Abstract

A recent MIS Quarterly article rightfully points out that service is an important part of the role of the information systems (IS) department and that most IS assessment measures have a product orientation (Pitt et al. 1995). The article went on to suggest the use of an IS-context-modified version of the SERVQUAL instrument to assess the quality of the services supplied by an information services provider (Parasuraman et al. 1985, 1988, 1991). However, a number of problems with the SERVQUAL instrument have been discussed in the literature (e.g., Babakus and Boller 1992; Carman 1990; Cronin and Taylor 1992, 1994; Teas 1993). This article reviews that literature and discusses some of the implications for measuring service quality in the information systems context. Findings indicate that SERVQUAL suffers from a number of conceptual and empirical difficulties. Conceptual difficulties include the operationalization of perceived service quality as a difference or gap score, the ambiguity of the expectations construct, and the unsuitability of using a single measure of service quality across different industries. Empirical problems, which may be linked to the use of difference scores, include reduced reliability, poor convergent validity, and poor predictive validity. This suggests that (1) some alternative to difference scores is preferable and should be utilized; (2) if used, caution should be exercised in the interpretation of IS-SERVQUAL difference scores; and (3) further work is needed in the development of measures for assessing the quality of IS services.

Keywords: IS management, evaluation, measurement, service quality, user attitudes, user expectations

ISRL Categories: A10104, A10109, A104, E10206.03, GB02, GB07

---

1Robert Zmud was the accepting senior editor for this paper.
Research Note

Introduction

Due to the growth of outsourcing, end-user-controlled information assets, joint ventures, and other alternative mechanisms by which organizations are meeting their need for information systems services, IS managers are increasingly concerned about improving the perceived (as well as actual) service quality of the IS function (Kettinger and Lee 1994). In recent years, the use of SERVQUAL-based instruments has become increasingly popular with information systems (IS) researchers. However, a review of the literature suggests that the use of such instruments may result in a number of measurement problems.

A recent article makes several important contributions to the assessment and evaluation of the effectiveness of information systems (IS) departments in organizations (Pitt et al. 1995). The article:

1. Points out that, although service is an important part of the role of the IS department, most IS assessment measures have a product orientation.

2. Proposes an extension of the categorization of IS success measures (DeLone and McLean 1992) to include service quality.

3. Proposes the use of the SERVQUAL instrument from marketing (Parasuraman et al. 1985, 1988, 1991) to operationalize the IS service quality construct and modify the wording of the instrument to better accommodate its use in an IS context.

4. Adapts and augments a theory regarding the determinants of service quality expectations to an IS context and offers ideas for future research.

A number of studies, however, identify potential difficulties with the SERVQUAL instrument (e.g., Babakus and Boller 1992; Carman 1990; Cronin and Taylor 1992, 1994; Teas 1993, 1994). This research note reviews some of the literature regarding difficulties with the SERVQUAL instrument in general and examines the implications of these difficulties to the use of the instrument in an IS context.

The SERVQUAL Instrument: Problems Identified in the Literature

The difficulties with the SERVQUAL instrument identified in the literature can be grouped into two main categories: (1) conceptual and (2) empirical; although, the boundary between them blurs because they are closely inter-related. The conceptual problems center around (1) the use of two separate instruments, one for each of two constructs (i.e., perceptions and expectations), to operationalize a third conceptually distinct construct (i.e., perceived service quality) that is itself the result of a complex psychological process; (2) the ambiguity of the expectations construct; and (3) the suitability of using a single instrument to measure service quality across different industries (i.e., content validity). The empirical problems are, by and large, the result of these conceptual difficulties, most notably the use of difference scores, in conjunction with the a theoretical nature of the process used in the construction of the original five dimensions of service quality. The empirical difficulties most often attributed to the SERVQUAL instrument include low reliability, unstable dimensionality, and poor convergent validity. A review of these conceptual and empirical problems should serve to caution those who wish to use SERVQUAL to measure the service quality of an information systems provider.

Conceptual difficulties with SERVQUAL

Subtraction as a “Simulation” of a Psychological Process

Many of the difficulties associated with the SERVQUAL instrument stem from the opera-
tionalization of a service quality construct that is theoretically grounded in a discrepancy or gap model. In conceptualizing service quality, Parasuraman et al. (1985, 1988, 1991, 1994b) use the “service quality model,” which posits that one's perception of service quality is the result of an evaluation process whereby “the customer compares . . . the perceived service against the expected service” (Grönroos 1984, p. 37).

