh and Ryan's (1992) modified replication in the hotel sector reported 78.6 per cent. Bouman and van der Wiele's (1992) modified replication in car servicing reported 40.7 per cent only. Generally, the modified scales tended to produce higher levels of variance extracted. The higher the variance extracted, the more valid is the measure.

Conclusion

SERVQUAL has undoubtedly had a major impact on the business and academic communities.

This review has identified a number of theoretical and operational issues which should concern users of the instrument. Since the most serious of these are concerned with face validity and construct validity, this conclusion briefly reviews the nature and significance of validity.

Face validity is concerned with the extent to which a scale appears to measure what it purports to measure.

Construct validity generally:

is used to refer to the vertical correspondence between a construct which is at an unobservable, conceptual level and a purported measure of it which is at an operational level. In an ideal sense, the term means that a measure assesses the magnitude and direction of (1) *all* of the characteristics and (2) *only* the characteristics of the construct it is purported to assess (Peter, 1981, emphases added).

In particular, the concerns about the adoption of an inappropriate paradigm, the gaps model, SERVQUAL's process orientation, and SERVQUAL's dimensionality (the four theoretical criticisms as listed earlier) are construct validity issues.

Critical face and construct validity questions which SERVQUAL researchers face are: Do consumers actually evaluate SQ in terms of expectations and perceptions? Do the five RATER dimensions incorporate the full range of SQ attributes? Do consumers incorporate "outcome" evaluations into their assessments of SQ?

Construct validity is itself a composite of several forms of validity: nomological validity, convergent validity and discriminant validity.

Nomological validity is the extent to which a measure correlates in theoretically predictable ways with measures of different but related constructs. SQ is one of a number of apparently interrelated constructs whose precise alignment has yet to be explored. Included in the nomological net are customer (dis)satisfaction, customer retention and defection, behavioural intention, attitude to service provider or organization, and service provider or organization choice. Some research into these questions has been published (Parasuraman *et al.*, 1991b; Richard and Allaway, 1993) but the relationships have yet to be explored fully.

Convergent validity is the extent to which a scale correlates with other measures of the same construct. A high level of intercorrelation between items comprising each RATER dimension would indicate high convergent validity

internal to SERVQUAL. A high level of correlation between SERVQUAL scores and a different, reliable and valid measure of SQ, would indicate a high level of external convergent validity. Discriminant validity is the extent to which a measure does *not* correlate with other measures from which it is purported to differ. If SQ evaluations were composed of five distinct RATER dimensions, one would expect little correlation between the five factors. SERVQUAL's dimensionality would be regarded as more stable if individual items loaded on to the dimensions to which they belong.

Issues of face and construct validity are of overriding importance in the development of instruments such as SERVQUAL. The operational criticisms are evidently less significant than the theoretical criticisms, and pose less of a threat to validity. The theoretical criticisms raised in this article are of such moment that the validity of the instrument must be called into question.

Despite these shortcomings, SERVQUAL seems to be moving rapidly towards institutionalized status. As Rust and Zahorik (1993) have observed, "the general SERVQUAL dimensions ... should probably be put on any first pass as a list of attributes of service".

These criticisms indicate that there is still a need for fundamental research. There are still doubts about whether customers routinely assess SQ in terms of Expectations and Perceptions; there are doubts about the utility and appropriateness of the disconfirmation paradigm; there are doubts about the dimensionality of SQ; there are doubts about the universality of the five RATER dimensions. These are serious concerns which are not only significant for users of SERVQUAL but for all those who wish to understand better the concept of SQ.

Directions for future research

This review has raised several conceptual and operational difficulties surrounding SERVQUAL which are yet to be resolved. The following represent a set of questions which SQ researchers should address:

- (1) Do consumers always evaluate SQ in terms of expectations and perceptions? What other forms of SQ evaluation are there?
- (2) What form do customer expectations take and how best, if at all, are they measured? Are expectations common across a class of service providers?
- (3) Do attitude-based measures of SQ perform better than the disconfirmation model? Which attitudinal measure is most useful?
- (4) Is it advantageous to integrate outcome evaluations into SQ measurement and how best can this be done?
- (5) Is the predictive validity of P measures of service quality better than that of P – E measures?
- (6) What are the relationships between SQ, customer satisfaction, behavioural intention, purchase behaviour, market share, word-of-mouth and customer retention?

