Adapting the SERVQUAL Scale to Hospital Services: An Empirical Investigation

Emin Babakus and W. Glynn Mangold

Defining and measuring the quality of service has been a major challenge for health care marketers. A comprehensive service quality measurement scale (SERVQUAL) is empirically evaluated for its potential usefulness in a hospital service environment. Active participation by hospital management helped to address practical and user-related aspects of the assessment. The completed expectations and perceptions scales met various criteria for reliability and validity. Suggestions are provided for the managerial use of the scale, and a number of future research issues are identified.

Evidence in both the manufacturing and services industries indicates that quality is a key determinant of market share and return on investment as well as cost reduction (Anderson and Zeithaml 1984; Parasuraman, Zeithaml, and Berry 1985). Two forms of quality are relevant to service-providing organizations: technical quality and functional quality (Grönroos 1984). Technical quality in the health care environment, also referred to as quality in fact, is defined primarily on the basis of the technical accuracy of the diagnoses and procedures. Various techniques for measuring technical quality have been proposed and are currently in use in health care organizations (Joint Commission for Accreditation of Health Care Organizations 1987). Because

---

Address correspondence and requests for reprints to Emin Babakus, Ph.D., Associate Professor, Department of Marketing, Memphis State University, Memphis, TN 38152. W. Glynn Mangold, Ph.D. is Associate Professor in the Department of Marketing at the University of Louisville, Louisville, KY. This article, submitted to *Health Services Research* on March 22, 1990, was revised and accepted for publication on January 17, 1991.

this information is not generally available to the consuming public, knowledge of the technical quality of health care services remains within the purview of health care professionals and administrators (Bopp 1990).

Functional quality refers to the manner in which the health care service is delivered to the patient. Since patients are often unable to accurately assess the technical quality of a health care service, functional quality is usually the primary determinant of patients' quality perceptions (Donabedian 1980, 1982; Kovner and Smits 1978). There is growing evidence to suggest that this perceived quality is the single most important variable influencing consumers' value perceptions. These value perceptions, in turn, affect consumers' intentions to purchase products or services (Bolton and Drew 1988; Zeithaml 1988).

Research suggests that service organizations share various commonalities in the service delivery process both within and across industries (Zeithaml, Berry, and Parasuraman 1988). For example, the intangible nature of services dictates that, unlike products, most services are produced and consumed at the same time. This characteristic increases the importance of the provider-consumer relationship as well as the potential for variation in service quality.

Consequently, identifying techniques that enhance service quality perceptions in one industry may enable researchers to develop generalizations applicable to other industries as well (Parasuraman, Zeithaml, and Berry 1985, 1986). The development and scientific examination of such generalizations, however, requires the use of standard measurement instruments that are applicable across the service industries studied (Heise 1974; Price and Mueller 1986). Parasuraman, Zeithaml, and Berry (1988) recently developed such a measurement tool, SERVQUAL, to be used in a variety of service industries. While SERVQUAL has been tested in a number of service settings, its applicability to the hospital environment has not yet been assessed.

Therefore, the purpose of this article is to report the results of a study that examined the usefulness of the SERVQUAL scale for assessing patients' perceptions of service quality in the hospital environment. At the practical level, the representativeness of the SERVQUAL items as they relate to hospital services was assessed. In addition to content appropriateness, the length of the scale was a major consideration for the population under study, in this case former patients of a hospital. The scale was subjected to extensive reliability and validity assessment. The potential usefulness of the study results were enhanced by the fact that health care practitioners were actively involved in the research process.

BACKGROUND

The SERVQUAL scale was developed based on a marketing perspective with the support of the Marketing Science Institute (Parasuraman, Zeithaml, and Berry 1986). Its purpose was to provide an instrument for measuring service quality that would apply across a broad range of services with minor modifications in the scale. SERVQUAL provides a foundation for a growing body of research that pertains to the creation of quality among service industries.

The developers of the scale contend that, while each service industry is unique in some aspects, there are five dimensions of service quality that are applicable to service-providing organizations in general. These dimensions are: (1) tangibles—physical facilities, equipment, and appearance of personnel; (2) reliability—ability to perform the promised service dependably and accurately; (3) responsiveness—willingness to help customers and provide prompt service; (4) assurance—knowledge and courtesy of employees and their ability to inspire trust and confidence; and (5) empathy—caring, the individualized attention the firm provides its customers (Parasuraman, Zeithaml, and Berry 1988).

The scale was developed and tested across four service environments: banking, credit card services, repair and maintenance, and long distance telephone services. In its final form, SERVQUAL contains 22 pairs of items. Half of these items are intended to measure consumers' expected level of service for a particular industry (expectations). The other 22 matching items are intended to measure consumer perceptions of the present level of service provided by a particular organization (perceptions). Both sets of items are presented in seven-point Likert response format, with the anchors "strongly agree" and "strongly disagree." Service quality is measured on the basis of the difference scores by subtracting expectation scores from the corresponding perception scores.

