Assessing The Fit And Stability Of Alternative Measures Of Service Quality

Marla Royne Stafford, (E-mail: stafford@cobaf.unt.edu), University of North Texas,
Victor Prybutok, (E-mail: Prybutok @cobaf.unt.edu), University of North Texas
Brenda P. Wells, (E-mail: wells@cobaf.unt.edu), University of North Texas
Leon Kappelman, (E-mail: Kappelma@cobaf.unt.edu), University of North Texas

Abstract

Despite continued research, no single approach to assessing service quality has emerged as a standard, and reconciliation of the differences among these approaches remains quixotic. This paper focuses on the empirical assessment of different approaches to measuring service quality. Furthermore, this work examines the stability of these approaches by comparing the different measures among three companies within the same industry. Two of these approaches (difference scores and perceptions only) are discussed in the existing literature. The third approach (log of the ratio) is introduced in this work because of its potential quantitative advantage for assessing service quality while preserving the comparison between perceptions and expectations. A fourth approach (ratio) is included for completeness, Despite the proposed advantages of the log of the ratio approach, results suggest that the difference score performs the best. Moreover, the five-factor solution originally posited by Parasuraman, Zeithaml and Berry (1988) provides a better fit than the one-factor solution advanced by Cronin and Taylor (1992). While these findings substantiate Parasuraman, Zeithaml and Berry's approach, the strength of the model fit still suggests a need for continued development of measures to effectively assess service quality.

Introduction

Service quality has emerged as an important concern over the past 10 years because organizations attempt to achieve a competitive advantage through improved service. Once equated with courtesy or friendliness, service quality is now generally acknowledged as a multi-dimensional, complex concept that demands academic inquiry and methodological examination.

Readers with comments or questions are encouraged to contact the authors via e-mail.

In response to the need for a measure of evaluating service quality, Parasuraman, Zeithaml, and Berry (1988) developed SERVQUAL, a 22-item measure comprising five dimensions (tangibles, empathy, assurance, reliability and responsiveness) that proposes to operationalize service quality as the mathematical difference between measured expectations and measured perceptions. Paired questions are used to obtain a difference score by subtracting the expectation measure for each individual facet of service quality from the perception measure for that facet. The resulting difference score is used to assess the achieved level of service quality for that particular item.

Since its conception, SERVQUAL has been used and applied in several studies examining service quality across various industries (e.g., Dabholkar, Thorpe, and Rentz 1996; Bolton and Drew 1991a, 1991b; Brown and Swartz 1989). However, despite the long-standing use of SERVQUAL much of the attention afforded the instrument has been negative, with topics ranging from unstable dimensionality to variance restriction. Further, the documented use of SERVQUAL by numerous organizations demonstrates that providing guidance to address the debated issues is of practical significance. As a result of the need for such quantitative assessment of service quality in conjunction with the instrument's identified problems, alternative measures of service quality were proposed. For example, Cronin and Taylor (1992) advanced the use of SERVPERF, a perceptions-only measure, suggesting that the elimination of the expectation score reduces potential threats to reliability and validity associated with the use of difference scores.

Although the use of expectation scores may present problems, there are also advantages, such as richness of data (Parasuraman, Zeithaml, and Berry 1993). For example, expectation scores provide a better understanding as to why a person might perceive lower levels of service quality. That is, an above average perceptions measure may indicate positive service quality, but if the customer expected an even higher level of service quality for that particular item, the provider will not have an accurate understanding of the consumer's true perception of service quality. Further, in a comparative assessment of three scales for measuring service quality, Parasuraman, Zeithaml, and Berry (1994b) found that a better assessment of service quality can be obtained through the use of both expectations and perceptions as compared to perceptions alone.

In an effort to address some of the concerns that are raised through calculation of a difference, we develop and investigate an alternative methodology that allows comparisons of expectation and perceptions. Rather than calculate the difference score, the log (natural log) of the ratio measure is proposed and explored as a means of analyzing service quality. The log of the ratio (lnr) is obtained by taking the natural log of the quantity obtained from the perception score divided by the expectation score. As with the difference score, the log of the ratio (lnr) of perception to expectation allows comparison of performance to expectation on a by-item basis.

The purpose of this paper is to empirically assess the use of these different approaches (difference scores, perception scores, and ln(r) scores) for measuring service quality. For completeness, the ratio approach is also included in the analysis. The analysis is conducted with insurance service provider data because service quality is identified as a critical issue facing the nearly $250 billion industry (CPCU Society 1996). This data also allows comparison of the stability of the different methodological approaches because the data represent three separate (competing) companies within the same industry. These companies present a unique opportunity for assessing the stability of SERVQUAL because they are reasonably homogeneous in size (three of the largest in the state) and focus (insurance industry). Further, each of these companies has historically used the same measure (complaint ratio) to evaluate service quality and direct improvement actions. Such homogeneity is unique to published SERVQUAL studies and supports comparing the stability of the models using the different methodologies.

We begin by discussing the alternative methods for measuring service quality, and presenting several arguments associated with those approaches that have been debated in the literature. We then present an empirical study that investigates the use of these four methods in evaluating service quality.

Difference Scores: Criticisms and Problems

A difference score is calculated by subtracting one measure from another. For example, the SERVQUAL instrument operationalizes the theorized cognitive "gap" score by subtracting an expectations score from a perceptions score. Because the Black Box theory of consumer behavior suggests that internal cognitive evaluation processes cannot be fully traced (Wilkie 1994), we cannot be certain that consumers are able to accurately evaluate service quality by subtracting an expectation from a perception. Parasuraman, Zeithaml, and Berry's SERVQUAL instrument (1988, 1991) measures both the expected level of service and the experienced level of service; service quality scores are then calculated as the difference between these two measures. These scores are generally referred to as expectation (E), perception (P), and the difference, where the difference = P minus E. Although the gap concept itself is theoretically grounded, the implicit assumption that this difference score accurately represents the outcome of this cognitive process is debated in the literature. Moreover, the valid and reliable direct measurement of one's perception of service quality that is the outcome of this cognitive evaluation process seems no more difficult than the valid and reliable direct measurement of one's expected and experienced levels of service quality.

Further, the use of difference scores is questionable when using them as measures of psychological constructs (Lord 1958; Johns 1981; Peter, Churchill, and Brown 1993; Wall and Payne 1973). In fact, several problems associated with the difference score paradigm have been identified by scholars (cf. Carman 1990; Teas 1993; Peter, Churchill and Brown 1993). Among those problems are the following: 1) unstable dimensionality; 2) the ambiguous definition of the "expectations" construct; 3) reliability; and 4) variance restrictions. In addition, the calculation of the difference assumes that both scales are providing interval (or ratio) data and this is not necessarily the case.

Unstable Dimensionality of the SERVQUAL Instrument

The SERVQUAL measure is based on a 10-dimensional model of service quality grounded in both relevant literature as well as the extensive use of both executive and focus group interviews. Parasuraman, Zeithaml, and Berry (1988) (hereafter, referred to as PZB) began the development process with a problematically small sample size of 200 across five industries. Initially, 97 paired questions (i.e., one for expectation and one for perception), or about 10 question pairs for each of the 10 dimensions, were used to calculate difference scores (i.e., P minus E). Forty-three question pairs were first dropped on the basis of within-dimension Cronbach (Cronbach 1951) coefficient alphas, reducing the pool to 54 question pairs. 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 dimensions contained items representing seven of the original 10 dimensions, identified in PZB (1985). Cronbach and Meehl (1955), however, cautioned those who interpret those factors as real dimensions; further proof of these dimensions is necessary, and such proof must rely on more than just empirical evidence (Bynner 1988; Galletta and Lederer 1989).

