SERVQUAL: A multiple- Item Scale for measuring consumer perceptions of service quality

Article *in* Journal of Retailing · January 1988

CITATIONS
11,555

READS
136,312

3 authors:

A Parsu Parasuraman
University of Miami
156 PUBLICATIONS **64,742** CITATIONS
SEE PROFILE

Valarie A. Zeithaml
University of North Carolina at Chapel Hill
114 PUBLICATIONS **85,438** CITATIONS
SEE PROFILE

Leonard L Berry
Texas A&M University
146 PUBLICATIONS **65,237** CITATIONS
SEE PROFILE

Volume 64 Number 1 Spring 1988

SERVQUAL: A Multiple-Item Scale for Measuring Consumer Perceptions of Service Quality

A. PARASURAMAN
Foley's/Federated Professor of Retailing
and Marketing Studies
Texas A & M University
College Station, Texas

VALARIE A. ZEITHAML
Associate Professor
of Marketing
Duke University
Raleigh-Durham, N. Carolina

LEONARD L. BERRY
Foley's/Federated Professor of Retailing
and Marketing Studies
Texas A & M University
College Station, Texas

This paper describes the development of a 22-item instrument (called SERVQUAL) for assessing customer perceptions of service quality in service and retailing organizations. After a discussion of the conceptualization and operationalization of the service quality construct, the procedures used in constructing and refining a multiple-item scale to measure the construct are described. Evidence of the scale's reliability, factor structure, and validity on the basis of analyzing data from four independent samples is presented next. The paper concludes with a discussion of potential applications of the scale.

Intensifying competition and rapid deregulation have led many service and retail businesses to seek profitable ways to differentiate themselves. One strategy that has been related to success in these businesses is the delivery of high service quality (Rudie and Wansley 1985; Thompson,

The research reported in this article was made possible by a grant from the Marketing Science Institute, Cambridge, MA.

DeSouza, and Gale 1985). Delivering superior service quality appears to be a prerequisite for success, if not survival, of such businesses in the 1980s and beyond.

Unlike goods quality, which can be measured objectively by such indicators as durability and number of defects (Crosby 1979; Garvin 1983), service quality is an abstract and elusive construct because of three features unique to services: intangibility, heterogeneity, and inseparability of production and consumption (Parasuraman, Zeithaml, and Berry 1985). In the absence of objective measures, an appropriate approach for assessing the quality of a firm's service is to measure consumers' perceptions of quality. As yet, however, no quantitative yardstick is available for gauging these perceptions.

The purpose of this article is twofold: (1) to describe the development of a multiple-item scale for measuring service quality (called SERVQUAL) and (2) to discuss the scale's properties and potential applications. The basic steps employed in constructing the scale closely parallel procedures recommended in Churchill's (1979) paradigm for developing better measures of marketing constructs. Figure 1 provides an overview of the steps.

This article is divided into five sections. The first section delimits the domain of the service-quality construct and describes the generation of scale items (Steps 1, 2, and 3 in Figure 1). The second section presents the data-collection and scale-purification procedures (Steps 4 through 9), while the third section provides an evaluation of the scale's reliability and factor structure (Step 10). The next section deals with assessment of the scale's validity (Step 11). The final section discusses potential applications of the scale.

DOMAIN OF THE SERVICE-QUALITY CONSTRUCT

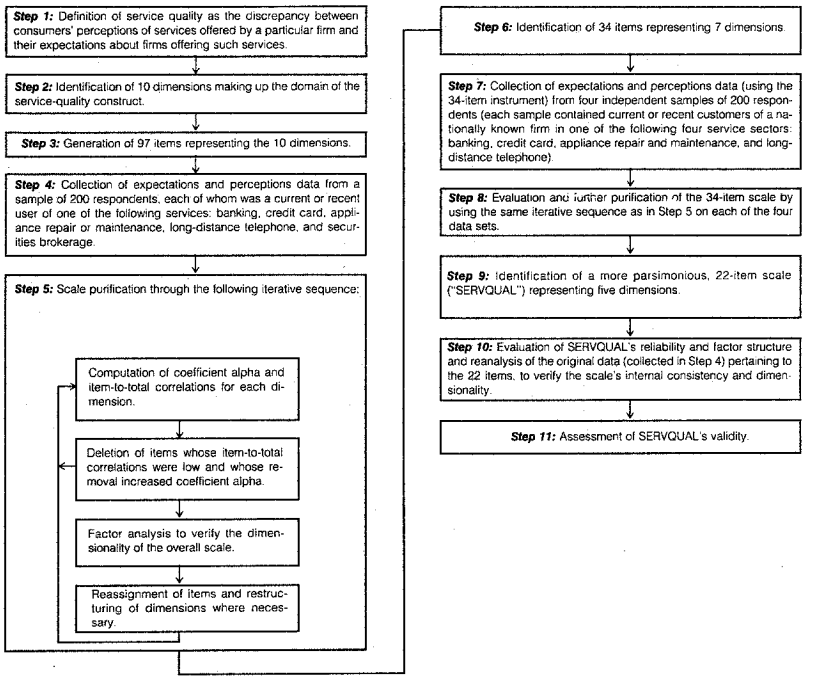
In deploring the inadequacy of measurement procedures used in the marketing discipline Jacoby (1978) wrote:

> Many of our measures are developed at the whim of a researcher with nary a thought given to whether or not it is meaningfully related to an explicit conceptual statement of the phenomena or variable in question. In most instances, our concepts have no identity apart from the instrument or procedures used to measure them. (p. 92)

The need for scale development to be preceded by, and rooted in, a sound conceptual specification of the construct being scaled has been emphasized by other scholars as well (e.g., Churchill 1979; Peter 1981). The

FIGURE 1

Summary of Steps Employed in Developing the Service-Quality Scale



coneptual foundation for the SERVQUAL scale was derived from the works of a handful of researchers who have examined the meaning of service quality (Sasser, Olsen, and Wyckoff 1978; Gronroos 1982; Lehtinen and Lehtinen 1982) and from a comprehensive qualitative research study that defined service quality and illuminated the dimensions along which consumers perceive and evaluate service quality (Parasuraman, Zeithaml, and Berry 1985).

Conceptualization of Service Quality

The construct of quality as conceptualized in the services literature and as measured by SERVQUAL, the scale that is the focus of this article, involves perceived quality. Perceived quality is the consumer's judgment about an entity's overall excellence or superiority (Zeithaml 1987). It differs from objective quality (as defined by, for example, Garvin 1983 and Hjorth-Anderson 1984); it is a form of attitude, related but not equivalent to satisfaction, and results from a comparison of expectations with perceptions of performance.