Putting service quality into operation as a difference or "gap" score is a consistent extension of the theoretical work of Parasuraman and his colleagues on the determinants of service quality. It is unique in the sense that the definition of the construct is based on the difference between expectations and perceptions. The construct is differentiated from consumer satisfaction in a way that defines the expectations/perceptions "gap" as an enduring perception about the overall excellence of a particular firm. This approach to defining and measuring service quality as the difference between expectations and perceptions

is a major departure from previous scale development efforts in health care services (e.g., Bopp 1990; Casarreal, Mills, and Plant 1986; Ware and Snyder 1975). While it should be pointed out that SERVQUAL is intended to measure functional quality rather than technical quality, this limitation is inherent in the fact that the technical aspects of the delivery process are, in most cases, industry specific (e.g., health care versus banking services). Hence, SERVQUAL can help researchers to identify general principles of functional service quality and to test the effectiveness of a given model among service-providing industries.

RESEARCH DESIGN AND METHODOLOGY

PRELIMINARY EVALUATION OF SERVQUAL

The present study was designed with the cooperation of a multihospital corporation. A midsized hospital, located in the southern part of the United States, was chosen as the pilot service organization. The management team was up-to-date with the current literature and emphasized the pragmatic aspects of the research. Their active involvement helped in the assessment of content validity and ensured that the research instrument would be of practical significance.

As the developers of SERVQUAL have pointed out, SERVQUAL "can be adapted or supplemented to fit the characteristics or specific research needs of a particular organization." (Parasuraman, Zeithaml, and Berry 1988, 31). Therefore, an initial evaluation of SERVQUAL was undertaken before the main data collection effort began, to ensure that SERVQUAL was tailored to the research needs characteristic of the hospital environment. Input was provided by both academicians (two marketing faculty) and the management team (marketing research director and two managers). Decisions to modify the instrument were based on the relevancy of the questions to hospital services and on the ability of patients to respond to the questions without experiencing confusion or undue frustration.

The initial meeting between the researchers and the management team resulted in a decision to discard several items that were not relevant to the hospital environment or that could lead to invalid responses. For example, the statement, "Customers should be willing to wait a little while to get appointments with these firms," was eliminated because the response obviously would depend on the nature of the illness.

The scale was then administered to a small sample of patients to gather further input. The pretest indicated that respondents perceived some of the items included in the scale to be redundant. Because this redundancy led to frustration and low response rates, the researchers and the management team agreed to reduce the number of items further. SERVQUAL items that showed high content validity were retained in each subscale. The final scale consisted of 15 pairs of matching expectation/perception items representing all five dimensions of service quality (see Appendix A for the list of retained items).

The preliminary test also indicated that the mixture of negatively and positively worded statements created confusion and frustration on the part of respondents. While it is theoretically appropriate to have such a mix of items (Churchill 1979; Likert 1932), the practice itself does not guarantee the prevention of yea-saying or nay-saying tendencies. For this particular population, it was believed that the confusion and inaccurate responses resulting from the use of negatively worded statements would adversely affect the quantity and the quality of the data. Therefore, the negatively worded statements contained in the research instrument were converted to positive connotations.

The decision not to use mixed connotations draws support from the literature also. For instance, the factor structure of role conflict and ambiguity scales (Rizzo, House, and Lirtzman 1970) has been attributed to the effects of wording (Howell et al. 1988). A more recent study indicates that the frequency distributions of the negatively worded SERVQUAL items were bimodal while the distributions of positive item scores were unimodal (Vogels, Lemmink, and Kasper 1989). Such results are indications of response quality problems resulting from the use of both negatively and positively worded questions, and are consistent with our pretest results.

Finally, a five-point Likert response format (ranging from "strongly agree = 5" to "strongly disagree = 1") was adopted instead of the original seven-point scale format. This modification was based on the management team's experience with previous surveys, which indicated that the five-point format would reduce the frustration level of the respondent patients, and would thereby increase the response rate and the quality of the responses.

SAMPLING AND THE DATA COLLECTION PROCEDURES

Questionnaires were mailed to 2,036 patients who had been discharged from the hospital within the previous 13 months. Thirty-seven ques-

tionnaires were returned by the postal service because of address problems. There were 443 returns from the remaining 1,999 questionnaires, yielding a 22 percent response rate. Most discharged patients receive a large volume of correspondence from the hospital relating to billing, insurance payments, and other matters. Because the management team believed this deluge of correspondence should be minimized to the greatest extent possible, no follow-up attempts were made. While the response rate in the current study may be considered low for attempting to make inferences for an entire target population, it is adequate for scale development and testing purposes (Press and Ganey 1989). In addition, a comparison of the respondents' gender and age composition and insurance status with those of the targeted population revealed no significant differences between the groups.

