Volume 67 Number 4 Winter 1991

Refinement and Reassessment of the SERVQUAL Scale

A. PARASURAMAN
Federated Professor of Marketing
Texas A & M University
College Station, Texas

LEONARD L. BERRY
JCPenney Chair of Retailing Studies
Texas A & M University
College Station, Texas

VALARIE A. ZEITHAML
Associate Professor
Duke University
Durham, North Carolina

In a previous article we presented SERVQUAL, a multiple-item scale for measuring service quality (Parasuraman, Zeithaml, and Berry 1988). In the present article, we discuss findings from a follow-up study in which we refined SERVQUAL and replicated it in five different customer samples. We also compare our findings with those of other researchers who have recently employed and evaluated SERVQUAL. On the basis of insights from this comparative discussion, we offer directions for future SERVQUAL research and applications.

In an article in the Spring 1988 issue of the *Journal of Retailing*, we discussed the development, testing, and potential applications of SERVQUAL, an instrument for measuring customer perceptions of service quality (Parasuraman, Zeithaml, and Berry—hereafter, PZB—1988). To our knowledge, many proprietary service quality studies have productively used SERVQUAL. A number of published studies have also invoked the SERVQUAL framework (Crompton and Mackay 1989; Webster 1989; Woodside, Frey, and Daly 1989; Johnson, Dotson, and Dunlap 1988) and have assessed the scale's reliability and validity (Babakus and Boller 1991; Brensinger and Lambert 1990; Carman 1990; Finn and Lamb 1991).

The authors gratefully acknowledge the financial support and cooperation provided for this study by the Marketing Science Institute and five of its corporate sponsors.

Copyright © 2001 All Rights Reserved

The objectives of this paper are to: (i) describe a multi-sector study in which we refined our original SERVQUAL instrument and reexamined its reliability and validity; (ii) provide a comparative discussion of insights from our study and those from other SERVQUAL-replication studies; and (iii) offer directions for future research and SERVQUAL use.

In the following sections we first describe the study in which we replicated SERVQUAL. We then discuss refinements made to SERVQUAL and our rationale for them. Next, we examine the reliability, factor structure, and validity of the revised instrument. Finally, we discuss SERVQUAL-related issues emerging from the various replications and offer suggestions for future applications of the instrument.

THE STUDY

The findings discussed in this article are from a study in which customer assessments of service quality were measured for three types of services: telephone repair, retail banking, and insurance. With the exception of retail banking, these services are different from the services originally studied to develop SERVQUAL (PZB 1988). In the present study, SERVQUAL was used to measure the service quality of five nationally-known companies—one telephone company, two insurance companies and two banks. Following a pretest and refinement, to be discussed in the next section, a questionnaire containing the modified instrument was mailed to about 1,800 to 1,900 randomly chosen customers of each company. A reminder post card was sent two weeks after the questionnaire was mailed. The number of complete questionnaires returned ranged from 290 to 487 across the five companies (response rates ranged from 17 to 25 percent). The aggregate sample contained 1,936 customers for a combined response rate of 21 percent. Demographic profiles of the five customer samples were reviewed by managers in the respective companies and considered to be representative of their customer bases.

REFINEMENT OF SERVQUAL

The original version of SERVQUAL (PZB 1988) consists of two sections: a 22-item section to measure customers' service expectations of companies within a specific sector (e.g., banking) and a corresponding 22-item section to measure customers' perceptions of a particular company in that sector. This original version of SERVQUAL was pretested through a mail survey of a regional sample of 300 customers of the telephone

Copyright © 2001 All Rights Reserved

company that participated in our study. Completed questionnaires were returned by 68 customers for a response rate of 23 percent.

Using the pretest data, means and standard deviations for the individual SERVQUAL items were computed, as were reliability coefficients (coefficient alphas) for the perception-minus-expectation gap scores making up the five SERVQUAL dimensions of tangibles, reliability, responsiveness, assurance, and empathy. Examination of these pretest results suggested several refinements as discussed next.

The means for a majority of the 22 items in the expectations section of SERVQUAL were above a "6" on the 7-point, "Strongly Disagree" to "Strongly Agree" scale (the overall mean expectation score was 6.22). These high mean values were not totally unanticipated because the items were intended to measure customers' *normative* expectations. For example, one expectation statement from the original instrument read: "Telephone companies should keep their records accurately." Recognizing that the "should" terminology in this and other expectation statements might be contributing to unrealistically high expectation scores, a slightly different wording was adopted. The revised wording focused on what customers would expect from companies delivering excellent service. To illustrate, the aforementioned statement was modified to read: "Excellent telephone companies will insist on error-free records." The instructions pertaining to the expectations items were also appropriately changed. The perceptions section of SERVQUAL remained unchanged except for minor wording changes in several items to make them parallel to the revised expectations items.

Of the 22 SERVQUAL items included in the pretest questionnaire, sixteen were worded positively and six were worded negatively. The pretest results indicated that the negatively worded items may be problematic for several reasons. First, the standard deviations for the negatively worded expectations items were consistently higher (mean standard deviation of 2.07) than for the positively worded expectations items (mean standard deviation of 0.77). The wider variation for the negatively worded items implied that respondents may have been confused by those items. Second, managers in the five companies who reviewed the preliminary questionnaire felt that negatively worded expectations statements were awkward and not as meaningful as the positively worded items. Third, the reliability coefficients (Cronbach's alphas), reported in the second column of Table 1, were consistently lower than in the original SERVQUAL study (Table 2, PZB 1988) for responsiveness and empathy—the two dimensions that included all the negatively worded items. For these reasons, negatively

Copyright © 2001 All Rights Reserved

TABLE 1

Reliability Coefficients (Alphas) for the SERVQUAL Dimensions

			Final Study Results				
Dimension	No. of Items	Pretest Results	Tel. Co.	Ins. Co. 1	Ins. Co. 2	Bank 1	Bank 2
Tangibles	4	.60	.83	.80	.84	.85	.86
Reliability	5	.85	.88	.92	.92	.92	.88
Responsiveness	4	.61	.91	.92	.93	.92	.88
Assurance	4	.81	.89	.87	.91	.90	.87
Empathy	5	.66	.87	.85	.89	.88	.87

worded items were all changed to a positive format in the final questionnaire.

Two new items—one each under tangibles and assurance—were substituted for two original items to more fully capture the dimensions and to incorporate suggestions from managers who reviewed the pretest questionnaire. An original tangibles item, "The appearance of the physical facilities of telephone companies should be in keeping with the type of services provided" (an item that was felt to be confusing and also redundant with another item dealing with appearance of physical facilities), was replaced with, "Materials associated with the service (such as pamphlets or statements) will be visually appealing in an excellent telephone company." The original SERVQUAL instrument had no item pertaining to the appearance of communication materials, an omission that managers believed should be corrected by adding the new item. Also, an original assurance item, "Telephone company employees should get adequate support from their companies to do their jobs well," was replaced with, "Employees in excellent telephone companies will have the knowledge to answer customer questions," because customers could more meaningfully evaluate knowledge of employees than behind-the-scenes support. The original SERVQUAL instrument had no item focusing directly on employee knowledge. Apart from the above changes, minor wording adjustments were made in a few items to incorporate feedback from company managers and to clarify the items (e.g., "up-to-date equipment" was changed to "modern-looking equipment" in an item under tangibles). The expectation and perception sections of the modified SERVQUAL instrument are shown in the appendix.