Several studies have demonstrated that the five dimensions claimed for SERVQUAL are unstable. Babakus and Boller (1992) reported no clear dimensional structure; more specifically, their confirmatory factor analysis of the proposed five-dimensional model provided a poor fit. Rather, a principal-axis factor analysis followed by an oblique rotation suggested a two-dimensional structure.

Other research also indicates the instability of the SERVQUAL dimensions. For example, Brensinger and Lambert's (1990) results suggested a four-dimensional solution. The work of Carman (1990) and Finn and Lamb (1991) also failed to replicate Parasuraman, Zeithaml, and Berry's (1988) five-dimension model. Finn and Lamb reported a poor fit to the model, while Carman's (1990) work across four different service industries concluded that seven or eight of Parasuraman, Zeithaml, and Berry's original 10 dimensions were represented by the 22 paired-item SERVQUAL instrument. Conversely, the work of Cronin and Taylor (1992; 1994) argue that no substantial evidence exists to justify anything beyond a summed index. After a LISREL-based analysis failed to confirm Parasuraman, Zeithaml, and Berry's five-factor model, Cronin and Taylor concluded that SERVQUAL may be treated as unidimensional since "all of the items loaded predictably on a single factor with the exception of item 19 (i.e., personal attention)" (p. 61). Thus, SERVPERF was conceived.

Even the creators of SERVQUAL were unable to confirm their original five dimensions in a later (1991) study across five companies in three industries. Although results did suggest five factors, they were not the same five reported in the 1988 study. In the 1991 study, the tangibles dimension split into two factors, while the responsiveness and assurance dimensions loaded strongly on a single factor, and a six-factor solution was required to differentiate between them (Parasuraman, Zeithaml, and Berry 1991).

In short, research attempting to replicate the original five SERVQUAL dimensions has been unsuccessful. It is possible that the inability to replicate these dimensions is a function of cross industry investigation. That is, although PZB argue that SERVQUAL transcends industry boundaries, it may require extensive industry modification to ensure relevancy and validity. However, the inability to replicate the dimensionality may also be a function of another SERVQUAL concern: the ambiguity of the expectations construct.

Ambiguity of the Expectations Construct

Teas (1994) noted that SERVQUAL expectations have been defined in numerous ways, including 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, Zeithaml and Berry, 1988; 1991; 1994a; Zeithaml, Berry, and Parasuraman, 1993). These multiple definitions and corresponding operationalizations of expectations in the SERVQUAL literature result in multiple interpretations (Teas 1994).

More specifically, Teas (1993) identified three different interpretations of expectations based on 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 Oliver's (1980) disconfirmed expectations model of consumer satisfaction. Yet, this interpretation is inconsistent with PZB's (1988) definition of service quality. Consequently, there appears to be some difference in the definition of expectation based on whether or not customer satisfaction or service quality is being measured. Furthermore, expectations themselves are a function of personal experience and could vary widely, based on that experience. Collectively, individual experience, or the lack thereof, can establish an average level of expectation that varies widely from industry to industry, but probably not among companies within the same industry. This expected homogeneity within an industry lends greater importance to this research study and our unique data set. A second interpretation of expectations is as a measure of attribute importance. Based on this interpretation, resulting gap scores exhibit an inverse relationship between attribute importance and perceived service quality.

Yet another interpretation noted by Teas is the "classic ideal point" (Teas 1993) concept. PZB described this when they stated, "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 displease the customer (e.g., friendliness of a salesperson in a retail store)" (Parasuraman, Zeithaml, and Berry, 1994a, p. 116). This is similar to the finding reported by Doll and Torkzadeh (1989) that satisfaction scores were highest when actual user participation was congruent with the user's need for participation, rather than merely maximized. This third interpretation of expectation results in an inverse of the relationship between SERVQUAL calculated as perceptions minus expectations (P-E) and perceived SERVQUAL (P only) for all values when perception scores are greater than expectation scores (i.e., P>E).

A final interpretation of expectations is consistent with the SERVQUAL model. This is the *vector attribute* interpretation, or "one on which a customer's ideal point is at an infinite level" (Parasuraman, Zeithaml, and Berry, 1994a, p. 116). Unfortunately, as the proportion of extreme responses (e.g., 7 on a 7 point scale) increases, the expectation scores become less useful, as an increasing proportion of the variation in gap-based SERVQUAL scores is due only to changes in perceptions scores.

Reliability

Previous research suggests that Cronbach's (1951) alpha, a well-accepted measure for estimating instrument reliability, is inappropriate for difference scores (e.g., Lord 1958; Cronbach and Furby 1970; Wall and Payne 1973; Johns 1981; Prakash and Lounsbury 1983; Peter, Churchill and Brown 1993; Edwards 1995) because the reliability of such a measure is dependent on the reliability of the original components and the correlation between them. The proper formula for calculating the reliability of a difference score (r_D) is:

$$r_D = \sigma_1^2 r_{11} + \sigma_2^2 r_{22} - 2r_{12}\sigma_1\sigma_2 / \sigma_1^2 + \sigma_2^2 - 2r_{12}\sigma_1\sigma_2 \quad (1)$$

where r_{11} and r_{22} are the reliabilities of the two component scores, σ_1^2 and σ_2^2 are the variances of the component scores, and r_{12} is the correlation between the component scores (Johns 1981).

This formula illustrates that as the correlation of the component scores increases, the reliability of the difference score decreases. An example of this relationship is offered by Johns (1981). 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. Although the average reliability of the two components (.70) is considered by some to be acceptable, (cf. Pitt, Watson, and Kavan 1995; Nunnally 1978), the correlation between the original components reduces the reliability of the difference score to a level that most scholars would consider unacceptable (Peter, Churchill, and Brown 1993). Through the use of Cronbach's alpha, researchers tend to overestimate the reliabilities of difference scores particularly when the component scores are highly correlated as is the case with SERVQUAL (Peter, Churchill, and Brown 1993).

Variance Restrictions

Yet another problem associated with SERVQUAL scores is the potential restriction of the variance of the difference scores. This problem arises when one of the component scores is consistently higher than the other (Peter, Churchill, and Brown 1993), as is the case with SERVQUAL. For example, Teas (1993) reported that 62% of the responses for the expectations scores were 7 on a 7-point scale.

Given that the expectation component of SERVQUAL is usually higher than the perception component, those customers with low perception scores have a wider potential range on the difference score than do customers with high perception scores. This restriction in range can lead to a restriction in the variance of the difference scores. The variance restriction is systematic and problematic for many types of statistical analyses (Prakash and Lounsbury 1983; Peter, Churchill, and Brown 1993).

Performance-only Measures

In an attempt to alleviate many of the problems associated with difference scores, several scholars have argued for performance-only based measures, where a linear sum score of the individual items comprising the perceptions of service quality is used (e.g., Bolton and Drew 1991a, 1991b; Churchill and Surprenant 1981; Cronin and Taylor 1992; Woodruff, Cadotte and Jenkins 1983). However, critics maintain that the perceptions-only paradigm is not as superior as believed (Parasuraman, Zeithaml, and Berry 1993), although PZB (1994a) concede that the use of perceptions-only scores dominates in practice.

Not surprisingly, the strongest critics of the perceptions-only model are Parasuraman, Zeithaml, and Berry, originators of SERVQUAL. PZB (1993; 1994a) argue that the problems associated with SERVQUAL, and eliminated by performance-only measures, are not completely valid. The authors also contend that there is no conceptual reason that a customer's general expectation scores are correlated with company-specific perceptions and that correlations between expectations and perceptions are simply an artifact of shared method variance; hence, correlations are not likely to be high. PZB further argue that strong reliabilities have been demonstrated with the use of SERVQUAL difference scores, and that Cronin and Taylor (1992) use reliability (coefficient alpha) incorrectly as a measure of unidimensionality. This latter contention is well supported in the literature (cf. Anderson and Gerbing 1982; Gerbing and Anderson 1988).