Perceived quality versus objective quality. Researchers (Garvin 1983; Dodds and Monroe 1984; Holbrook and Corfman 1985; Jacoby and Olson 1985; Zeithaml 1987) have emphasized the difference between objective and perceived quality. Holbrook and Corfman (1985), for example, note that consumers do not use the term quality in the same way as researchers and marketers, who define it conceptually. The conceptual meaning distinguishes between mechanistic and humanistic quality: "mechanistic (quality) involves an objective aspect or feature of a thing or event; humanistic (quality) involves the subjective response of people to objects and is therefore a highly relativistic phenomenon that differs between judges" (Holbrook and Corfman 1985, p. 33). Garvin (1983) discusses five approaches to defining quality, including two (product-based and manufacturing-based) that refer to objective quality and one (user-based) that parallels perceived quality.

Quality as attitude. Olshavsky (1985) views quality as a form of overall evaluation of a product, similar in many ways to attitude. Holbrook concurs, suggesting that quality acts as a relatively global value judgment. Exploratory research conducted by Parasuraman, Zeithaml, and Berry (1985) supports the notion that service quality is an overall evaluation similar to attitude. The researchers conducted a total of twelve focus group interviews with current or recent consumers of four different services—retail banking, credit card, securities brokerage, and product repair and maintenance. The discussions centered on issues such as the

meaning of quality in the context of the service in question, the characteristics the service and its provider should possess in order to project a high-quality image, and the criteria customers use in evaluating service quality. Comparison of the findings from the focus groups revealed that, regardless of the type of service, customers used basically the same general criteria in arriving at an evaluative judgement about service quality.

Quality versus satisfaction. Oliver (1981) summarizes current thinking on satisfaction in the following definition: "[satisfaction is a] summary psychological state resulting when the emotion surrounding disconfirmed expectations is coupled with the consumer's prior feelings about the consumption experience" (p. 27). This and other definitions (e.g., Howard and Sheth 1969; Hunt 1979) and most all measures of satisfaction relate to a specific transaction. Oliver (1981) summarizes the transaction-specific nature of satisfaction, and differentiates it from attitude, as follows:

> Attitude is the consumer's relatively enduring affective orientation for a product, store, or process (e.g., customer service) while satisfaction is the emotional reaction following a disconfirmation experience which acts on the base attitude level and is consumption-specific. Attitude is therefore measured in terms more general to product or store and is less situationally oriented. (p. 42)

Consistent with the distinction between attitude and satisfaction, is a distinction between service quality and satisfaction: perceived service quality is a global judgment, or attitude, relating to the superiority of the service, whereas satisfaction is related to a specific transaction. Indeed, in the twelve focus group interviews included in the exploratory research conducted by Parasuraman, Zeithaml, and Berry (1985), respondents gave several illustrations of instances when they were satisfied with a specific service but did not feel the service firm was of high quality. In this way, the two constructs are related, in that incidents of satisfaction over time result in perceptions of service quality. In Oliver's (1981) words, "satisfaction soon decays into one's overall attitude toward purchasing products."

Expectations compared to perceptions. The writings of Sasser, Olsen, and Wyckoff (1978); Gronroos (1982); and Lehtinen and Lehtinen (1982), and the extensive focus group interviews conducted by Parasuraman, Zeithaml, and Berry (1985), unambiguously support the notion that service quality, as perceived by consumers, stems from a comparison of what they feel service firms should offer (i.e., from their expectations) with their perceptions of the performance of firms providing the services.

Perceived service quality is therefore viewed as the degree and direction of discrepancy between consumers' perceptions and expectations.

The term "expectations" as used in the service quality literature differs from the way it is used in the consumer satisfaction literature. Specifically, in the satisfaction literature, expectations are viewed as *predictions* made by consumers about what is likely to happen during an impending transaction or exchange. For instance, according to Oliver (1981), "It is generally agreed that expectations are consumer-defined probabilities of the occurrence of positive and negative events if the consumer engages in some behavior" (p. 33). In contrast, in the service quality literature, expectations are viewed as desires or wants of consumers, i.e., what they feel a service provider *should* offer rather than *would* offer.

Dimensions of service quality. Exploratory research of Parasuraman, Zeithaml, and Berry (1985) revealed that the criteria used by consumers in assessing service quality fit 10 potentially overlapping dimensions. These dimensions were tangibles, reliability, responsiveness, communication, credibility, security, competence, courtesy, understanding/knowing the customer, and access (a description of the dimensions can be found in Parasuraman, Zeithaml, and Berry 1985, p. 47). These 10 dimensions and their descriptions served as the basic structure of the service-quality domain from which items were derived for the SERVQUAL scale.

Generation of Scale Items

Items representing various facets of the 10 service-quality dimensions were generated to form the initial item pool for the SERVQUAL instrument. This process resulted in the generation of 97 items (approximately 10 items per dimension). Each item was recast into two statements—one to measure expectations about firms in general within the service category being investigated and the other to measure perceptions about the particular firm whose service quality was being assessed. Roughly half of the statement pairs were worded positively and the rest were worded negatively, in accordance with recommended procedures for scale development (Churchill 1979). A seven-point scale ranging from "Strongly Agree" (7) to "Strongly Disagree" (1), with no verbal labels for scale points 2 through 6, accompanied each statement (scale values were reversed for negatively worded statements prior to data analysis). The expectation statements were grouped together and formed the first half of the instrument. The corresponding perception statements formed the second half. An abbreviated version of the instrument, containing a set of expectation statements (labeled as E's) and a corresponding set of perception

statements (labeled as P's), along with directions for responding to them, is included in the appendix. Negatively worded statements are identified by a minus sign within parentheses in the appendix.

DATA COLLECTION AND SCALE PURIFICATION

The 97-item instrument was subjected to two stages of data collection and refinement. The first stage focused on: (1) condensing the instrument by retaining only those items capable of discriminating well across respondents having differing quality perceptions about firms in several categories, and (2) examining the dimensionality of the scale and establishing the reliabilities of its components. The second stage was primarily confirmatory in nature and involved re-evaluating the condensed scale's dimensionality and reliability by analyzing fresh data from four independent samples. Some further refinements to the scale occurred in this stage.

Data Collection, First Stage

Data for initial refinement of the 97-item instrument were gathered from a quota sample of 200 adult respondents (25 years of age or older) recruited by a marketing research firm in a shopping mall in a large metropolitan area in the Southwest. The sample size of 200 was chosen because other scale developers in the marketing area had used similar sample sizes to purify initial instruments containing about the same number of items as the 97-item instrument (e.g., Churchill, Ford, and Walker 1974; Saxe and Weitz 1982). The sample was about equally divided between males and females. Furthermore, the respondents were spread across five different service categories—appliance repair and maintenance, retail banking, long-distance telephone, securities brokerage, and credit cards. These categories were chosen to represent a broad cross-section of services that varied along key dimensions used by Lovelock (1980, 1983) to classify services. For each service category, a quota of 40 recent users of the service was established. To qualify for the study, respondents had to have used the service in question during the past three months.