Copyright © 2001 All Rights Reserved

In our original study, estimates of the importance of the five SERVQUAL dimensions were derived indirectly in the form of regression weights (Table 6, PZB 1988). However, direct measures of the importance of various service attributes are also useful, particularly for combining individual attribute ratings to obtain a composite, weighted estimate of overall service quality. For this reason, we ascertained the relative importance of the five dimensions in this study by asking customers to allocate a total of 100 points across the dimensions according to how important they considered each to be. This point-allocation question listed descriptive definitions of the five dimensions without naming them. The exact wording of this question appears in the appendix.

REASSESSMENT OF SERVQUAL'S PSYCHOMETRIC PROPERTIES

We evaluated the reliability, factor structure, and validity of the refined SERVQUAL scale on the same criteria as in our original study (PZB 1988) and through certain additional analyses. For ease of comparison, the following discussion of our findings is presented in a format similar to that in PZB (1988).

Reliability and Factor Structure

As shown by Table 1, the reliability coefficients for the perception-minus-expectation gap scores for the five SERVQUAL dimensions are consistently high across the various samples, thereby indicating high internal consistency among items within each dimension. Moreover, *every* alpha value obtained for each dimension in the final study is higher than the corresponding values in the pretest and the original study (Table 2, PZB 1988). Thus, the refinements made to SERVQUAL seem to have improved the cohesiveness of the items under each dimension.

To verify the dimensionality of the 22 items in the revised SERVQUAL, perception-minus-expectation gap scores for these items were factor analyzed. Because SERVQUAL was hypothesized to have a five-dimensional structure, the initial analysis was constrained *a priori* to five factors. As in the original study (PZB 1988), the five-factor solution was subjected to oblique rotation. The rotated factor loading matrices for the five individual samples and the combined sample are shown in Table 2. The format of this table, including the labeling and ordering of the gap-score items (i.e., the Qs), is identical to that of Table 3 in PZB (1988). The item numbers in Table 2 correspond to those of the expectations and perceptions statements in the appendix.

Copyright © 2001 All Rights Reserved

The general pattern of loadings in Table 2 is similar across samples, suggesting that the factor structure of the gap scores generated by the revised SERVQUAL scale is fairly stable. However, this factor structure differs from the one obtained in the original study (Table 3, PZB 1988) in two ways. First, the four items under tangibles consistently break into *two* factors, with Q1 and Q2 (pertaining to equipment and physical facilities) forming one factor, and Q3 and Q4 (pertaining to employees and communication materials) forming another factor. Second, the responsiveness and assurance dimensions show considerable overlap and load on the same factor. The greater dimensional overlap in the present study was also evidenced by somewhat higher intercorrelations among the five factors extracted: the average pairwise correlations between factors following oblique rotation ranged from .35 to .39 for the factor-loading matrices in Table 2, compared to a range of .21 to .26 for the factor-loading matrices in Table 3 of PZB (1988).[1]

To further explore these differences in factor structures, the expectations ratings (Es) and perceptions ratings (Ps) were factor analyzed separately. As in the factor analyses of the gap scores (Qs), five factors were extracted and subjected to oblique rotation. The resulting factor loading matrices for the combined sample are shown in Table 3 (factor analyses of E and P ratings from the individual company samples consistently yielded factor-loading patterns similar to those in Table 3).

The factor-loading matrix for the expectations items in Table 3 suggests a four-factor structure, with relatively low loadings for all items on the fifth factor (F5). Interestingly, unlike in the factor pattern for the Q scores (Table 2), the four items under tangibles (E1–E4) do not split into two subdimensions. The structure of customers' service expectations along various facets of tangibles seems to be homogeneous. Therefore, the apparent reason for the two-factor split of the Q scores for tangibles is that customers *perceive and evaluate* the appearance of physical facilities/equipment (items 1 and 2) somewhat differently than they do the appear-

[1] A plausible explanation for the difference in dimensional distinctiveness between the original and revised SERVQUAL scales is the conversion of negatively worded items to a positive format. All the negatively worded items were under responsiveness and empathy and therefore could have contributed to the uniqueness of the factors representing them in the original study. However, if rewording of negative items was the sole or even principal reason for the greater dimensional overlap in the revised scale, the empathy items (which, like the responsiveness items, were all negatively worded in the original scale) should also have high factor loadings on the common responsiveness-assurance factor (F4). But this is not the case as the factor-loading patterns in Table 2 show. The empathy items by and large form a distinct factor (F5).

Copyright © 2001 All Rights Reserved

TABLE 2

Factor Loading Matrices Following Oblique Rotation of Five-Factor Solutions for Q Scores[a]

	Factor Loadings														
	Telephone Co.					Ins. Co. 1					Ins. Co. 2				
Items	F1	F2	F3	F4	F5	F1	F2	F3	F4	F5	F1	F2	F3	F4	F5
Tangibles															
Q1	84	—	—	—	—	91	—	—	—	—	90	—	—	—	—
Q2	71	—	—	—	—	78	—	—	—	—	79	—	—	—	—
Q3	—	90	—	—	—	—	86	—	—	—	—	78	—	—	—
Q4	—	85	—	—	—	—	76	—	—	—	—	91	—	—	—
Reliability															
Q5	—	—	77	—	—	—	—	87	—	—	—	—	88	—	—
Q6	—	—	32	33	—	—	—	40	34	—	—	—	64	—	—
Q7	—	—	49	—	—	—	—	78	—	—	—	—	75	—	—
Q8	—	—	96	—	—	—	—	95	—	—	—	—	93	—	—
Q9	—	—	—	33	—	—	—	56	—	—	—	—	36	42	—
Responsiveness															
Q10	—	—	72	—	—	—	—	31	35	—	—	—	52	—	—
Q11	—	—	57	—	—	—	—	48	40	—	—	—	40	—	33
Q12	—	—	—	54	36	—	—	—	58	—	—	—	—	47	—
Q13	—	—	—	42	27	—	—	—	53	—	—	—	—	72	—
Assurance															
Q14	—	—	—	59	—	—	—	—	58	—	—	—	—	76	—
Q15	—	—	—	67	—	—	—	—	62	—	—	—	—	84	—
Q16	—	—	—	90	—	—	—	—	81	—	—	—	—	68	—
Q17	—	—	—	61	—	—	—	—	64	—	—	—	—	76	—
Empathy															
Q18	—	—	—	—	58	—	—	—	—	70	—	—	—	—	47
Q19	—	—	—	—	49	—	—	—	—	53	—	—	—	—	32
Q20	—	—	—	—	87	—	—	—	—	81	—	—	—	—	88
Q21	—	—	32	—	34	—	—	—	43	34	—	—	—	39	28
Q22	—	—	27	—	29	—	—	—	49	—	—	—	—	30	36