However, PZB do agree that variance restriction is a concern of SERVQUAL and concede that a perceptions-only measure is superior on this front. Nevertheless, they question whether an increased ability to explain variance is worth the trade-off in the richness and fullness of the data provided by SERVQUAL. Further, several studies conducted by PZB suggest that perceptions alone may lead to different (and possibly incorrect) implications than provided by difference scores. In short, PZB argue that clear superiority for the perceptions score has not been demonstrated.

A Proposed Alternative: Log of the Ratio

That customers' expectations exceed their perceptions is consistent with the service quality construct. Naturally, an exceptional company's performance may exceed expectations for a period, but in the longer term such a company is likely to raise expectations and still have a valid measure through comparison of expectation and perception. Unfortunately, use of the difference scores for validation of constructs associated with this service quality is problematic, and such problems are well documented. As an alternative, comparison between expectation and perception could be achieved through the use of the ratio of perception over expectation (P/E). However, such ratios also contain documented variance problems that preclude their direct use (Neter and Wasserman 1974). To simultaneously address concerns that are raised by comparison of expectations with perceptions through calculation of a difference and the problems associated with the ratio, we propose using the natural log of the ratio of the perception score for an item divided by the expectation score for the same item. For comparison purposes, however, we have included models using both the ratio and the log of the ratio. While the log of the ratio approach allows comparison of an item's perception with its expectation, we believe it also provides a better approach to management of the variance restriction problem that is associated with the difference scores. While the variance restriction problem with differences is well documented (Peter, Churchill, and Brown 1993; Teas 1993), there exists a paucity of suggestions for addressing the problem in a manner that allows inclusion of the expectation. However, utilization of ratios allows the use of the log transformation that addresses the variance restriction problem associated with proportions. This log transformation is a documented approach to linearize a multiplicative model such as is obtained via the ratio (Neter and Wasserman 1974). As with the difference score, the log of the ratio (lnr) of perception to expectation allows comparison of performance to expectation on a by-item basis. In this ratio model:

Y = Performance / Expectations
or (2)
Y = P/E

Taking the $\log_n$ provides the following:

$$Y' = \log_n{}^{y} = \log_n(P/E) \quad (3)$$
$$Y' = \log_n{}^{P} - \log_n{}^{E} \quad (4)$$

To verify that this approach improved the variance restriction problem by items, histograms from the resulting data collection were examined for the performance, the difference, the ratio and the log of the ratio measurement. Visual examination of these histograms showed that the log of the ratio provided the best distribution of responses (the most normalized data). While such graphics do not constitute statistical evidence, they were deemed sufficient to warrant further investigation of the proposed method.

Method

A questionnaire was developed utilizing the 22 items comprising the original SERVQUAL instrument; both measures of expectations and perceptions were included. Following the approach of PZB (1988), respondents were instructed to indicate their expectations of service quality performance for the insurance industry, in general, for the 22 expectation items. They were then asked to rate their perceptions of their individual insurance company's performance on each of these 22 items. All measures used a seven-point scale. Demographic information was also included on the survey.

Sample and Data Collection

Because insurance companies sell essentially the same state-mandated insurance policy, an important area for competition is service quality provided to the insured. The scope of the insurer's service begins with the selling of the policy, and continues until the policy is either not renewed or canceled. It is documented, however, that the major purpose of insurance is indemnification of losses (Cummins, et. al. 1974). The consumer purchases insurance in anticipation of an unfortunate event, and only when such an event occurs does the consumer receive the full benefit of the insurer's services. The claim for damages is the most important stage of the insurance transaction, and is the point at which service quality should be assessed. Therefore, a random sample of consumers in a southwestern state filing a first-party claim with their auto insurer during one single month was selected. The sample was restricted to first-party claims to avoid the potential adversarial nature of third-party liability claims that could confound results.

The sample was obtained through the cooperation of three large insurance companies. However, as a condition of participation in this research, the three companies were assured of complete anonymity, and thus are not identified by name.

A total of 2093 packets containing the questionnaire, a cover letter and a postage-paid envelope for returning the completed surveys were mailed. Company sample sizes varied:

Company 1 provided 647 names, Company 2 provided 696 names, and Company 3 provided 750 names. Sixty packets were returned due to address problems. After a follow-up mailing, and elimination of incomplete questionnaires, the final sample sizes were 178, 164 and 241, respectively, and usable response rates for each of the companies were 27.5 percent, 23.4 percent and 32.1 percent, respectively. The total usable sample of 583 resulted in an overall final response rate of about 28.7 percent, which is consistent with the response rate of previous studies in the area of service quality (e.g., Babakus and Boller 1992; Parasuraman, Zeithaml, and Berry 1991). Demographic characteristics of the respondents are presented in Table 1. These characteristics are presented in total and by each company.

Analysis and Results

To assess the fit and stability of each of the four potential measures for assessing service quality, LISREL 8.03 was used for confirmatory factor analysis. The analysis required a confirmatory factor analysis (CFA) for each of the possible approaches for each of the three companies. Moreover, the question of dimensionality was examined by assessing whether or not one or five dimensions provided the better fitting model for each of the analytical approaches. Therefore, 24 different models were required to complete our analysis: both the one-dimension and five-dimension versions of difference scores, perceptions only, the ratio, and the log of the ratio for three different companies. Results of these models are presented in Tables 2, 3 and 4.

Table 1
Respondent Demographics

Age	**Company 1 n=178**	**Company 2 n=164**	**Company 3 n=241**	**All Cos. n=583**
18-25	2.3%	4.3%	4.6%	3.8%
26-35	14.3%	23.5%	15.4%	17.3%
36-45	25.1%	22.8%	19.9%	22.3%
46-55	21.7%	22.2%	27.0%	24.0%
56-65	10.3%	11.7%	12.0%	11.4%
Over 65	26.3%	15.4%	21.2%	21.2%
Annual Income	**Company 1 n=178**	**Company 2 n=164**	**Company 3 n=241**	**All Cos. n=583**
Under $15,000	6.5%	9.7%	4.4%	6.5%
$15,001- $25,000	12.9%	14.9%	13.6%	13.8%
$25,001- $35,000	10.6%	21.4%	14.5%	15.2%
$35,001- $45,000	14.7%	12.3%	15.4%	14.3%
$45,001- $55,000	15.3%	10.4%	12.3%	12.7%
$55,001- $65,000	12.4%	8.4%	9.6%	10.1%
$65,001- $75,000	10.0%	7.8%	7.5%	8.3%
Over $75,000	17.6%	14.9%	22.8%	19.0%
Sex	**Company 1 n=178**	**Company 2 n=164**	**Company 3 n=241**	**All Cos. n=583**
Male	62.6%	56.2%	60.1%	59.8%
Female	37.4%	43.8%	39.9%	40.2%
Education	**Company 1 n=178**	**Company 2 n=164**	**Company 3 n=241**	**All Cos. n=583**
Some High School	4.6%	4.4%	2.5%	3.7%
High School Graduate	33.9%	34.4%	27.3%	31.3%
Technical/Trade School	12.1%	15.0%	12.2%	12.9%
College Graduate	24.1%	33.1%	30.7%	29.4%
Graduate Work	7.5%	6.9%	10.5%	8.6%
Completed Graduate Studies	17.8%	6.3%	16.8%	14.2%

NOTE: Due to rounding errors and respondent omissions, percentages do not always sum to exactly 100.0%.