Rather than develop an instrument to directly measure the perception of service quality that is the outcome of this cognitive evaluation process, the SERVQUAL instrument (Parasuraman et al. 1988, 1991) separately measures the expected level of service and the experienced level of service. Then service quality scores are calculated as the difference between these two measures. These three sets of scores are commonly referred to as expectation (E), perception (P), and SERVQUAL (whereby, SERVQUAL = P - E). Although not without precedent, and certainly worthy of fair empirical evaluation, the implicit assumption that subtraction accurately portrays this cognitive process seems overly simplistic. Even if one fully accepts this discrepancy model of experiences vis-à-vis expectations as indicative of the general process whereby one arrives at an evaluation of a service experience, the notion that the specific mechanism is merely subtraction does not logically follow. The use of differences was, and remains, an operational decision. Regrettably, it does not appear to have been a particularly good one. The direct measurement of one's perception of service quality that is the outcome of this cognitive evaluation process seems more likely to yield a valid and reliable outcome. If the discrepancy is what one wants to measure, then one should measure it directly.

Ambiguity of the “Expectations” Construct

Teas (1994) notes that SERVQUAL expectations are variously defined as desires, wants, what a service provider should possess, normative expectations, ideal standards, desired service, and the level of service a customer hopes to receive (e.g., Parasuraman et al. 1985, 1988, 1991; Zeithaml et al. 1993). These multiple definitions and corresponding operationalizations of “expectations” in the SERVQUAL literature result in a concept that is loosely defined and open to multiple interpretations (Teas). Yet even when concise definitions are provided, various interpretations of the expectations construct can result in potentially serious measurement validity problems.

The conceptualization of “expectations” consistent with the SERVQUAL model is the vector attribute interpretation—“that is one on which a customer's ideal point is at an infinite level” (Parasuraman et al. 1994, p. 116). Unfortunately, as the proportion of extreme responses (e.g., seven on a seven point scale) increases, the expectation scores become less useful as an increasing proportion of the variation in difference-based SERVQUAL scores is due only to changes in perception scores.

Teas (1993) found three different interpretations of “expectations” derived from an analysis of follow-up questions to an administration of the SERVQUAL questionnaire. One interpretation of expectations is as a forecast or prediction. The forecast interpretation of expectations cannot be discriminated from the disconfirmed expectations model of consumer satisfaction (Oliver 1980). This interpretation is inconsistent with the definition of service quality put forth by Parasaraman et al. (1988) and results in a discriminant validity problem with respect to consumer satisfaction. A second interpretation of expectations is as a measure of attribute importance. When respondents use this interpretation, the resulting perception-minus-expectation scores exhibit an inverse relationship between attribute importance and perceived service quality, all other things being equal.

The third interpretation identified by Teas (1993) is the “classic ideal point” concept. Parasuraman et al. (1994) describe this when they note that “the P-E [i.e., perceptions-minus-expectations] specification could be problematic when a service attribute is a classic ideal point attribute—that is one on which a customer’s ideal point is at a finite level and therefore, performance beyond which will dis-
Research Note

please the customer (e.g., friendliness of a salesperson in a retail store)" (p. 116). This interpretation of expectations results in an inverse of the relationship between the SERVQUAL score, calculated as perception minus expectation (P - E), and actual service quality for all values when perception scores are greater than expectation scores (i.e., P > E). This interpretation is consistent with the finding that user satisfaction scores were highest when actual user participation was in congruence with the user's need for participation, rather than merely maximized (Doll and Torkzadeh 1989).

These various interpretations of the "expectation" construct lead to a number of measurement problems. The findings suggest that a considerable portion of the variance in the SERVQUAL instrument is the result of measurement error induced by respondent's varying interpretations of the "expectations" construct (Teas 1993).