Copyright © 2001 All Rights Reserved

TABLE 2 Cont'd

	Factor Loadings														
	Bank 1					Bank 2					All Cos.				
Items	F1	F2	F3	F4	F5	F1	F2	F3	F4	F5	F1	F2	F3	F4	F5
Tangibles															
Q1	99	—	—	—	—	85	—	—	—	—	91	—	—	—	—
Q2	52	29	—	—	—	79	—	—	—	—	73	—	—	—	—
Q3	—	81	—	—	—	—	85	—	—	—	—	83	—	—	—
Q4	—	88	—	—	—	—	79	—	—	—	—	87	—	—	—
Reliability															
Q5	—	—	76	—	—	—	—	79	—	—	—	—	84	—	31
Q6	30	—	46	—	—	—	—	27	—	59	—	—	39	—	—
Q7	—	—	92	—	—	—	—	64	—	—	—	—	72	—	—
Q8	—	—	85	—	—	—	—	83	—	—	—	—	90	—	—
Q9	—	—	67	—	—	—	—	57	—	—	—	—	52	—	—
Responsiveness															
Q10	—	—	—	50	—	—	—	34	40	—	—	—	36	36	—
Q11	—	—	—	87	—	—	—	—	68	—	—	—	—	58	—
Q12	—	—	—	97	—	—	—	—	74	—	—	—	—	80	—
Q13	—	—	—	82	—	—	—	—	83	—	—	—	—	81	—
Assurance															
Q14	—	—	—	56	—	—	—	—	76	—	—	—	—	79	—
Q15	—	—	44	—	30	—	—	37	48	—	—	—	—	63	—
Q16	—	—	—	60	—	—	—	—	62	—	—	—	—	84	—
Q17	—	—	—	61	—	—	—	—	61	—	—	—	—	72	—
Empathy															
Q18	—	—	—	—	79	—	—	—	—	84	—	—	—	—	78
Q19	—	—	—	—	39	—	—	—	—	35	—	—	—	—	54
Q20	—	—	—	—	89	—	—	—	—	73	—	—	—	—	91
Q21	—	—	—	—	69	—	—	—	—	74	—	—	—	—	58
Q22	—	—	28	—	51	—	—	—	—	57	—	—	—	—	49

[a] All numbers in the table are magnitudes of factor loadings multiplied by 100. Loadings of .25 or less are not shown. The percentage of variance explained by the five factors in the Tel. Co., Ins. Co. 1, Ins. Co. 2, Bank 1, Bank 2, and the combined sample were 67.2%, 68.3%, 70.9%, 71.6%, 66.9%, and 67.9%, respectively.

Copyright © 2001 All Rights Reserved

TABLE 3

Factor Loading Matrices Following Oblique Rotation of Five-Factor Solutions for E and P Scores[a]

Items	Factor Loadings F1	F2	F3	F4	F5	Items	Factor Loadings F1	F2	F3	F4	F5
Tangibles						**Tangibles**					
E1	79	05	03	01	14	P1	86	01	02	04	00
E2	94	03	09	00	05	P2	86	02	02	06	04
E3	63	05	16	03	14	P3	24	01	23	02	47
E4	74	01	05	00	11	P4	34	16	04	00	42
Reliability						**Reliability**					
E5	02	83	05	02	03	P5	01	88	05	07	03
E6	00	77	03	06	09	P6	02	34	16	46	07
E7	02	50	20	11	25	P7	07	72	15	03	09
E8	03	80	09	00	09	P8	01	93	03	01	04
E9	04	28	27	15	21	P9	19	51	17	05	06
Responsiveness						**Responsivenes**					
E10	05	02	59	13	12	P10	04	55	10	17	26
E11	04	13	81	03	01	P11	09	36	24	31	20
E12	03	05	84	00	07	P12	04	05	48	41	11
E13	01	10	84	04	01	P13	05	10	43	40	11
Assurance						**Assurance**					
E14	07	07	70	04	06	P14	03	14	61	21	03
E15	03	14	60	04	04	P15	14	26	63	03	07
E16	06	03	77	08	11	P16	04	06	65	15	17
E17	01	00	79	04	08	P17	07	16	53	13	08
Empathy						**Empathy**					
E18	04	08	02	78	03	P18	01	04	04	81	01
E19	08	07	00	69	08	P19	11	07	07	49	12
E20	01	02	03	84	11	P20	03	08	04	88	05
E21	02	21	16	50	13	P21	11	13	09	67	13
E22	08	08	30	38	03	P22	11	10	14	63	06

[a] All numbers in the table are magnitudes of factor loadings multiplied by 100. The percentage variance explained by the five factors in the expectation (E) and perception (P) data sets were 64.7% and 72.5%, respectively.

Copyright © 2001 All Rights Reserved

ance of employees/communication materials (items 3 and 4). The factor loading matrix for the perceptions items in Table 3 shows evidence of this perceptual difference: while the four items representing tangibles cluster together under F1, items P3 and P4 have stronger loadings on F5 and dominate that factor.

Another insight from the factor loading matrix for the Es in Table 3 is that the items making up responsiveness and assurance load strongly and unambiguously on the same factor (F3). Thus, customers' expectations on these two dimensions form a homogeneous structure which is distinct from the other three dimensions. The factor loading matrix for the Ps shows that customers' perceptual structures for responsiveness and assurance also overlap. However, this overlap is much less pronounced because several responsiveness items also have high loadings on the factors representing reliability (F2) and empathy (F4).

The overall factor loading pattern for expectations is similar to the pattern for perceptions except that the latter is not quite as clear. The more diffused factor pattern for perceptions was also reflected by higher inter-factor correlations. The average pairwise correlations between factors was .51 for the perceptions matrix, compared to .35 for the expectations matrix.

The preceding factor analyses were constrained to five-factor solutions because of the clear five-dimensional structure observed in the original SERVQUAL study (PZB 1988). However, in the present study, the items under tangibles consistently split into two distinct dimensions in the factor analyses of the gap scores (Qs) and the perception ratings (Ps). This finding raised the possibility that the overlap between responsiveness and assurance might be an artifact of the five-factor constraint imposed on the analyses. In other words, since tangibles was "using up" two factors, the items under the remaining four dimensions had only three factors on which to load, thereby making some degree of overlap between the dimensions inevitable. To explore this possibility, the data on the Qs and Ps were reanalyzed to extract six factors instead of five. The factor loading matrices for the combined sample are shown in Table 4 (factor patterns for the individual company samples were similar to those in Table 4).

The results in Table 4 provide partial support for the speculation that the apparent overlap between responsiveness and assurance might have stemmed from the five-factor constraint in the previous analyses. While the two dimensions still overlap in Table 4, responsiveness exhibits a degree of distinctiveness (represented by F4) that was absent in the five-factor solutions.