Table 2 - Overall Goodness of Fit Measures

Perception Scores								
	Company 1		Company 2		Company 3		Standard Deviation	
Measure*	1 dimens.	5 dimens.s	1 dimens.	5 dimens.s	1 dimens.	5 dimens.	1 dimens.	5 dimens.
χ^2	1094.83	694.64	877.15	493.18	1237.70	811.74	---	---
$p(\chi^2)$	0.00	0.00	0.00	0.00	0.00	0.00	---	---
GFI	0.61	0.72	0.66	0.79	0.64	0.74	0.025	0.036
AGFI	0.53	0.65	0.59	0.73	0.56	0.67	0.030	0.042
RMR	0.09	0.08	0.08	0.06	0.07	0.09	0.009	0.015
SRMR	0.09	0.08	0.08	0.06	0.07	0.09	0.009	0.015
RMSEA	0.15	0.12	0.14	0.10	0.14	0.11	0.006	0.010
p (RMSEA < .05)	0.00	0.00	0.00	0.00	0.00	0.00	---	---
NFI	0.68	0.80	0.73	0.85	0.76	0.84	0.040	0.026
NNFI	0.69	0.82	0.76	0.89	0.77	0.85	0.044	0.035
Difference Scores								
	Company 1		Company 2		Company 3		Standard Deviation	
Measure*	1 dimens.	5 dimens.	1 dimens.	5 dimens.	1 dimens.	5 dimens.	1 dimens.	5 dimens.
χ^2	875.48	487.69	728.03	364.08	964.72	518.25	---	---
$p(\chi^2)$	0.00	0.00	0.00	0.00	0.00	0.00	---	---
GFI	0.67	0.80	0.68	0.83	0.70	0.83	0.015	0.017
AGFI	0.60	0.75	0.62	0.79	0.64	0.78	0.020	0.021
RMR	0.10	0.09	0.09	0.06	0.08	0.07	0.009	0.014
SRMR	0.10	0.09	0.09	0.06	0.08	0.07	0.009	0.014
RMSEA	0.13	0.09	0.12	0.07	0.12	0.08	0.006	0.010
p (RMSEA < .05)	0.00	0.00	0.00	0.00	0.00	0.00	---	---
NFI	0.62	0.79	0.74	0.87	0.72	0.85	0.064	0.042
NNFI	0.65	0.84	0.78	0.93	0.74	0.89	0.067	0.045
Ratios								
	Company 1		Company 2		Company 3		Standard Deviation	
Measure*	1 dimens.	5 dimens.	1 dimens.	5 dimens.	1 dimens.	5 dimens.	1 dimens.	5 dimens.
χ^2	1621.83	739.90	1349.68	652.71	1437.69	700.62	---	---
$p(\chi^2)$	0.00	0.00	0.00	0.00	0.00	0.00	---	---
GFI	0.42	0.76	0.43	0.76	0.63	0.81	0.118	0.029
AGFI	0.30	0.70	0.32	0.64	0.55	0.75	0.139	0.055
RMR	0.22	0.13	0.21	0.15	0.13	0.10	0.049	0.025
SRMR	0.22	0.13	0.21	0.15	0.13	0.10	0.049	0.025
RMSEA	0.20	0.12	0.18	0.12	0.16	0.10	0.020	0.012
p (RMSEA < .05)	0.00	0.00	0.00	0.00	0.00	0.00	---	---
NFI	0.30	0.68	0.58	0.80	0.49	0.73	0.143	0.060
NNFI	0.25	0.70	0.57	0.82	0.43	0.76	0.160	0.060
Log of Ratios								
	Company 1		Company 2		Company 3		Standard Deviation	
Measure*	1 dimens.	5 dimens.	1 dimens.	5 dimens.	1 dimens.	5 dimens.	1 dimens.	5 dimens.
χ^2	1154.11	670.24	945.13	552.90	1111.76	1631.90	---	---
$p(\chi^2)$	0.00	0.00	0.00	0.00	0.00	0.00	---	---
GFI	0.60	0.76	0.63	0.77	0.68	0.79	0.040	0.015
AGFI	0.52	0.69	0.55	0.71	0.61	0.73	0.046	0.020
RMR	0.13	0.11	0.10	0.08	0.09	0.08	0.022	0.018
SRMR	0.13	0.11	0.10	0.08	0.09	0.08	0.022	0.018
RMSEA	0.16	0.12	0.15	0.10	0.13	0.11	0.015	0.010
p(RMSEA <.05)	0.00	0.00	0.00	0.00	0.00	0.00	---	---
NFI	0.51	0.71	0.66	0.80	0.69	0.81	0.096	0.055
NNFI	0.51	0.74	0.68	0.84	0.70	0.80	0.104	0.050

*GFI = Goodness of Fit Index, AGFI = Adjusted Goodness of Fit Index, RMR = Root Mean Square Residual, RMSEA = Root Mean Square