- (7) What is the role of context in determining E and P evaluations? What context-markers do consumers employ?
- (8) Are analytical context markers such as tangibility and consumer involvement helpful in advancing SQ theory?
 - Do evaluative criteria in intangible-dominant services (e.g. consulting) differ from those in tangible-dominant services (e.g. hotels)?
 - How does involvement influence the evaluation of SQ?
- (9) How do customers integrate transaction-specific or MOT-specific evaluations of SQ? To what extent are some MOTs more influential in the final evaluation than others?
- (10) What are the relationships between the five RATER factors? How stable are those relationships across context?
- (11) What is the most appropriate scale format for collecting valid and reliable SQ data?
- (12) To what extent can customers correctly classify items into their a priori dimensions?

Answers to questions such as these would help improve our understanding of the service quality construct and assess the value of the SERVQUAL instrument. Even in its present state SERVQUAL is a helpful operationalization of a somewhat nebulous construct.

Many of these questions require contextually sensitive qualitative research. The first question, "Do consumers always evaluate SQ in terms of expectations and perceptions?", is perhaps best approached through in-depth case analyses of particular service encounters. The formation of expectations implies a consumer who accumulates and processes information about a class of service providers. This would appear to make prima facie sense for high-cost, high-risk services, e.g. if purchasing a weekend break to celebrate 25 years of wedded bliss. Is it as likely that expectations high in cognitive content would be formed for a low-cost, low-risk service such as a hot drink from a coffee shop? The role of context appears to have been repressed or subjugated in the present body of SERVQUAL research. Context needs to be recovered.

Other questions lend themselves to multisectoral comparative analyses. For example, the question, "Is the predictive validity of P-measures of SQ better than that of P – E measures?", is perhaps best approached in multi-sectoral study which thoroughly tests the predictive performance of P and P – E SQ measures.

Pursuit of this research agenda would surely strengthen our understanding of the meaning, measurement and management of service quality. Parasuraman, Zeithaml and Berry have undoubtedly done a splendid job of marketing SERVQUAL's measurement and management technologies. It remains to be seen whether its dominance will remain unchallenged.

Notes

1. The mnemonic RATER is a helpful aide m emoire, where R = reliability, A = assurance, T = tangibles, E = empathy and R = responsiveness.
2. Richard and Allaway's (1993) research was largely focused on testing SERVQUAL's predictive validity. Parasuraman *et al.* (1991b) have also tested the predictive validity of the modified SERVQUAL scale. Customers in five samples were asked three questions: Have you recently had a service problem with the company? If you have experienced a problem, was it resolved to your satisfaction? Would you recommend the service firm to a friend? It was hypothesized that positive answers to these questions would be correlated negatively, positively and positively, respectively, with higher perceived SQ scores. All results were statistically significant in the hypothesized direction, lending support to the predictive validity of the instrument.
3. Babakus and Boller (1992) have expressed concern that it is unclear whether SERVQUAL is measuring a number of distinct constructs or a single, global, more abstract variable.
4. Cronin and Taylor (1992), following a test of SERVQUAL in four classes of service firm, conclude in stark contrast that "the five-component structure proposed by Parasuraman, Zeithaml and Berry (1988) for their SERVQUAL scale is not confirmed".
5. Babakus and Mangold's (1992) research into hospital SQ identified three factors within the expectations data, accounting for 56.2 per cent of the variance in the item scores, two factors within the perceptions data (70.6 per cent) and "no meaningful factor structure" within the difference or gaps data.
6. Customer kindness, that is "the front office personnel's approach to the customer and his problems, regardless of the service delivered", was the only factor to have a significant relationship with future car servicing intentions, future car purchase intentions, and word-of-mouth recommendation.
7. Carman's Security factor is composed of Credibility, Security and Competence. Parasuraman *et al.* (1988) had incorporated these three components, together with Communication and Courtesy, into the factor Assurance.
8. For a discussion of construct, convergent and discriminant validity see Churchill (1979) and Peter (1981).