The findings from the various factor analyses suggest the following key

Copyright © 2001 All Rights Reserved

TABLE 4

Factor Loading Matrices Following Oblique Rotation of Six-Factor Solutions for P and Q Scores[a]

Items	Factor Loadings F1	F2	F3	F4	F5	F6	Items	Factor Loadings F1	F2	F3	F4	F5	F6
Tangibles							**Tangibles**						
P1	89	01	00	02	01	00	Q1	90	05	00	01	01	04
P2	93	02	04	01	02	01	Q2	76	10	02	02	02	05
P3	00	77	02	00	07	03	Q3	00	86	06	01	09	00
P4	17	60	12	00	07	01	Q4	04	86	08	01	07	02
Reliability							**Reliability**						
P5	01	01	82	11	04	04	Q5	02	02	83	06	01	05
P6	01	03	35	03	14	47	Q6	02	09	39	02	19	03
P7	02	01	74	08	13	04	Q7	03	03	73	09	07	05
P8	01	05	88	09	03	00	Q8	00	05	90	07	01	00
P9	09	09	55	15	13	14	Q9	06	01	53	16	09	14
Responsiveness							**Responsiveness**						
P10	02	19	47	32	04	05	Q10	02	07	36	19	37	08
P11	06	07	28	45	18	13	Q11	01	01	24	31	65	03
P12	07	07	02	27	40	32	Q12	01	03	01	13	79	14
P13	10	01	05	34	37	27	Q13	04	04	02	12	79	10
Assurance							**Assurance**						
P14	08	09	16	14	50	19	Q14	03	07	10	07	73	08
P15	11	07	31	03	53	03	Q15	06	01	24	22	60	03
P16	00	31	04	08	50	17	Q16	04	08	11	11	79	06
P17	11	13	16	15	43	11	Q17	01	04	13	07	67	02
Empathy							**Empathy**						
P18	03	04	05	01	01	85	Q18	01	02	01	03	04	79
P19	12	10	03	11	07	43	Q19	04	05	03	04	04	53
P20	02	10	09	01	00	90	Q20	02	03	09	03	02	91
P21	13	11	14	01	09	66	Q21	01	07	16	11	10	63
P22	11	03	11	03	12	62	Q22	06	03	14	04	20	52

[a] All numbers in the table are magnitudes of factor loadings multiplied by 100. The percentage variance explained by the six factors in the perception (P) and gap score (Q) data sets were 74.1% and 69.3%, respectively.

Copyright © 2001 All Rights Reserved

conclusions: (1) tangibles, which was unidimensional in the original scale, splits into two subdimensions in the revised scale—one pertaining to physical facilities/equipment and another pertaining to employees/communication materials; (2) the degree of overlap among dimensions (as reflected by interfactor correlations) is somewhat higher in the revised scale; and (3) responsiveness and assurance are virtually indistinguishable in the five-factor solutions; however, they do seem distinct in the six-factor solutions, which allow sufficient "degrees of freedom" for the dimensions other than tangibles to display their uniqueness. Collectively these conclusions suggest that while the interdimensional overlap in the refined SERVQUAL scale is somewhat greater than in the original scale, the refinement still reflects the basic five-dimensional structure of the original scale with one key exception—namely, the dichotomization of tangibles into two subdimensions.

Results from the question that asked customers to allocate 100 points across the five SERVQUAL dimensions are summarized in Table 5. The allocation patterns are virtually identical for the different customer samples and suggest that the relative importance of the SERVQUAL dimensions are stable across settings. Moreover, paired-sample t-tests comparing the points allocated to responsiveness and assurance—the two dimensions displaying the most overlap in the factor analyses—showed a statistically significant difference favoring responsiveness in every sample. Thus, customers apparently do distinguish between these two dimensions. This finding offers additional, if indirect, support for treating the two dimensions as being distinct.

TABLE 5

Relative Importance of SERVQUAL Dimensions

	Mean Number of Points Allocated Out of 100 Points[a]					
Dimension	Tel. Co.	Ins. Co. 1	Ins. Co. 2	Bank 1	Bank 2	All Cos.
Tangibles	12	10	11	11	11	11
Reliability	34	33	29	31	32	32
Responsiveness	24	22	23	23	22	23
Assurance	17	19	20	20	19	19
Empathy	16	16	18	17	16	17

[a] Numbers in some columns do not add exactly to 100 due to rounding.

Copyright © 2001 All Rights Reserved

Validity

The high reliability and consistent factor structure of the revised SERVQUAL scale across five independent samples support the scale's trait validity (Campbell 1960). Such evidence, while necessary, is not sufficient for establishing a scale's construct validity (Churchill 1979). For this reason, several additional analyses were performed to assess the revised SERVQUAL scale's validity more directly.

The association between the SERVQUAL gap scores and customers' ratings on a separate measure of service quality was examined. Each customer sample rated the overall service quality (OSQ) of the corresponding company on a ten-point scale, anchored at the ends by "extremely poor" (scale value of 1) and "extremely good" (scale value of 10). [The original study (PZB 1988) employed just a four-category scale to obtain customers' overall quality ratings.] The OSQ ratings were regressed on the SERVQUAL gap scores along the five dimensions. The R^2 values shown in Table 6 provide strong evidence of the gap scores' ability to explain the variance in the OSQ ratings. In fact, the R^2 values in Table 6 are all higher

TABLE 6

Regression Analysis of OSQ vs. SERVQUAL Scores for the Five Dimensions[a]

Independent Variables	Standardized Regression Coefficients				
	Tel. Co.	Ins. Co. 1	Ins. Co. 2	Bank 1	Bank 2
Tangibles	.03	.05	.02	.01	−.09**
Reliability	.51***	.38***	.19**	.39***	.29***
Responsiveness	.17**	.10	.04	.24***	.31***
Assurance	.32***	.37***	.45***	.14*	.15**
Empathy	.23***	.02	.21***	.13**	.15**
Adjusted R^2 Value	.57***	.66***	.71***	.68***	.58**

[a] Dependent Variable: OSQ (10-point scale).
* Significant at $p < .1$.
** Significant at $p < .05$.
*** Significant at $p < .01$.

Copyright © 2001 All Rights Reserved

than the largest value of .52 reported in the original study (Table 6, PZB 1988). This high degree of convergence between the revised SERVQUAL scale and a separate measure of service quality (OSQ) supports the scale's construct validity. Moreover, while high multicollinearity (the average correlation between independent variables was .58) prevents unambiguous interpretation of the regression coefficients, the general pattern of these coefficients—reliability having the strongest coefficients, assurance and responsiveness having the next strongest coefficients, and empathy and tangibles having the weakest coefficients in four of the five samples—is similar to the pattern of coefficients reported in Table 6 of PZB 1988.

The revised SERVQUAL scale's validity was further assessed by examining whether the measured construct was associated empirically with measures of conceptually related variables. Customers in each sample indicated in separate questions: (1) whether they had experienced a recent service problem with the company; (2) if they had experienced a problem, whether it was resolved to their satisfaction; and (3) whether they would recommend the service firm to a friend. Table 7 summarizes the SERVQUAL scores for customers grouped according to their responses to these three questions.

TABLE 7

Weighted Average SERVQUAL Scores for Respondents Segmented According to the Variables Problem, Resolved, and Recommend[a]

	Problem?		Resolved?		Recommend?	
Company	Yes	No	Yes	No	Yes	No
Tel. Co.	−1.60	−0.75	−1.26	−2.38	−0.64	−2.85
	(159)	(219)	(99)	(49)	(295)	(74)
Ins. Co. 1	−1.98	−0.70	−1.72	−2.73	−1.00	−2.59
	(147)	(184)	(91)	(46)	(284)	(42)
Ins. Co. 2	−1.72	−0.28	−1.50	−2.54	−0.39	−2.89
	(80)	(180)	(43)	(30)	(235)	(32)
Bank 1	−2.14	−0.92	−1.69	−2.67	−0.85	−2.84
	(128)	(209)	(61)	(61)	(244)	(89)
Bank 2	−1.85	−0.55	−1.41	−2.46	−0.61	−2.55
	(130)	(312)	(66)	(53)	(383)	(58)

[a] Numbers shown in parentheses are sample sizes. The average SERVQUAL scores under "Yes" and "No" are significantly different (at $p \leq .01$) in every instance.