Three separate types of expectations have been described (Boulding et al. 1993): (1) the will expectation, what the customer believes will happen in their next service encounter; (2) the should expectation, what the customer believes should happen in the next service encounter; and (3) an ideal expectation, what a customer wants in an ideal sense. The ideal interpretation of expectation is often used in the SERVQUAL literature (Boulding et al. 1993). Boulding et al. (1993) differentiate between should and ideal expectations by stating that what customers think should happen may change as a result of what they have been told to expect by the service provider, as well as what the consumer views as reasonable and feasible based on what they have been told and their experience with the firm or a competitor's service. In contrast, an ideal expectation may "be unrelated to what is reasonable/feasible and/or what the service provider tells the customer to expect" (Boulding et al. 1993, p. 9). A series of experiments demonstrated results that were incompatible with the gap model of service quality (Boulding et al. 1993). Instead, the results demonstrate that service quality is influenced only by perceptions. Moreover, the results indicate that perceptions are influenced by both will and should expectations, but in opposite directions. Increasing will expectations leads to a higher perception of service quality whereas an increasing expectation of what should be delivered during a service encounter will actually decrease the ultimate perception of the quality of the service provided (Boulding et al. 1993). Not only do these findings fail to support the gap model of service quality, but these results also demonstrate the wildly varying impact of different interpretations of the expectations construct.

Different methods to operationalize "expectations" in developing their IS versions of SERVQUAL have been used (Pitt et al. 1995; Kettinger and Lee 1994). One study used the instructions to the survey to urge the respondents to "think about the kind of IS unit that would deliver excellent quality of service" (Pitt et al. 1995). The items then take a form such as: E1 They will have up-to-date hardware and software. Whereas the second study (Kettinger and Lee 1994) used the form: E1 Excellent college computing services will have up-to-date equipment.

Recall that some respondents to SERVQUAL were found to interpret expectations as forecasts or predictions (Teas 1993). This interpretation corresponds closely with the will expectation (Boulding et al. 1993). It is easy to see how this interpretation might be formed especially with the "They will" phrasing (Pitt et al. 1995). Unfortunately, the impact of the will expectation on perceptions of service quality is opposite from that intended by the SERVQUAL authors and the (P-E) or gap model of service quality (Boulding et al. 1993).

In summary, a review of the literature indicates that respondents to SERVQUAL may have numerous interpretations of the expectations construct and that these various interpretations have different and even opposite impacts on perceptions of service quality. Moreover, some of the findings demonstrate that expectations influence only perceptions and that perceptions alone directly influence overall service quality (Boulding et al. 1993). These findings fail to support the (P-E) gap model of service quality.
quality and indicate that the use of the expectations construct as operationalized by SERVQUAL-based instruments is problematic.

Applicability of SERVQUAL Across Industries

Another often mentioned conceptual problem with SERVQUAL concerns the applicability of a single instrument for measuring service quality across different industries. Several researchers have articulated their concerns on this issue. A study of SERVQUAL across four different industries found it necessary to add as many as 13 additional items to the instrument in order to adequately capture the service quality construct in various settings, while at the same time dropping as many as 14 items from the original instrument based on the results of factor analysis (Carman 1990). The conclusion arrived at was that considerable customization was required to accommodate differences in service settings. Another study attempted to utilize SERVQUAL in the banking industry (Brown et al. 1993). The authors were struck by the omission of items which they thought a priori would be critical to subject's evaluation of service quality. They concluded that it takes more than simple adaptation of the SERVQUAL items to effectively address service quality across diverse settings. A study of service quality for the retail sector also concluded that utilizing a single measure of service quality across industries is not feasible (Dabholkar et al. 1996).

Researchers of service quality in the information systems context appear to lack consensus on this issue. Pitt et al. (1995) state that they could not discern any unique features of IS that make the standard SERVQUAL dimensions inappropriate nor could they discern any dimensions with some meaning of service quality in the IS domain that had been excluded from SERVQUAL. Kettinger and Lee (1994), however, found that SERVQUAL should be used as a supplement to the UIS (Baroudi and Orlikowski 1988) because that instrument also contains items that are important determinants of IS service quality. Their findings suggest that neither the UIS nor SERVQUAL alone can capture all of the factors which contribute to perceived service quality in the IS domain. For example, items contained in the UIS include the degree of training provided to users by the IS staff, the level of communication between the users and the IS staff, and the time required for new systems development and implementation, all of which possess strong face validity as determinants of IS service quality. In addition, Kettinger and Lee dropped the entire tangibles dimension from their version of SERVQUAL based on the results of confirmatory factor analysis. These finding contradict the belief that all dimensions of SERVQUAL are relevant and that there are of no unique features of the IS domain not included in the standard SERVQUAL instrument (Pitt et al. 1995). It is difficult to argue that items concerning the manner of dress of IS employees and the visual attractiveness of IS facilities (i.e., tangibles) should be retained as important factors in the IS domain while issues such as training, communication, and time to complete new systems development are excluded. We agree that using a single measure of service quality across industries is not feasible (Dabholkar et al. 1996) and therefore future research should involve the development of industry-specific measures of service quality.