Table 3
Factor Loadings for Five (5) Dimension Models

	Item	λx	Company 1 Perc	Company 1 Diff	Company 1 Ratio	Company 1 Log	Company 2 Perc	Company 2 Diff	Company 2 Ratio	Company 2 Log	Company 3 Perc	Company 3 Diff	Company 3 Ratio	Company 3 Log
Tangibility	EQUIP	λ_{x11}	.84	.72	.77	.71	.77	.63	.59	.62	.81	.51	.48	.53
			(13.51)	(10.29)	(11.87)	(10.43)	(11.34)	(8.19)	(7.28)	(8.00)	(14.61)	(7.40)	(6.68)	(11.97)
	APPEAL	λ_{x21}	.86	.86	.97	.93	.82	.74	.48	.71	.88	.73	.78	.66
			(14.00)	(13.18)	(17.19)	(15.00)	(12.48)	(10.04)	(5.90)	(9.47)	(16.46)	(11.14)	(10.62)	(15.26)
	NEAT	λ_{x31}	.82	.75	.76	.70	.83	.79	.86	.79	.75	.74	.64	.72
			(13.05)	(11.01)	(11.77)	(10.11)	(12.53)	(11.08)	(11.04)	(10.68)	(13.10)	(11.40)	(8.93)	(16.67)
	FACIL	λ_{x41}	.87	.69	.84	.71	.89	.76	.56	.71	.64	.60	.33	.56
			(14.11)	(9.85)	(13.65)	(10.33)	(13.95)	(10.56)	(6.88)	(9.39)	(10.49)	(8.93)	(4.45)	(12.54)
Reliability	PROMISE	λ_{x52}	.87	.88	.89	.95	.88	.85	.41	.82	.91	.89	.90	.90
			(14.37)	(14.58)	(14.68)	(16.79)	(14.14)	(13.36)	(5.44)	(12.68)	(18.33)	(17.48)	(17.76)	(27.91)
	PROBE	λ_{x62}	.82	.78	.75	.73	.86	.85	.44	.83	.84	.79	.78	.80
			(13.06)	(11.97)	(11.45)	(11.08)	(13.58)	(13.40)	(5.84)	(12.80)	(15.91)	(14.35)	(14.16)	(22.96)
	DEPEND	λ_{x72}	.79	-12.270	.78	.69	.88	.90	.98	.89	.88	.86	.80	.85
			(12.46)		(11.98)	(10.20)	(14.31)	(14.64)	(17.50)	(14.40)	(17.19)	(16.64)	(14.77)	(25.10)
	TIME	λ_{x82}	.86	.86	.87	.90	.89	.90	.97	.90	.93	.90	.91	.91
			(14.14)	(14.00)	(14.11)	(15.15)	(14.54)	(14.65)	(16.84)	(14.76)	(18.92)	(17.80)	(17.85)	(28.08)
	ACCUR	λ_{x92}	.51	.46	.44	.30	.67	.71	.94	.67	.64	.65	.63	.56
			(7.08)	(6.20)	(5.88)	(3.98)	(9.65)	(10.19)	(15.96)	(9.52)	(10.94)	(10.98)	(10.67)	(14.47)
Responsiveness	PERFORM	λ_{x103}	.76	.62	.47	.61	.85	.74	.48	.72	.83	.76	.61	.76
			(11.81)	(8.68)	(5.96)	(8.40)	(13.37)	(10.87)	(6.21)	(10.35)	(15.67)	(13.39)	(9.41)	(20.65)
	PROMPT	λ_{x113}	.88	.69	.26	.62	.92	.85	.70	.84	.90	.87	.82	.88
			(14.61)	(10.05)	(3.23)	(8.70)	(15.32)	(13.32)	(9.68)	(12.79)	(17.84)	(16.53)	(13.73)	(25.83)
	HELP	λ_{x123}	.75	.69	.59	.64	.72	.73	.67	.68	.77	.72	.61	.69
			(11.55)	(10.05)	(7.37)	(8.92)	(10.49)	(10.63)	(9.04)	(9.62)	(13.87)	(12.53)	(9.50)	(18.37)
	BUSY	λ_{x133}	.66	.62	.41	.59	.67	.69	.48	.67	.60	.63	.64	.59
			(9.65)	(8.81)	(5.09)	(8.17)	(9.59)	(9.79)	(6.20)	(9.31)	(10.01)	(10.47)	(10.01)	(15.00)
Assurance	TRUST	λ_{x144}	.86	.83	.78	.76	.80	.83	.94	.80	.83	.79	.40	.74
			(13.81)	(12.66)	(10.85)	(11.01)	(11.94)	(12.59)	(15.90)	(11.92)	(15.42)	(14.25)	(6.42)	(20.23)
	SAFE	λ_{x154}	.88	.83	.39	.73	.90	.91	.97	.93	.89	.89	.99	.93
			(14.32)	(12.63)	(4.85)	(10.30)	(14.31)	(14.68)	(17.04)	(14.94)	(17.21)	(16.92)	(19.18)	(28.30)
	POLITE	λ_{x164}	.71	.64	.65	.71	.71	.74	.95	.67	.83	.75	.85	.78
			(10.55)	(9.04)	(8.83)	(9.95)	(10.05)	(10.71)	(16.19)	(9.40)	(15.39)	(13.18)	(15.21)	(21.77)
	SUPPORT	λ_{x174}	.52	.38	-.05	.22	.49	.39	-.02	.17	.68	.65	.45	.62
			(7.18)	(4.94)	(-.60)	(2.71)	(6.44)	(5.02)	(.29)	(2.08)	(11.74)	(10.78)	(7.27)	(16.00)
Empathy	ATTENT	λ_{x185}	.85	.70	.49	.63	.88	.89	.90	.88	.87	.82	.78	.80
			(13.94)	(10.30)	(6.58)	(8.74)	(14.08)	(14.34)	(14.56)	(14.02)	(16.69)	(15.07)	(12.39)	(22.44)
	EMPATT	λ_{x195}	.87	-12.200	.89	.77	.93	.93	.97	.95	.89	.78	.50	.72
			(14.25)		(14.47)	(11.48)	(15.58)	(15.45)	(16.47)	(16.02)	(17.63)	(13.91)	(7.33)	(19.23)
	NEEDS	λ_{x205}	.68	.68	.54	.64	.71	.73	.47	.73	.87	.75	.53	.79
			(10.03)	(9.78)	(7.36)	(8.97)	(10.33)	(10.61)	(6.25)	(10.60)	(16.82)	(13.18)	(7.91)	(21.83)
	HEART	λ_{x215}	.81	.78	.89	.80	.79	.74	.57	.74	.83	.77	.46	.77
			(12.93)	(11.85)	(14.37)	(12.05)	(11.83)	(10.92)	(7.87)	(10.90)	(15.57)	(13.54)	(6.76)	(21.07)
	HOURS	λ_{x225}	.70	.62	.46	.63	.63	.66	.41	.63	.50	.51	.53	.49
			(10.45)	(8.78)	(6.15)	(8.79)	(8.81)	(9.30)	(5.37)	(8.78)	(8.14)	(8.08)	(7.81)	(11.84)

Table 2 displays the overall goodness-of-fit measures for each of the models, while Tables 3 and 4 show the factor loadings (lambda coefficients) and associated t-values.

Although Table 2 suggests that none of our models offer a particularly strong fit, it does allow comparison among the companies as well as the four approaches. Further, while we could have pursued changes in correlations or error estimates that would have improved the fit of our models, these changes would have been problematic for purposes of comparisons. The optimization of the fits would either result in different approaches for each of the models or would have been one that optimized one model more than the others. Such an approach would not be consistent with the intent of this work.

The ratio was used as an alternative approach for comparing the perceptions and expectations on an item-by-item basis. As with the difference scores, the calculation of the ratio assumes that both measurements (perceptions and expectations) are at least interval scale. As noted earlier, the logarithmic transformation of this ratio was used to avoid the variance restriction problem that is associated with ratio and difference measures. As expected, the logs of the ratio models were superior to the ratio models. For example, for the ratio models, the goodness-of-fit indices (GFIs) were as low as .42, and the normed fit indices (NFIs) were as low as .30. Only two adjusted goodness-of-fit indices (AGFIs) exceeded .70. Clearly, the ratio models were the poorest across all three companies, and thus, we believe do not warrant further discussion.

In spite of the justification for the log of the ratio approach, Table 2 shows that the difference scores with five dimensions provides the best fitting model across all three companies. The GFIs for the three companies for this model range from .80 to .83, and the AGFIs range from .75 to .79. The NFIs are in the range of .79 to .87, and the non-normed fit indices (NNFIs) range from .84 to .93.

Further, there seems to be little difference in the variation exhibited among the company to company values for a given model. To examine this variation, the standard deviations of the fit measures were calculated and presented to allow comparison among the models for stability across companies. (The χ^2 standard deviation was not calculated because comparison among models with different degrees of freedom is inappropriate.) These standard deviations reveal little difference among the different models using this criteria, although the largest deviations are produced by the ratio models followed by the ln(r) models. For both the NNFI and the NFI, these deviations are clearly highest for the one-dimension model, although they are evident for the five-dimension models, as well. Such results suggest that the ratio and the ln(r) models have the least stability of the different service quality measures. However, the variation identified by both the perceptions and difference models are quite similar, and relative to the ratio and log of the ratio models, fairly minimal. Thus, there appears to be little difference in the stability, or lack thereof, of the perception and difference measures across companies.

Interestingly, while the perceptions-only one-dimension models fit poorly (AGFIs ranging from .53 to .59, and NFIs from .68 to .76), the five-dimension perceptions-only models were much stronger with AGFIs ranging from .65 to .72, NFIs from .80 to .85, and NNFIs from .82 to .89. Thus, while advocates of the perceptions only approach argue for a one-dimension measure, results of the current study suggest that no matter which analytical approach is being utilized, a multi-dimensional conceptualization of service quality is evident. Tables 3 and 4 present the factor loadings for each of the models on the original SERVQUAL for both the one- and five-dimension options. These tables show that no one model consistently provides the best fit to the original five dimensions of SERVQUAL. In fact, results suggest that the highest loadings in