Screened and qualified respondents self administered a two-part questionnaire consisting of a 97-statement expectations part followed by a 97-statement perceptions part. For the first part, respondents were instructed to indicate the level of service that should be offered by firms within the service category in question. For the second part, respondents were first asked to name a firm (within the service category) that they had used and with which they were most familiar. Respondents were then instructed to express their perceptions about the firm.

Scale Purification, First Stage

The 97-item instrument was refined by analyzing pooled data (i.e., data from all five service categories considered together). The pooling of data was deliberate and appropriate because the basic purpose of this research stage was to develop a concise instrument that would be reliable and meaningful in assessing quality in a variety of service sectors. In other words, the purpose was to produce a scale that would have general applicability.

Purification of the instrument began with the computation of coefficient alpha (Cronbach 1951), in accordance with Churchill's (1979) recommendation. Because of the multidimensionality of the service-quality construct, coefficient alpha was computed separately for the 10 dimensions to ascertain the extent to which items making up each dimension shared a common core.

The raw data used in computing coefficient alpha (and in subsequent analyses) were in the form of difference scores. Specifically, for each item a difference score Q (representing perceived quality along that item) was defined as $Q = P - E$, where P and E are the ratings on the corresponding perception and expectation statements, respectively. The idea of using difference scores in purifying a multiple-item scale is not new. This approach has been used in developing scales for measuring constructs such as role conflict (Ford, Walker, and Churchill 1975).

The values of coefficient alpha ranged from .55 to .78 across the 10 dimensions and suggested that deletion of certain items from each dimension would improve the alpha values. The criterion used in deciding whether to delete an item was the item's corrected item-to-total correlation (i.e., correlation between the score on the item and the sum of scores on *all other* items making up the dimension to which the item was assigned). The corrected item-to-total correlations were plotted in descending order for each dimension. Items with very low correlations and/or those whose correlations produced a sharp drop in the plotted pattern were discarded. Recomputation of alpha values for the reduced sets of statements and examination of the new corrected item-to-total correlations led to further deletion of items whose elimination improved the corresponding alpha values. The iterative sequence of computing alphas and item-to-total correlations, followed by deletion of items, was repeated several times and resulted in a set of 54 items, with alpha values ranging from .72 to .83 across the 10 dimensions.

Examining the dimensionality of the 54-item scale was the next task in this stage of scale purification and was accomplished by factor analyzing

the difference scores on the 54 items. The principal axis factoring procedure (Harman 1967) was used and the analysis was constrained *a priori* to 10 factors. When the 10-factor solution was rotated orthogonally, no clear factor pattern emerged. Many of the items had high loadings on several factors, thereby implying that the factors may not be independent of one another. Moreover, some degree of overlap among the 10 conceptual dimensions was anticipated by the researchers who initially identified and labeled the dimensions (Parasuraman, Zeithaml, and Berry 1985). Therefore the 10-factor solution was subjected to oblique rotation (using the OBLIMIN procedure in SPSS-X) to allow for intercorrelations among the dimensions and to facilitate easy interpretation.

The oblique rotation produced a factor-loading matrix that was by and large easy to interpret. However, several items still had high loadings on more than one factor. When such items were removed from the factor-loading matrix, several factors themselves became meaningless because they had near-zero correlations with the remaining items, thereby suggesting a reduction in the presumed dimensionality of the service-quality domain. Furthermore, the highest loadings of a few of the remaining items were on factors to which they were not originally assigned. In other words, the factor loadings suggested reassignment of some items.

The deletion of certain items (and the resultant reduction in the total number of factors or clusters of items) and the reassignment of certain others necessitated the recomputation of alphas and item-to-total correlations and the reexamination of the factor structure of the reduced item pool. This iterative sequence of analyses (Step 5 in Figure 1) was repeated a few times and resulted in a final pool of 34 items representing seven distinct dimensions. The alpha values and factor loadings pertaining to the 34-item instrument are summarized in Table 1.

As shown in Table 1, five of the original 10 dimensions—tangibles, reliability, responsiveness, understanding/knowing customers, and access—remained distinct. The remaining five dimensions—communication, credibility, security, competence, and courtesy—collapsed into two distinct dimensions (D4 and D5), each consisting of items from several of the original five dimensions. The average pairwise correlation among the seven factors following oblique rotation was .27. This relatively low correlation, along with the relatively high factor loadings shown in Table 1, suggested that service quality might have seven fairly unique facets.

The high alpha values indicated good internal consistency among items within each dimension. Moreover, the combined reliability for the 34-item scale, computed by using the formula for the reliability of linear combinations (Nunnally 1978), was quite high (.94). Therefore, the 34-item in-

TABLE 1

Summary of Results from First Stage of Scale Purification

Dimension	Label	Reliability Coefficients (Alphas)	Number of Items	Factor Loadings of Items on Dimensions to Which They Belong[a]
Tangibles	D1	.72	4	63 75 62 47
Reliability	D2	.83	5	74 56 73 71 47
Responsiveness	D3	.84	5	60 73 59 76 66
Communication Credibility Security Competence Courtesy	D4	.79	4	35 53 66 56
	D5	.85	7	41 62 47 50 75 52 54
Understanding/ Knowing Customers	D6	.85	4	80 76 62 77

Access	D7	.78	5	57
				50
				75
				52
				71
Reliability of Linear Combination (Total-Scale Reliability)		.94		

[a] Numbers are the magnitudes of the factor loadings multiplied by 100. The loadings of items on dimensions to which they did not belong were all less than .3. The percentage of variance extracted by the seven factors was 61.7%.

strument was considered to be ready for further testing with data from new samples.

Data Collection, Second Stage

To further evaluate the 34-item scale and its psychometric properties, data were collected pertaining to the service quality of four nationally-known firms: a bank, a credit-card company, a firm offering appliance repair and maintenance services, and a long-distance telephone company. For each firm, an indpendent shopping-mall sample of 200 customers 25 years-of-age or older were recruited by a marketing research firm in a major metropolitan area in the East. To qualify for the study, respondents had to have used the services of the firm in question within the past three months. Each sample was divided about equally between males and females. As in the first stage of data collection, questionnaires were self-administered by qualified respondents.