Copyright © 2001 All Rights Reserved

The perception-minus-expectation gap scores in Table 7 are *weighted* average SERVQUAL scores. The weighted score for each respondent was obtained by multiplying the respondent's mean gap score for each dimension by the dimension's relative importance weight and summing the results across the five dimensions. Less negative gap scores in Table 7 imply higher levels of perceived service quality—i.e., customers' perceptions come closer to matching expectations. Respondents answering no to the "Problem?" question, yes to the "Resolved?" question and yes to the "Recommend?" question should perceive higher service quality than other respondents. As Table 7 shows, all results are statistically significant in the hypothesized direction, providing additional support for the validity of the refined SERVQUAL scale.

COMPARISON WITH OTHER STUDIES USING SERVQUAL

The SERVQUAL scale has been used in a variety of proprietary and published studies to assess customer perceptions of service quality. While not all these studies have formally examined the scale's psychometric properties, four recent studies have done so (Babakus and Boller 1991; Brensinger and Lambert 1990; Carman 1990; Finn and Lamb 1991). Table 8 provides a comparative summary.

Table 8, in addition to summarizing the contexts and procedures used in the various studies, reveals areas of consensus as well as unresolved issues regarding SERVQUAL's psychometric properties. The across-study comparison also suggests directions for future SERVQUAL research and applications.

Reliability

The Cronbach's alpha reliability coefficients for the five SERVQUAL dimensions are similar across studies and are at least of the same order of magnitude as those reported in PZB (1988). These findings support the internal cohesiveness of the scale items forming each dimension. Furthermore, the fact that the reliability coefficients obtained in the current study are higher than those obtained in the other studies (including PZB 1988) suggests that the wording changes incorporated in the current study have improved SERVQUAL. The conversion of negatively worded items to positive items—a refinement consistent with recommendations made by Babakus and Boller (1991) and Carman (1990) based on problems with negatively worded items in their studies—seems to have been especially beneficial.

Copyright © 2001 All Rights Reserved

TABLE 8

Comparison of Current Study with Other SERVQUAL-Replication Studies

Study	Current Study	Babakus and Boller (1991)	Brensinger and Lambert (1990)	Carman (1990)	Finn and Lamb (1991)
Data Collection					
Study sample(s)	Customers of a telephone co., two insurance cos., and two banks	Customers of an electric and gas utility co.	Purchasers of motor carrier services in business-to-business markets	Customers of a dental school patient clinic, a business school placement center, a tire store and a hospital	Customers of four retail store types: "stores like K-mart, Wal-Mart, etc.; JC Penney, Sears, etc.; Dillard's, Foley's, etc.; and Saks, Neiman-Marcus, etc."
Sample size	Ranged from 290 to 487 across companies	689	170	Ranged from 74 to 600+ across settings	Ranged from 58 to 69 across settings
Questionnaire format	Similar to PZB (1988) format with separate expectations and perceptions sections	Similar to PZB (1988)	Similar to PZB (1988)	Similar to PZB (1988) in the placement-center setting; only perceptions measured in the other three settings	Similar to PZB (1988)

Copyright © 2001 All Rights Reserved

TABLE 8 Cont'd

Study	Current Study	Babakus and Boller (1991)	Brensinger and Lambert (1990)	Carman (1990)	Finn and Lamb (1991)
Major wording changes	Negatively worded questions changed to positive form; "companies should" terminology in expectations section replaced with "excellent companies will" terminology	No major changes	No major changes	No major changes in the SERVQUAL items retained; however, several of the new items added were transaction-specific (rather than general attitude statements as in the original SERVQUAL)	No major changes
Original SERVQUAL items retained	Twenty items (two remaining original items were replaced by new items)	All 22 items	All 22 items	Ranged from 10 to 17 across settings	All 22 items
Response scale	7-point scale	7-point scale	7-point scale	7-point scale	5-point scale
Questionnaire administration	Mail survey	Mail survey	Mail survey	Self-administered by respondents on-site	Telephone survey
Data Analysis					
Procedure for assessing factor structure	Principal-axis factor analysis followed by oblique rotation	Principal-axis factor analysis followed by oblique rotation; LISREL confirmatory	Principal-axis factor analysis followed by oblique rotation	Principal-axis factor analysis followed by oblique rotation	LISREL confirmatory factor analysis of 5-dimensional measurement model

Copyright © 2001 All Rights Reserved

TABLE 8 Cont'd

Study	Current Study	Babakus and Boller (1991)	Brensinger and Lambert (1990)	Carman (1990)	Finn and Lamb (1991)
		factor analysis of 5-dimensional measurement model			
Basis for initial number of factors extracted	PZB's (1988) 5-dimensional structure	PZB's (1988) 5-dimensional structure	PZB's (1988) 5-dimensional structure	Factors with eigenvalues greater than one	PZB's (1988) 5-dimensional structure
Findings					
Reliability coefficients (Cronbach's alphas)	.80 to .93	.67 to .83	.64 to .88	Mean of .75 (across 35 scales derived through factor analysis)	.59 to .83
Final number of dimensions	Five (six if "tangibles" is split into two) dimensions	No clear 5-dimensional factor structure; LISREL model fit poor; 2-dimensional structure (with one representing negative items and the other positive items) most viable solution	Five dimensions as in PZB (1988); however, only four factors retained based on eigenvalues/ greater-than-one criterion	Between six and eight dimensions depending on setting	LISREL model fit for 5-dimensional structure poor (no alternative factor structures examined)

Copyright © 2001 All Rights Reserved

TABLE 8 Cont'd

Study	Current Study	Babakus and Boller (1991)	Brensinger and Lambert (1990)	Carman (1990)	Finn and Lamb (1991)
Predictive/concurrent validity	Q (i.e., P-E) scores on the five dimensions explain .57 to .71 of variance in overall quality scores measured on a 10-pt scale; Q scores also related as hypothesized to presence of service problems, satisfactory problem resolution and willingness to recommend	Total Q scores (across all 22 items) correlate .59 with overall quality scores measured on a 4-pt scale (however, total P scores correlate .66 with the same measure); correlations of Q and P scores with satisfactory complaint resolution are .58 and .60, respectively	Q scores on the five dimensions explain: .39 of variance in 4-pt. overall quality scale; .43 of variance in 100-pt. overall performance scale; .08 of variance in customer's "share of business" received by motor carriers	Not examined	Not examined

Copyright © 2001 All Rights Reserved

Validity

Several different forms of validity can serve as criteria for assessing the psychometric soundness of a scale: face validity, convergent validity, discriminant validity, and predictive or concurrent validity. Sufficient information (not all of which is summarized in Table 8) is available from the current study and at least one of the four other replication studies to perform a comparative evaluation of SERVQUAL on each of these forms of validity.

Face validity. SERVQUAL's face validity, a subjective criterion reflecting the extent to which scale items are meaningful and appear to represent the construct being measured, was explicitly assessed *a priori* in the current study and two others (Babakus and Boller 1991; Carman 1990). In the current study, feedback from executives (in each of the five participating companies) who reviewed the questionnaire confirmed that SERVQUAL—with minor wording changes in a few items—had face validity. Likewise, Babakus and Boller (1991) confirmed the suitability of SERVQUAL for the utility industry through preliminary discussions with 25 utility customers and extensive interviews with company executives and technical personnel. In contrast, Carman's (1990) initial assessment of the scale resulted in his using a subset of the original 22 items (ranging from 10 in the dental clinic setting to 17 in the tire store and placement center setting). Brensinger and Lambert (1990) and Finn and Lamb (1991) do not explicitly discuss SERVQUAL's face validity. However, the fact that all 22 SERVQUAL items were used in both studies implies support for the meaningfulness of the items in the settings involved. Thus, with few exceptions, the SERVQUAL items appear to be appropriate for assessing service quality in a variety of settings.