References and further reading

- Andersson, T.D. (1992), "Another model of service quality: a model of causes and effects of service quality tested on a case within the restaurant industry", in Kunst, P. and Lemmink, J. (Eds), *Quality Management in Service*, van Gorcum, The Netherlands, pp. 41-58.
- Aubry, C.A. and Zimble, D.A. (1983), "The banking industry: quality costs and improvement", *Quality Progress*, December, pp. 16-20.
- Babakus, E. and Boller, G.W. (1992), "An empirical assessment of the SERVQUAL scale", *Journal of Business Research*, Vol. 24, pp. 253-68.
- Babakus, E. and Inhofe, M. (1991), "The role of expectations and attribute importance in the measurement of service quality", in Gilly, M.C. *et al.* (Eds), *Proceedings of the Summer Educators' Conference*, American Marketing Association, Chicago, IL, pp. 142-4.
- Babakus, E. and Mangold, W.G. (1992), "Adapting the SERVQUAL scale to hospital services: an empirical investigation", *Health Services Research*, Vol. 26 No. 2, February, pp. 767-86.
- Babakus, E., Pedrick, D.L. and Inhofe, M. (1993b), "Empirical examination of a direct measure of perceived service quality using SERVQUAL items", unpublished manuscript, Memphis State University, TN.
- Babakus, E., Pedrick, D.L. and Richardson, A. (1993a), "Measuring perceived service quality within the airline catering service industry", unpublished manuscript, Memphis State University, TN.

- Baker, J.A. and Lamb, C.W. Jr (1993), "Managing architectural design service quality", *Journal of Professional Services Marketing*, Vol. 10 No. 1, pp. 89-106.
- Bolton, R.N. and Drew, J.H. (1991), "A multistage model of customers' assessment of service quality and value", *Journal of Consumer Research*, Vol. 17, March, pp. 375-84.
- Boulding, W., Kalra, A., Staelin, R. and Zeithaml, V.A. (1993), "A dynamic process model of service quality: from expectations to behavioral intentions", *Journal of Marketing Research*, Vol. 30, February, pp. 7-27.
- Bouman, M. and van der Wiele, T. (1992), "Measuring service quality in the car service industry: building and testing an instrument", *International Journal of Service Industry Management*, Vol. 3 No. 4, pp. 4-16.
- Brown, T.J., Churchill, G.A. and Peter, J.P. (1993), "Improving the measurement of service quality", *Journal of Retailing*, Vol. 69 No. 1, Spring, pp. 127-39.
- Buzzell, R.D. and Gale, B.T. (1987), *The PIMS Principles*, Free Press, New York, NY.
- Carman, J.M. (1990), "Consumer perceptions of service quality: an assessment of the SERVQUAL dimensions", *Journal of Retailing*, Vol. 66 No. 1, Spring, pp. 33-5.
- Churchill, G.A. (1979), "A paradigm for developing better measures of marketing constructs", *Journal of Marketing Research*, Vol. 19, February, pp. 64-73.
- Churchill, G.A. and Surprenant, C. (1982), "An investigation into the determinants of customer satisfaction", *Journal of Marketing Research*, Vol. 19, pp. 491-504.
- Clow, K.E. and Vorhies, D.E. (1993), "Building a competitive advantage for service firms", *Journal of Services Marketing*, Vol. 7 No. 1, pp. 22-3.
- Cronin, J.J. Jr and Taylor, S.A. (1992), "Measuring service quality: a reexamination and extension", *Journal of Marketing*, Vol. 56, July, pp. 55-68.
- Cronin, J.J. Jr and Taylor, S.A. (1994), "SERVPERF versus SERVQUAL: reconciling performance-based and perceptions-minus expectations measurement of service quality", *Journal of Marketing*, Vol. 58, January, pp. 125-31.
- Crosby, P.B. (1979), *Quality Is Free*, McGraw-Hill, New York, NY.
- Fick, G.R. and Ritchie, J.R.B. (1991), "Measuring service quality in the travel and tourism industry", *Journal of Travel Research*, Vol. 30 No. 2, Autumn, pp. 2-9.
- Ford, J.W., Joseph, M. and Joseph, B. (1993), "Service quality in higher education: a comparison of universities in the United States and New Zealand using SERVQUAL", unpublished manuscript, Old Dominion University, Norfolk, VA.
- Fornell, C. and Larcker D.F. (1981), "Evaluating structural equation models with unobservable variables and measurement error", *Journal of Marketing Research*, Vol. 18, February, pp. 39-50.
- Fort, M. (1993), "Customer defined attributes of service quality in the IBM mid-range computer software industry", unpublished MBA dissertation, Manchester Business School, Manchester.
- Freeman, K.D. and Dart, J. (1993), "Measuring the perceived quality of professional business services", *Journal of Professional Services Marketing*, Vol. 9 No. 1, pp. 27-47.
- Gagliano, K.B. and Hathcote, J. (1994), "Customer expectations and perceptions of service quality in apparel retailing", *Journal of Services Marketing*, Vol. 8 No. 1, pp. 60-9.
- Grönroos, C. (1982), *Strategic Management and Marketing in the Service Sector*, Swedish School of Economics and Business Administration, Helsinki.
- Grönroos, C. (1984), "A service quality model and its marketing implications", *European Journal of Marketing*, Vol. 18, pp. 36-44.
- Grönroos, C. (1993), "Toward a third phase in service quality research: challenges and future directions", in Swartz, T.A., Bowen, D.E. and Brown, S.W. (Eds), *Advances in Services Marketing and Management*, Vol. 2, JAI Press, Greenwich, CT, pp. 49-64.
- Hardie, B.G.S., Johnson, E.J. and Fader, P.S. (1992), "Modelling loss aversion and reference dependence effects on brand choice", working paper, Wharton School, University of Pennsylvania, PA.