Scale Purification, Second Stage

A major objective of this stage was to evaluate the robustness of the 34-item scale when used to measure the service quality of the four firms. Therefore the data from each of the four samples were analyzed separately to obtain alpha values (along with corrected item-to-total correlations) and a factor-loading matrix following oblique rotation of a seven-factor solution. The results from each sample facilitated cross-validation of the results from the other samples.

The results of the four sets of analyses were quite consistent, but differed somewhat from the first-stage findings summarized in Table 1. Spe-

cifically, two differences emerged. First, the corrected item-to-total correlations for several items (particularly among items making up the dimensions labeled D4 and D7 in Table 1) and the alphas for the corresponding dimensions were lower than those obtained from the first stage. Second, the factor-loading matrices obtained from all four analyses showed much greater overlap between dimensions D4 and D5, and between dimensions D6 and D7. Because these differences occurred consistently across four independent samples and data sets, further purification of the 34-item scale was deemed necessary.

A few items with relatively low item-to-total correlations were deleted. Furthermore, as suggested by the factor analyses, the items remaining in D4 and D5, as well as those in D6 and D7, were combined to form two separate dimensions. For each sample, alpha values were recomputed for the reduced set of five dimensions and a factor analysis (involving extraction of five factors followed by oblique rotation) was performed. In examining the results of these analyses, an iterative sequence similar to the one shown in Step 5 in Figure 1 was followed. This procedure resulted in a refined scale ("SERVQUAL") with 22 items spread among five dimensions (D1, D2, D3, a combination of D4 and D5, and a combination of D6 and D7). The expectation and perception statements in the final SERVQUAL instrument are shown in the appendix.

An examination of the content of the final items making up each of SERVQUAL's five dimensions (three original and two combined dimensions) suggested the following labels and concise definitions for the dimensions:

Tangibles:	Physical facilities, equipment, and appearance of personnel
Reliability:	Ability to perform the promised service dependably and accurately
Responsiveness:	Willingness to help customers and provide prompt service
Assurance:	Knowledge and courtesy of employees and their ability to inspire trust and confidence
Empathy:	Caring, individualized attention the firm provides its customers

The last two dimensions (assurance and empathy) contain items representing seven original dimensions—communication, credibility, security, competence, courtesy, understanding/knowing customers, and access—that did not remain distinct after the two stages of scale purification.

Therefore, while SERVQUAL has only five distinct dimensions, they capture facets of all 10 originally conceptualized dimensions.

SERVQUAL'S RELIABILITY AND FACTOR STRUCTURE

Table 2 shows the component and total reliabilities of SERVQUAL for each of the four samples. The reliabilities are consistently high across all four samples, with the possible exception of a couple of values pertaining to the tangible dimension. The total-scale reliability (i.e., reliability of linear combination) is close to .9 in each of the four instances.

Results of the factor analyses of data from the four samples are summarized in Table 3. The overall patterns of factor loadings are remarkably similar across the four independent sets of results. With few exceptions, items assigned to each dimension consistently have high loadings on only one of the five factors extracted. The distinctiveness of SERVQUAL's five dimensions implied by the results in Table 3 was further supported by relatively low intercorrelations among the five factors—the average pairwise correlations between factors following oblique rotation were .21, .24, .26, and .23 for the bank, credit card, repair and maintenance, and long-distance telephone samples, respectively.[1]

As an additional verification of the reliabilities and factor structure of SERVQUAL, the first-stage data set that resulted in the 34-item instrument with seven dimensions was reanalyzed after deleting the 12 items that dropped out during the second stage of scale purification. The results of this reanalysis are summarized in Table 4 and reconfirm the high reliabilities and dimensional distinctiveness of the scale. The average pairwise correlation among the five factors following oblique rotation was .35.

It is worth noting that the interative procedure used to refine the initial instrument was guided by empirical criteria and by the goal of obtaining a concise scale whose items would be meaningful to a variety of service firms. The reliabilities and factor structures indicate that the final 22-item scale and its five dimensions have sound and stable psychometric properties. Moreover, by design, the iterative procedure retained only those items that are common and relevant to all service firms included in the study. However, by the same token, this procedure may have deleted certain "good" items relevant to some but not all firms. Therefore, while SERVQUAL can be used in its present form to assess and compare service

[1] Complete matrices of the interfactor correlations can be obtained from the first author.

TABLE 2

Internal Consistencies of the Five Service-Quality Dimensions Following Second Stage of Scale Purification

Dimension	Label	Number of Items	Reliability Coefficients (Alphas)[a] B	CC	R&M	LDT	Items[b]
Tangibles	F1	4	.52	.62	.64	.64	Q1 Q2 Q3 Q4
Reliability	F2	5	.80	.78	.84	.74	Q5 Q6 Q7 Q8 Q9
Responsiveness	F3	4	.72	.69	.76	.70	Q10 Q11 Q12 Q13
Assurance	F4	4	.84	.80	.87	.84	Q14 Q15 Q16 Q17
Empathy	F5	5	.71	.80	.72	.76	Q18 Q19 Q20 Q21 Q22
Reliability of Linear Combination (Total-Scale Reliability)			.87	.89	.90	.88	

[a] B = Bank; CC = Credit Card Company; R&M = Repair and Maintenance Company; LDT = Long-Distance Telephone Company

[b] The item numbers correspond to those of the expectation and perception statements in the appendix.

TABLE 3

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

	FACTOR LOADINGS																			
	Bank					Credit Card Co.					Repair & Maintenance Co.					L-D Telephone Co.				
Items	F1	F2	F3	F4	F5	F1	F2	F3	F4	F5	F1	F2	F3	F4	F5	F1	F2	F3	F4	F5
Q1	34	28	—	—	—	36	—	35	—	—	34	—	—	—	—	42	—	—	—	—
Q2	64	—	—	—	—	70	—	—	—	—	70	—	—	—	—	72	—	—	—	—
Q3	39	—	—	28	—	52	—	—	—	—	53	—	—	—	—	51	—	—	—	—
Q4	28	—	—	28	—	52	—	—	—	—	65	—	—	—	—	59	—	—	30	—
Q5	—	72	—	—	—	—	54	—	—	—	—	73	—	—	—	—	52	—	—	—
Q6	—	63	—	—	—	—	43	27	—	—	—	51	—	—	—	—	40	—	—	—
Q7	—	71	—	—	—	—	87	—	—	—	—	84	—	—	—	—	79	—	—	—
Q8	—	80	—	—	—	—	83	—	—	—	—	88	—	—	—	—	59	—	—	—
Q9	—	39	—	—	—	—	49	—	—	—	—	29	—	30	—	—	54	—	—	—

