IJSIM
7,4

4

Received September 1995
Revised May 1996

A comparison of question scales used for measuring customer satisfaction

Peter J. Danaher
Department of Marketing, University of Auckland, New Zealand, and
Vanessa Haddrell
Department of Marketing, University of Auckland, New Zealand

Introduction

Customer satisfaction surveys are now increasingly administered by many service industries ranging from dentists to airlines. This increase has been fuelled by the increasing growth of service industries compared with manufacturing industries and the demand for total quality management for both industries. Along with the increased research into customer satisfaction has been a corresponding increase in the diversity of measurement scales used in customer satisfaction surveys (Devlin *et al.*, 1993). A review of the literature revealed over 40 different scales used to measure consumer product or service satisfaction, from retail to airline settings (Haddrell, 1994). With such a large number of alternative scales available, many in everyday use, it is timely to compare some of the common scales to evaluate their merits. This is the purpose of this paper.

Almost all studies of customer satisfaction share a common characteristic that has been noted by a number of researchers (Oliver, 1981; Peterson and Wilson, 1992; Westbrook, 1980b; Yi, 1990). That is, "virtually all self-reports of customer satisfaction possess a distribution in which a majority of the responses indicate that customers are satisfied and the distribution itself is negatively skewed" (Peterson and Wilson, 1992, p. 62). Peterson and Wilson (1992) offered various explanations for this phenomenon. These include:

(1) individuals could in fact be very satisfied with what they consume and purchase;

(2) satisfaction is caused by factors (such as expectations, performance), therefore its antecedents may influence the shape and level of the observed distributions (see also Bearden and Teel, 1983; Cadotte *et al.*, 1987; Oliver, 1980, 1981);

(3) satisfaction may possess a distribution that is different from what we know as a normal distribution; and

(4) the level and shape of customer satisfaction ratings are a function of the research methodologies employed, for example, response rate bias,

International Journal of Service Industry Management, Vol. 7 No. 4, 1996, pp. 4-26. © MCB University Press, 0956-4233

question form, question context and collection mode bias, and the characteristics of individuals.

Peterson and Wilson (1992, p. 69) concluded that the skewness of customer satisfaction self-reports is due to all four factors above, with particular causal importance placed on the research methodologies employed. They state "it is not clear what customer satisfaction ratings are measuring … perhaps it is not possible to measure 'true satisfaction'. True satisfaction is probably so intertwined with both intrapersonal and methodological considerations that it may never be possible to disentangle them".

This skewness in the distribution of satisfaction ratings causes difficulties in the analysis, especially if multiple regression (which assumes symmetry in the error distribution) is used to determine the relative importance of components of a service delivery process (Danaher and Mattsson, 1994; Rust *et al.*, 1994). It also causes difficulties in interpretation as it appears that most customers are satisfied, yet customer loyalty is still not necessarily strong even for these apparently satisfied customers.

We study some of the issues raised in points (3) and (4), namely, the shape of the distribution of satisfaction scores, and, in particular, the type of measurement scale used to measure customer satisfaction.

In the next section we review the scales that have been used to measure customer satisfaction. Following this, we outline the criteria we use to evaluate the measurement scales. We then discuss the method we use to compare scales, and report the results of the study. Finally, we summarize our findings and make conclusions about the relative strengths and weaknesses of the scales we compare.

Scales used in customer satisfaction surveys

Selecting the scales

There are two broad types of scale used in customer satisfaction surveys, single- and multi-item scales. Many researchers have used simple single-item scales (generally having 2-9 points) to reflect "very satisfied" to "very dissatisfied" responses (Andreasen and Best, 1977; Oliver, 1977; Olshavsky and Miller, 1972; Westbrook, 1980a). Although single-item scales are simple, they do have at least two faults. First, the single-item scale cannot provide information on components and cannot assess various dimensions separately, therefore it may not capture the complexity of customer satisfaction entirely. Second, it is very difficult to assess reliability with a single-item measure, with the only reliability estimate being that of a test-retest format (Yi, 1990). Westbrook (1980b) compared the test-retest reliability of some single-item scales in several product and service industries. The three scales examined were the delighted-terrible, percentage satisfied and needs ("met extremely well" to "met extremely poorly") scales, with the delighted-terrible scale emerging as the most reliable.

Recent customer satisfaction studies have tended to use multi-item measures of customer satisfaction (Bearden and Teel, 1983; Churchill and Surprenant,

1982; Danaher and Mattsson, 1994; Oliver, 1980; Rust and Zahorik, 1993; Rust *et al.*, 1994). Here, survey respondents are not just asked to give an overall evaluation of their satisfaction with the service but are also asked to rate the key components of the service process. Westbrook and Oliver (1981) studied the reliabilities of a number of multi-item scales – verbal, graphic, semantic differential, Likert and inferential scales. They found that, on average, the semantic differential scale had the highest reliability. Other studies have also shown multi-item measures to be more reliable compared with their single-item counterparts (Bearden and Teel, 1983; Churchill and Surprenant, 1982; Oliver, 1980). Therefore, we use multi-item measures of customer satisfaction in our study.

An audit of marketing research textbooks and journals revealed a large number of scales commonly used in consumer research, including rank order, constant sum, graphical, Likert, semantic differential, paired comparison and stapel scales. In the customer satisfaction measurement setting these scales can be grouped into three broad categories. They are: performance scales, such as poor, fair, good and excellent; disconfirmation scales, such as worse than expected to better than expected; and satisfaction scales, such as very dissatisfied to very satisfied (Devlin *et al.*, 1993; Haddrell, 1994).

Devlin *et al.* (1993) and Rust *et al.* (1994, pp. 61-2) have recommended disconfirmation scales in preference to performance and satisfaction scales for at least three reasons. The first is that they incorporate the well known disconfirmation paradigm into the customer satisfaction setting (Cadotte *et al.*, 1987). In a sense they compress Parasuraman *et al.*'s (1988) two-stage SERVQUAL measurement of both expectation and perception into one succinct question. A respondent simply rates the service with respect to their own expectations. The second attractive feature of this scale is that "it can be shown mathematically that comparison with expectations will correlate higher with customer retention than either a quality question or a satisfaction question" (Rust *et al.*, 1994, p. 61). The last attractive feature of the disconfirmation scale is that it can drastically reduce the asymmetry in the measured perceived service. This is because not all customers who rate the service highly, for example, as good or excellent on a poor to excellent scale, will also say they perceive the service as "better than expected".

Other studies which have compared customer satisfaction measurement scales include Devlin *et al.* (1993), who made some suggestions on the relative merits of performance, satisfaction and disconfirmation scales on the basis of their experience at Bellcore. Westbrook (1980b) introduced a delighted-terrible scale from the sociology literature and recommended it on the basis of its reliability and symmetry properties. However, he did not compare the delighted-terrible scale with scales that are now commonly used in customer satisfaction surveys. Finally, Parasuraman *et al.* (1994) recently compared three question formats for measuring service quality, as opposed to customer satisfaction (see Boulding *et al.*, 1993; Cronin and Taylor, 1992; Oliver, 1993a; Parasuraman *et al.*, 1991; Taylor, 1995; and Taylor and Baker, 1994 for a

clarification of the distinction of these two concepts). Of these three studies, only Westbrook's asked the same respondents to rate satisfaction with different scales and this was for a single item scale, namely, overall product evaluation. The other studies used different respondents for each scale format. Clearly it is preferable to obtain comparative information from the same respondents to eliminate between-subjects variability, which is the method used in our study.