Convergent validity. This pertains to the extent to which scale items assumed to represent a construct do in fact "converge" on the same construct. The reliability of a scale as measured by coefficient alpha reflects the degree of cohesiveness among the scale items and is therefore an indirect indicator of convergent validity. As already mentioned, coefficient alpha values for the five SERVQUAL dimensions are fairly high in all studies.

A more stringent test of convergent validity is whether scale items expected to load together in a factor analysis actually do so. The factor-loading patterns in the current study and in the Brensinger and Lambert (1990) study are similar to that obtained in PZB (1988) and attest to SERVQUAL's convergent validity. The only major exception is that the four tangibles items split into two separate dimensions in the current study.

Copyright © 2001 All Rights Reserved

Evidence of convergent validity as reflected by the factor-loading patterns in the Babakus and Boller (1991) and Carman (1990) studies is weaker because several SERVQUAL items had low loadings on the dimensions they were supposed to represent. Finn and Lamb (1991) report overall fit statistics for the LISREL measurement model, but do not provide a factor-loading matrix. Therefore an assessment of convergent validity in their study by examining factor loadings is not possible.

Discriminant Validity. The findings from the replication studies differ the most from the original study with respect to SERVQUAL's discriminant validity—the extent to which SERVQUAL has five distinct dimensions. Collectively, the replication studies imply greater overlap among the SERVQUAL dimensions—especially among responsiveness, assurance, and empathy—than implied by the original study. However, as summarized in Table 8, the number of distinct dimensions based solely on the factor analysis results is not the same across studies. It varies from two in the Babakus and Boller (1991) study to eight in one of the four settings studied by Carman (1990).

To some extent this variation across studies may be due to differences in data collection and analysis procedures (see Table 8). However, there is another, perhaps more plausible, explanation: Respondents may indeed consider the SERVQUAL dimensions to be conceptually unique (as demonstrated by the distinct and consistent pattern of relative importance ratings reported for the five dimensions in Table 5 of the current study); however, if their evaluations of a specific company on individual scale items are similar *across* dimensions, fewer than five dimensions will result as in the Babakus and Boller (1991) study. Alternatively, if their evaluations of a company on scale items *within* a dimension are sufficiently distinct, more than five dimensions will result as in Carman's (1990) study (the addition of several new items in each of Carman's settings may also account for this result). In other words, differences in the number of 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. At a *general* level, the five-dimensional structure of SERVQUAL may still serve as a meaningful conceptual framework for summarizing the criteria customers use in assessing service quality. Verifying the conceptual meaningfulness of the SERVQUAL framework has an important implication for future research which we explore in the next section.

Predictive or Concurrent Validity. This facet of SERVQUAL's validity refers to the extent to which SERVQUAL scores are associated as

Copyright © 2001 All Rights Reserved

hypothesized with other conceptually related measures. Predictive/concurrent validity was examined in the current study as well as in the Babakus and Boller (1991) and Brensinger and Lambert (1991) studies. As summarized in the last row of Table 8, SERVQUAL performs fairly well in this regard, with two exceptions.

One exception is that the SERVQUAL scores received by motor carriers in the Brensinger and Lambert (1991) study accounted for only 8 percent of the variance in the share of customers' business obtained by those carriers. However, as Brensinger and Lambert point out in their paper, this result is not surprising because factors other than service quality influence market share. Furthermore, although the authors' survey asked customers to evaluate up to three motor carriers that they used most frequently, the data analysis was limited to the first carrier evaluated. Presumably, the first carrier listed and evaluated by each customer was also the one with the highest share of the customer's business. If so, insufficient variation in market shares in the analysis sample may have contributed to the low observed correlation between the SERVQUAL scores and market shares.

The second exception is that in the Babakus and Boller (1991) study, perception scores by themselves have stronger correlations with other dependent measures (e.g., overall quality) than do the SERVQUAL scores (i.e., perception minus expectation scores). Likewise, rerunning the regression analyses in the current study with just the perception scores as independent variables produced R^2 values ranging from .72 to .81 (in contrast to the range of .57 to .71 obtained by using SERVQUAL scores as independent variables). These results call into question the empirical usefulness of the expectations data, an issue that Babakus and Boller (1991) and Carman (1990) raise in their papers. These authors also raise psychometric concerns about the appropriateness of using measures defined as difference scores in multivariate analyses.

In summary, the collective findings from the various replications by and large provide consistent support for the reliability, face validity and predictive/concurrent validity for the SERVQUAL scores on the five dimensions. Factor-analysis results pertaining to the convergent validity of the items representing each dimension are mixed because in several studies the highest loadings for some items were on different dimensions than those in PZB (1988). Support for the discriminant validity of SERVQUAL as reflected by the factor-loading patterns and the number of factors retained is inconsistent across studies. Finally, the usefulness of the expectations scores and the appropriateness of analyzing gap scores remain unresolved.

The current study's findings and general insights from the across-study

Copyright © 2001 All Rights Reserved

comparisons have important implications for service quality researchers and SERVQUAL users. These implications are explored in the next two sections.

DIRECTIONS FOR FUTURE SERVQUAL-RELATED RESEARCH

Though the SERVQUAL dimensions represent five conceptually distinct facets of service quality, they are also interrelated, as evidenced by the need for oblique rotations of factor solutions in the various studies to obtain the most interpretable factor patterns. One fruitful area for future research is to explore the nature and causes of these interrelationships. The factor patterns in Table 4 show that the responsiveness items, while exhibiting some degree of distinctiveness, have sizeable crossloadings on other dimensions, particularly assurance and reliability. In light of these findings, is responsiveness an *antecedent* of assurance and reliability? In other words, by demonstrating responsiveness—by being consistently prompt and willing to assist customers—can a service provider *enhance* customers' impressions about the firm's reliability and trustworthiness? Alternatively, to what extent can customer goodwill earned through superior responsiveness counteract negative feelings when service problems occur or customers' confidence in the firm is shaken? Research directed at these and other questions focusing on the nature of the interrelationships among the dimensions can contribute to our understanding of service quality.

An intriguing finding from the five replications in the current study is that tangibles consistently split into two subdimensions—one focusing on equipment and facilities and the other focusing on personnel and communication materials. The item pertaining to communication materials was not used in our original study or in the other replication studies. The addition of this item, in place of a deleted item on physical facilities, appears to be responsible for the split. Interestingly, as discussed earlier, the split seems to be driven by differences in the structure of perceptions rather than expectations (see Table 3). Will this pattern of findings hold in settings different from the ones included in the current study? If so, what accounts for the apparent difference in how customers perceive the two subcategories of tangibles? Is there a difference in the relative importance to customers of the two subcategories? Research focusing on these issues would be useful.