One section of the questionnaire contained the modified SERVQUAL scale, with 15 statements relating to patients' expectations on the quality of the service that hospitals should offer and 15 corresponding items relating to their perceptions of the quality of service actually delivered. This simultaneous administration of expectations and perceptions statements is consistent with the methodology employed by the developers of SERVQUAL. The instrument also contained a question about patients' overall perceptions of hospital quality, and a question about whether or not they intended to return to the same hospital if a need were to arise. The overall quality perceptions statement was measured on a five-point scale with end points labeled "very good" and "very poor." The statement was phrased as follows: "In terms of the quality of care received, what is your overall impression of _____ hospital?" The intention-to-return statement, measured on a five-point "strongly agree"-"strongly disagree" scale, was worded as follows: "If I were to find myself in the same situation I was in when I went to _____ hospital, I would want to receive my treatment there again." A final set of questions pertained to the respondents' demographic characteristics.

DATA ANALYSIS AND RESULTS

The adequacy of the SERVQUAL scale for assessing patients' perceptions of service quality in the hospital environment was examined in accordance with the recommendations provided in the recent measurement literature (e.g., Anderson and Gerbing 1988; Bagozzi 1981; Bagozzi and Yi 1988; Churchill 1979). Therefore, the analyses conducted related to the scale's reliability; underlying dimensionality; and

convergent, discriminant, and nomological validity. Reliability assessments were based on the internal consistency of the items (using the coefficient alpha) representing the same dimension of service quality as well as the overall scale. Validity assessment was based on correlation and factor analyses.

INITIAL RESULTS

An initial description of the data revealed that, even though a "don't know" option was not provided, item nonresponses on the perceptions portion of the scale were common. No such tendency appeared on the expectations part of the scale. These results suggest that patients may have a clear idea on desirable levels of service attributes, but that actual service performance becomes difficult to assess either because of the time lapse or the unique nature of the service experience.

Reliability

Items for each subscale (e.g., tangibles) were subjected to reliability assessment. Corrected item-to-total correlations were also examined; that is, the scores for an item and the summated scores of the rest of the items comprising a subscale (e.g., the subscale measuring the tangibles dimension of service quality) were correlated. The coefficient alpha values for the expectations subscales were .587, .677, .715, .801, and .495 for tangibles, reliability, responsiveness, assurance, and empathy, respectively.

Of the individual expectation items, only one had a correlation with the total scores that was lower than the .35 cut-off value suggested by Saxe and Weitz (1982). This item, "Hospitals should have up-to-date equipment," had a correlation of .289 with the total scores. The rest of the item-to-total correlations for the expectations scale ranged from .358 to .652. Item analysis results for both the expectations and perceptions scores are presented in Appendix A. Appendix A also contains item means and standard deviations.

Coefficient alpha values for the perceptions subscales were .782, .759, .903, .892, and .874 for tangibles, reliability, responsiveness, assurance, and empathy, respectively. None of the item-to-total correlations for the perception items were less than the .35 cut-off value. Item-to-total correlations in the perceptions subscales ranged from .486 to .870.

Reliabilities for linear combinations of the five subscales were also computed to assess the overall internal consistency of the expectations and perceptions measures (Nunnally 1978). The overall coefficient

alpha values were .897 and .964 for the expectations and perceptions scores, respectively. These values suggest that both measures exhibit desirable levels of internal consistency at the aggregate level.

Exploratory Factor Analysis

In order to examine the dimensionality of the scale, each one of the correlation matrixes (i.e., correlations of difference scores, expectations, and perceptions) was factor analyzed separately. Data on expectations produced three factors with eigenvalues greater than 1.0, accounting for 56.2 percent of the variation in item scores. The first factor accounted for more than 41 percent of the variability, and all 15 items loaded more heavily on this factor. The results from an oblique rotation did not show any meaningful patterns in terms of dimensionality. Therefore, the first factor from the initial solution was considered as a viable underlying factor. Factor analysis results for expectations are provided in Appendix B.

Factor analysis results for the perceptions data also appear in Appendix B. Two factors, with eigenvalues greater than 1.0, accounted for 70.2 percent of the variation. Rotation results did not identify any conceptually meaningful dimensions. The initial solution indicated that a single factor adequately summarizes the data on the basis of factor loadings and variance explained.

Factor analysis of the difference scores did not provide a clear picture of any meaningful factor structure. A fourth factor analysis was also conducted using expectations and perceptions items together. The results identified two distinct factors representing expectations and perceptions. All items heavily loaded on the appropriate factor. Factor loadings from the combined solution did not differ in any significant way from those obtained with separate analyses of the expectations and perceptions.

CONFIRMATORY FACTOR ANALYSIS

To further address the dimensionality and the convergent and discriminant validity issues, a confirmatory factor analysis framework was used. In their revised service quality model, Parasuraman, Zeithaml, and Berry (1988) proposed that the service quality construct can be measured using its five dimensions as reflective indicators. This framework suggests a single-factor measurement model with five observable variables. Each observable variable is a composite score obtained from the subscales. This approach enables the five dimensions of the service quality expectations and perceptions to remain intact. Although our

Table 1: Sample Correlation Matrix of Composite Service Quality Indicators Formed on the Basis of A Priori Dimensions (N = 330)