Conceptual relationship among the scales

Before proceeding we must address some potential conceptual problems in measuring customer satisfaction using performance, disconfirmation and satisfaction scales. It might be argued that, for instance, using a performance scale to measure customer satisfaction is bound to be problematic because performance is different from satisfaction. Hence, of course the performance scale will be inferior to the satisfaction scale for measuring customer satisfaction. While performance, disconfirmation and satisfaction are somewhat distinct concepts, they are strongly related. Several studies have shown that service performance drives disconfirmation, which, in turn, drives satisfaction (Oliver, 1980, 1981; Tse and Wilton, 1988; Yi, 1990). Furthermore, it is possible for performance to drive satisfaction directly, without disconfirmation acting as an intermediary (Bolton and Drew, 1991; Churchill and Surprenant, 1982; Oliver, 1993b; Tse and Wilton, 1988).

In addition to the growing body of evidence showing the performance, disconfirmation and satisfaction relationships, Oliver (1993a, p. 80) posits a model in which performance drives "ideal disconfirmation", which drives service quality, which impacts on satisfaction. Therefore service quality may be an intermediary between disconfirmation and satisfaction. Given that service quality and customer satisfaction are distinct concepts (Oliver, 1993a; Parasuraman *et al.*, 1991; Taylor, 1995; Taylor and Baker, 1994), it could be argued that the performance and disconfirmation scales measure service quality rather than customer satisfaction. Therefore, any differences among the performance, disconfirmation and satisfaction scales may be due to the already established theoretical differences between service quality and customer satisfaction. However, the model proposed by Oliver (1993a) has not been empirically validated so the importance of service quality as an intermediary between disconfirmation and satisfaction is not yet established, whereas strong relationships have been established among performance, disconfirmation and satisfaction. Hence, we do not consider potential effects of service quality to be influential in our measurement scale comparison. Moreover, in this study almost all the measurements are made on specific service transactions. Hence the effect of service quality, which is more of a summary concept, are likely to be negligible. Also Oliver (1993a) noted that service quality perceptions do not necessarily require experience with the service whereas satisfaction evaluation does require the service to be experienced, as is the case in our study.

Criteria for evaluating measurement scales

Reliability

Reliability is the "extent to which a measurement is free of variable errors" (Tull and Hawkins, 1987, p. 272). In particular we are looking for consistency among the scales used to evaluate customer satisfaction. Peter (1979) states that the most popular measure of reliability is Cronbach's coefficient alpha, which "should be routinely calculated to assess the quality of measure" (Churchill, 1991, p. 498). In this study Cronbach's alpha is used to assess the reliability of measurement scales with multi-point items relating to the components of a service process.

Validity

The validity of a measurement instrument is defined as "the extent to which difference in scores on it reflect true differences among individuals on the characteristic we seek to measure, rather than constant or random errors" (Sellitz *et al.*, 1976, p. 169). Specifically, we examine convergent, discriminant and predictive validity.

Convergent validity is the extent of agreement among scales measuring the same concept, for example, satisfaction with a hotel room. Conversely, discriminant validity looks at the disagreement among scales used to measure concepts that are not expected to be related, for instance, satisfaction with the hotel valet and the hotel room cleanliness. Finally, predictive validity (also known as criterion validity) examines the relationship of concepts where it is expected that some concepts have an impact on a subsequent behavioural measure. This is particularly relevant for customer satisfaction surveys where high satisfaction with the important components of a service process is expected to result in a higher tendency to reuse the service or recommend it to others (Rust and Zahorik, 1993; Rust *et al.*, 1994).

Skewness

Peterson and Wilson (1992) and Westbrook (1980b) noted that commonly used satisfaction and performance rating scales gave skewed distributions, with the majority of customers rating the service towards the higher end of the scale. The major concern with such a phenomenon is that a considerable proportion of these seemingly satisfied respondents may not, in fact, be very loyal. Therefore, although a firm may appear to rate highly in a customer satisfaction survey, customer loyalty need not be high. For instance, using the four-point scale excellent, good, fair or poor, Kordupleski *et al.* (1993) gave evidence which shows that customers who rate the service as good are only 50 per cent likely to reuse the service, whereas those who rate the service as excellent are 80 per cent likely to reuse the service. This has led to the AT&T motto "Good is not good enough". Hence, there is a need to use a scale which discriminates between customers properly to assess their true loyalty.

Relationship to open-ended responses (face validity)
Most customer satisfaction surveys allow respondents to make comments about their service experience. These comments can be categorized into negative, positive or both. In turn, the type of comment can be compared with the respondent's overall evaluation of the service received. It would be expected that those who made positive comments would rate the service highly and vice versa. This is a good test of a scale's face validity since some respondents may not feel they have been able to articulate their service perception adequately on a categorical scale, whereas they can communicate their feeling in detail in an open-ended question, especially if they have received poor service. Westbrook (1980b) also used this as a measure of scale performance.

Managerial value
Given that the purpose of customer satisfaction surveys is to assist management to improve their service standards, it is imperative that they receive consistent advice via the survey instrument. For instance, regression analysis is frequently used to determine the relative importance of attributes that comprise the service process (Danaher and Mattsson, 1994; Parsuraman *et al.*, 1988; Rust and Zahorik, 1993; Rust *et al.*, 1994). Management may decide they want to find the service attributes most strongly related to customer loyalty and good word of mouth. Here "likeliness to recommend" and "likeliness to return" ratings can be regressed on the ratings for the service attributes to find which ones are most important. It would be frustrating to obtain conflicting advice from these two separate regressions. Alternatively, overall satisfaction can be regressed on service attribute ratings. Again it would be comforting for management to find that the same attributes come out as being important as in the two previous regressions, otherwise it would be difficult to decide how best to proceed with a quality improvement programme.

Method

Questionnaire development
The setting for our study was a downtown hotel. In keeping with current customer satisfaction research, we designed the questionnaire to cover the key components of the hotel stay (Barsky and LaBagh, 1992; Cadotte and Turgeon, 1988; Danaher and Mattsson, 1994). The purpose of this questionnaire format is to recreate the entire service experience for a respondent by arranging the questions so that respondents are asked to rate the service components in the order they encounter each component. For hotels the main components are check-in, the room, the restaurant and check-out (Barsky and LaBagh, 1992; Danaher and Mattsson, 1994). Although this questionnaire format may suffer from primacy and recency effects, it has been shown to have high predictive validity on overall satisfaction measures in many service industries (Danaher and Gallagher, 1996; Danaher and Mattsson, 1994; Rust and Zahorik, 1993; Rust *et al.*, 1994).