- Hedvall, M.-B. and Paltschik, M. (1989), "An investigation in, and generation of, service quality concepts", in Avlonitis, G.J. *et al.* (Eds), *Marketing Thought and Practice in the 1990s*, European Marketing Academy, Athens, pp. 473-83.
- Higgins, L.F., Ferguson, J.M. and Winston, J.M. (1991), "Understanding and assessing service quality in health maintenance organizations", *Health Marketing Quarterly*, Vol. 9, Nos 1-2, pp. 5-20.
- Iacobucci, D., Grayson, K.A. and Omstrom, A.L. (1994), "The calculus of service quality and customer satisfaction: theoretical and empirical differentiation and integration", in Swartz, T.A., Bowen, D.E. and Brown, S.W. (Eds), *Advances in Services Marketing and Management*, Vol. 3, JAI Press, Greenwich, CT, pp. 1-68.
- Johns, N. (1993), "Quality management in the hospitality industry, part 3: recent developments", *International Journal of Contemporary Hospitality Management*, Vol. 5 No. 1, pp. 10-15.
- Juran, J.M. (1951), *Quality Control Handbook*, McGraw-Hill, New York, NY.
- Kahneman, D. and Miller, D.T. (1986), "Norm theory: comparing reality to its alternatives", *Psychological Review*, Vol. 93, pp. 136-53.
- Kong, R. and Mayo, M.C. (1993), "Measuring service quality in the business-to-business context", *Journal of Business and Industrial Marketing*, Vol. 8 No. 2, pp. 5-15.
- Kwon, W. and Lee, T.J. (1994), "Measuring service quality in Singapore retail banking", *Singapore Management Review*, Vol. 16 No. 2, July, pp. 1-24.
- Leblanc, G. and Nguyen, N. (1988), "Customers' perceptions of service quality in financial institutions", *International Journal of Bank Marketing*, Vol. 6 No. 4, pp. 7-18.
- Lehtinen, J.R. and Lehtinen, O. (1982), "Service quality: a study of quality dimensions", unpublished working paper, Service Management Institute, Helsinki.
- Lewis, B.R. (1993), "Service quality measurement", *Marketing Intelligence and Planning*, Vol. 11 No. 4, pp. 4-12.
- Lord, F.M. (1963), "Elementary models for measuring change", in Harris, C.W. (Ed.), *Problems in Measuring Change*, University of Wisconsin Press, Madison, WI, pp. 22-38.
- McElwee, G. and Redman, T. (1993), "Upward appraisal in practice: an illustrative example using the QUALED scale", *Education and Training*, Vol. 35 No. 2, December, pp. 27-31.
- Mangold, G.W. and Babakus, E. (1991), "Service quality: the front-stage perspective vs the back stage perspective", *Journal of Services Marketing*, Vol. 5 No. 4, Autumn, pp. 59-70.
- Masser, W.J. (1957), "The quality manager and quality costs", *Industrial Quality Control*, Vol. 14, pp. 5-8.
- Oliver, R.L. (1980), "A cognitive model of the antecedents and consequences of satisfaction decisions", *Journal of Marketing Research*, Vol. 17, November, pp. 460-9.
- Oliver, R.L. (1993), "A conceptual model of service quality and service satisfaction: compatible goals, different concepts", in Swartz, T.A., Bowen, D.E. and Brown, S.W. (Eds), *Advances in Services Marketing and Management*, Vol. 2, JAI Press, Greenwich, CT, pp. 65-85.
- Parasuraman, A., Berry, L.L. and Zeithaml, V.A. (1990), *An Empirical Examination of Relationships in an Extended Service Quality Model*, Marketing Science Institute, Cambridge, MA.
- Parasuraman, A., Berry, L.L. and Zeithaml, V.A. (1991a), "Perceived service quality as a customer-based performance measure: an empirical examination of organizational barriers using an extended service quality model", *Human Resource Management*, Vol. 30 No. 3, Autumn, pp. 335-64.
- Parasuraman, A., Zeithaml, V. and Berry, L.L. (1985), "A conceptual model of service quality and its implications for future research", *Journal of Marketing*, Vol. 49, Autumn, pp. 41-50.
- Parasuraman, A., Zeithaml, V. and Berry, L.L. (1986), "SERVQUAL: a multiple-item scale for measuring customer perceptions of service quality", Report No. 86-108, Marketing Science Institute, Cambridge, MA.