	*QP1**	*QP2*	*QP3*	*QP4*	*QP5*	*QE1**	*QE2*	*QE3*	*QE4*	*QE5*
Perceptions										
Tangibles (QP1)	.600†									
Reliability (QP2)	.660	.883								
Responsiveness (QP3)	.620	.795	.994							
Assurance (QP4)	.706	.769	.859	.830						
Empathy (QP5)	.639	.769	.859	.902	1.025					
Expectations										
Tangibles (QE1)	.407	.325	.343	.391	.343	.428				
Reliability (QE2)	.252	.206	.198	.215	.161	.610	.412			
Responsiveness (QE3)	.307	.232	.253	.228	.184	.461	.605	.524		
Assurance (QE4)	.416	.326	.283	.344	.305	.581	.599	.626	.375	
Empathy (QE5)	.327	.222	.173	.251	.243	.441	.558	.552	.579	.549

*QPs and QEs are composite quality indicators obtained from the initial items in each subscale. For instance, QP1 was computed by dividing the summated scores of P1, P2, and P3 (items measuring perceptions on tangibles) by 3. All correlations are significant beyond the .05 level.

†Diagonal entries (underlined) are standard deviations.

initial results did not identify the proposed five dimensions, this could be due to the highly correlated nature of the five dimensions of service quality. The composite single indicators were created on the basis of the a priori dimensions. Such practices are common in the literature to reduce model complexities (cf. Bagozzi 1980; Joachimsthaler and Lastovicka 1984; Jöreskog 1978; Michaels, Day, and Joachimsthaler 1987).

Consequently, two new correlation matrixes were constructed on the basis of composite indicators. Each subscale (consisting of multiple items) was converted to a single composite score that represented one of the five dimensions of service quality. This procedure was followed separately for both the perceptions and the expectations scales. As in the original item-scoring format, this resulted in mean subscale scores ranging from 1 to 5. For example, the scores for the three items representing the "tangibles" dimension (see Appendix A) were summated and the resulting score was divided by three for each respondent. Descriptive statistics on these newly generated indicators of the perceptions and expectations constructs were now comparable within each construct as well as between constructs. Table 1 provides the correlation matrixes (and standard deviations).

An examination of the correlations in Table 1 reveals that Bagoz-

Table 2: LISREL Estimates of the Measurement Model for Perceptions and Expectations Scales (N = 330)*

Parameter†	*Estimate*	*Std. Error*	t-*Value*
LXP1	.718	.048	14.92
LXP2	.831	.045	18.42
LXP3	.910	.043	21.39
LXP4	.952	.041	23.16
LXP5	.939	.042	22.61
LXE1	.705	.051	13.91
LXE2	.783	.049	16.07
LXE3	.750	.050	15.12
LXE4	.811	.048	16.91
LXE5	.702	.051	13.83
ϕ	.385	.052	7.40
Fit Statistics			
χ^2_{34} = 159.90			
p-Value = .000			
GFI = .916			
AGFI = .864			
RMR = .064			
NFI‡ = .937			

*Sample size was reduced due to deletion of missing observations on lists.

†LXP1 to LXP5 represent factor loadings of the respective perception indicators, LXE1 to LXE5 represent factor loadings of the expectation indicators, and ϕ is the correlation between the two underlying factors.

‡The value of the chi-square statistic for the null model (df = 45) was 2528.50, which was used to compute the NFI.

zi's (1981) rules for convergence and discrimination in measurement are met. That is, correlations for items representing the same construct (e.g., expectations indicators) were uniformly high, compared to correlations between items representing different constructs (i.e., cross-correlations). Although it is easy to distinguish between these measures since they are not expected to be related (as it would be easier to distinguish between apples and oranges than between different kinds of apples), these results provide the initial justification for a two-construct measurement model.

The new correlation matrixes were used as input to confirmatory factor analysis using the LISREL program (Jöreskog and Sorbom 1986). A summary of the results is presented in Table 2. Table 2 contains estimates of factor loadings (LXPs and LXEs), the correlation between the two constructs (ϕ), a set of overall fit statistics, standard errors, and t-statistics. The overall fit measures, the goodness-of-fit statistic (GFI), adjusted goodness-of-fit statistic (AGFI), root mean

squared residual (RMR), and the normed fit index (NFI) (Bentler and Bonett 1980), are all useful measures in assessing the quality of the hypothesized measurement model.

An examination of factor loading estimates reveals that each one is significant (based on *t*-values). Anderson and Gerbing (1988) suggest that significant pattern coefficients (loading estimates) provide evidence for convergent validity. An additional criterion beyond that of statistical significance is the proportion of variance in each indicator accounted for by the hypothesized underlying construct. Fornell and Larcker (1981) proposed that at least 50 percent of the variability in each indicator should be accounted for by the underlying factor (which corresponds to a minimum loading estimate of .707). This latter criterion, which is a highly demanding one, has been used recently in the development of a consumer ethnocentrism scale (Shimp and Sharma 1987). The results in Table 2 indicate that this criterion is met in most cases. Finally, a 95 percent confidence interval for the correlation between underlying expectations and perceptions does not contain the value 1.0, providing additional evidence on discriminant validity.