We have seen that there are a large number of scales to choose from for customer satisfaction measurement, but they fall into three main categories of performance, disconfirmation and satisfaction. Therefore, we now need to select some scales which represent each of these categories. We planned to ask each respondent to rate every service component on each of the scales but anticipated that using too many scales might introduce a severe respondent fatigue effect. To overcome this difficulty and keep the questionnaire brief and simple, thereby helping to increase the response rate, we used only three scales in the questionnaire (one for each of the scale type categories). Having just one scale represent each scale type creates a confounding of scales and scale types. However, the alternative is to use two scales to represent each scale type, giving six scales in total. This would be far too many for a respondent to evaluate for each service attribute. As can be seen in the sequel, the actual scales used patently belong to one of three scale type categories. Hence, the confounding of scales and scale types is minimized.

For the performance measure we used the six-point scale excellent, very good, good, poor, very poor and terrible (adapted from Churchill and Surprenant, 1982). The disconfirmation measure was also adapted from Churchill and Surprenant (1982), being the three point scale "it was better than I thought it would be", "it was just as I expected" and "it was better than I thought it would be". Another popular disconfirmation scale, which we might equally have used, is the three-point "better than expected" to "worse than expected" scale (Rust *et al.*, 1994). For the satisfaction measure we could have chosen either the scale "very satisfied" to "very dissatisfied" or a graphical scale with happy and sad faces, for example. We chose to implement the graphical scale used by Churchill and Surprenant (1982) and Danaher and Mattsson (1994) both to give some variety to the questionnaire layout and because graphical scales permit many response points without cluttering up the scale with words. The particular graphical scale used was an 11-point scale anchored at three places, with a happy face at the first point, a neutral face at the sixth point and an unhappy face at the 11th point (see also Danaher and Mattsson, 1994). An excerpt from the questionnaire, showing the exact wording and layout, is given in the Appendix.

Although many other scales could have been chosen for this study we felt the three selected were distinct enough to convey to respondents that we were not asking the same question repeatedly and these scales clearly represent the three scale categories of performance, disconfirmation and satisfaction as studied by numerous researchers (Bearden and Teel, 1983; Bolton and Drew, 1991; Boulding *et al.*, 1993; Cronin and Taylor, 1992; Danaher and Mattsson, 1994; Parasuraman *et al.* 1988, 1991; Teas, 1993; Rust and Zahorik, 1993; Rust *et al.*, 1994).

Following the service attribute part of the questionnaire respondents were asked to rate their overall experience at the hotel on all three measurement scales. Hence, there were three measurements made of the check-in, room, restaurant, check-out and overall experience. After the overall evaluation

respondents were asked "How likely would you be to return to the hotel, if the opportunity arose?" and "How likely would you be to recommend the hotel to someone you know?" The response categories for these last two questions were, very likely, quite likely, not very likely and not at all likely. Demographic questions such as age, gender, purpose and length of stay were also asked at the end of the questionnaire. Finally, respondents were given ample space to make any further written comments. The entire questionnaire fitted on to a single A4 sheet and took only a few minutes to complete.

Scale and response order effects

In the customer satisfaction literature, the disconfirmation of expectations paradigm has been the focus of several studies (Cadotte *et al.* 1987; Oliver, 1977; 1980). There has been considerable debate on the interaction of the three concepts of performance, disconfirmation and satisfaction (Bearden and Teel 1983; Bolton and Drew, 1991; Boulding *et al.*, 1993; Cronin and Taylor, 1992; Parasuraman *et al.*, 1988, 1991). The emerging consensus appears to be that performance drives disconfirmation, which in turn drives satisfaction (Oliver, 1993a). We felt that this causality ordering may have an impact on the responses to our questionnaire, as we originally had our three scales presented in the order of performance, disconfirmation and satisfaction, which is precisely the order of theorized causality. So as not to confound the effects of known service quality relationships and scale presentation order we arranged for half of our questionnaires to have the three scales in the order performance, disconfirmation and satisfaction and the other half in reverse order to this.

In addition to scale presentation order effects, there may be an effect due to order of the response categories for a particular scale (Babakus and Boller, 1992; Sudman, 1982). That is, the excellent to terrible scale may produce different results if the scale begins with terrible and ends with excellent. We therefore interchanged the order of the response categories so that half of the time they went from favourable (such as excellent) to unfavourable (such as terrible) and the other half of the questionnaires had the response category order reversed. In each case all three scale types had the same order, as Parasuraman *et al.* (1991) reported that respondents seemed to make too many errors when response category orders were mixed up within the same questionnaire.

The combination of the alternative scale order and response category order formats created four versions of the questionnaire shown in Table I.

Version	Scale order	Response order
1	Performance, disconfirmation, satisfaction	Favourable to unfavourable
2	Performance, disconfirmation, satisfaction	Unfavourable to favourable
3	Satisfaction, disconfirmation, performance	Favourable to unfavourable
4	Satisfaction, disconfirmation, performance	Unfavourable to favourable

Table I.
Scale and response category orders

The sample

The hotel chosen for this study was one of five first-class hotels located in the central city region of a city with over one million people. It is a member of a large international hotel chain, with a reputation for good service. As the hotel was undergoing a major refurbishment, only the guests staying in the 75 rooms that had been recently remodelled were eligible for the survey. This ensured homogeneity of the rooms across respondents.

The survey period was the months of May and June, 1994. About 20 questionnaires were distributed to guests each day by front desk staff along with a covering letter signed by the general manager. This letter detailed the purpose of the research study and encouraged guest participation. A sealable envelope, addressed to the general manager and marked "confidential" was attached to the questionnaire with the intention that the questionnaire be handed back to front desk staff after check-out. Every third guest in the eligible rooms was asked to complete the survey. Guests from the eligible rooms were reminded about the questionnaire at check-out, although the front desk staff did not know if a particular guest had been selected for the survey.

Some 1,000 questionnaires were distributed with 185 being returned, but only 171 were usable. The number of questionnaires returned for versions 1-4 were 54, 23, 44 and 50 respectively. Although the response rate (17.1 per cent) is low by most survey standards, it does compare favourably with other hotel guest surveys with no incentives offered. For instance, Barsky and Huxley (1992) report a survey of hotel guests where the response rate was 7 per cent. As it happened, in Barsky and Huxley's study, three levels of incentive were used simultaneously, resulting in response rates of 100, 29 and 7 per cent respectively, for generous, moderate and no incentives. As would be expected, the respondents in the group receiving no incentive had less favourable opinions of the hotel than those in the group with 100 per cent response rate. This is probably because the small subset of people who responded in the without incentive group had some problem with their stay and they saw the questionnaire as an opportunity to articulate these problems. We would expect a similar effect in our survey, but it must be remembered that the primary purpose of our study was not to measure customer satisfaction levels, but rather to compare measurement scales. There is not much reason to expect people who had a problem with their stay to have relatively different evaluations of the three measurement scales from the pool of all eligible hotel guests. Naturally, guests with a grievance will probably rate the hotel lower on the satisfaction scale but they will probably be lower on all three scales than a guest who had no problems with their stay.

Results

In this section we present the results of our survey and compare the measurement scales, using the criteria outlined previously.