- Parasuraman, A., Zeithaml, V. and Berry, L.L. (1988), "SERVQUAL: a multiple-item scale for measuring consumer perceptions of service quality", *Journal of Retailing*, Vol. 64, Spring, pp. 12-40.
- Parasuraman, A., Zeithaml, V. and Berry, L.L. (1991b), "Refinement and reassessment of the SERVQUAL scale", *Journal of Retailing*, Vol. 67 No. 4, pp. 420-50.
- Parasuraman, A., Zeithaml, V. and Berry, L.L. (1993), "Research note: more on improving service quality measurement", *Journal of Retailing*, Vol. 69 No. 1, Spring, pp. 140-7.
- Parasuraman, A., Zeithaml, V. and Berry, L.L. (1994), "Reassessment of expectations as a comparison standard in measuring service quality: implications for future research", *Journal of Marketing*, Vol. 58, January, pp. 111-24.
- Peter, J.P. (1981), "Construct validity: a review of basic issues and marketing practices", *Journal of Marketing Research*, Vol. 18, May, pp. 133-45.
- Reichheld, F.F. and Sasser, W.E. Jr (1990), "Zero defections: quality comes to service", *Harvard Business Review*, September-October, pp. 105-11.
- Reidenbach, R.E. and Sandifer-Smallwood, B. (1990), "Exploring perceptions of hospital operations by a modified SERVQUAL approach", *Journal of Health Care Marketing*, Vol. 10 No. 4, December, pp. 47-55.
- Richard, M.D. and Allaway, A.W. (1993), "Service quality attributes and choice behavior", *Journal of Service Marketing*, Vol. 7 No. 1, pp. 59-68.
- Rigotti, S. and Pitt, L. (1992), "SERVQUAL as a measuring instrument for service provider gaps in business schools", *Management Research News*, Vol. 15 No. 3, pp. 9-17.
- Rust, R.T. and Zahorik, A.J. (1993), "Customer satisfaction, customer retention and market share", *Journal of Retailing*, Vol. 69 No. 2, Summer, pp. 193-215.
- Saleh, F. and Ryan, C. (1992), "Analysing service quality in the hospitality industry using the SERVQUAL model", *Services Industries Journal*, Vol. 11 No. 3, pp. 324-43.
- Scott, D. and Shieff, D. (1993), "Service quality components and group criteria in local government", *International Journal of Service Industry Management*, Vol. 4 No. 4, pp. 42-53.
- Soliman, A.A. (1992), "Assessing the quality of health care", *Health Care Marketing*, Vol. 10 Nos 1-2, pp. 121-41.
- Spreng, R.A. and Singh, A.K. (1993), "An empirical assessment of the SERVQUAL scale and the relationship between service quality and satisfaction", unpublished manuscript, Michigan State University, TN.
- Taylor, S.A., Sharland, A., Cronin, A.A. Jr and Bullard, W. (1993), "Recreational quality in the international setting", *International Journal of Service Industries Management*, Vol. 4 No. 4, pp. 68-88.
- Teas, K.R. (1993a), "Expectations, performance evaluation and consumers' perceptions of quality", *Journal of Marketing*, Vol. 57 No. 4, pp. 18-24.
- Teas, K.R. (1993b), "Consumer expectations and the measurement of perceived service quality", *Journal of Professional Services Marketing*, Vol. 8 No. 2, pp. 33-53.
- Teas, K.R. (1994), "Expectations as a comparison standard in measuring service quality: an assessment of a reassessment", *Journal of Marketing*, Vol. 58, January, pp. 132-9.
- Vandamme, R. and Leunis, J. (1993), "Development of a multiple-item scale for measuring hospital service quality", *International Journal of Service Industry Management*, Vol. 4 No. 3, pp. 30-49.
- Walbridge, S.W. and Delene, L.M. (1993), "Measuring physician attitudes of service quality", *Journal of Health Care Marketing*, Vol. 13 No. 4, Winter, pp. 6-15.
- Wason, P.J. and Johnson-Laird, P.N. (1972), *Psychology of Reasoning: Structure and Content*, B.T. Batsford, London.
- Woodruff, R.B., Cadotte, E.R. and Jenkins, R.L. (1983), "Modeling consumer satisfaction processes using experience-based norms", *Journal of Marketing Research*, Vol. 20, pp. 296-304.
- Wong, S.M. and Perry, C. (1991), "Customer service strategies in financial retailing", *International Journal of Bank Marketing*, Vol. 9 No. 3, pp. 11-16.