ADDITIONAL VALIDATION

Attitude theory (Fishbein and Ajzen 1975) predicts that perceived service quality should have a significant direct impact on behavioral intentions as evidence of nomological validity (Peter 1981). The service quality construct is also expected to correlate strongly with overall quality perceptions. These predictions should provide a basis for further validation.

Individual item scores for perceptions and expectations, and difference scores (perceptions minus expectations) were summed to obtain overall scores for each respondent. These scores were then correlated with a single-item behavioral intentions scale (return to the hospital in case of another illness) and another single-item measure of overall quality perceptions. Both items were Likert scales with a five-point response format where a higher score indicated a more favorable response. The resulting correlation matrix is presented in Table 3. All of the correlations are statistically significant in the predicted direction. While these results may be interpreted as additional evidence for the validity of the scale, they also raise an interesting question on the role of the expectations component of the scale. Of particular interest are the correlation coefficients underlined in Table 3. Behavioral intentions correlated with perceptions and difference scores almost identi-

Table 3: Correlations of Summated Scores of Perceptions, Expectations, and Difference Scores with Intentions and Overall Ratings (N = 330)

	1	2	3	4	5
1. Perceptions	___				
2. Expectations	.377	___			
3. Difference scores	.886	-.096‡	___		
4. Intentions*	.759	.135	.749	___	
5. Overall ratings†	.829	.130	.827	.859	___

*Intentions to return to the same hospital if a need arises was measured with a single-item scale.

†Overall service quality ratings of hospital services was measured using a single-item scale.

‡Not significant. All other correlations are significant beyond the .05 level.

cally (.759 and .749). The same was true for the correlations of overall ratings with perceptions and difference scores (.829 and .827).

These empirical findings suggest that the "expectations" or "desired level" scores may not be contributing to the strength of the relationship between service quality and the third variables (intention to return in case of another illness and overall quality perceptions) beyond that already contributed by the perceptions scores. This finding points out the need for further investigation of the validity of such difference scores, an issue also raised in the early measurement literature (e.g., Cronbach and Furby 1970; Herman and Hulin 1973; Wall and Payne 1973).

DISCUSSION AND IMPLICATIONS

Because of the ever stronger emphasis on cost containment, changing consumer attitudes, and stiff competition, many of the successful hospitals of the next decade will position themselves as "high-quality" health care providers. Even those hospitals that do not seek a high-quality position will find it necessary to define, monitor, and improve the quality of the services they provide. Technical quality alone, however, will not lead to increased revenues and facility utilization.

The expectations and perceptions scales in the modified SERVQUAL instrument have emerged as unidimensional measures with excellent internal consistency reliabilities. The two scales exhibit adequate validity as separate measures of (1) patients' expectations of

hospital services and (2) their perceptions of the subject hospital's performance level. These results indicate that the scales can be successfully used to assess the magnitude of the gap between patient perceptions and expectations. For this particular purpose, the modified SERVQUAL appears to be a concise and practical instrument useful for monitoring expectations and perceptions.

Researchers agree that consumers utilize product or service attributes to build higher-order abstractions leading to such constructs as quality and value (Grönroos 1984; Olson and Reynolds 1983; Parasuraman, Zeithaml, and Berry 1985; Zeithaml 1988). Zeithaml (1988) suggests that while single attributes may be product- or service-specific, their combination at a higher level of abstraction can be generalizable to product and service categories. Parasuraman and his co-workers have taken a major challenge in attempting to define and capture this abstraction for the marketing of services.

In the SERVQUAL scale, this combination of attributes at a higher level of abstraction was manifested in the form of five dimensions of service quality. Other researchers, however, have conceptualized the service quality construct with different numbers of dimensions (see for example, Gronroos 1984; Hedvall and Paltschik 1989; Lehtinen and Lehtinen 1982, for two- or three-dimensional definitions). In the present study, the modified SERVQUAL scale produced unidimensional measures of expectations and perceptions. These observations indicate the need for further work on the dimensionality and abstraction level of the construct.

The development and use of standardized measurement scales that enable researchers to compare the results of studies across industries is desirable. This study attempted to determine the applicability of such a scale. In the process, a number of changes were imposed on the original scale as a consequence of suggestions from the literature, practitioners, and former patients. Some of the results may be due to these changes, because any time an original scale is revised in any way, a new variable may be defined.

In the present study, based on the pretest results and suggestions from the literature, all items were worded positively. This places a limitation on the study because a balanced mix of negative and positive items can uncover data quality problems at an early stage provided that item responses are carefully scrutinized before further analyses (see Churchill 1979). There is a need to compare the results from such a mix with those from all-positive connotations.

CONCLUSION

It has been argued that extensive scrutiny adds to the validity of the measures (Peter and Churchill 1986). We believe the results of the present study make a contribution to that end. SERVQUAL, a standard instrument for measuring functional service quality, is reliable and valid in the hospital environment and in a variety of other service industries. It will enable researchers to test the effectiveness of quality-enhancing techniques and actions across a range of industries and to develop generalizations about these actions and methods. For example, researchers may believe that a given recognition and reward system will improve service-providing employees' attitudes on the responsiveness dimension of service quality. The amount of improvement can be measured across a range of services, and those organizations in which the technique is most useful can be identified and generalizations developed.