Q10	—	—	37	—	—	—	—	43	—	26	—	—	56	—	—	—	—	39	—	—
Q11	—	—	55	—	—	—	—	48	—	—	—	—	52	—	—	—	—	43	—	—
Q12	—	—	62	—	—	—	—	54	—	—	—	—	74	—	—	—	—	92	—	—
Q13	—	—	69	—	—	—	—	33	—	—	—	—	71	—	—	—	—	53	—	—
Q14	—	—	—	68	—	—	—	—	65	—	—	—	—	86	—	—	—	—	69	—
Q15	—	—	—	84	—	—	—	—	76	—	—	—	—	89	—	—	—	—	81	—
Q16	—	—	—	72	—	—	—	—	73	—	—	—	—	65	—	—	—	—	61	—
Q17	—	—	—	64	—	—	—	—	61	—	—	—	—	64	—	—	—	—	66	—
Q18	—	—	—	—	37	—	—	—	—	64	—	—	—	—	42	—	—	—	—	59
Q19	—	—	—	—	48	—	—	—	—	72	—	—	—	—	61	—	—	—	—	79
Q20	—	—	—	—	41	—	—	—	—	63	—	28	34	—	46	—	—	—	—	55
Q21	—	—	—	—	33	—	—	—	—	59	—	—	—	—	32	—	—	—	—	36
Q22	—	—	—	—	68	—	—	—	—	64	—	—	—	—	61	—	—	—	—	59

[a] All numbers in the table are magnitudes of the factor loadings multiplied by 100. Loadings that are .25 or less are not shown. The percentage of variance extracted by the five factors in the bank, credit card, repair and maintenance, and long-distance telephone samples were 56.0%, 57.5%, 61.6%, and 56.2%, respectively.

quality across a wide variety of firms or units within a firm, appropriate adaptation of the instrument may be desirable when only a single service is investigated. Specifically, items under each of the five dimensions can be suitably reworded and/or augmented to make them more germane to the context in which the instrument is to be used.

ASSESSMENT OF SERVQUAL'S VALIDITY

SERVQUAL's high reliabilities and consistent factor structures across several independent samples provide support for its trait validity (Campbell 1960; Peter 1981). However, while high reliabilities and internal consistencies are necessary conditions for a scale's construct validity—the extent to which a scale fully and unambiguously captures the underlying, unobservable, construct it is intended to measure—they are not sufficient (Churchill 1979). The scale must satisfy certain other conceptual and empirical criteria to be considered as having good construct validity.

The basic conceptual criterion pertaining to construct validity is face or content validity. (Does the scale appear to measure what it is supposed to? Do the scale items capture key facets of the unobservable construct being measured?) Assessing a scale's content validity is necessarily qualitative rather than quantitative. It involves examining two aspects: (1) the thoroughness with which the construct to be scaled and its domain were explicated and (2) the extent to which the scale items represent the construct's domain. As discussed in earlier sections, the procedures used in developing SERVQUAL satisfied both these evaluative requirements. Therefore the scale can be considered to possess content validity.

The scale's validity was also assessed empirically by examining its convergent validity—i.e., the association between SERVQUAL scores and responses to a question that asked customers to provide an overall quality rating of the firm they were evaluating. Respondents in the second stage of data collection rated the service firm's overall quality (referred to hereafter as "Overall Q") by checking one of four categories—excellent, good, fair, poor. The correspondence between the Overall Q ratings and the SERVQUAL scores was examined using one-way ANOVA. The treatment variable in the ANOVA's was Overall Q—with three categories instead of four because very few respondents checked "poor," thereby necessitating creation of a combined "fair/poor" category. The dependent variable was the average difference score (i.e., perception-minus-expectation score) on each SERVQUAL dimension as well as on the total SERVQUAL scale (separate ANOVA's were conducted for each dimension and for the total scale). Significant ANOVA results were investigated further

TABLE 4

Reanalysis of First-Stage Data for the Five-Dimensional Scale

Dimension	Label	Number of Items	Reliability Coefficients (Alphas)	Items	Factor Loadings of Items on Dimensions to Which They Belong[a]
Tangibles	F1	4	.72	Q1	69
				Q2	68
				Q3	64
				Q4	51
Reliability	F2	5	.83	Q5	75
				Q6	63
				Q7	71
				Q8	75
				Q9	50
Responsiveness	F3	4	.82	Q10	51
				Q11	77
				Q12	66
				Q13	86
Assurance	F4	4	.81	Q14	38
				Q15	72
				Q16	80
				Q17	45
Empathy	F5	5	.86	Q18	78
				Q19	81
				Q20	59
				Q21	71
				Q22	68
Reliability of Linear Combination (Total-Scale Reliability)			.92		

[a] Numbers are magnitudes of the factor loadings multiplied by 100. The loadings of items on dimensions to which they did not belong were all less than .3. The percentage of variance extracted by the five factors was 63.2%.

using Duncan's multiple range test to identify significant differences across the Overall Q categories. The results of these analyses for each of the four samples are summarized in Table 5 under the heading "Overall Q".

The numbers reported in Table 5 are average SERVQUAL scores

within each Overall Q category, measured on a −6 to +6 scale on which the higher (less negative) the score, the higher is the level of perceived service quality. In each of the four samples, the combined SERVQUAL score for those in the "excellent" category is significantly higher (less negative) than for those in the "good" category. Furthermore, respondents in the "good" category have a significantly higher combined SERVQUAL score than those in the "fair/poor" category. A similar pattern of findings is evident for the scores on the individual SERVQUAL dimensions as well. The strength and persistence of the linkage between the Overall Q categories and the SERVQUAL scores across four independent samples offer strong support for SERVQUAL's convergent validity.

SERVQUAL's validity was further assessed by examining whether the construct measured by it was empirically associated with measures of other conceptually related variables. Respondents in each sample answered two general questions that provided measures of variables (labeled "Recommend" and "Problem" in Table 5) which one could expect to be related conceptually to perceived service quality: (1) whether the respondents would recommend the service firm to a friend and (2) whether they had ever reported a problem with the services they received from the firm. Respondents answering yes to the first (Recommend) question and no to the second (Problem) question could be hypothesized to perceive higher service quality than other respondents. As Table 5 shows, the results are consistent with this hypothesis. These findings provide additional support for SERVQUAL's validity.

APPLICATIONS OF SERVQUAL

It is difficult to identify any retailer that offers no services whatsoever. Some retailers offer facilitating services, such as sales assistance and delivery, to help sell goods. Some retailers sell services directly, in addition to offering facilitating services. Some retailers sell only services. Quality of service is an important issue for all of these retailers. Competing goods retailers (department stores, supermarkets) may sell many identical products and quality of service is a primary means of competitive differentiation. Retailers that sell only services (telephone companies, airlines) have little to offer if their service is poor (Berry 1986).

