

A Comprehensive Framework for Service Quality: An Investigation of Critical Conceptual and Measurement Issues Through a Longitudinal Study

PRATIBHA A. DABHOLKAR
University of Tennessee, Knoxville

C. DAVID SHEPHERD
Kennesaw State University

DAYLE I. THORPE
University of Tennessee, Knoxville

This study finds that factors relevant to service quality are better conceived as its antecedents rather than its components and that customer satisfaction strongly mediates the effect of service quality on behavioral intentions. The article discusses the application of this chronological framework in understanding and predicting service quality and its consequences. The study also finds that perceptions and measured disconfirmation offer several advantages over computed disconfirmation (i.e., difference scores), and that a cross-sectional measurement design for service quality is preferred to a longitudinal design. The article discusses the implications of these findings for practitioners and for future research on service quality.

Marketers realize that to retain customers, and to survive and grow, they must provide a high quality of service. Consequently, academic and managerial interest in service quality has been evident in the services marketing literature for the past several years. Critical conceptual and measurement issues related to service quality have been raised and researchers have begun to address them. However, past studies have found varying results

Pratibha A. Dabholkar is Associate Professor of Marketing, University of Tennessee, Knoxville, Department of Marketing, Logistics, and Transportation, 307 Stokely Management Center, Knoxville, Tennessee 37996 (e-mail: pratibha@utk.edu). C. David Shepherd is Associate Professor of Marketing, Kennesaw State University, Department of Marketing and Professional Sales, 1000 Chastain Road, Kennesaw, GA 30144. Dayle I. Thorpe is Director, Executive MBA Program, University of Tennessee, Knoxville, Department of Marketing, Logistics, and Transportation, 310 Stokely Management Center, Knoxville, Tennessee 37996.

Journal of Retailing, Volume 76(2) pp. 139–173, ISSN: 0022-4359
Copyright © 2000 by New York University. All rights of reproduction in any form reserved.

with respect to some issues and other issues have not yet been addressed. This paper proposes and tests a comprehensive framework of service quality in an attempt to address these critical conceptual and measurement issues.

CONCEPTUAL ISSUES IN SERVICE QUALITY

Despite much research on service quality, the bulk of the literature continues to conceptualize factors associated with service quality (e.g., reliability, responsiveness) as dimensions or components of the construct rather than as antecedents to the consumer's overall evaluation of service quality. Often in building a literature stream, constructs are first defined in terms of components. Later, as the literature develops, some of these components are viewed as antecedents to offer greater understanding of the phenomenon under study. For example, satisfaction was originally defined as disconfirmation (Miller, 1976), but later, disconfirmation was viewed as an antecedent to satisfaction (Oliver, 1981). Next, satisfaction was equated with emotion (Westbrook, 1983), and later as the construct was "fine tuned," emotion was viewed as an antecedent to satisfaction (Westbrook and Oliver, 1991). Similarly, viewing related dimensions as antecedents to service quality is a natural step in the progression of the service quality construct. As such, it should increase our understanding of service quality evaluations and of the role of antecedents in forming these evaluations.

Further, direct measures of overall service quality serve as better predictors of behavioral intentions than a value of service quality computed from measured dimensions. In recent years, researchers have begun to study the consequences of service quality (e.g., Boulding et al., 1993; Cronin and Taylor, 1992; Zeithaml, Berry, and Parasuraman, 1996) and have found support for a positive relationship in general. However, no study has looked at the relative predictive power of the components model of service quality versus that of the antecedents model of service quality.

Another conceptual issue relates to the role of customer satisfaction within the framework of service quality predicting behavioral intentions. Empirical studies have, for the most part, not addressed the differential effects of service quality and customer satisfaction on behavioral consequences. Exceptions are studies by Taylor and Baker (1994), Gotlieb, Grewal, and Brown (1994), Dabholkar (1995a), and Bansal and Taylor (1997) that have examined the differential effects of the two constructs. However, the findings are somewhat different across these studies and more research is needed to investigate the possible mediating role of customer satisfaction in the relationship between service quality and behavioral intentions.

The conceptual issues discussed above can be summarized as follows. First, are relevant factors related to service quality better conceived as its components or its antecedents? Which format increases our understanding of service quality and which offers greater predictive power? Second, what role does customer satisfaction play in the framework of service quality predicting behavioral consequences? Does it add greater predictive power over that of service quality? If so, does it act as an independent determinant or does it mediate the effect of service quality on behavior?