Empirical difficulties with the SERVQUAL instrument

A difference score is created by subtracting the measure of one construct from the measure of another in an attempt to create a measure of a third distinct construct. For example, in scoring the SERVQUAL instrument, an expectation score is subtracted from a perception score to create such a "gap" measure of service quality. Even if one assumes that the discrepancy theory is correct and that these are the only (or at least, the last) two inputs into this cognitive process, it still raises the question: Can calculated difference scores operationalize the outcome of a cognitive discrepancy? It appears that several problems with the use of difference scores make them a
poor measure of psychological constructs (e.g., Edwards 1995; Johns 1981; Lord 1958; Peter et al. 1993; Wall and Payne 1973). Among the difficulties related to the use of difference measures discussed in the literature are low reliability, unstable dimensionality, and poor predictive and convergent validities.

Reliability Problems With Difference Scores

Many studies demonstrate that Cronbach's (1951) alpha, a widely-used method of estimating instrument reliability, is inappropriate for difference scores (e.g., Cronbach and Furby 1970; Edwards 1995; Johns 1981; Lord 1958; Peter et al. 1993; Prakash and Lounsbury 1983; Wall and Payne 1973). This is because the reliability of a difference score is dependent on the reliability of the component scores and the correlation between them. The correct formula for calculating the reliability of a difference score (rD) is:

\[ r_D = \frac{\sigma_{D}^2}{\sigma_{D}^2 + \sigma_1^2 + \sigma_2^2 - 2r_{12}\sigma_1\sigma_2} \]

where \( r_{11} \) and \( r_{22} \) are the reliabilities of the two component scores, \( \sigma_1^2 \) and \( \sigma_2^2 \) are the variances of the component scores, and \( r_{12} \) is the correlation between the component scores (Johns 1981).

This formula shows that as the correlation of the component scores increases, the reliability of the difference scores is decreased. An example was provided where the reliability of the difference score formed by subtracting one component from another with an average reliability of .70, and a correlation of .40, is only .50 (Johns 1981). Thus, while the average reliability of the two components is .70, which is considered acceptable (Pitt et al. 1995; cf., Nunnally 1978), the correlation between the components reduces the reliability of the difference score to a level that most researchers would consider unacceptable (Peter et al. 1993).

An example of the overestimation for the reliability caused by the misuse of Cronbach's alpha can be found in the analysis of service quality for a computer manufacturer (Parasuraman et al. 1994a; see Table 1). Note that Cronbach's alpha consistently overestimates the actual reliability for the difference scores of each dimension (column 2). Also note that the use of the correct formula for calculating the reliability of a difference score has demonstrated that the actual reliabilities for the SERVQUAL dimensions may be as much as .10 lower than reported by researchers incorrectly using Cronbach's alpha. In addition, these findings show that the non-difference, direct response method results in consistently higher reliability scores than the (P-E) difference method of scoring.

These results have important implications for the IS-SERVQUAL (Pitt et al. 1995).

Table 1. Reliability of SERVQUAL: The Misuse of Cronbach's Alpha

<table>
<thead>
<tr>
<th>A Priori Dimensions</th>
<th>Cronbachs' α (Non-Difference)</th>
<th>Cronbachs' α (Difference)</th>
<th>Johns' α for Differences (Difference)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tangibles</td>
<td>.83</td>
<td>.75</td>
<td>.65</td>
</tr>
<tr>
<td>Reliability</td>
<td>.91</td>
<td>.87</td>
<td>.83</td>
</tr>
<tr>
<td>Responsiveness</td>
<td>.87</td>
<td>.84</td>
<td>.81</td>
</tr>
<tr>
<td>Assurance</td>
<td>.86</td>
<td>.81</td>
<td>.71</td>
</tr>
<tr>
<td>Empathy</td>
<td>.90</td>
<td>.85</td>
<td>.81</td>
</tr>
</tbody>
</table>

Note: Difference scores calculated as perception minus expectation (P - E).
Cronbach’s alpha, which consistently overestimates the reliability of difference scores, was used incorrectly. Even when using the inflated alpha scores, Pitt et al. note that two of three reliability measures for the tangibles dimension fall below the 0.70 level required for commercial applications. Had they utilized the appropriate modified alpha, they may have concluded that the tangibles dimension is not reliable in the IS context, a finding which would have been consistent with the results of Kettinger and Lee (1994).