Table 4
Factor Loadings for One (1) Dimension Models

Item	λx	Company 1				Company 2				Company 3			
		Perc	Diff	Ratio	Log	Perc	Diff	Ratio	Log	Perc	Diff	Ratio	Log
		.57	.43	.91	.42	.51	.38	.18	.32	.46	.33	.29	.34
EQUIP	λ_{x11}	(8.23)	(5.78)	(15.38)	(5.60)	(6.85)	(4.99)	(2.29)	(4.17)	(7.49)	(5.15)	(4.35)	(5.26)
		.56	.23	.83	.13	.51	.23	.02	.15	.54	.20	.13	.25
APPEAL	λ_{x21}	(8.04)	(3.05)	(13.22)	(1.68)	(6.88)	(2.92)	(.21)	(1.95)	(8.95)	(3.03)	(1.93)	(3.81)
		.64	.42	.78	.33	.52	.37	.10	.31	.58	.24	.09	.26
NEAT	λ_{x31}	(9.41)	(5.70)	(12.23)	(4.35)	(7.10)	(4.87)	(1.27)	(3.99)	(9.79)	(3.66)	(1.41)	(3.94)
		.55	.18	.73	.03	.56	.34	.10	.26	.67	.31	.18	.45
FACIL	λ_{x41}	(7.84)	(2.30)	(11.07)	(0.34)	(7.74)	(4.38)	(1.25)	(3.29)	(11.67)	(4.88)	(2.71)	(7.14)
		.79	.79	.21	.84	.85	.83	.42	.82	.86	.83	.87	.85
PROMISE	λ_{x52}	(12.56)	(12.53)	(2.72)	(13.58)	(13.60)	(12.86)	(5.51)	(12.68)	(16.60)	(15.74)	(16.73)	(16.31)
		.82	.76	.18	.76	.85	.85	.44	.83	.86	.80	.79	.80
PROBE	λ_{x62}	(13.23)	(11.68)	(2.35)	(11.58)	(13.52)	(13.34)	(5.90)	(12.85)	(16.81)	(14.71)	(14.37)	(14.86)
		.75	.75	.18	.66	.87	.85	.98	.83	.85	.83	.79	.83
DEPEND	λ_{x72}	(11.72)	(11.48)	(2.32)	(9.70)	(13.96)	(13.42)	(17.26)	(12.95)	(16.51)	(15.80)	(14.44)	(15.55)
		.79	.80	.18	.81	.86	.84	.96	.83	.86	.83	.87	.85
TIME	λ_{x82}	(12.56)	(12.54)	(2.37)	(12.89)	(13.63)	(13.09)	(16.72)	(12.99)	(16.73)	(15.74)	(16.59)	(16.17)
		.57	.50	.22	.40	.69	.69	.94	.65	.64	.64	.64	.56
ACCUR	λ_{x92}	(8.12)	(6.99)	(2.84)	(5.34)	(9.89)	(9.96)	(15.96)	(9.22)	(11.03)	(10.92)	(10.67)	(9.29)
		.76	.61	.48	.59	.81	.72	.39	.68	.79	.68	.40	.64
PERFORM	λ_{x103}	(11.82)	(8.75)	(6.58)	(8.45)	(12.62)	(10.47)	(5.08)	(9.73)	(14.59)	(11.70)	(6.20)	(10.83)
		.85	.68	.09	.62	.89	.82	.29	.79	.85	.79	.53	.76
PROMPT	λ_{x113}	(14.06)	(10.13)	(1.21)	(8.96)	(14.56)	(12.76)	(3.80)	(11.96)	(16.33)	(14.64)	(8.62)	(13.79)
		.73	.66	.16	.60	.74	.74	.34	.70	.79	.72	.34	.71
HELP	λ_{x123}	(11.24)	(9.66)	(2.01)	(8.57)	(11.04)	(10.86)	(4.49)	(10.05)	(14.60)	(12.76)	(5.20)	(12.47)
		.64	.58	.15	.55	.67	.66	.15	.62	.59	.56	.37	.49
BUSY	λ_{x133}	(9.45)	(8.25)	(1.88)	(7.71)	(9.61)	(9.47)	(1.95)	(8.64)	(9.99)	(9.24)	(5.68)	(7.91)
		.70	.61	-.14	.52	.69	.72	.93	.69	.76	.68	.32	.63
TRUST	λ_{x144}	(10.51)	(8.86)	(-1.79)	(7.21)	(9.97)	(10.53)	(15.59)	(9.85)	(13.75)	(11.75)	(4.86)	(10.71)
		.71	.60	.06	.53	.80	.80	.96	.80	.83	.74	.42	.77
SAFE	λ_{x154}	(10.77)	(8.68)	(.75)	(7.32)	(12.19)	(12.29)	(16.75)	(12.21)	(15.66)	(13.35)	(6.60)	(13.93)
		.74	.66	.12	.65	.64	.66	.93	.56	.80	.66	.29	.65
POLITE	λ_{x164}	(11.37)	(9.66)	(1.49)	(9.46)	(9.07)	(9.33)	(15.86)	(7.71)	(14.87)	(11.47)	(4.36)	(11.22)
		.54	.46	.86	.37	.53	.47	-.02	.33	.69	.64	.35	.59
SUPPORT	λ_{x174}	(7.73)	(6.32)	(14.03)	(4.98)	(7.15)	(6.29)	(-.24)	(4.24)	(12.19)	(10.98)	(5.37)	(9.80)
		.81	.64	.06	.53	.78	.81	.19	.77	.80	.75	.56	.74
ATTENT	λ_{x185}	(13.13)	(9.32)	(.80)	(7.34)	(11.86)	(12.51)	(2.42)	(11.55)	(14.88)	(13.47)	(9.20)	(13.31)
		.83	.69	.18	.63	.85	.84	.28	.84	.82	.72	.32	.66
EMPATT	λ_{x195}	(13.53)	(10.31)	(2.34)	(9.12)	(13.40)	(13.31)	(3.69)	(13.06)	(15.59)	(12.83)	(4.86)	(11.36)
		.64	.57	.15	.48	.70	.68	.17	.67	.83	.67	.39	.69
NEEDS	λ_{x205}	(9.39)	(8.03)	(1.96)	(6.51)	(10.10)	(9.72)	(2.11)	(9.51)	(15.82)	(11.63)	(6.02)	(12.07)
		.79	.73	.20	.70	.81	.72	.22	.72	.82	.69	.34	.68
HEART	λ_{x215}	(12.57)	(11.03)	(2.63)	(10.47)	(12.42)	(10.50)	(2.84)	(10.62)	(15.34)	(11.97)	(5.17)	(11.92)
		.70	.59	.11	.55	.61	.66	.24	.64	.51	.44	.28	.41
HOURS	λ_{x225}	(10.67)	(8.50)	(1.43)	(7.78)	(8.54)	(9.48)	(3.13)	(8.94)	(8.40)	(7.10)	(4.29)	(6.54)

Table 5
χ^2 Difference Tests
One-Dimension Models vs. Five-Dimension Models

		Perceptions	Difference	Ratio	Log of Ratio
Company 1					
	1-dimension model	1094.83	875.48	1621.83	1154.11
	5-dimension model	694.64	487.69	739.90	670.24
	$\chi^2_d(10)$ =	400.19*	387.79*	881.93*	483.87*
Company 2					
	1-dimension model	877.15	728.03	349.68	945.13
	5-dimension model	493.18	364.08	652.71	552.90
	$\chi^2_d(10)$ =	383.97*	363.95*	696.97*	392.23*
Company 3					
	1-dimension model	1237.70	964.72	1437.69	1111.76
	5-dimension model	811.74	518.25	700.62	1631.90
	$\chi^2_d(10)$ =	425.96*	446.47*	737.07*	520.14

* $p < .0001$

both the one and five-factor model are frequently, though not always, associated with the perceptions. For example, Table 3 shows that for Company 2, the difference model has the highest loadings on several of the items in the assurance and empathy dimensions, offering further relative support for the use of this approach. Table 3 also shows that almost all of the reported t-values are significant using the criteria that $t > 2$ is significant (Byrne 1989). These high loadings and significant t-values for the five-dimension model suggest that differences may be most appropriate for fitting the theorized relationship. Furthermore, the marginal difference in stability across the models would not warrant the use of an alternative approach such as the log of the ratios.

Additional support for the dimensionality of the original SERVQUAL scale is also provided, relative to the one-dimension model. For the five-dimension models, all of the individual items load significantly ($t > 2$) on their respective dimensions, albeit some of the t-values are barely above 2.0. For the one-dimension models, a few of the indicators were not significant for the ln(r) models. These same items (APPEAL and FACIL) do load significantly when the model is specified as five dimensions. Thus, the five dimension model is better than the one dimension model. However, the five dimension model is not a particularly strong fit.