MEASUREMENT ISSUES IN SERVICE QUALITY

Over the last several years, there have been a variety of discussions in the literature on different issues related to service quality measurement. A major debate has focused on whether service quality should be measured as perceptions or as disconfirmation (Cronin and Taylor, 1992, 1994; Parasuraman, Zeithaml, and Berry, 1994; Teas, 1993, 1994). Those who favor the former approach (e.g., Cronin and Taylor, 1992) suggest that perceptions of service quality more closely match customer evaluations of the service provided. Parasuraman, Zeithaml, and Berry (1994) counter that measuring service quality as disconfirmation (i.e., the difference between perceptions and expectations) is valid and further, it allows service providers to identify gaps in the service provided. A rigorous comparison of these alternate measures should offer further insights on this issue.

A second major issue related to service quality measurement is also tied to disconfirmation. When expectations and perceptions are measured separately (whether in a longitudinal or cross-sectional design), the computed difference scores for disconfirmation have problems of reliability, discriminant validity, and variance restriction (Brown, Churchill, and Peter, 1993; Peter, Churchill, and Brown, 1993). These researchers suggest that instead of computing difference scores through separate measures of expectations and perceptions, the difference should be measured directly to overcome the measurement problems. Parasuraman, Berry, and Zeithaml (1993) argue that the difference score format is reliable and valid and should not be abandoned. However, Parasuraman, Berry, and Zeithaml (1991) do call for further comparisons of the two-part measures versus the direct disconfirmation measures. A comparison of these alternative measures, in a research design that rigorously captures computed disconfirmation, may help resolve the issue.

The third and related measurement issue is the validity of a cross-sectional versus a longitudinal research design for measuring service quality. (A longitudinal design is defined as one where expectations are measured before the service is delivered and perceptions are captured after the service.) For researchers using the disconfirmation approach, it is not clear whether both expectations and perceptions should be measured in a cross-sectional study or whether a longitudinal design is more appropriate. Interestingly, both groups—disconfirmation and perceptions proponents—have used cross-sectional research designs to measure expectations and perceptions after the service has been provided. An exception is a study by Boulding et al. (1993) that does measure expectations before the service; however, it is not a true longitudinal study in that it uses a computer simulation with a hypothetical situation to create a longitudinal effect. A test of a longitudinal versus a cross-sectional design is needed in a field study to determine whether the original expectations of customers prior to service delivery need to be captured.

The measurement issues discussed above can be summarized as follows. First, are perception measures superior to disconfirmation measures? Second, is measured disconfirmation superior to computed disconfirmation? Third, is a cross-sectional design adequate or does a longitudinal design offer significant advantage?

This paper addresses both sets of issues—conceptual and methodological—summarized above. A comprehensive model of service quality is developed and includes an

examination of its antecedents, consequences, and mediators to provide a deeper understanding of conceptual issues related to service quality. Further, this is the first study to employ a true, longitudinal setting, such as envisioned in early conceptions of service quality, and it allows a more rigorous comparison of critical measurement issues involving service quality.

DEVELOPING A COMPREHENSIVE FRAMEWORK FOR SERVICE QUALITY

Factors as Components versus Antecedents

Traditional service quality models (Grönroos, 1978; Parasuraman, Zeithaml, and Berry, 1988) have treated relevant factors related to service quality (e.g., reliability, responsiveness) as components of service quality. The majority of the studies that followed have done the same. Service quality is not viewed as a separate construct, but rather the components are summed to obtain an estimate of service quality. Parasuraman, Zeithaml, and Berry (1988) write that "an overall measure of quality (can be obtained) in the form of an average score across all five dimensions" (p. 31). This view of service quality fails to capture the effect of the relevant factors (referred to as *dimensions*¹ in the literature) as antecedents of service quality and also fails to capture customers' overall evaluations of service quality as a separate, multi-item construct.²