SERVQUAL is a concise multiple-item scale with good reliability and validity that retailers can use to better understand the service expectations and perceptions of consumers and, as a result, improve service. The instrument has been designed to be applicable across a broad spectrum of services. As such, it provides a basic skeleton through its expectations/

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

SERVQUAL is most valuable when it is used periodically to track service quality trends, and when it is used in conjunction with other forms of service quality measurement. A retailer, for example, would learn a great deal about its service quality and what needs to be done to improve it by administering both SERVQUAL and an employee survey three or four times a year, plus systematically soliciting and analyzing customer suggestions and complaints. The employee survey should include questions concerning perceived impediments to better service, e.g., what is the biggest problem you face trying to deliver high-quality service to your customers? If you could be president for a day, what one change would you make in the company to improve quality of service?

SERVQUAL can be used to assess a given firm's quality along each of the five service dimensions by averaging the difference scores on items making up the dimension. It can also provide an overall measure of service quality in the form of an average score across all five dimensions. Because meaningful responses to the perception statements require respondents to have some knowledge of or experience with the firm being researched, SERVQUAL is limited to current or past customers of that firm. Within this constraint, a variety of potential applications are available.

One potential application of SERVQUAL is to determine the *relative importance* of the five dimensions in influencing customers' overall quality perceptions. An approach for doing this is to regress the overall quality perception scores on the SERVQUAL scores for the individual dimensions. The results of such a regression analysis for the four companies in the present study are shown in Table 6 (the dependent variable was Overall Q, coded as excellent = 4, good = 3, fair = 2, and poor = 1).

The adjusted R^2 values are statistically significant in all four cases and are also quite respectable, particularly in view of the fact that the dependent variable had only four categories, and the first three accounted for most of the responses. A striking result in terms of the relative importance of the five dimensions in predicting overall quality is that reliability is consistently the most critical dimension. Assurance is the second most important dimension in all four cases. Tangibles is more important in the case of the bank than in the other three firms, while the reverse is true for responsiveness. Empathy is the least important dimension in all four cases. However, the relatively small magnitudes of the regression coeffi-

TABLE 5

Significant Differences in Mean Scale Values for Respondents—Segmented According to the Variables Overall Q, Recommend, and Problem[a]

Bank							
Individual Scale Dimensions	Overall Q			Recommend		Problem	
	Excellent	Good	Fair/Poor	Yes	No	Yes	No
Tangibles	−0.04[b]	−0.52[c]	−1.08[d]	−0.41[b]	−0.98[c]	−0.75[b]	−0.45[b]
Reliability	−0.25[b]	−0.96[c]	−2.30[d]	−0.82[b]	−2.21[c]	−1.55[b]	−0.92[c]
Responsiveness	−0.32[b]	−0.97[c]	−1.54[c]	−0.74[b]	−1.81[c]	−1.22[b]	−0.84[b]
Assurance	−0.49[b]	−1.03[c]	−1.98[d]	−0.88[b]	−2.12[c]	−1.52[b]	−0.96[c]
Empathy	−0.30[b]	−1.02[c]	−1.52[c]	−0.76[b]	−1.88[c]	−1.07[b]	−0.91[b]
Combined scale	−0.22[b]	−0.92[c]	−1.61[d]	−0.72[b]	−1.77[c]	−1.22[b]	−0.80[c]
Sample size	46	112	40	164	33	47	151
Credit Card Company							
Individual Scale Dimensions	Overall Q			Recommend		Problem	
	Excellent	Good	Fair/Poor	Yes	No	Yes	No
Tangibles	0.06[b]	−0.61[c]	−0.79[c]	−0.39[b]	−0.80[b]	−0.76[b]	−0.29[c]
Reliability	−0.42[b]	−0.94[c]	−2.32[d]	−0.82[b]	−2.50[c]	−1.42[b]	−0.82[c]
Responsiveness	−0.08[b]	−1.13[c]	−1.71[c]	−0.75[b]	−2.59[c]	−1.31[b]	−0.77[c]
Assurance	−0.59[b]	−1.31[c]	−2.29[d]	−1.08[b]	−2.83[c]	−1.49[b]	−1.15[b]
Empathy	−0.50[b]	−1.38[c]	−1.94[c]	−1.03[b]	−2.77[c]	−1.62[b]	−1.01[b]
Combined Scale	−0.32[b]	−1.10[c]	−1.79[d]	−0.83[b]	−2.27[c]	−1.29[b]	−0.83[c]
Sample Size	60	112	28	183	17	50	149

Repair and Maintenance Company

Individual Scale Dimensions	Overall Q			Recommend		Problem	
	Excellent	Good	Fair/Poor	Yes	No	Yes	No
Tangibles	-0.15^{b}	$-0.40^{b,c}$	-0.86^{c}	-0.36^{b}	-0.85^{b}	-0.58^{b}	-0.34^{b}
Reliability	-0.48^{b}	-1.30^{c}	-3.20^{d}	-1.14^{b}	-3.48^{c}	-2.14^{b}	-1.18^{c}
Responsiveness	-0.08^{b}	-1.08^{c}	-2.41^{d}	-0.83^{b}	-2.54^{c}	-1.71^{b}	-0.80^{c}
Assurance	-0.33^{b}	-1.35^{c}	-2.84^{d}	-1.16^{b}	-2.91^{c}	-2.04^{b}	-1.13^{c}
Empathy	0.15^{b}	-1.11^{c}	-2.17^{d}	-0.85^{b}	-2.19^{c}	-1.67^{b}	-0.74^{c}
Combined Scale	-0.16^{b}	-1.07^{c}	-2.30^{d}	-0.88^{b}	-2.40^{c}	-1.65^{b}	-0.85^{c}
Sample Size	45	114	40	168	30	65	132

Long-Distance Telephone Company

Individual Scale Dimensions	Overall Q			Recommend		Problem	
	Excellent	Good	Fair/Poor	Yes	No	Yes	No
Tangibles	-0.08^{b}	-0.44^{c}	-0.50^{c}	-0.26^{b}	-0.95^{c}	-0.42^{b}	-0.26^{b}
Reliability	-0.45^{b}	-1.42^{c}	-2.53^{d}	-1.05^{b}	-2.71^{c}	-1.54^{b}	-1.03^{c}
Responsiveness	-0.30^{b}	-1.43^{c}	-1.90^{c}	-1.00^{b}	-2.03^{c}	-1.46^{b}	-0.86^{c}
Assurance	-0.39^{b}	-1.45^{c}	-2.10^{d}	-1.00^{b}	-2.64^{c}	-1.62^{b}	-0.87^{c}
Empathy	-0.33^{b}	-1.19^{c}	-2.10^{d}	-0.86^{b}	-2.34^{c}	-1.16^{b}	-0.90^{b}
Combined Scale	-0.30^{b}	-1.15^{c}	-1.83^{d}	-0.83^{b}	-2.13^{c}	-1.24^{b}	-0.76^{c}
Sample Size	69	104	25	178	19	78	120

[a] Numbers are *mean values* on a scale ranging from -6 to $+6$, on which zero implies that consumer perceptions and expectations coincide, negative values imply that perceptions fall short of expectations, and positive values imply that perceptions exceed expectations.