A review of the literature clearly indicates that by utilizing Cronbach’s alpha, researchers tend to overestimate the reliabilities of difference scores especially when the component scores are highly correlated: Such is the case with the SERVQUAL instrument (Peter et al. 1993).

Predictive and Convergent Validity Issues With Difference Scores

Another problem with the SERVQUAL instrument concerns the poor predictive and convergent validities of the measure. Convergent validity is concerned with the extent to which multiple measures of the same construct agree with each other (Cambell and Fiske 1959). Predictive validity refers to the extent to which scores of one construct are empirically related to scores of other conceptually-related constructs (Bagozzi et al. 1992; Kappelman 1995; Parasuraman et al. 1991).

One study reported that perceptions-only SERVQUAL scores had higher correlations with an overall service quality measure (i.e., convergent measure) and with complaint resolution scores (i.e., the predictive measure) than did the perception-minus-expectation difference scores used with SERVQUAL (Babakus and Boller 1992). A different study performed regression analyses in which an overall single-question service quality rating was regressed separately on both difference scores (i.e., perception minus expectation) and perception-only scores (Parasuraman et al. 1991). The perception-only SERVQUAL scores produced higher adjusted r-squared values (ranging from .72 to .81) compared to the SERVQUAL difference scores (ranging from .51 to .71).

The predictive validity of difference scores, a non-difference direct response score, and the perceptions only scores for SERVQUAL in the context of a financial institution has been compared (Brown et al. 1993). Correlation analysis was performed between the various scores and a three-item behavioral intentions scale. Behavioral intentions include such concepts as whether the customer would recommend the financial institution to a friend or whether they would consider the financial institution first when seeking new services. The results of the study show that both the perceptions only (.31) and direct response (.32) formats demonstrated higher correlations with the behavioral intentions scale than did the traditional difference score (.26).

The superior predictive and convergent validity of perception-only scores was confirmed (Cronin and Taylor 1992). Those results indicated higher adjusted r-squared values for perception-only scores across four different industries. The perception component of the perception-minus-expectation score consistently performs better as a predictor of overall service quality than the difference score itself (Babakus and Boller 1992; Boulding et al. 1993; Cronin and Taylor 1992; Parasuraman et al. 1991).

Unstable Dimensionality of the SERVQUAL Instrument

The unstable nature of the factor structure for the SERVQUAL instrument may be related to the atheoretical process by which the original dimensions were defined. The SERVQUAL questionnaire is based on a multi-dimensional model (i.e., theory) of service quality. A 10 dimensional model of service quality based on a review of the service quality literature and the extensive use of both executive and focus group interviews was developed (Parasuraman et al. 1985). During instrument development, Parasuraman et al. (1988)
Table 2. Unstable Dimensionality of SERVQUAL