Scale and response order effects

Recall that we hypothesized that the order of presentation of the three scales may have an effect on the responses due to the known causal relationship among the performance, disconfirmation and satisfaction concepts. In addition, we hypothesized that the order of presentation of the scale responses may be important. To test these hypotheses we administered four versions of the questionnaire. An appropriate analysis for testing scale and response order effect is a two-way ANOVA, where the first factor is scale order, having two levels, performance through satisfaction and satisfaction through performance. The second factor is response order, also having two levels, favourable to unfavourable and unfavourable to favourable response level anchors.

We conducted two-way ANOVAs with these factors for each of the service attribute questions from check-in to check-out, as well as for the overall evaluation questions. These ANOVAs were also conducted for each of the three scale types for each service attribute, giving 15 ANOVAs in total. None of these ANOVAs showed either a significant scale or response order (or interaction) effect. Therefore, for the remainder of this analysis we combine all four questionnaire versions as if they were the same (with the appropriate recodings of course).

Reliability

We now examine the reliability of the scales. Table II gives the Cronbach alpha values for the three scales, for each of the service attribute components and for the overall evaluation. When all three scales are examined simultaneously, the alphas range from 0.64 to 0.71, being of moderate reliability. We also calculated Cronbach's alpha for the three scales compared two at a time. The alphas for the pairings performance/satisfaction and performance/disconfirmation are not very different from those when all three scales are compared. However, there is a substantial drop in reliability for the pairing disconfirmation/satisfaction, indicating that these two scales do not appear to have much in common (Nunnally, 1967, p. 210).

		Cronbach's alpha		
Service attribute	All three scales	Performance and satisfaction scales	Performance and disconfirmation scales	Disconfirmation and satisfaction scales
Check-in	0.67	0.70	0.65	0.40
Room	0.71	0.74	0.66	0.44
Restaurant	0.71	0.74	0.68	0.44
Check-out	0.64	0.66	0.64	0.36
Overall evaluation	0.70	0.71	0.72	0.43

Table II.
Cronbach's alpha values for various combinations of the scales

Convergent and discriminant validity

A simple method for looking at scale convergent and discriminant validity is to examine the correlations between scale ratings for the same service attribute, where correlations would be expected to be high, and across different service attributes, where correlations between intuitively independent attributes should be low. Table III gives the pairwise correlation matrix among all the variables used in the survey. Variables are grouped in the same order in which they appeared in the questionnaire. Figures in italics indicate the submatrices corresponding to the three scales used for each service attribute, and for the overall evaluation. Clearly, the correlations within these submatrices are all very high (being significant at the 0.01 per cent significance level or lower). Hence we can be confident of having convergent validity. However, it is worth noting that in every submatrix, the performance and satisfaction scales have higher pairwise correlation than for the disconfirmation scale with either the performance or satisfaction scales.

To assess discriminant validity we first determine which hotel attributes we would, a priori, not expect to be correlated. In the case of a hotel all of the attributes might be somewhat related, but, for example, it would not be expected that the check-in and restaurant processes will have necessarily correlated customer satisfaction ratings. On the other hand, the check-in and check-out processes would be expected to have similar ratings given that the staff, the type of encounter and location within the hotel are the same. In summary, we would expect independence of the check-in, room and restaurant attributes, while the check-out attribute should depend on check-in but not the room and restaurant.

Table III shows that if we compare the check-in and room submatrices, only four of the nine pairwise correlations are significant at the 1 per cent level (we chose 1 per cent rather than 5 per cent as the significance level due to the large number of hypothesis tests conducted, given the known pitfalls of multiple comparisons). It is important to note that these four significant correlations are between the performance and satisfaction scales. The non-significant correlations are those between the ratings on the disconfirmation scale and either of the other two scales. An identical pattern emerges when comparing check-in and restaurant attributes, the room and restaurant and restaurant and check-out attributes. That is, the performance and satisfaction scales show significant correlations where none would be expected, but the disconfirmation scale ratings show no significant overlap with any other scale ratings from an unrelated service process. The only exception to this is that room and check-out ratings show significant correlations among all the rating scales. Finally, as expected, the check-in and check-out ratings are significantly correlated for all three measurement scales.

In summary, only the disconfirmation scale appears to have discriminant validity, while all three scales have convergent validity.

	Chkinperf	Chkindisc	Chkinsat	Roomperf	Roomdisc	Roomsat	Restperf	Restdisc	Restsat	Chkoutperf	Chkoutdisc	Chkoutsat	Overperf	Overdisc	Oversat	Recommend	Return
Chkinperf	*1.000*	*0.565**	*0.774**	0.443*	0.189	0.392*	0.398*	0.124	0.351*	0.642*	0.306*	0.592*	0.667*	0.367*	0.571*	0.584*	0.549*
Chkindisc	*0.565**	*1.000*	*0.575**	0.182	0.169	0.153	0.163	0.172	0.127	0.382*	0.379*	0.357*	0.370*	0.415*	0.310*	0.371*	0.386*
Chkinsat	*0.774**	*0.575**	*1.000*	0.399*	0.141	0.457*	0.313*	0.052	0.373*	0.519*	0.306*	0.647*	0.559*	0.319*	0.645*	0.504*	0.491*
Roomperf	0.443*	0.182	0.399*	*1.000*	*0.553**	*0.819**	0.351*	0.103	0.377*	0.412*	0.337*	0.496*	0.641*	0.445*	0.637*	0.610*	0.579*
Roomdisc	0.189	0.169	0.141	*0.553**	*1.000*	*0.592**	0.153	0.127	0.233	0.265*	0.337*	0.301*	0.397*	0.532*	0.391*	0.413*	0.458*
Roomsat	0.392*	0.153	0.457*	*0.819**	*0.592**	*1.000*	0.287*	0.041	0.373*	0.451*	0.351*	0.575*	0.593*	0.447*	0.723*	0.584*	0.554*
Restperf	0.398*	0.163	0.313*	0.351*	0.153	0.287*	*1.000*	*0.582**	*0.814**	0.372*	0.143	0.408*	0.509*	0.337*	0.456*	0.465*	0.427*
Restdisc	0.124	0.172	0.052	0.103	0.127	0.041	*0.582**	*1.000*	*0.584**	0.153	0.179	0.258*	0.187*	0.333*	0.203	0.253*	0.195
Restsat	0.351*	0.127	0.373*	0.377*	0.233*	0.373*	*0.814**	*0.584**	*1.000*	0.299*	0.136	0.491*	0.439*	0.365*	0.519*	0.432*	0.393*
Chkoutperf	0.642*	0.382*	0.519*	0.412*	0.265*	0.451*	0.372*	0.153	0.299*	*1.000*	*0.588**	*0.717**	0.606*	0.388*	0.614*	0.566*	0.526*
Chkoutdisc	0.306*	0.379*	0.306*	0.337*	0.337*	0.351*	0.143	0.179	0.136	*0.588**	*1.000*	*0.587**	0.423*	0.502*	0.459*	0.472*	0.475*
Chkoutsat	0.592*	0.357*	0.647*	0.496*	0.301*	0.575*	0.408*	0.258*	0.491*	*0.717**	*0.587**	*1.000*	0.589*	0.398*	0.796*	0.581*	0.551*
Overperf	0.667*	0.370*	0.559*	0.641*	0.397*	0.593*	0.509*	0.187	0.439*	0.606*	0.423*	0.589*	*1.000*	*0.643*	*0.773*	0.743*	0.728*
Overdisc	0.367*	0.415*	0.319*	0.445*	0.532*	0.447*	0.337*	0.333*	0.365*	0.388*	0.502*	0.398*	*0.643*	*1.000*	*0.599*	0.621*	0.622*
Oversat	0.571*	0.310*	0.645*	0.637*	0.391*	0.723*	0.456*	0.203	0.519*	0.614*	0.459*	0.796*	*0.773*	*0.599*	*1.000*	0.733*	0.706*
Recommend	0.584*	0.371*	0.504*	0.610*	0.413*	0.584*	0.465*	0.253	0.432*	0.566*	0.472*	0.581*	0.743*	0.621*	0.733*	1.000	0.866*
Return	0.549*	0.386*	0.491*	0.579*	0.458*	0.554*	0.427*	0.195	0.393*	0.526*	0.475*	0.551*	0.728*	0.622*	0.706*	0.866*	1.000