SERVQUAL also provides hospital administrators with a tool for the measurement of functional quality in their own organizations. Deficient scores on one or more SERVQUAL dimensions will normally signal the existence of a deeper underlying problem in the organization. For example, assume that SERVQUAL indicates that patients do not perceive hospital employees as being willing to help. The low score on this aspect of quality may be symptomatic of deeper problems that center on the organization's ability to hire and retain high-quality employees, to evaluate and reward superior performance, or to provide adequate training. Likewise, billing inaccuracies may be symptomatic of staffing problems that prevent insurance claims from being filed promptly and payments from being recorded accurately. Therefore, one of SERVQUAL's major contributions to the health care industry will be its ability to identify symptoms and to provide a starting point for the examination of underlying problems that inhibit the provision of quality services.

The measurement of patient expectations as well as perceptions provides a valuable dimension of insight into the process by which the quality of health care service is evaluated. Administrators should understand the areas in which expectations are particularly high so that the service delivery process can be tailored to meet those expectations (Parasuraman, Zeithaml, and Berry 1985). Similarly, in order to identify and correct service quality problems quickly, administrators should understand patients' perceptions of the quality of service delivered and the manner in which expectations and perceptions are balanced. In

addition, the scale can also be used to measure the views of hospital managers and employees as they think patients perceive the quality of the service. This can be done easily by changing the instructions portion of the scale. Hence, the existence of another potential gap, the gap between the provider's view and the customer's view, can be assessed and monitored (Parasuraman, Zeithaml, and Berry 1985).

Finally, it should be pointed out that SERVQUAL is designed to measure functional quality only (defined as the manner in which the health care service is delivered to the patient). However, functional quality in a health care setting cannot be sustained without accurate diagnoses and procedures. Such technical quality is the focus of research that is being conducted by a number of organizations, including the Joint Commission for Accreditation of Healthcare Organizations (JCAHO). For the long-run success of a health care organization, both functional and technical quality have to be monitored and managed effectively.

APPENDIX A

Expectations Scale Item-to-Total Correlations, Reliabilities, and Item Means and Standard Deviations (N = 330)

Items in Each Dimension	*Item-to-Total Correlations*	$\bar{X}$	s.d.
Tangibles (α = .587)			
E1. Hospitals should have up-to-date equipment.	.289	4.86	.41
E2. Hospitals' physical facilities should be visually appealing.	.476	4.34	.74
E3. Hospital employees should appear neat.	.499	4.72	.54
Reliability (α = .677)			
E4. Hospitals should provide their services at the time they promise to do so.	.506	4.72	.52
E5. When patients have problems, hospital employees should be sympathetic and reassuring.	.496	4.62	.61
E6. Hospitals should be accurate in their billing.	.497	4.80	.44
Responsiveness (α = .715)			
E7. Hospital employees should tell patients exactly when services will be performed.	.532	4.55	.64
E8. It is realistic for patients to expect prompt service from hospital employees.	.577	4.37	.72
E9. Hospital employees should always be willing to help patients.	.505	4.61	.60

Continued

Assurance (α = .801)			
E10. Patients should be able to feel safe in their interactions with hospital employees.	.652	4.78	.47
E11. Hospital employees should be knowledgeable.	.641	4.81	.45
E12. Hospital employees should be polite.	.613	4.73	.52
E13. Hospital employees should get adequate support from their employers to do their jobs well.	.556	4.82	.45
Empathy (α = .495)			
E14. Hospital employees should be expected to give patients personal attention.	.358	4.25	.80
E15. It is realistic to expect hospitals to have their patients' best interests at heart.	.358	4.77	.52
Overall Scale (α = .897)*			

*Coefficient alpha for the overall scale (as a linear combination of subscales).

Perceptions Scale Item-to-Total Correlations, Reliabilities, and Item Means and Standard Deviations (N = 330)

Items in Each Dimension	*Item-to-Total Correlations*	$\bar{X}$	s.d.
Tangibles (α = .782)			
P1. XYZ has up-do-date equipment.	.530	4.53	.72
P2. XYZ's physical facilities are visually appealing.	.685	4.51	.69
P3. XYZ's employees appear neat.	.649	4.47	.74
Reliability (α = .759)			
P4. XYZ provides its services at the time it promises to do so.	.655	4.05	1.08
P5. When patients have problems, XYZ's employees are sympathetic and reassuring.	.644	4.20	.99
P6. XYZ is accurate in its billing.	.486	4.24	1.15
Responsiveness (α = .903)			
P7. XYZ employees tell patients exactly when services will be performed.	.767	4.03	1.02
P8. Patients receive prompt service from XYZ's employees.	.870	3.93	1.21
P9. XYZ's employees are always willing to help patients.	.806	4.19	1.01
Assurance (α = .892)			
P10. Patients feel safe in their interactions with XYZ's employees.	.826	4.25	.98
P11. XYZ's employees are knowledgeable.	.844	4.28	.87
P12. XYZ's employees are polite.	.741	4.39	.88
P13. Employees get adequate support from XYZ to do their jobs well.	.670	4.39	.88

Empathy (α = .874)			
P14. XYZ's employees give patients personal attention.	.776	4.11	1.06
P15. XYZ has patients' best interests at heart.	.776	4.14	1.11
Overall Scale (α = .964)*			

*Coefficient alpha for the overall scale (as a linear combination of subscales).