Moreover, χ^2 difference tests (presented in Table 5) indicate that the χ^2 is significantly lower for each five-dimension model as compared to the corresponding one-dimension model. Such results indicate that the five-dimension model offers the better fit then the one-dimension model.

The reliability issue can be examined through Table 6 which presents Cronbach's alpha for all individual models, as well as individual dimensions. A general trend is noted which indicates that, overall, reliabilities are the highest for the perceptions-only models. This is an interesting finding given the previous research suggesting that reliabilities of difference scores may be overestimated with the use of Cronbach's alpha (Peter, Churchill, and Brown 1993). Thus, even if the current reliabilities are overestimations, they are still lower than the reliabilities provided by the perceptions-only models. Finally, reliabilities are the lowest for the models using the ratio followed by the ln(r), although the reliabilities for the ln(r) models do not fall below well-accepted measures of reliability (Nunnally 1978).

Discussion

Results of this research provide support for the relative validity of the difference score approach for measuring service quality, in comparison to the two other approaches examined here. Such relative validity is not an endorsement of the traditional SERVQUAL model because our data show that even these models were weak. Although our models could have been improved, as noted earlier, such optimization is not consistent with the intended comparisons. In short, however, the assessment of the models examined here shows that despite the proposed advantages, the ln(r) approach results in models that are not only less stable than both the perceptions and differences measures, but the ln(r) models offer the weakest indicators [with the exception of the ratio models] suggesting that their use is inappropriate. Additionally, the goodness-of-fit indicators also show that difference scores provide slightly better models than the perceptions-only model.

Moreover, the goodness-of-fit statistics and the factor loadings indicate that five dimensions provide better fitting models than one dimension. Thus, the current study offers relative support for the use of the SERVQUAL dimensions as a measure of assessing service quality. That is, as compared to three other alternative measures of service quality, the five-dimension SERVQUAL seems to offer the best fit and the most stability. These results provide some sup-

Table 6
Reliabilities
(Cronbach's Alpha)

	Co. 1 (n = 178)	Co. 2 (n = 164)	Co. 3 (n = 241)
Perceptions			
22 items	.95	.96	.96
Tangibility (4 items)	.91	.89	.84
Reliability (5 items)	.87	.92	.92
Responsiveness (4 items)	.83	.86	.86
Assurance (4 items)	.81	.81	.87
Empathy (5 items)	.88	.89	.88
Differences			
22 items	.92	.95	.93
Tangibility (4 items)	.84	.82	.73
Reliability (5 items)	.85	.92	.91
Responsiveness (4 items)	.75	.83	.83
Assurance (4 items)	.74	.80	.85
Empathy (5 items)	.83	.88	.83
Ratio			
22 items	.81	.88	.81
Tangibility (4 items)	.90	.67	.62
Reliability (5 items)	.84	.88	.90
Responsiveness (4 items)	.46	.59	.74
Assurance (4 items)	.28	.84	.75
Empathy (5 items)	.76	.73	.68
Log of Ratio			
22 items	.90	.94	.93
Tangibility (4 items)	.84	.79	.68
Reliability (5 items)	.82	.91	.90
Responsiveness (4 items)	.70	.81	.81
Assurance (4 items)	.66	.73	.84
Empathy (5 items)	.81	.88	.82

port for the original factor structure posited by PZB in their introduction of SERVQUAL. This support, however, must be taken in the context in which it is provided: relative to the other models tested in this study, and without the release of potentially correlated error. Further, while the perceptions-only measure produced higher reliabilities, the Cronbach's alphas associated with the SERVQUAL dimensions and overall models more than exceed suggested minimum levels (Nunnally 1978).

The use of three companies within one industry may explain the stability of the SERVQUAL model. That is, expectations of different companies within the same industry are likely to be consistent whereas expectations probably vary across industries, and this consistency may contribute to the stability exhibited by the SERVQUAL models in the present study. This implies that the unstable dimensionality of SERVQUAL in previous studies may be due to differences across industries. Such findings lend additional support for the appropriateness of SERVQUAL, with the relevant modifications, for use within a particular industry.

This stability among companies within a single industry shows that SERVQUAL difference scores are most appropriate for comparison within a homogeneous group, suggesting that monitoring service quality changes over time within a single company may be best achieved with the difference scores. However, it would be appropriate to test temporal stability before utilizing the instrument to document change.

Finally, it is important to note that using either differences or ratios is inappropriate as a solo measure when variation among companies exists. However, Table 2 shows that this is not a problem in the current study because the company to company variation is small among these three companies. If the variation were not small, determining if the difference among companies was the result of a difference in perception, expectations, or both would require further analysis. This analysis would require comparison of the components of the difference or ratio among the companies. However, in our data such exploration was unnecessary because of the small observed variation.

Implications, Conclusions, and Suggestions for Future Research

It is important to note that this research effort did not examine other potential factor structures for the three models. Thus, while the five-factor approach was deemed relatively superior to the one-dimension model, it is possible that other factor structures might produce better fitting models. Moreover, the use of five dimensions for the perceptions-only model has received limited attention in the existing literature, and the current research suggests that this model may have some value. Hence, future research should further examine potential factors that may comprise a perceptions only approach.

Thus, despite the considerable discourse on the topic and the current findings, the issue of whether the perceptions-only approach or the difference score approach is superior is still not resolved to the satisfaction of these researchers. To our knowledge, this is the first study with a comparison of several different measures of service quality within one industry, and thus, the findings provide important information for the measurement of service quality. The current study provides empirical support for the relative superiority of difference scores in assessing service quality. Although the log of the ratio may offer the potential statistical advantage associated with reduced variance, while maintaining the comparison between expectations and perceptions, the current research fails to provide support for the use of this approach.

Therefore, a contribution of this research is empirical support for the use of difference scores (i.e. SERVQUAL) as compared to three alternative measures of service quality. Given the importance of service quality in industry, the academic arguments surrounding the measurement of the construct, and the prevalence of perceptions-only measures in practical application, our findings provide critical implications and insights for both academic researchers and practitioners.

First, in spite of the support for difference scores among companies, there is considerable work necessary to understand the most appropriate measure of service quality. Our findings, therefore, should be used as an initial step for additional investigations directly comparing SERVQUAL and SERVPERF. The five-dimension perceptions-only model, arising here as a potentially stable and well-fitting model, deserves extensive evaluation to determine its usefulness. Second, other industries should be utilized to reexamine the ln(r) approach before discarding its potential usefulness. Given its attractive statistical properties, the log deserves further investigation.

From a practitioner standpoint, our findings may be even more important because the use of perceptions-only measures seems to be the standard in industry application. Given that the current findings indicate a comparatively superior fit and stability for the difference models, these companies are failing to capture true impressions of service quality.

Additionally, our findings show that the five-dimension model is relatively superior to the unidimensional model. Consequently, companies that do utilize SERVQUAL should conceptualize their service quality model using the five dimensions. If properly modeled to the specific company, these dimensions can provide detailed information on the particular areas within a company that are performing well and identify those areas that need quality improvement.

Furthermore, when using the five-dimension model, our work shows the relative stability of the difference approach. Thus, future research should examine alternate measures in other industries to see if this relative support for the five-dimension gap model is consistent across other applications. If this stability is evident within other industries, then practitioners should certainly consider reevaluating their current tools. Finally, a longitudinal study examining one company over time with the different approaches would also be useful to assess the stability and fit of the different models.