Note: *significant at the 1 per cent level

Table III.
Correlation matrix of ratings for hotel attributes, overall evaluation and likeliness to recommend and return

Predictive validity

As outlined above, the particular format used for this questionnaire was intended to re-create the key components of the entire hotel stay. Therefore, we would expect that respondents' overall evaluation of their stay would be related to their evaluation of their experiences with the component service attributes. For instance, high satisfaction with the room should also result in high overall satisfaction with the hotel (Danaher and Mattsson, 1994; Rust *et al.*, 1994). To test for predictive validity we regressed the overall evaluation scores on the scores for the four service attributes, for each of the three scales. For example, the overall performance rating was regressed on the ratings for check-in to check-out using just the performance scale. The adjusted R^2s for these three regressions were 72.0, 59.6 and 82.2 per cent respectively for the performance, disconfirmation and satisfaction scales. The satisfaction scale, therefore, has the highest predictive validity, while the disconfirmation scale has the lowest. Even so, the 59.6 per cent adjusted R^2 for the disconfirmation scale compares favourably with adjusted R^2s from other customer satisfaction regression analyses (Danaher and Gallagher, 1996; Danaher and Mattsson, 1994; Rust *et al.*, 1994).

It may not be too surprising that performance and satisfaction scales have higher predictive validity than the disconfirmation scale, as Parasuraman *et al.* (1994, p. 228) note that "a plausible explanation for the consistently superior predictive power of the perceptions-only ratings is that in the regressions involving these ratings both the dependent and independent variables are perceptions-only ratings, in contrast to other regressions wherein the independent variables are disconfirmation ratings. In other words, the higher R^2 values could be an artifact of the common method variance".

We are also keen to examine customer repurchase intentions and word of mouth, as this has been shown to be related to market share (Danaher and Rust, 1996). We measured these effects with the likeliness to return and recommend questions (Woodside *et al.* 1989) considered these as proxies for customer loyalty). Table III gives the correlations between likeliness to return and recommend and the overall evaluation for the three scales. It can be seen that the correlations with the likeliness to recommend variable are about the same for the performance and satisfaction scales (0.743 and 0.733), but the disconfirmation correlation is lower at 0.621 (Churchill and Surprenant (1982) and Liljander and Strandvik (1992) also found that the correlation between the overall evaluation ratings and purchase probability was lowest for the disconfirmation scale compared with the performance and satisfaction scales). Similar figures are evident for the overall evaluation/likeliness to return correlations, where, once again, the disconfirmation scale is not as strongly correlated with this word of mouth measure. This is a potential weakness of the disconfirmation scale. However, it must be remembered that correlation is a measure of the linear relationship between two variables and the fact that the disconfirmation scale has only three points compared with the performance and satisfaction scale's six and 11 points respectively, may be affecting the

linear relationship between the disconfirmation and loyalty measures. We illustrate this effect in Table IV, where we have generated a hypothetical, but realistic, bivariate distribution between overall evaluation using the disconfirmation scale and likeliness to return. The hypothetical distribution is not very different from that observed in this study. Table IV comprises a three-point disconfirmation scale, and a five-point scale (see Devlin *et al.*, 1993). The two bivariate distributions have been arranged so that the marginal distributions for the likeliness to return variable are identical and so that the average rating on the disconfirmation scale is zero in both distributions. The correlation between likeliness to return and overall evaluation is 0.603 for the three-point scale but rises to 0.783 for the five-point scale. Hence, increasing the number of scale points in the disconfirmation scale, while maintaining the same likeliness to return profile, has increased the correlation between these two variables. Devlin *et al.* (1993) also comment on the loss of predictive power for scales with only a small number of points. Therefore, the lower correlation between likeliness to return and overall disconfirmation ratings, compared with the performance and satisfaction scales, is very likely to be due to the smaller number of points in the disconfirmation scale. A similar effect is apparent in the lower adjusted R^2s for the regression analyses mentioned earlier. It is evident that the five-point disconfirmation scale, as recommended by Devlin *et al.* (1993), should have better predictive validity than the three-point scale used in this study.

		Likeliness to return				
	Codes	Very unlikely 1	Unlikely 2	Likely 3	Very likely 4	Total
Three-point scale						
Worse than expected	–1	1/16	1/16	0	0	2/16
About as expected	0	0	4/16	6/16	2/16	12/16
Better than expected	1	0	0	1/16	1/16	2/16
Total		1/16	5/16	7/16	3/16	1
Five-point scale						
Much worse than expected	–2	1/16	0	0	0	1/16
Worse than expected	–1	0	2/16	0	0	2/16
About as expected	0	0	3/16	6/16	1/16	10/16
Better than expected	1	0	0	1/16	1/16	2/16
Much better than expected	2	0	0	0	1/16	1/16
Total		1/16	5/16	7/16	3/16	1

Note: The marginal distribution of the likeliness to return variable is the same for both distributions. The correlation between these two variables is 0.60 for the three-point scale and 0.78 for the five-point scale

Table IV.
Hypothetical bivariate distribution of the overall disconfirmation and likeliness to return variables for three-point and five-point disconfirmation scales

Skewness

Peterson and Wilson (1992) noted that almost all customer satisfaction ratings have skewed distributions, with respondents tending to select the favourable end of the scale. While this may be comforting for management, there is still uneasiness since apparently satisfied customers may not have very strong loyalty (Kordupleski *et al.*, 1993). Westbrook (1980b) also noted this effect and suggested calculating the coefficient of skewness, which measures symmetry around the mean. Table V gives the skewness coefficients for the three scales for each service attribute. A symmetric distribution will have a zero value for the skewness coefficient. Clearly, only the disconfirmation scale has skewness values near zero, while the performance and satisfaction scales are skewed to the left on all the attributes. The satisfaction scale is particularly badly skewed (bar graphs of these distributions confirm the symmetry of the disconfirmation scale and the asymmetry of the performance and satisfaction scales).