APPENDIX B

Maximum Likelihood Exploratory Factor Analysis Results on Expectations Scale

Expectations Scale Items	*Factor Loadings*		
	Factor 1	*Factor 2*	*Factor 3*
E1	.482	-.009	.178
E2	.497	.297	.187
E3	.610	.126	.298
E4	.614	.221	.113
E5	.644	.152	-.103
E6	.564	.156	.292
E7	.603	.149	-.146
E8	.620	.280	-.301
E9	.650	-.136	-.330
E10	.716	-.305	-.001
E11	.664	-.303	.038
E12	.718	-.078	.051
E13	.627	-.141	.156
E14	.576	.180	-.179
E15	.580	-.215	.078
Eigenvalue	6.203	1.146	1.083
Variance explained	41.4%	7.6%	7.2%

Maximum Likelihood Exploratory Factor Analysis Results on Perceptions Scale

Perceptions Scale Items	*Factor Loadings*	
	Factor 1	*Factor 2*
P1	.524	.437
P2	.585	.393
P3	.701	.293
P4	.744	.127
P5	.844	.000
P6	.496	.316
P7	.794	.123
P8	.859	-.114
P9	.909	-.219
P10	.876	.044
P11	.856	.063
P12	.811	-.085
P13	.762	.120
P14	.874	-.167
P15	.887	.042
Eigenvalue	9.478	1.055
Variance explained	63.2%	7.0%

REFERENCES

Anderson, C., and C. P. Zeithaml. "Stage of the Product Life Cycle, Business Strategy and Business Performance." *Academy of Management Journal* 27 (March 1984): 5-24.

Anderson, J. C., and D. W. Gerbing. "Structural Equation Modeling in Practice: A Review and Recommended Two-Step Approach." *Psychological Bulletin* 103, no. 3 (1988): 411-23.

Bagozzi, R. P. "Evaluating Structural Equation Models with Unobservable Variables and Measurement Error: A Comment." *Journal of Marketing Research* 18 (August 1981): 375-81.

———. "Performance and Satisfaction in an Industrial Salesforce: An Examination of Their Antecedents and Simultaneity." *Journal of Marketing* 44 (Spring 1980): 65-77.

Bagozzi, R. P., and Y. Yi. "On the Evaluation of Structural Equation Models." *Journal of the Academy of Marketing Science* 16, no. 1 (Spring 1988): 74 -94.

Bentler, P. M., and P. G. Bonett. "Significance Tests and Goodness-of-Fit in the Analysis of Covariance Structures." *Psychological Bulletin* 88, no. 3 (1980): 588-606.

Bolton, R. N., and J. H. Drew. "A Model of Perceived Service Value." *Technical Note* 88-420.1. Waltham, MA: GTE Laboratories, 1988.

Bopp, K. D. "How Patients Evaluate the Quality of Ambulatory Medical Encounters: A Marketing Perspective." *Journal of Health Care Marketing* 10, no. 1 (March 1990): 6-15.

Casarreal, K. M., J. I. Mills, and M. A. Plant. "Improving Service through Patient Surveys in a Multihospital Organization." *Hospital & Health Services Administration* 31, no. 2 (March/April 1986): 41-52.

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

Cronbach, L. J., and L. Furby. "How We Should Measure 'Change'—Or Should We?" *Psychological Bulletin* 74, no. 1 (1970): 68-80.

Donabedian, A. *Explorations in Quality Assessment and Monitoring.* Vol. 1, *The Definition of Quality and Approaches to Its Assessment.* Ann Arbor, MI: Health Administration Press, 1980.

———. *Explorations in Quality Assessment and Monitoring.* Vol. 2, *The Criteria and Standards of Quality.* Ann Arbor, MI: Health Administration Press, 1982.

Fishbein, M., and I. Ajzen. *Beliefs, Attitudes, Intention, and Behavior: An Introduction to Theory and Research.* Reading, MA: Addison-Wesley Publishing Company, 1975.

Fornell, C., and D. F. Larcker. "Evaluating Structural Equation Models with Unobservable Variables and Measurement Error." *Journal of Marketing Research* 18 (February 1981): 39-50.

Grönroos, C. "A Service Quality Model and Its Marketing Implications." *European Journal of Marketing* 18, no. 4 (1984): 36-44.

Hedvall, M., and M. Paltschik. "An Investigation in and Generation of Service Quality Concepts." In *Marketing Thought and Practice in the 1990's.*

Edited by G. J. Avlonitis, N. K. Papavasiliou, and A. G. Kouremenos. Athens, Greece: European Marketing Academy, 1989.