<table>
<thead>
<tr>
<th>Study</th>
<th>Instrument</th>
<th>Analysis</th>
<th>Factor Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carman (1990)</td>
<td>Four modified SERVQUALs using 12-21 of the original items</td>
<td>Principal axis factor analysis with oblique rotation</td>
<td>Five to nine factors</td>
</tr>
<tr>
<td>Bresinger and Lambert (1990)</td>
<td>Original 22 items</td>
<td>Principal axis factor analysis with oblique rotation</td>
<td>Four factors with eigenvalues &gt; 1</td>
</tr>
<tr>
<td>Parasuraman, Zeithaml, and Berry (1991)</td>
<td>Original 22 items</td>
<td>Principal axis factor analysis with oblique rotation</td>
<td>Five factors, but different from a priori model. <strong>Tangibles</strong> dimension split into two factors, while <strong>responsiveness</strong> and <strong>assurance</strong> dimensions loaded on a single factor.</td>
</tr>
<tr>
<td>Finn and Lamb (1991)</td>
<td>Original 22 items</td>
<td>LISREL confirmatory factor analysis</td>
<td>Five-factor model had poor fit.</td>
</tr>
<tr>
<td>Babakus and Boller (1991)</td>
<td>Original 22 items</td>
<td>(1) Principal axis factor analysis with oblique rotation. (2) Confirmatory factor analysis</td>
<td>(1) Five-factor model not supported (2) Two factors</td>
</tr>
<tr>
<td>Cronin and Taylor (1992)</td>
<td>Original 22 items</td>
<td>Principal axis factor analysis with oblique rotation</td>
<td>Unidimensional structure</td>
</tr>
<tr>
<td>Van Dyke and Popelka (1993)</td>
<td>19 of original 22 items</td>
<td>Principal axis factor analysis with oblique rotation</td>
<td>Unidimensional structure</td>
</tr>
<tr>
<td>Pitt, Watson, and Kavan (1995)</td>
<td>Original 22 items</td>
<td>Principal components and maximum likelihood with verimax rotation</td>
<td>(1) Financial institution seven-factor model with <strong>tangibles</strong> and <strong>empathy</strong> split into two. (2) Consulting firm five-factors, none matching the original. (3) Information systems service firm—three-factor model.</td>
</tr>
<tr>
<td>Kettinger and Lee (1994)</td>
<td>Original 22 items</td>
<td>LISREL confirmatory factor analysis</td>
<td>Four-factor model, <strong>tangibles</strong> dimension dropped.</td>
</tr>
<tr>
<td>Kettinger, Lee, and Lee (1995)</td>
<td>Original 22 items</td>
<td>Principal axis factor analysis with oblique rotation</td>
<td>(1) Korea, three-factor model, <strong>tangibles</strong> retained. (2) Hong Kong, four-factor model, <strong>tangibles</strong> retained.</td>
</tr>
</tbody>
</table>

*Measured information systems service quality.
SERVQUAL Concerns

began with 97 paired questions (i.e., one for expectation and one for perception). Items (i.e., question pairs) were first dropped on the basis of within-dimension Cronbach coefficient alphas, reducing the pool to 54 question pairs. More items were then dropped or re-assigned based on oblique-rotation factor loadings and within-dimension Cronbach coefficient alphas resulting in a 34 paired-item instrument with a proposed seven-dimensional structure. A second data collection and analysis with this "revised" definition and operationalization of service quality resulted in the 22 paired-item SERVQUAL instrument with a proposed five-dimensional structure. Two of these five dimensions contained items representing seven of the original 10 dimensions. We are cautioned, however, that those who wish to interpret factors as real dimensions should a substantial burden of proof (Cronbach and Meehl 1955). Moreover, such proof must rely on more than just empirical evidence (e.g., Bynner 1988; Galletta and Lederer 1989).

The results of several studies have demonstrated that the five dimensions claimed for the SERVQUAL instrument are unstable (see Table 2). SERVQUAL studies in the information systems domain have also demonstrated the unstable dimensionality of the SERVQUAL instrument. The service quality of IS services was measured in three different industries, a financial institution, a consulting firm, and an information systems service business (Pitt et al. 1995). Factor analysis was conducted using principal components and maximum likelihood methods with varimax rotation for a range of models. Analysis indicated differing factor structures for each type of firm. Analysis of the results for a financial institution indicated a seven-factor model with both the tangibles and empathy dimensions split into two. These breakdowns should not be surprising. Pitt et al. note that "up-to-date hardware and software" are quite distinct from physical appearances in the IS domain. The empathy dimension was created by the original SERVQUAL authors from two distinctly different constructs, namely understanding and access, which were combined due to the factor loadings alone, without regard to underlying theory. Not only did IS-SERVQUAL not match the proposed model, its factor structure varied across settings. Analysis of the data from the consulting firm resulted in a five-factor model although none of these matched the original a priori factors. The factor analysis of the information systems business data resulted in the extraction of only three factors.

LISREL confirmatory factor analysis was used on SERVQUAL data collected from users (i.e., students) of a college computing services department (Kettinger and Lee 1994). Analysis of this data resulted in a four factor solution. The entire tangibles dimension was dropped. An IS version of SERVQUAL was used in a cross-national study (Kettinger et al. 1995). Results of exploratory common factor analysis with oblique rotation indicated a three-factor model from a Korean sample and a four-factor model was extracted from a Hong Kong data set. The tangibles dimension was retained in the analysis of both of the Asian samples.