Table V.
Coefficients of skewness for the three scales for each attribute

Service attribute	Skewness coefficient		
	Performance scale	Disconfirmation scale	Satisfaction scale
Check-in	–0.75	–0.10	–0.95
Room	–0.95	0.01	–0.89
Restaurant	–0.49	0	–0.74
Check-out	–0.36	–0.39	–0.92
Overall evaluation	–0.51	–0.15	–0.79

Relationship to open-ended responses (face validity)

Recall that we gave respondents the opportunity to submit written comments on their questionnaire. Some 95 respondents took advantage of this opportunity. We coded these comments into positive, negative or both types, there being 17, 65 and 13 of these types respectively. It may seem disturbing that 65 of 171 (38 per cent) respondents gave negative comments, but remember that the response rate was quite low and we suspect that guests who had a problem with their stay tended to return their questionnaires. The negative comments ranged from "the bed was too firm" to "staff needed a few lessons in courtesy, particularly reception staff".

As the number of positive and both positive and negative commenting respondents is small we shall not analyse these two groups in detail. However, we can make some quantitative assessment of the 65 respondents who reported just negative comments in relation to their overall evaluation of the hotel. Table VI gives the distribution of high ratings for the three scales among those people who wrote only negative comments. It shows that over 30 per cent of these respondents still gave their overall evaluation as good or excellent on the performance scale and that over 35 per cent still rated their overall evaluation

as high as 9 to 11 on the 11-point satisfaction scale. In contrast, only 4.6 per cent of these seemingly unhappy guests rated their overall evaluation as better than expected on the disconfirmation scale. Given that we would expect people who make negative comments not to be very satisfied with their stay, it is surprising that the performance and satisfaction scales report such a high level of overall customer satisfaction. Only the disconfirmation scale seems to have captured the negative feelings of these guests. Of course, there is a wide range of severity of these negative comments and some of these negative experiences articulated in the open-ended comments may not be very important to guests in respect of their overall evaluation of the hotel. Nonetheless, the fact that these people went to the extra effort of writing a comment is indication enough that a negative experience might have had an impact on their overall evaluation.

	Scale		
	Performance	Disconfirmation	Satisfaction
Percentage of high overall ratings	30.8	4.6	35.4
Scale levels	Excellent/good	Better than expected	9-11 on 11-point scale
Percentage of low overall ratings	3.1	32.4	7.7
Scale levels	Poor/terrible	Worse than expected	1-3 on 11-point scale

Table VI.
Percentage of the 65 respondents making negative comments who rated the hotel high or low on the three scales

Table VI also examines the distribution of low ratings among those making negative comments. This shows precisely the reverse pattern of the high ratings. Surprisingly, of the guests making negative comments only a handful (3.1 per cent and 7.7 per cent respectively) give low overall ratings on the performance and satisfaction scales. On the other hand, 32.4 per cent of guests making negative comments rated the hotel as worse than expected overall. This is more in keeping with what we would anticipate from these seemingly unhappy guests.

Managerial value

One of the benefits that management see in customer satisfaction surveys is in providing information on how best to proceed with a quality improvement programme (Danaher and Gallagher, 1996; Kordupleski *et al.*, 1993; Rust *et al.*, 1994). As discussed above, a popular technique for providing this information is to regress the overall evaluation ratings on the ratings for the service attributes. The relative importance of the attributes can be assessed by one of several methods. One is to compare the magnitude of the regression coefficients or the standardized coefficients (betas). However, Bring (1994) has shown that the beta coefficients may not give a very reliable measure of the relative importance of regression independent variables, especially in the presence of multicollinearity (see also Danaher and Mattsson, 1994; Rust and Zahorik, 1993; Rust *et al.*, 1994). Instead, Bring (1994) recommends using the magnitude of

each independent variable's *t*-statistic as an indicator of relative importance, which is what we use in the sequel.

Table VII gives the regression models for 15 separate regression analyses. Here, for instance, the overall evaluation based on the performance scale is regressed first on the ratings for the four service attributes check-in to check-out using just the ratings on the performance scale for these attributes. The adjusted R^2 for this particular regression is 72 per cent, being reasonably high. The largest *t*-statistic corresponds to the check-in attribute, followed by the room, check-out and restaurant in decreasing order of importance. The rankings of just the significant attributes are given alongside the *t*-statistics. These rankings could be affected by multicollinearity, if such effects were present in the data. We tested for multicollinearity using the variance inflation criterion and found that the largest variance inflation factor across the 15 regressions in Table VII was only 2.3, being much smaller than the recommended cutoff of five (Hair *et al.*, 1995, p. 127). Therefore our results should not be affected by multicollinearity.

The next regression in Table VII (reading down column three) retains the overall performance rating as the dependent variable but now the independent variables are the ratings for the four attributes based on the disconfirmation scale. Here the decreasing order of importance for the service attributes is the room, check-out and check-in, with the restaurant not a significant attribute. When the attribute ratings using the satisfaction scale become the independent variables the rankings are similar to those of the disconfirmation scale except that the relative importance of check-in and check-out are interchanged.

Although the regression of overall performance scale against the performance scale attribute ratings produces the highest adjusted R^2, the face validity of the rankings of the attributes is questionable. Several previous studies in hotel settings have consistently shown the room to be the most important factor in determining overall customer satisfaction (Cadotte and Turgeon, 1988; Danaher and Mattsson, 1994), yet this regression analysis claims that check-in is the most important attribute. Given that the check-in experience is usually very brief and often hassle-free it is particularly difficult to justify its dominance over the room as the attribute which impacts mostly on the overall evaluation. The room is, after all, the primary purpose of a hotel stay. A similar anomaly occurs with the regression of overall evaluation using the satisfaction scale on the attributes rated with the satisfaction scale. Here the check-out attribute is, surprisingly, more important than the room.

We also regressed the likelihood of returning and recommending ratings on the service attributes, as management are usually keen to find factors which relate strongly to customer loyalty and word of mouth (Danaher and Rust, 1996; Rust *et al.*, 1994). All three scale types rank the room as the most important attribute, while the restaurant is never significant. There is less agreement on the relative importance of check-in and check-out, with the satisfaction scale claiming neither are important in predicting likeliness to recommend.

Scale type	Attribute	Dependent variable: Overall performance *t*-stat	Overall performance Rank	Overall disconfirmation *t*-stat	Overall disconfirmation Rank	Overall satisfaction *t*-stat	Overall satisfaction Rank	Likelihood of returning *t*-stat	Likelihood of returning Rank	Likely to recommend *t*-stat	Likely to recommend Rank
Performance	Check-in	5.70*	1	2.17*	2	3.85*	3	2.76*	2	2.98*	2
	Room	4.08*	2	3.58*	1	5.59*	1	3.62*	1	4.10*	1
	Restaurant	2.91*	4	0.40	–	2.24*	4	1.76	–	1.91	-
	Check-out	3.24*	3	1.86	–	3.88*	2	2.18*	3	2.00*	3
	Adjusted R^2	72.0 %		40.3%		70.7%		50.5%		53.6%	
Disconfirmation	Check-in	2.71*	3	4.25*	2	2.05*	3	2.90*	3	2.34*	3
	Room	3.72*	1	6.62*	1	4.74*	1	4.98*	1	5.01*	1
	Restaurant	0.77	–	1.00	–	1.10	–	0.30	–	1.19	-
	Check-out	2.91*	2	4.11*	3	3.59*	2	2.91*	2	2.68*	2
	Adjusted R^2	35.2%		59.6%		42.1%		42.3%		40.6%	
Satisfaction	Check-in	3.26*	2	1.01	–	4.17*	3	2.06*	3	1.85	-
	Room	3.87*	1	4.96*	1	5.74*	2	3.66*	1	4.36*	1
	Restaurant	0.89	–	0.32	–	1.40	–	0.43	–	0.96	
	Check-out	2.73*	3	1.43	–	6.73*	1	2.36*	2	1.64	-
	Adjusted R^2	60.1%		45.8%		82.2%		49.7%		49.9%	