A major inconsistency in the findings from the various replication studies pertains to the factor structure and hence discriminant validity of SERVQUAL. As previously discussed, respondents' rating a *specific* com-

Copyright © 2001 All Rights Reserved

pany similarly on SERVQUAL items pertaining to different dimensions—rather than respondents' inability to distinguish conceptually between the dimensions *in general*—is a plausible explanation for the diffused factor patterns. To illustrate, consider the following SERVQUAL items:

- "Employees at XYZ give you prompt service" (a *responsiveness* item)
- "Employees of XYZ have the knowledge to answer your questions" (an *assurance* item)
- "XYZ has operating hours convenient to all its customers" (an *empathy* item)

If customers happen to rate XYZ the same or even similarly on these items, it does not necessarily mean that customers consider the items to be part of the same dimension. Yet, because of high intercorrelations among the three sets of ratings, the items are likely to load on the same factor when the ratings are factor analyzed. Therefore, whether an unclear factor pattern obtained through analyzing company-specific ratings necessarily implies poor discriminant validity for the general SERVQUAL dimensions is debatable. An intriguing approach for assessing SERVQUAL's discriminant validity—and an important direction for future research—is to give customers definitions of the five dimensions and ask them to place each SERVQUAL item under the most appropriate definition based solely on item content (rather than on a specific company's performance on each item). The proportions of customers "correctly" sorting the items into the five dimensions would be indicative of the scale's discriminant validity. The pattern of correct and incorrect classifications could also reveal potentially confusing items and the consequent need to reword the items and/or the dimensional definitions.

Another fruitful and critical area for future research is the measurement of expectations and the related issue of computing perception-minus-expectation gap scores. Babakus and Boller (1991) and Carman (1990) discuss this subject and make several useful suggestions that are worthy of additional research. We would like to add to their discussion by drawing upon our experiences in using the two-part SERVQUAL instrument. There are practical as well as theoretical aspects to the pros and cons of measuring expectations and perceptions separately and then computing gap scores.

From a practical standpoint, while the two parts of SERVQUAL may seem repetitive to some respondents and does increase the questionnaire's length, we have not encountered problems with respondents' not understanding the distinction between the parts. Moreover, managers in com-

Copyright © 2001 All Rights Reserved

panies for which we have conducted SERVQUAL surveys find the gap-score format for reporting results to be of diagnostic value. Using the two-part SERVQUAL over time allows managers to track the trend in expectations as well as in perceptions. In short, based on our extensive experience with SERVQUAL, the managerial appeal and usefulness of reporting service quality shortfalls as gap scores more than compensate for the increased survey length.

From a theoretical perspective, the appropriateness of using difference scores in multivariate analyses has been questioned on grounds that such scores might suffer from low reliability and validity (e.g., Prakash 1984). Babakus and Boller (1991) and Carman (1990) echo this concern. However, findings from the various replication studies indicate that the gap scores along the five SERVQUAL dimensions possess adequate reliability as measured by Cronbach's alpha. Moreover, the three studies that examined SERVQUAL's predictive/concurrent validity are also fairly supportive of the gap scores, although from a strictly empirical standpoint the perception scores alone perform better. The major inconsistencies across replication studies pertain to the factor structures of the gap scores. While the current study and the Brensinger and Lambert (1991) study are similar to PZB (1988) in this regard, the other studies are not. Therefore, support for SERVQUAL gap scores' discriminant validity, and to some extent convergent validity, is mixed.

Though the collective evidence from the replication studies does not fully endorse the two-part measurement approach, it does not suggest abandoning the approach either. Rather, it calls for comparative studies wherein the two-part measurement approach is evaluated against more direct measures of the perception-expectation difference using separate but equivalent customer samples. One direct-measurement approach is to ask customers to indicate on the same rating scale where they would place a high-quality service company and where they would place XYZ company. Another approach is to ask customers to treat the high end of the rating scale (e.g., 7 on a 7-point scale) as representing the level of service they would expect from a high-quality company and then rate XYZ on the same scale. A third approach is to anchor the scale's midpoint as the service level expected of a high-quality company and ask customers to rate the extent to which XYZ exceeds or falls short of their expectations. These alternative measurement approaches would require appropriate wording changes in SERVQUAL's directions and items, although the item content would be the same. Comparative evaluations of the different approaches within the same studies would be valuable in resolving the measurement-related issues raised by previous SERVQUAL studies.

Copyright © 2001 All Rights Reserved

GUIDELINES FOR USING SERVQUAL

We stated in our original study that SERVQUAL is a generic instrument with good reliability and validity and broad applicability. Having now refined and reassessed SERVQUAL, we continue to feel confident of its usefulness. Nevertheless, the combined evidence from various replication studies raises important issues that call for further research as discussed in the preceding section. In the meantime, the refined SERVQUAL scale can be used to assess service quality, provided potential users are cognizant of the nature, scope, and limitations of the instrument.

The purpose of SERVQUAL is to serve as a diagnostic methodology for uncovering *broad* areas of a company's service quality shortfalls and strengths. SERVQUAL's dimensions and items represent *core* evaluation criteria that transcend specific companies and industries, as implied by the systematic, multi-stage, and iterative process that produced the instrument (PZB 1988). As such, the following guidelines and caveats, in addition to those in our original article, should be helpful in ensuring the most appropriate and effective use of SERVQUAL.

First, since SERVQUAL is the basic "skeleton" underlying service quality, it should be used in its entirety as much as possible. While minor modifications in the wording of items to adapt them to a specific setting are appropriate, deletion of items could affect the integrity of the scale and cast doubt on whether the reduced scale fully captures service quality.

Second, context-specific items can be used to supplement SERVQUAL. However, the new items should be similar in form to existing SERVQUAL items (e.g., they should be general rather than transaction specific). Moreover, each new item, based on its content, should be classified under the most appropriate SERVQUAL dimension to facilitate computation of the average gap score for each dimension. Although items that do not fit under any of the five dimensions (e.g., items about customers' perceptions of a service's cost) may be useful to include in the survey questionnaire, such items should be treated separately in analyzing the survey data since they do not fall under the conceptual domain of service quality.

Finally, the use of SERVQUAL can fruitfully be supplemented with additional qualitative or quantitative research to uncover the causes underlying the key problem areas or gaps identified by a SERVQUAL study. SERVQUAL is a useful starting point, not the final answer, for assessing and improving service quality. Its standard five-dimensional structure serves as a meaningful framework for tracking a firm's service quality performance over time and comparing it against the performance of competitors. As we stated in our original article, "SERVQUAL is most valu-

Copyright © 2001 All Rights Reserved

able when it is used periodically to track service quality trends, and when it is used in conjunction with other forms of service quality measurement." (PZB 1988, p. 31)

APPENDIX: MODIFIED SERVQUAL INSTRUMENT

Expectations Section

DIRECTIONS: Based on your experiences as a customer of telephone repair services, please think about the kind of telephone company that would deliver excellent quality of repair service. Think about the kind of telephone company with which you would be pleased to do business. Please show the extent to which you think such a telephone company would possess the feature described by each statement. If you feel a feature is *not at all essential* for excellent telephone companies such as the one you have in mind, circle the number "1". If you feel a feature is *absolutely essential* for excellent telephone companies, circle "7". If your feelings are less strong, circle one of the numbers in the middle. There are no right or wrong answers—all we are interested in is a number that truly reflects your feelings regarding telephone companies that would deliver excellent quality of service.