[b,c,d] Means with the *same* superscripts are not significantly different. Means with *different* superscripts are significantly different.

TABLE 6

Relative Importance of the Five Dimensions in Predicting Overall Quality

Dimension	Standardized Slope Coefficient	Significance Level of Slope[a]	Adusted R^2
Bank			
Tangibles	.13	.07	.28 ($p < .00$)
Reliability	.39	.00	
Responsiveness	.07	.35	
Assurance	.13	.09	
Empathy	.01	.89	
Credit Card Co.			
Tangibles	.07	.26	.27 ($p < .00$)
Reliability	.33	.00	
Responsiveness	.12	.11	
Assurance	.17	.02	
Empathy	.04	.58	
Repair & Maintenance Co.			
Tangibles	.04	.48	.52 ($p < .00$)
Reliability	.54	.00	
Responsiveness	.11	.09	
Assurance	.16	.02	
Empathy	.01	.81	
L-D Telephone Co.			
Tangibles	.08	.17	.37 ($p < .00$)
Reliability	.45	.00	
Responsiveness	.12	.09	
Assurance	.15	.03	
Empathy	.02	.78	

[a] Significance levels are for two-tailed tests.

cients for empathy and their lack of statistical significance should be interpreted with caution because empathy did have a statistically significant simple correlation with overall quality, ranging from .20 in the case of the bank to .40 in the case of the repair and maintenance company. Empathy also had significant correlations of the same order of magnitude with reli-

ability and assurance (the two most important dimensions), thereby implying that its importance in the regression analyses may have been masked somewhat by possible multicollinearity. Therefore, while empathy is apparently the least important of the five SERVQUAL dimensions, it is by no means *un*important.

Another application of the instrument is its use in categorizing a firm's customers into several perceived-quality segments (e.g., high, medium, and low) on the basis of their individual SERVQUAL scores. These segments then can be analyzed on the basis of (1) demographic, psychographic and/or other profiles; (2) the relative importance of the five dimensions in influencing service quality perceptions; and (3) the reasons behind the perceptions reported. For example, suppose a department store found that a large number of SERVQUAL respondents falling in the "medium" perceived-quality group fit its prime target market based on demographic and psychographic criteria. Suppose further that reliability and assurance were found to be the most important quality dimensions and, based on perception-expectation gap scores for items concerning these dimensions, the items relating to record-keeping accuracy and behavior of contact personnel revealed the biggest gaps. With these data, the department store's management would understand better what needs to be done to improve its image in the eyes of a very important group—customers within the firm's prime target market who give the firm "medium" service quality scores and who are in position to either respond to improved service from the firm or defect to the competition.

SERVQUAL can also be used by multi-unit retail companies to track the level of service provided by each store in the chain. By asking respondents to indicate the particular store in the chain with which they are most familiar, and to provide perception responses for that unit, the researcher can compare each store's average SERVQUAL score with the scores from other stores. Service quality scores can then be a factor in store manager performance appraisals and compensation, among other uses. Also, SERVQUAL scores for the individual stores can be used to group outlets into several clusters with varying quality images. A careful examination of the characteristics of the stores in the different clusters may reveal key attributes that facilitate—or hinder—the delivery of high quality service.

A retailer can also use SERVQUAL to assess its service performance relative to its principal competitors. The two-section format of the instrument, with separate expectation and perception sections, makes it convenient to measure the quality of several firms simply by including a set of perception statements for each firm. The expectations section does not have to be repeated for each firm. For example, a supermarket chain could

include its two principal competitors in a total market survey, asking respondents to provide perception ratings for each of the companies with which they have shopping experience. A retailer that uses SERVQUAL to identify the most salient service quality dimensions for its target markets, and to compare itself to the competition in terms of strengths and weaknesses on these particular dimensions, will certainly have a sense of what its priorities should be with regard to service quality.

In summary, SERVQUAL has a variety of potential applications. It can help a wide range of service and retailing organizations in assessing consumer expectations about and perceptions of service quality. It can also help in pinpointing areas requiring managerial attention and action to improve service quality. In addition, we hope the availability of this instrument will stimulate much-needed empirical research focusing on service quality and its antecedents and consequences.

REFERENCES

Berry, Leonard L. (1986), "Retail Businesses Are Service Businesses," *Journal of Retailing,* **62** (Spring), 3–6.

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

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

Churchill, Gilbert A., Jr., Neil M. Ford, and Orville C. Walker, Jr. (1974), "Measuring the Job Satisfaction of Industrial Salesmen," *Journal of Marketing Research,* **11** (August), 254–260.

Cronbach, Lee J. (1951), "Coefficient Alpha and the Internal Structure of Tests," *Psychometrika,* **16** (October), 297–334.

Crosby, Philip B. (1979), *Quality is Free: The Art of Making Quality Certain,* New York: New American Library.

Dodds, William B. and Kent B. Monroe (1984), "The Effect of Brand and Price Information on Subjective Product Evaluations," *Advances in Consumer Research XII.*

Ford, Neil M., Orville C. Walker, Jr., and Gilbert A. Churchill, Jr. (1975), "Expectation-Specific Measures of the Intersender Conflict and Role Ambiguity Experienced by Industrial Salesmen," *Journal of Business Research,* **3** (April), 95–112.

Garvin, David A. (1983), "Quality on the Line," *Harvard Business Review,* **61** (September-October), 65–73.

Gronroos, Christian (1982), *Strategic Management and Marketing in the Service Sector,* Helsingfors: Swedish School of Economics and Business Administration.

Harman, Harry H. (1967), *Modern Factor Analysis,* 2nd ed., Chicago: The University of Chicago Press.

Hjorth-Anderson, Chr. (1984), "The Concept of Quality and the Efficiency of Markets for Consumer Products," *Journal of Consumer Research,* **11** (September), 708–718.

Holbrook, Morris B. and Kim P. Corfman (1985), "Quality and Value in the Consumption Experience: Phaldrus Rides Again," in *Perceived Quality,* J. Jacoby and J. Olson (eds.), Lexington, Massachusetts: Lexington Books, 31–57.