Notes: Rankings of the relative importance of each of these attributes on the basis of these *t*-statistics are also given (when the *t*-statistics are significant). The sample size for each regression is 171

*Significant at the 5 per cent level

Table VII.
T-statistics for regressions of the overall evaluation and the likeliness to return and recommend ratings on process attributes for all three scale types

To assess the consistency of the attribute importance rankings across the five regression dependent variables for the different scales we now read horizontally across Table VII. For the performance scale, check-in ranks between one and three, the room ranks first in four of five cases, the restaurant is significant in two of five cases, while check-out ranks two or three and is not significant on one occasion. Similar inconsistencies arise for the satisfaction scale. In contrast, there is remarkable consistency for the disconfirmation scale, with the room being the most important attribute in all five cases, followed by check-out then check-in for four out of five cases. The restaurant is never significant when the independent regression variables are measured with the disconfirmation scale. There is only one instance when the rankings of the check-in and check-out are not three and two respectively, and here the *t*-statistics are very close, being 4.25 and 4.11.

One final observation from Table VII concerns the adjusted R^2 values. Notice that the highest adjusted R^2 value, when overall evaluation is measured on the performance scale, corresponds to the situation where the attribute ratings are also measured using the performance scale (adjusted R^2 =72 per cent). Similarly when the overall evaluation is based on the disconfirmation scale, the highest adjusted R^2 (59.6 per cent) occurs when the attributes are also measured on the disconfirmation scale. Again, this effect also occurs for the satisfaction scale. Part of the reason for this effect may be what Parasuraman *et al.* (1994) refer to as "common method variance", where the strongest associations occur when both service attributes and overall evaluation are measured using the same method, be it a performance, disconfirmation or satisfaction scale.

Summary and Conclusion

In this paper we have contrasted performance, disconfirmation and satisfaction scales on six criteria to determine their relative merits for measuring customer satisfaction. These criteria were reliability, convergent and discriminant validity, predictive validity, skewness, face validity and managerial value. In all but one of these criteria, namely, predictive validity, the disconfirmation scale was superior to both the performance and satisfaction scales. We were able to attribute the inferior predictive validity of the disconfirmation scale to the small number of points (three) and show that a five-point disconfirmation scale should have higher predictive validity.

Our results agree with the assertions by Devlin *et al.* (1993) and Rust *et al.* (1994) that the disconfirmation scale is a preferred method for measuring customer satisfaction (although none of these researchers compared several scales simultaneously on the same people, as we have done). In particular, we agree with Devlin *et al.* (1993) that a five-point disconfirmation scale would be an improvement over the three-point scale if high predictive validity is essential. One drawback of the five-point disconfirmation scale (as noted by Devlin *et al.* (1993) could be its use in telephone surveys, where respondents might have to be continually reminded of five rather than three scale points, thereby

increasing survey length and difficulty. In contrast, the three-point disconfirmation scale is easy to remember. Further empirical work replicating our study, but with a five-point disconfirmation scale is required.

As Peterson and Wilson (1992) and Yi (1990) noted, performance and satisfaction scales are almost always skewed, with the majority of respondents rating the perceived service highly. The frustrating aspect of this phenomenon is that apparently satisfied customers do not necessarily remain loyal. We also observed very skewed performance and satisfaction rating distributions, but not so for the disconfirmation scale. In fact, the average overall evaluation scores for the performance, disconfirmation and satisfaction scales were 3.5, 1.9 and 7.0 respectively. The respective scale midpoints are 3.5, 2 and 6. Given that 38 per cent of the questionnaire respondents wrote negative comments about the hotel service we would expect a below average overall evaluation, but only the disconfirmation scale appears to capture it, as does the performance scale to a lesser extent.

Having determined that the disconfirmation scale appears to be better than both the performance and satisfaction scales, it is now interesting to see if there is much difference between the performance and satisfaction scales. First, we saw earlier that there is moderately high reliability (using Cronbach's alpha) between the disconfirmation and performance scale, but not so for the disconfirmation and satisfaction scales. This indicates that the satisfaction scale is measuring quite a different construct to the disconfirmation scale. Second, neither the performance nor the satisfaction scales showed discriminant validity, although both did have high convergent validity. Third, both scales showed high predictive validity. Fourth, the satisfaction scale was much more skewed than the performance scale. Fifth, both scales showed poor face validity among the subset of respondents who wrote negative comments, with these respondents tending to avoid the lower extremes of the performance and satisfaction scales. Finally, for managerial value both the performance and satisfaction scales gave inconsistent service attribute importance rankings across a range of regression independent variables. In summary, when deciding between the performance and satisfaction scales, the performance scale is better on two of the criteria, with these two scales performing similarly for the other four criteria.

A limitation of our study is that our survey had a low response rate and probably over-represented slightly unhappy customers. Even though we do not expect this to have an impact on the results, given that we were primarily interested in comparing scales, it would be interesting to replicate our results for a sample which is more representative of a service firm's customers.

In this article we have contrasted three commonly used scale types for measuring customer satisfaction. Although the disconfirmation scale emerged as the best scale in our study, it remains to test our findings in other service settings. In addition, other disconfirmation scales with different wordings or more scale points should also be tested.

References

Andreasen, A.R. and Best, A. (1977), "Consumers complain – does business respond?", *Harvard Business Review*, Vol. 55, July-August, pp. 93-101.

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

Barsky, J.D. and Huxley, S.J. (1992), "A customer survey tool: using the quality sample", *The Cornell Hotel and Restaurant Association Quarterly*, December, pp. 18-25.

Barsky, J.D. and LaBagh, R. (1992), "A strategy for customer satisfaction", *The Cornell Hotel and Restaurant Association Quarterly*, October, pp. 32-40.

Bearden, W.O. and Teel, J.E. (1983), "Selected determinants of consumer satisfaction and complaint reports", *Journal of Marketing Research*, Vol. XX, February, pp. 21-8.

Bolton, R.N. and Drew, J.H. (1991), "A multistage model of customers' assessments of service quality and value", *Journal of Consumer Research,* Vol. 17, March, pp. 375-85.

Boulding, W., Kalra, A., Staelin, R. and Zeithaml, V.A. (1993), "A dynamic process model of service quality: from expectations to intentions", *Journal of Marketing Research*, Vol. 30 February, pp. 71-84.

Bring, J. (1994), "How to standardize regression coefficients", *American Statistician*, Vol. 48 No. 3, August, pp. 209-13.