The unstable dimensionality of SERVQUAL demonstrated in many domains, including information services, is not just a statistical curiosity. The scoring procedure for SERVQUAL calls for averaging the P-E gap scores within each dimension (Parasuraman et al. 1988). Thus a high expectation coupled with a low perception for one item would be canceled by a low expectation and high perception for another item within the same dimension. This scoring method is only appropriate if all of the items in that dimension are interchangeable. This type of analysis would be justified if SERVQUAL demonstrated a clear and consistent dimensional structure. However, given the unstable number and pattern of the factor structures, averaging groups of items to calculate separate scores for each dimension cannot be justified. Therefore, for scoring purposes, each item should be treated individually and not as part of some a priori dimension.
Discussion and Conclusions

In summary, numerous problems with the original SERVQUAL instrument are described in the literature (e.g., Babakus and Boller 1992; Carman 1990; Cronin and Taylor 1992, 1994; Teas 1993, 1994). The evidence suggests that difference scores, like the SERVQUAL perception-minus-expectation calculation, tend to exhibit reduced reliability, poor discriminant validity, spurious correlations, and restricted variance problems (Edwards 1995; Peter et al. 1993). The fact that the perception component of the difference score exhibits better reliability, convergent validity, and predictive validity than the perception-minus-expectation difference score itself calls into question the empirical and practical usefulness of both the expectations scores as well as the difference scores (Babakus and Boller 1992; Cronin and Taylor 1992; Parasuraman et al. 1994). Moreover, inconsistent definitions and/or interpretations of the "expectation" construct lead to a number of problems. The findings of Teas (1993) suggest that a considerable portion of the variance in SERVQUAL scores is the result of measurement error induced by respondents' varying interpretations of the expectations construct. In addition, since expectations, as well as perceptions, are subject to revision based on experience, concerns regarding the temporal reliability of SERVQUAL difference scores are raised. Furthermore, the dimensionality of the SERVQUAL instrument is problematic.

It was reported that an analysis of IS-SERVQUAL difference scores resulted in either three, five, or seven factors depending on the industry (Pitt et al. 1995). A portion of the instability in the dimensionality of SERVQUAL can be traced to the development of the original instrument (i.e., Parasuraman et al. 1988). Given these problems, users of the SERVQUAL instrument should be cautioned to assess the dimensionality implicit in their specific data set in order to determine whether the hypothesized five-factor structure that has been proposed (Parasuraman et al. 1988, 1991) is supported in their particular domain. Moreover, if the item elimination and dimensional collapsing utilized in the development of SERVQUAL has resulted in a 22 paired-item instrument that in fact does not measure all of the theoretical dimensions of the service quality construct (i.e., content validity), then the use of linear sums of those items for purposes of measuring overall service quality is problematic as well (Galletta and Lederer 1989).

Many of the difficulties identified with the SERVQUAL instrument also apply to the IS-modified versions of the SERVQUAL instrument used by Pitt et al. (1995) and by Kettinger and Lee (1994). It appears that the IS-SERVQUAL instrument, utilizing difference scores, is neither a reliable nor a valid measurement for operationalizing the service quality construct for an information systems services provider. The IS versions of the SERVQUAL instrument, much like the original instrument (Parasuraman et al. 1988), suffer from unstable dimensionality and are likely to exhibit relatively poor predictive and convergent validity, as well as reduced reliability when compared to non-difference scoring methods. The existing literature provides impressive evidence that the use of perception-minus-expectation difference scores is problematic.

This critique of the perceptions-minus-expectations gap score should not be interpreted as a claim that expectations are not important or that they should not be measured. On the contrary, evidence indicates that both should and will expectations are precursors to perceptions but that perceptions alone directly influence overall perceived service quality (Boulding et al. 1993). Our criticism is not with the concept of expectations per se, but rather the operationalization of service quality as a simple subtraction of an ambiguously defined expectations construct from the perceptions of the service actually delivered.

IS professionals have been known to raise expectations to an unrealistically high level in order to gain user commitment to new systems and technologies. This can make it much more difficult to deliver systems and services that will be perceived as successful. According to the model developed by Boulding et al. (1993), per-
ceived service quality can be increased by either improving actual performance or by managing expectations, specifically by reducing should expectations and/or increasing will expectations. These two different types of expectations are not differentiated by the traditional SERVQUAL gap scoring method. A better approach to understanding the impact of expectations on perceived service quality may be to measure will and should expectations separately and then compare them to a service quality measure that utilizes either a direct response or perceptions-only method of scoring.