Interestingly, several researchers have measured overall service quality directly. However, most (Babakus and Boller, 1992; Bolton and Drew, 1991a; Boulding et al., 1993; Cronin and Taylor, 1992; Parasuraman, Zeithaml, and Berry, 1988; Zeithaml, Berry, and Parasuraman, 1996) have used a single-item measure that makes it impossible to ascertain the reliability of this construct. Some of these studies (e.g., Zeithaml, Berry, and Parasuraman, 1996) have viewed overall service quality as merely an alternative way to measure service quality and have not conceptualized the factors as antecedents of service quality. Others studies (e.g., Babakus and Boller, 1992; Bolton and Drew, 1991a; Cronin and Taylor, 1992) have viewed the factors as antecedents but have tested this relationship with a single-item dependent variable. Only a few studies have used multi-item measures of overall service quality (e.g., Dabholkar, Thorpe, and Rentz, 1996; Spreng and Mackoy, 1996; Spreng and Singh, 1993; Taylor and Baker, 1994), but they have either not related this measure to the factors, or only tested the relationship with a correlation measure. No study to date has conducted a rigorous examination of the factors as antecedents and none has explored which model (factors as components versus factors as antecedents) is better for understanding service quality.

⑦ A related issue is the prediction of behavioral intentions. Past research has examined the influence of service quality on behavioral intentions (e.g., Boulding et al., 1993; Cronin and Taylor, 1992; Zeithaml, Berry, and Parasuraman, 1996), but no study has looked at the relative predictive power of the factors as components of service quality versus as antecedents of service quality. Being able to predict service quality evaluations and behavioral intentions is critical to managers. The antecedents framework would offer

additional insights into how customers view service quality and how this view predicts their behavior. As discussed earlier, a component-to-antecedent transition would be a natural progression in the development of constructs. In addition, this framework fits better with how people actually make evaluations of service delivery. Hence, it is expected that the antecedent model of service quality (where service quality mediates the influence of these factors on behavioral intentions) will be superior to the components model of service quality (where the factors represent service quality and therefore have a direct influence on behavioral intentions).

Proposition

Incomplete proposition!

- 1: Factors relevant to service quality act as antecedents to overall evaluations of service quality rather than acting as its components (i.e., the antecedents model of service quality is superior to the components model).

Role of Customer Satisfaction within the Service Quality Framework

Several researchers have raised the issue of whether service quality and customer satisfaction are the same or different constructs (Dabholkar, 1993, 1995b; Iacobucci, Grayson and Ostrom, 1994; Oliver, 1993). In fact, researchers have not always been able to separate the two constructs empirically. Spreng and Singh (1993) studied evaluations of service by banking customers, but failed to find discriminant validity between service quality and customer satisfaction. In a study of retail customers, Dabholkar (1995a) found the two constructs to be distinct for recent customers, but to overlap in meaning for long-term customers, as customer satisfaction evaluations grew increasingly cognitive over time. Bansal and Taylor (1997) found a very high correlation (0.96) between the two constructs for another study of banking customers, but reported that a χ^2 difference test found discriminant validity. Others have been able to separate service quality and customer satisfaction more easily, but possibly because one construct was defined at a transactional level and the other at a global level.

Assuming that the two concepts are distinct, at least under certain conditions (e.g., Dabholkar, 1995b; Iacobucci, Ostrom, and Grayson, 1995), a logical question relates to the order of their occurrence in the consumer's mind (and, hence, the causal link between them). Traditionally, researchers had suggested that customer satisfaction with a given service experience would lead to an overall evaluation/attitude about service quality over time (Bitner, 1990; Oliver, 1981; Parasuraman, Zeithmal, and Berry, 1988). More recently, the opposite view appears to have strong favor. Oliver (1993) first suggested that service quality would be antecedent to customer satisfaction regardless of whether these constructs were measured for a given experience or over time. Several researchers (e.g., Anderson and Sullivan, 1993; Spreng and Mackoy, 1996) have found empirical support for this model, wherein customer satisfaction is a consequence of service quality. The issue then becomes two-fold: does customer satisfaction have an incremental effect over

that of service quality on behavioral intentions, and is this effect an independent or a mediating one?

A number of researchers have addressed this issue, but findings have been mixed. In a study on hospital service quality and satisfaction after patient discharge, Gotlieb, Grewal, and Brown (1994) found that customer satisfaction mediates the effect of service quality on behavioral intentions. On the other hand, Bansal and Taylor (1997) failed to find such an effect in a study of banking customers. Whereas service quality did impact behavioral intentions in their study, customer satisfaction had no impact on intentions in the presence of service quality evaluation. In a study of retail customers, Dabholkar (1995a) found that customer satisfaction was a better predictor of behavioral intentions in the short term and had a differential effect over that of service quality. For long-term customers, however, there was no differential effect and either construct, service quality or customer satisfaction, was sufficient to predict behavioral intentions. Taylor and Baker (1994) found customer satisfaction to be a moderating variable in the relationship between service quality and repurchase intentions. Based on a study of four service industries (health care, recreation, airlines, and long-distance telephone service), the researchers concluded that the interaction of customer satisfaction with service quality increases the effect of repurchase intentions; no conclusions about causal sequence were drawn.