Cadotte, E.R. and Turgeon, N. (1988), "Key factors in guest satisfaction", *The Cornell Hotel and Restaurant Association Quarterly*, February, pp. 41-51.

Cadotte, E.R., Woodruff, R.B. and Jenkins R.L. (1987), "Expectations and norms in models of consumer satisfaction", *Journal of Marketing Research*, Vol. XXIV, August, pp. 305-14.

Churchill, G.A. Jr (1991), *Marketing Research: Methodological Foundations,* 5th ed., The Dryden Press, Hinsdale, IL; Orlando, FL; Chicago, IL.

Churchill, G.A. Jr and Surprenant, C. (1982), "An investigation into the determinants of customer satisfaction", *Journal of Marketing Research*, Vol. XIX, November, pp. 491-504.

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

Danaher, P.J. and Gallagher, R. (1996), "Measuring customer satisfaction in Telecom New Zealand", *European Journal of Marketing* (forthcoming).

Danaher, P.J. and Mattsson, J. (1994), "Customer satisfaction during the service delivery process", *European Journal of Marketing,* Vol. 28 No. 5, pp. 5-16.

Danaher, P.J. and Rust, R.T. (1996), "Indirect marketing benefits from service quality", *Quality Management Journal,* Vol. 3 No. 2, pp. 63-88.

Devlin, S.J., Dong, H.K. and Brown, M. (1993), "Selecting a scale for measuring quality", *Marketing Research: A Magazine of Management and Applications*, Vol. 5 No. 3, pp. 12-17.

Haddrell, V.E. (1994), "A comparison of question scales used for customer satisfaction measurement", unpublished Masters thesis, University of Auckland.

Hair, J.F., Anderson, R.E., Tatham, R.L. and Black, W.C. (1995), *Multivariate Data Analysis With Readings*, 4th ed., Prentice-Hall, Englewood Cliffs, NJ.

Kordupleski, R., Rust, R.T. and Zahorik, A.J. (1993), "Why improving quality doesn't improve quality (or whatever happened to marketing?)", *California Management Review*, Vol. 35, Spring, pp. 82-95.

Liljander, V. and Strandvik, T. (1992), "The relation between service quality, satisfaction and intentions", Unpublished Working Paper 43, Swedish School of Economics and Business Administration, Helsingfors, Finland.

Nunnally, J.C. (1967), *Psychometric Theory*, McGraw-Hill, New York, NY.

Oliver, R.L. (1977), "Effects of expectation and disconfirmation on postexposure product evaluations: an alternative interpretation", *Journal of Applied Psychology*, Vol. 62 No. 4, pp. 480-86.

Oliver, R.L. (1980), "A cognitive model of the antecedents and consequences of satisfaction decisions", *Journal of Marketing Research*, Vol. 17, September, pp. 46-9.

Oliver, R.L. (1981), "Measurement and evaluation of satisfaction processes in retailing setting", *Journal of Retailing*, Vol. 57, Fall, pp. 25-48.

Oliver, R.L. (1993a), "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.

Oliver, R.L. (1993b), "Cognitive, affective and attribute bases of the satisfaction response,", *Journal of Research*, Vol. 20 No. 2, pp. 418-30.

Olshavsky, R. and Miller, J. (1972), "Consumer expectations, product performance and perceived product quality", *Journal of Marketing Research*, Vol. 9, February, pp. 19-21.

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, Spring, pp. 12-40.

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

Parasuraman, A., Zeithaml, V. and Berry, L. (1994), "Alternative scales for measuring service quality: a comparative assessment based on psychomatric and diagnostic criteria", *Journal of Retailing*, Vol. 70 No. 3, Fall, pp. 201-30.

Peter, J.P. (1979), "Reliability: a review of psychometric basics and recent marketing practices", *Journal of Marketing Research*, Vol. 16, February, pp. 6-17.

Peterson, R.A. and Wilson, W.R. (1992), "Measuring customer satisfaction: fact and artifact", *Journal of the Academy of Marketing Science,* Vol. 20, Winter, pp. 61-71.

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.

Rust, R.T. and Zahorik, A.J. and Keiningham, T.L. (1994), *Return on Quality*, Probus Publishing, Chicago, IL.

Sellitz, C., Wrightsman L.L. and Cook, S.W. (1976), *Research Methods in Social Relations,* 3rd ed., Holt, Rinehart and Winston, New York, NY.

Sudman, S. (1982), *Asking Questions*, Jossey-Bass, San Francisco, CA.

Taylor, S.A. (1995), "Service quality and consumer attitudes: reconciling theory and measurement", in Swartz, T.A., Bowen, D.E. and Brown, S.W. (Eds), *Advances in Services Marketing and Management,* Vol. 4, JAI Press, Greenwich, CT.

Taylor, S.A. and Baker, T.L. (1994), "An assessment of the relationship between service quality and customer satisfaction in the formation of consumers' purchase intentions", *Journal of Retailing*, Vol. 70 No. 2, Summer, pp. 163-78.

Teas, R.K. (1993), "Expectations, performance evaluation and consumer's perceptions of quality", *Journal of Marketing*, Vol. 57, October, pp. 18-34.

Tse, D.K. and Wilton, P.C. (1988), "Models of consumer satisfaction formation: an extension", *Journal of Marketing Research*, Vol. 25, May, pp. 204-12.

Tull, D.S. and Hawkins, D.I. (1987), *Marketing Research: Measurement and Method*, 3rd ed., Macmillan, New York, NY.

Westbrook, R.A. (1980a), "Intrapersonal affective influences on customer satisfaction with products", *Journal of Consumer Research*, Vol. 7, June, pp. 49-54.

Westbrook, R.A. (1980b), "A rating scale for measuring product/service satisfaction", *Journal of Marketing*, Vol. 44, Fall, pp. 68-72.

Westbrook, R.A. and Oliver, R.L. (1981), "Developing better measures of customer satisfaction: some preliminary results", in Monroe, K.B. (Ed.), *Advances in Consumer Research,* Association for Consumer Research, Ann Arbor, MI, pp. 94-9.

Woodside, A.G., Frey, L.L. and Daly, R.T. (1989), "Linking service quality, customer satisfaction and behavioral intention", *Journal of Health Care Marketing,* Vol. 9 No. 4, December, pp. 5-17.

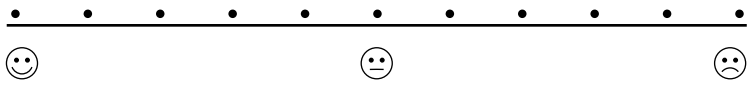
Yi, Y. (1990), "A critical review of customer satisfaction", in Zeithaml, V.A. (Ed.), *Review of Marketing*, 1990, American Marketing Association, Duke University.

Appendix: Excerpt from the questionnaire

When thinking about your hotel room:

a. Your overall satisfaction level with regard to your room was:
(Please indicate on scale below)

b. Your expectation of your room was:



Too low (it was better than I thought it would be)	Accurate (it was just as I expected)	Too High (It was worse than I thought it would be)

c. How would you rate the overall quality of your room?

Excellent	Very Good	Good	Poor	Very Poor	Terrible