Heise, D. R. "Some Issues in Sociological Measurement." In *Sociological Methodology.* Edited by H. Costner. San Francisco: Jossey-Bass, 1974.

Herman, J. B., and C. L. Hulin. "Managerial Satisfactions and Organizational Roles: An Investigation of Porter's Need Deficiency Scales." *Journal of Applied Psychology* 57, no. 2 (1973): 118-24.

Howell, R. D., J. B. Wilcox, D. N. Bellenger, and L. B. Chonko. "An Assessment of the Role Conflict and Ambiguity Scales." In *Efficiency and Effectiveness in Marketing.* Edited by G. Frazier, C. Ingene, D. Aaker, A. Ghosh, and T. Kinnear. Chicago, IL: American Marketing Association, 1988.

Joachimsthaler, E. A., and J. L. Lastovicka. "Optimal Stimulation Level—Exploratory Behavior Models." *Journal of Consumer Research* 11 (December 1984): 830-35.

Joint Commission for Accreditation of Healthcare Organizations. "Agenda for Change (Update)." *Newsletter* 1, no. 1 (September 1987): 1-4.

Jöreskog, K. G. "Structural Analysis of Covariance and Correlation Matrices." *Psychometrika* 43, no. 4 (1978): 443-77.

Jöreskog, K. G., and D. Sorbom. *LISREL VI: Analysis of Linear Structural Relationships by the Method of Maximum Likelihood.* Chicago, IL: National Educational Resources Inc., 1986.

Kovner, A. R., and H. L. Smits. "Point of View: Consumer Expectations of Ambulatory Care." *Health Care Management Review* 3, no. 1 (Winter 1978): 69-75.

Lehtinen, J. R., and O. Lehtinen. "Service Quality: A Study of Quality Dimensions." Unpublished working paper. Helsinki: Service Management Institute, 1982.

Likert, R. "A Technique for Measurement of Attitudes." *Archives of Psychology* monograph no. 140 (1932).

Michaels, R. E., R. L. Day, and E. A. Joachimsthaler. "Role Stress among Industrial Buyers: An Integrative Model." *Journal of Marketing* 51 (April 1987): 28-45.

Nunnally, C. J. *Psychometric Methods.* New York: Harper and Row, 1978.

Olson, J. C., and T. J. Reynolds. "Understanding Consumers' Cognitive Structures: Implications for Advertising Strategy." In *Advertising and Consumer Psychology.* Edited by L. Percy and A. Woodside. Lexington, MA: Lexington Books, 1983.

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

———. "SERVQUAL: A Multiple-Item Scale for Measuring Customer Perceptions of Service Quality." Technical Report. Report no. 86-108. Marketing Science Institute, Cambridge, MA, 1986.

———. "SERVQUAL: A Multiple-Item Scale for Measuring Customer Perceptions of Service Quality." *Journal of Retailing* 64, no. 1 (Spring 1988): 12-40.

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

Peter, J. P., and G. A. Churchill, Jr. "Relationships among Research Design

Choices and Psychometric Properties of Rating Scales: A Meta-Analysis." *Journal of Marketing Research* 23 (February 1986): 1-10.

Press, I., and R. F. Ganey. "The Mailout Questionnaire as the Practical Method of Choice in Patient Satisfaction Monitoring." *Journal of Health Care Marketing* 9, no. 1 (March 1989): 67-68.

Price, J. L., and C. W. Mueller. *Handbook of Organizational Measurement.* Marshfield, MA: Pitman Publishing Co., 1986.

Rizzo, J., R. J. House, and S. I. Lirtzman. "Role Conflict and Ambiguity in Complex Organizations." *Administrative Science Quarterly* 15, no. 2 (1970): 150-63.

Saxe, R., and B. A. Weitz. "The SOCO Scale: A Measure of the Customer Orientation of Salespeople." *Journal of Marketing Research* 19 (August 1982): 343-51.

Shimp, T. A., and S. Sharma. "Consumer Ethnocentrism: Construction and Validation of the CETSCALE." *Journal of Marketing Research* 24 (August 1987): 280-89.

Vogels, R., J. Lemmink, and H. Kasper. "Some Methodological Remarks on the SERVQUAL Model." In *Marketing Thought and Practice in the 1990s.* Edited by G. J. Avlonitis, N. K. Papavasiliou, and A. G. Kouremenos. Athens, Greece: European Marketing Academy, 1989.

Wall, T. D., and R. Payne. "Are Deficiency Scores Deficient?" *Journal of Applied Psychology* 58, no. 3 (1973): 322-26.

Ware, J. E., and M. K. Snyder. "Dimensions of Patient Attitudes Regarding Doctors and Medical Care." *Medical Care* 8, no. 8 (August 1975): 669-82.

Zeithaml, V. A. "Consumer Perceptions of Price, Quality, and Value: A Means-End Model and Synthesis of Evidence." *Journal of Marketing* 52 (July 1988): 2-22.

Zeithaml, V. A., L. L. Perry, and A. Parasuraman. "Communication and Control Processes in the Delivery of Service Quality." *Journal of Marketing* 52, no. 2 (April 1988): 35-48.