Prescriptions for the use of SERVQUAL

The numerous problems associated with the use of difference scores suggest the need for an alternative response format. One alternative is to use the perceptions-only method of scoring. A review of the literature (Babakus and Boller 1992; Boulding et al. 1993; Cronin and Taylor 1992; Parasuraman et al. 1991, 1994), indicates that perceptions-only scores are superior to the perception-minus-expectation difference scores in terms of reliability, convergent validity, and predictive validity. In addition, the use of perceptions-only scores reduces by 50% the number of items that must be answered and measured (44 items to 22). Moreover, the findings of Boulding et al. (1993) suggest that expectations are a precursor to perceptions and that perceptions alone directly influence service quality.

A second alternative, suggested by Carman (1990) and Babakus and Boller (1992), is to revise the wording of the SERVQUAL items into a format combining both expectations and perceptions into a single question. Such an approach would maintain the theoretical value of expectations and perceptions in assessing service quality, as well as reduce the number of questions to be answered by 50%. This direct response format holds promise for overcoming the inherent problems with calculated difference scores. Items with this format could be presented with anchors such as "falls far short of expectations" and "greatly exceeds expectations." One study indicates that such direct measures possess higher reliability and improved convergent and predictive validity when compared to difference scores (Parasuraman et al. 1994a).

Conclusion

Recognizing that we cannot manage what we cannot measure, the increasingly competitive market for IS services has emphasized the need to develop valid and reliable measures of the service quality of information systems services providers, both internal and external to the organization. An important contribution to this effort was made with the suggestion of a IS-modified version of the SERVQUAL instrument (Pitt et al. 1995). However, earlier studies raised several important questions concerning the SERVQUAL instrument (e.g., Babakus and Boller 1992; Carman 1990; Cronin and Taylor 1992, 1994; Peter et al. 1993; Teas 1993, 1994). A review of the literature suggests that the use of difference scores with the IS-SERVQUAL instrument results in neither a valid nor reliable measure of perceived IS service quality. Those choosing to use any version of the IS-SERVQUAL instrument are cautioned. Scoring problems aside, the consistently unstable dimensionality of the SERVQUAL and IS-SERVQUAL instruments intimates that further research is needed to determine the dimensions underlying the construct of service quality. Given the importance of the service quality concept in IS theory and practice, the development of improved measures of service quality for an information systems services provider deserves further theoretical and empirical research.

References

Bagozzi, R., Davis, F., and Warshaw, P. "Development and Test of a Theory of


Cronbach, L. J., and Furby, L. "How We Should Measure 'Change'–or Should We?" Psychological Bulletin (74), July 1970, pp. 74–73.


**About the Authors**

Thomas P. Van Dyke is an assistant professor of information systems and technologies at Weber State University and recently completed his a doctoral dissertation in business computer information systems at the University of North Texas. His current research interests include the effects of alternative presentation formats on biases and heuristics in human decision making and MIS evaluation and assessment. He has published articles in *The Journal of Computer Information Systems* and *Proceedings of the Decision Sciences Institute*.

Leon A. Kappelman is an associate professor of business computer information systems in the College of Business Administration at the University of North Texas and associate director of the Center for Quality and Productivity. After a successful career in industry, he received his Ph.D. (1990) in management information systems from Georgia State University. He has published over two dozen journal articles. His work has appeared in *Communications of the ACM*, *Journal of Management Information Systems*, *DATA BASE for Advances in Information Systems*, *Journal of Systems Management*, and *Journal of Computer Information Systems*.  

MIS Quarterly/June 1997 207
Victor R. Prybutok is the director of the University of North Texas Center for Quality and Productivity and an associate professor of management science in the Business Computer Information System Department. He has published articles in *Technometrics*, *Operations Research*, *Economic Quality Control*, and *Quality Progress*, as well as other journals and conference proceedings. Dr. Prybutok is a senior member of the American Society for Quality Control (ASQC), an ASQC certified quality engineer, a certified quality auditor, and a 1993 Texas quality award examiner. His current research interests include project management, assessment of quality programs, neural networks, and MIS evaluation and assessment.