-
- Wotruba, T.R. and Tyagi, P.K. (1991), "Met expectations and turnover in direct selling", *Journal of Marketing*, Vol. 55, pp. 24-35.
- Zahorik, A.J. and Rust R.T. (1992), "Modeling the impact of service quality of profitability: a review", in Swartz, T.A., Bowen, D.E. and Brown, S.W. (Eds), *Advances in Services Marketing and Management*, JAI Press, Greenwich, CT, pp. 49-64.
- Zeithaml, V.A. (1988), "Consumer perceptions of price, quality and value: a means-end model and synthesis of evidence", *Journal of Marketing*, Vol. 52, July, pp. 2-22.
- Zeithaml, V.A., Berry, L.L. and Parasuraman, A. (1991), "The nature and determinants of customer expectations of service", working paper 91-113, Marketing Science Institute, Cambridge, MA.
- Zeithaml, V.A., Berry, L.L. and Parasuraman, A. (1993), "The nature and determinants of customer expectation of service", *Journal of the Academy of Marketing Science*, Vol. 21 No. 1, pp. 1-12.
- Zeithaml, V.A., Parasuraman, A. and Berry, L.L. (1990), *Delivering Quality Service: Balancing Customer Perceptions and Expectations*, Free Press, New York, NY.
- Zeithaml, V.A., Parasuraman, A. and Berry, L.L. (1992), "Strategic positioning on the dimensions of service quality", in Swartz, T.A., Bowen, D.E. and Brown, S.W. (Eds), *Advances in Services Marketing and Management*, Vol. 2, JAI Press, Greenwich, CT, pp. 207-28.

Appendix. Ten components of service quality

- (1) *Reliability* involves consistency of performance and dependability. It also means that the firm performs the service right first time and honours its promises. Specifically, it may involve:
 - accuracy in billing;
 - performing the service at the designated time.
- (2) *Responsiveness* concerns the willingness or readiness of employees to provide service. It may involve:
 - mailing a transaction slip immediately;
 - calling the customer back quickly;
 - giving prompt service (e.g. setting up appointments quickly).
- (3) *Competence* means possession of the required skills and knowledge to perform the service. It involves:
 - knowledge and skill of the contact personnel;
 - knowledge and skill of operational support personnel;
 - research capability of the organization.
- (4) *Access* involves approachability and ease of contact. It may mean:
 - the service is easily accessible by telephone;
 - waiting time to receive service is not extensive;
 - convenient hours of operation and convenient location of service facility.
- (5) *Courtesy* involves politeness, respect, consideration, and friendliness of contact personnel (including receptionists, telephone operators, etc.). It includes:
 - consideration for the consumers property;
 - clean and neat appearance of public contact personnel.
- (6) *Communication* means keeping customers informed in language they can understand, and listening to them. It may mean that the company has to adjust its language for different customers. It may involve:
 - explaining the service itself and how much the service will cost;
 - explaining the trade-offs between service and cost;
 - assuring the consumer that a problem will be handled.

- (7) *Credibility* involves trustworthiness, believability, honesty. It involves having the customer's best interests at heart. Contributing to credibility are:
- company name and reputation;
 - personal characteristics of the contact personnel;
 - the degree of hard sell involved in interactions with the customer.
- (8) *Security* is the freedom from danger, risk, or doubt. It may involve:
- physical safety;
 - financial security and confidentiality.
- (9) *Understanding/knowing the customer* involves making the effort to understand the customer's needs. It involves:
- learning the customer's specific requirements;
 - providing individualized attention.
- (10) *Tangibles* include the physical evidence of the service:
- physical facilities and appearance of personnel;
 - tools or equipment used to provide the service;
 - physical representations of the service, such as a plastic credit card.