References

1. Anderson, J.C., & Gerbing, D.W, "Some methods for respecifying measurement models to obtain unidimensional construct measurement," *Journal of Marketing Research*, Vol. 19, November, pp. 453-460, 1982.
2. Babakus, E., & Boller, G.W., "An empirical assessment of the SERVQUAL scale," *Journal of Business Research*, Vol. 24, No. 3, pp. 253-268, 1992.
3. Bolton, R.N., & Drew, J.H., "A longitudinal analysis of the impact of service changes on customer attitudes," *Journal of Marketing*, Vol. 55, January, pp. 1-9, 1991a.
4. Bolton, R.N., & Drew, J.H., "A multistage model of customers' assessments of service quality and value," *Journal of Consumer Research*, Vol. 17, March, pp. 375-384, 1991b.
5. Brensinger, R.P., & Lambert, D.M., "Can the SERVQUAL scale be generalized to business-to-business services?" In *Knowledge Development in Marketing*, AMA's Summer Educators Conference Proceedings. Boston, MA, 289. 1990.
6. Brown, S.W., & Swartz, T.A., "A gap analysis of professional service quality," *Journal of Marketing*, Vol. 53, April, pp. 92-98, 1989.
7. Bynner, J., "Factor analysis and the construct indicator relationship," *Human Relations*, Vol. 41, No. 5, pp. 389-405, 1988.
8. Byrne, B.M. *A Primer of LISREL*. Springer-Verlag, New York, 1989.
9. Carman, J.M., "Consumer perceptions of service quality: An assessment of the SERVQUAL dimensions," *Journal of Retailing*, Vol. 66, No. 1, pp. 33-55,1990
10. Churchill, G.A., & Surprenant, C, "An investigation into the determinants of customer satisfaction," *Journal of Marketing Research*, Vol. 19, November, pp. 491-504, 1982.
11. *CPCU Society*, Vol. 14, No. 2. 1996.
12. Cronbach, L.J., "Coefficient alpha and the internal structure of tests," *Psychometrika*, Vol. 16, pp. 297-334, 1951.
13. Cronbach, L.J., & Furby, L., "How we should measure 'Change'--Or should we?" *Psychological Bulletin,* Vol. 74, July, pp. 74-73, 1970.
14. Cronbach, L.J., & Meehl, P., "Construct validity in psychological tests," *Psychological Bulletin* Vol. 21, pp. 281-302, 1955.
15. Cronin, J. J., & Taylor, S.A., "Measuring service quality: A reexamination and extension," *Journal of Marketing*, Vol. 56, July, pp. 55-68, 1992.
16. Cronin, J. J., & Taylor, S.A., "SERVPERF Versus SERVQUAL: Reconciling Performance-Based and Perceptions-Minus-Expectations Measurement of Service Quality," *Journal of Marketing*, Vol. 58, January, pp. 125-131, 1994.
17. Cummins, J.D., McGill, D.M. Winklevoss, H.E., & Zelten, R.A., 1974, *Consumer attitudes toward auto and homeowners insurance*. Wharton School, Philadelphia, PA., 1974.
18. Dabholkar, P.A., Thorpe, D.I., & Rentz, J.O. (1996). "A measure of service quality for retail stores: Scale development and validation," *Journal of the Academy of Marketing Science*, Vol. 24, Winter, pp. 3-16, 1996.
19. Doll, W.J., & Torkzadeh, G., "End-user computing involvement: Discrepancy model," *Management Science*, Vol. 35, No. 10, pp. 1151-1171, 1989.
20. Edwards, J.R.,. "Alternatives to difference scores as dependent variables in the study of congruence in organizational research," *Organizational Behavior and Human Decision Processes*, Vol. 64, No. 3, pp. 307-324, 1995.
21. Finn, D.W., & Lamb, C.W., "An evaluation of the SERVQUAL scales in a retailing setting." *Advances in Consumer Research,* Vol. 18, pp. 338-357, 1991.

22. Galletta, D. F., & Lederer. A. L., "Some cautions on the measurement of user information satisfaction," *Decision Sciences,* Vol. 20, Summer, pp. 419-439, 1989.
23. Gerbing, D.W., & Anderson, J.C., "An updated paradigm for scale development incorporating unidimensionality and its assessment," *Journal of Marketing Research,* Vol. 15, May, pp. 186-192, 1988.
24. Johns, G., "Difference score measures of organizational behavior variables: A critique," *Organizational Behavior and Human Performance,* Vol. 27, pp. 443-463, 1981.
25. Lord, F.M., "The utilization of unreliable difference scores," *Journal of Educational Psychology,* Vol. 49, No. 3, pp. 150-152, 1958.
26. Neter, J., & Wasserman, W., *Applied linear statistical models*. Richard D. Irwin, Chicago, 1974.
27. Nunnally, J.C., *Psychometric theory*. 2nd ed. McGraw-Hill, New York, 1978.
28. Oliver, R.L., "A cognitive model of the antecedents and consequences of satisfaction decisions," *Journal of Marketing Research,* Vol. 17, No. 3, pp. 460-469, 1980.
29. Parasuraman, A., Zeithaml, V.A., & Berry, L.L., "A conceptual model of service quality and its implications for future research," *Journal of Marketing,* Vol. 49, No. 4, pp. 41-50, 1985.
30. Parasuraman, A., Zeithaml, V.A., & Berry. L.L., "SERVQUAL: A multiple item scale for measuring consumer perceptions of service quality," *Journal of Retailing,* Vol. 64, No. 1, pp. 12-40, 1988.
31. Parasuraman, A., Zeithaml, V. A., & Berry. L. L., "Refinement and reassessment of the SERVQUAL scale," *Journal of Retailing,* Vol. 67, No. 4, pp. 420-450, 1991.
32. Parasuraman, A., Zeithaml, V. A., & Berry, L. L., "Reassessment of expectations as a comparison in measuring service quality: Implications for further research," *Journal of Marketing,* Vol. 58, No. 1, pp. 111-124, 1994a.
33. Parasuraman, A., Zeithaml, V.A., & Berry, L.L., "Alternative scales for measuring service quality: A comparative assessment based on psychometric and diagnostic criteria," *Journal of Retailing,* Vol. 70, No. 3, pp. 201-230, 1994b.
34. Parasuraman, A., Berry, L.L., & Zeithaml, V.A., "More on improving service quality measurement," *Journal of Retailing,* Vol. 69, No. 1, pp. 140-147, 1993.
35. Peter, J.P., Churchill, G.A., & Brown, T.J., "Caution in the use of difference scores in consumer research," *Journal of Consumer Research,* Vol. 19, No. 1, pp. 655-662, 1993.
36. Pitt, L.F., Watson, R.T., & Kavan, C.B., "Service quality: A measure of information systems effectiveness," *MIS Quarterly,* Vol. 19, June, pp. 173-187, 1995.
37. Prakash, V., & Lounsbury, J.W., "A reliability problem in the measurement of disconfirmation of expectations," In *Advances in consumer research,* Vol. 10. Eds. Bagozzi, R.P., & Tybout, A.M., pp. 244-249. Association for Consumer Research, Ann Arbor, MI, 1983.
38. Teas, R. K., "Expectations, performance evaluation and consumer's perception of quality," *Journal of Marketing,* Vol. 57, No. 4, pp.18-34, 1993.
39. Teas, R. K., "Expectations as a comparison standard in measuring service quality: An assessment of a reassessment," *Journal of Marketing,* Vol. 58, No. 1, pp. 132-139, 1994.
40. Wall, T.D., & Payne, R., "Are deficiency scores deficient?" *Journal of Applied Psychology,* Vol. 58, No. 3, pp. 322-326, 1973.
41. Wilkie, W. *Consumer Behavior,* 3rd ed., John Wiley & Sons, Inc., New York, 1994.
42. Woodruff, R. B., Cadotte, E. R., & Jen-

kins, R. L., "Modeling consumer satisfaction processes using experience-based norms," *Journal of Marketing Research,* Vol. 20, August, pp. 296-304, 1983.

43. Zeithaml, V., Berry, L. L., & Parasuraman, A., "The nature and determinants of customer expectations of service quality," *Journal of the Academy of Marketing Science,* Vol. 21, No. 1, pp. 1-12, 1993.