Note: Each of the statements was accompanied by a 7-point scale anchored at the ends by the labels "Strongly Disagree" (=1) and "Strongly Agree" (=7). Intermediate scale points were not labeled. Also, the headings (TANGIBLES, RELIABILITY, etc.), shown here to indicate which statements fall under each dimension, were not included in the actual questionnaire.

TANGIBLES

E1. Excellent telephone companies will have modern-looking equipment.
E2. The physical facilities at excellent telephone companies will be visually appealing.
E3. Employees of excellent telephone companies will be neat-appearing.
E4. Materials associated with the service (such as pamphlets or statements) will be visually appealing in an excellent telephone company.

RELIABILITY

E5. When excellent telephone companies promise to do something by a certain time, they will do so.

Copyright © 2001 All Rights Reserved

E6. When customers have a problem, excellent telephone companies will show a sincere interest in solving it.
E7. Excellent telephone companies will perform the service right the first time.
E8. Excellent telephone companies will provide their services at the time they promise to do so.
E9. Excellent telephone companies will insist on error-free records.

RESPONSIVENESS

E10. Employees of excellent telephone companies will tell customers exactly when services will be performed.
E11. Employees of excellent telephone companies will give prompt service to customers.
E12. Employees of excellent telephone companies will always be willing to help customers.
E13. Employees of excellent telephone companies will never be too busy to respond to customer requests.

ASSURANCE

E14. The behavior of employees of excellent telephone companies will instill confidence in customers.
E15. Customers of excellent telephone companies will feel safe in their transactions.
E16. Employees of excellent telephone companies will be consistently courteous with customers.
E17. Employees of excellent telephone companies will have the knowledge to answer customer questions.

EMPATHY

E18. Excellent telephone companies will give customers individual attention.
E19. Excellent telephone companies will have operating hours convenient to all their customers.
E20. Excellent telephone companies will have employees who give customers personal attention.
E21. Excellent telephone companies will have the customers' best interests at heart.
E22. The employees of excellent telephone companies will understand the specific needs of their customers.

Copyright © 2001 All Rights Reserved

Perceptions Section

DIRECTIONS: The following set of statements relate to your feelings about XYZ Telephone Company's repair service. For each statement, please show the extent to which you believe XYZ has the feature described by the statement. Once again, circling a "1" means that you strongly disagree that XYZ has that feature, and circling a "7" means that you strongly agree. You may circle any of the numbers in the middle that show how strong your feelings are. There are no right or wrong answers—all we are interested in is a number that best shows your perceptions about XYZ's repair service.

TANGIBLES

P1. XYZ has modern-looking equipment.
P2. XYZ's physical facilities are visually appealing.
P3. XYZ's employees are neat-appearing.
P4. Materials associated with the service (such as pamphlets or statements) are visually appealing at XYZ.

RELIABILITY

P5. When XYZ promises to do something by a certain time, it does so.
P6. When you have a problem, XYZ shows a sincere interest in solving it.
P7. XYZ performs the service right the first time.
P8. XYZ provides its services at the time it promises to do so.
P9. XYZ insists on error-free records.

RESPONSIVENESS

P10. Employees of XYZ tell you exactly when services will be performed.
P11. Employees of XYZ give you prompt service.
P12. Employees of XYZ are always willing to help you.
P13. Employees of XYZ are never too busy to respond to your requests.

ASSURANCE

P14. The behavior of employees of XYZ instills confidence in customers.
P15. You feel safe in your transactions with XYZ.
P16. Employees of XYZ are consistently courteous with you.
P17. Employees of XYZ have the knowledge to answer your questions.

EMPATHY

P18. XYZ gives you individual attention.
P19. XYZ has operating hours convenient to all its customers.

Copyright © 2001 All Rights Reserved

P20. XYZ has employees who give you personal attention.
P21. XYZ has your best interests at heart.
P22. Employees of XYZ understand your specific needs.

Point-Allocation Question

DIRECTIONS: Listed below are five features pertaining to telephone companies and the repair services they offer. We would like to know how important each of these features is to *you* when you evaluate a telephone company's quality of repair service. Please allocate a total of 100 points among the five features *according to how important each feature is to you*—the more important a feature is to you, the more points you should allocate to it. Please ensure that the points you allocate to the five features add up to 100.

1. The appearance of the telephone company's physical facilities, equipment, personnel, and communications materials. _____points
2. The ability of the telephone company to perform the promised service dependably and accurately. _____points
3. The willingness of the telephone company to help customers and provide prompt service. _____points
4. The knowledge and courtesy of the telephone company's employees and their ability to convey trust and confidence. _____points
5. The caring, individualized attention the telephone company provides its customers. _____points

TOTAL POINTS ALLOCATED 100 points

REFERENCES

Babakus, Emin and Gregory W. Boller (1991), "An Empirical Assessment of the SERVQUAL Scale," *Journal of Business Research* (forthcoming).

Brensinger, Ronald P. and Douglas M. Lambert (1990), "Can the SERVQUAL Scale be Generalized to Business-to-Business Services?" in *Knowledge Development in Marketing*, 1990 AMA's Summer Educators' Conference Proceedings, 289 (abstract), and full-length paper obtained from the authors.

Campbell, Donald T. (1960), "Recommendations for APA Test Standards Regarding Construct, Trait, or Discriminant Validity," *American Psychologist*, **15** (August), 546–53.

Carman, James M. (1990), "Consumer Perceptions of Service Quality: An Assessment of the SERVQUAL Dimensions," *Journal of Retailing*, **66** (Spring), 33–35.

Churchill, Gilbert A., Jr. (1979), "A Paradigm for Developing Better Measures of Marketing Constructs," *Journal of Marketing Research*, **16** (February), 64–73.

Crompton, John L. and Kelly J. Mackay (1989), "Users' Perceptions of the Relative Im-

Copyright © 2001 All Rights Reserved

portance of Service Quality Dimensions in Selected Public Recreation Programs," *Leisure Sciences,* **11,** 367–375.

Finn, David W. and Charles W. Lamb (1991), "An Evaluation of the SERVQUAL Scales in a Retailing Setting," in *Advances in Consumer Research,* Vol. **18** (forthcoming).

Johnson, Linda L., Michael J. Dotson, and B. J. Dunlap (1988), "Service Quality Determinants and Effectiveness in the Real Estate Brokerage Industry," *The Journal of Real Estate Research,* **3,** 21–36.

Parasuraman, A., Valarie A. Zeithaml, and Leonard L. Berry (1985), "A Conceptual Model of Service Quality and Its Implications for Future Research," *Journal of Marketing,* **49** (Fall), 41–50.

———(1988), "SERVQUAL: A Multiple-Item Scale for Measuring Customer Perceptions of Service Quality," *Journal of Retailing,* **64** (Spring), 12–40.

Prakash, Ved (1984), "Validity and Reliability of the Confirmation of Expectations Paradigm as a Determinant of Consumer Satisfaction," *Journal of the Academy of Marketing Science,* **12** (Fall), 63–76.

Webster, Cynthia (1989), "Can Consumers be Segmented on the Basis of Their Service Quality Expectations?" *The Journal of Services Marketing,* **3,** 35–53.

Woodside, Arch G., Lisa L. Frey, and Robert Timothy Daly (1989), "Linking Service Quality, Customer Satisfaction, and Behavioral Intention," *Journal of Health Care Marketing,* **9** (December), 5–17.

Copyright © 2001 All Rights Reserved