Howard, John and Jagdish Sheth (1969), *The Theory of Buyer Behavior,* New York: John Wiley and Sons.

Hunt, Keith (1979), *Conceptualization and Measurement of Consumer Satisfaction and Dissatisfaction,* Cambridge, Mass.: Marketing Science Institute.

Jacoby, Jacob (1978), "Consumer Research: A State of the Art Review," *Journal of Marketing,* **42** (April), 87–96.

Jacoby, J. and J. Olson, eds. (1985), *Perceived Quality,* Lexington, Massachusetts: Lexington Books.

Lehtinen, Uolevi and Jarmo R. Lehtinen (1982), "Service Quality: A Study of Quality Dimensions," unpublished working paper, Helsinki: Service Management Institute, Finland OY.

Lovelock, Christopher H. (1980), "Towards a Classification of Services," in *Emerging Perspectives on Services Marketing,* L. L. Berry, G. L. Shostack, and G. Upah (eds.), Chicago: American Marketing Association, 72–76.

——— (1983), "Classifying Services to Gain Strategic Marketing Insights," *Journal of Marketing,* **47** (Summer), 9–20.

Nunnally, Jum C. (1978), *Psychometric Theory,* 2nd ed., New York: McGraw-Hill Book Company.

Oliver, Richard (1981), "Measurement and Evaluation of Satisfaction Process in Retail Settings," *Journal of Retailing,* **57** (Fall), 25–48.

Olshavsky, Richard W. (1985), "Perceived Quality in Consumer Decision Making: An Integrated Theoretical Perspective," in *Perceived Quality,* J. Jacoby and J. Olson (eds.), Lexington, Massachusetts: Lexington Books.

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

Peter, J. Paul (1981), "Construct Validity: A Review of Basic Issues and Marketing Practices," *Journal of Marketing Research,* **18** (May), 133–45.

Rudie, Mary, J. and H. Brant Wansley (1985), "The Merrill Lynch Quality Program," in *Services Marketing in a Changing Environment,* Thomas Bloch, Gregory Upah, and Valarie A. Zeithaml (eds.), Chicago, IL: American Marketing Association.

Sasser, W. Earl, Jr., R. Paul Olsen, and D. Daryl Wyckoff (1978), *Management of Service Operations: Text and Cases,* Boston: Allyn & Bacon.

Saxe, Robert and Barton A. Weitz (1982), "The SOCO Scale: A Measure of the Customer Orientation of Salespeople," *Journal of Marketing,* **19** (August), 343–51.

Thompson, Phillip, Glenn DeSouza, and Bradley T. Gale (1985), *The Strategic Management of Service Quality,* Cambridge, MA: The Strategic Planning Institute, PIMSLETTER No. 33.

Zeithaml, Valarie (1987), *Defining and Relating Price, Perceived Quality, and Perceived Value,* Report No. 87-101, Cambridge, MA: Marketing Science Institute.

APPENDIX

THE SERVQUAL INSTRUMENT[a]

DIRECTIONS: This survey deals with your opinions of ——— services. Please show the extent to which you think firms offering ——— services should possess the features described by each statement. Do this by picking one of the seven numbers next to each statement. If you strongly agree that these firms should possess a feature, circle the number 7. If you strongly disagree that these firms should possess a feature, circle 1. If your feelings are not 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 best shows your expectations about firms offering ——— services.

E1. They should have up-to-date equipment.
E2. Their physical facilities should be visually appealing.
E3. Their employees should be well dressed and appear neat.
E4. The appearance of the physical facilities of these firms should be in keeping with the type of services provided.
E5. When these firms promise to do something by a certain time, they should do so.
E6. When customers have problems, these firms should be sympathetic and reassuring.
E7. These firms should be dependable.
E8. They should provide their services at the time they promise to do so.
E9. They should keep their records accurately.
E10. They shouldn't be expected to tell customers exactly when services will be performed. (−)[b]
E11. It is not realistic for customers to expect prompt service from employees of these firms. (−)
E12. Their employees don't always have to be willing to help customers. (−)
E13. It is okay if they are too busy to respond to customer requests promptly. (−)
E14. Customers should be able to trust employees of these firms.
E15. Customers should be able to feel safe in their transactions with these firms' employees.
E16. Their employees should be polite.

E17. Their employees should get adequate support from these firms to do their jobs well.
E18. These firms should not be expected to give customers individual attention. (−)
E19. Employees of these firms cannot be expected to give customers personal attention. (−)
E20. It is unrealistic to expect employees to know what the needs of their customers are. (−)
E21. It is unrealistic to expect these firms to have their customers' best interests at heart. (−)
E22. They shouldn't be expected to have operating hours convenient to all their customers. (−)

DIRECTIONS: The following set of statements relate to your feelings about XYZ. For each statement, please show the extent to which you believe XYZ has the feature described by the statement. Once again, circling a 7 means that you strongly agree that XYZ has that feature, and circling a 1 means that you strongly disagree. 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.

P1. XYZ has up-to-date equipment.
P2. XYZ's physical facilities are visually appealing.
P3. XYZ's employees are well dressed and appear neat.
P4. The appearance of the physical facilities of XYZ is in keeping with the type of services provided.
P5. When XYZ promises to do something by a certain time, it does so.
P6. When you have problems, XYZ is sympathetic and reassuring.
P7. XYZ is dependable.
P8. XYZ provides its services at the time it promises to do so.
P9. XYZ keeps its records accurately.
P10. XYZ does not tell customers exactly when services will be performed. (−)
P11. You do not receive prompt service from XYZ's employees. (−)
P12. Employees of XYZ are not always willing to help customers. (−)
P13. Employees of XYZ are too busy to respond to customer requests promptly. (−)
P14. You can trust employees of XYZ.
P15. You feel safe in your transactions with XYZ's employees.
P16. Employees of XYZ are polite.

P17. Employees get adequate support from XYZ to do their jobs well.
P18. XYZ does not give you individual attention. (−)
P19. Employees of XYZ do not give you personal attention. (−)
P20. Employees of XYZ do not know what your needs are. (−)
P21. XYZ does not have your best interests at heart. (−)
P22. XYZ does not have operating hours convenient to all their customers. (−)

[a] A seven-point scale ranging from "Strongly Agree" (7) to "Strongly Disagree" (1), with no verbal labels for the intermediate scale points (i.e., 2 through 6), accompanied each statement. Also, the statements were in random order in the questionnaire. A complete listing of the 34-item instrument used in the second stage of data collection can be obtained from the first author.

[b] Ratings on these statements were reverse-scored prior to data analysis.