Clearly, more research is needed on this issue. At the same time, the mediating role of customer satisfaction (on the effect of service quality on behavioral intentions) is more natural to the way people make evaluations. Evaluating the factors to judge service quality, deciding if one is satisfied, and then making a decision about patronizing (and recommending) the service in the future is a logical sequence for a large number of services where emotion is not too high and performance is not often outside the zone of indifference (Dabholkar, 1995b). Hence it is expected that customer satisfaction will have a mediating role on behavioral intentions rather than an effect independent of service quality. (To conduct a thorough comparison of all possible alternative models (Kline, 1998), a second alternative model of service quality as a mediator is tested against the proposed model.)

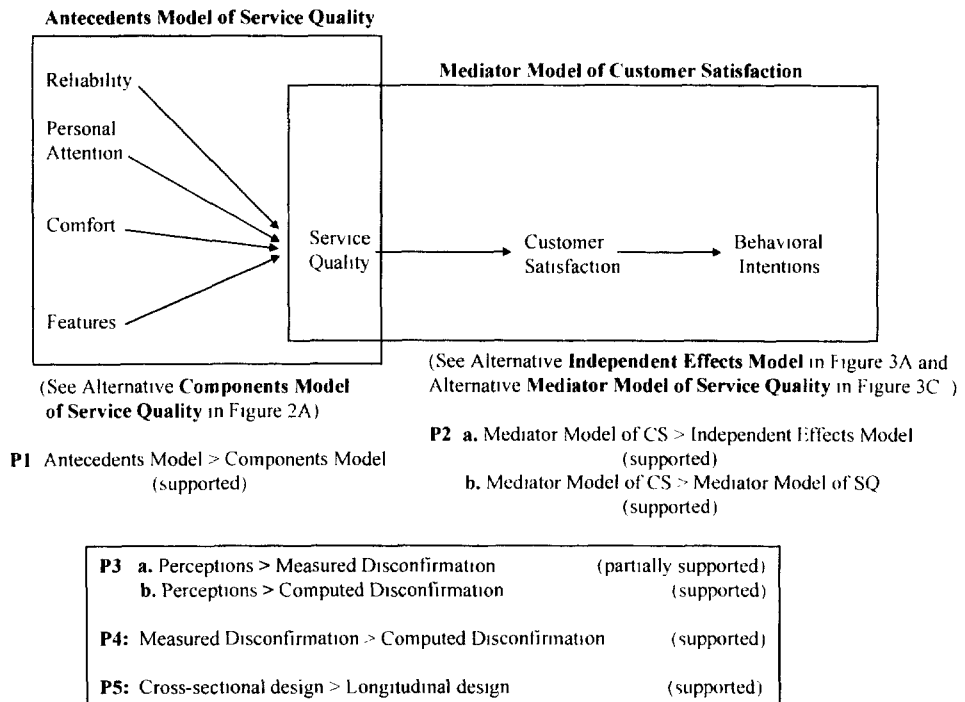
Proposition

- 2a: *Customer satisfaction (CS) mediates the influence of service quality on behavioral intentions rather than both variables having an independent effect on behavioral intentions (i.e., the mediator model of CS is superior to the independent effects model).*

Proposition

- 2b: *Customer satisfaction (CS) mediates the influence of service quality on behavioral intentions rather service quality (SQ) mediating the influence of customer satisfaction on behavioral intentions (i.e., the mediator model of CS is superior to the mediator model of SQ).*

(A necessary corollary to these propositions is that service quality and customer satisfaction are distinct constructs.)

**FIGURE 1**

A Comprehensive Framework for the Antecedents and Consequences of Service Quality with Customer Satisfaction as a Mediator

The comprehensive framework proposed for understanding service quality, its antecedents, and its consequences is presented in Figure 1.F1 The proposed conceptual models (P1 and P2) have links to Figs. 2 and 3 to allow comparison with alternative models.

Disconfirmation versus Perceptions

Traditionally, service quality has been conceptualized as a disconfirmation process (Grönroos, 1984; Lewis and Booms, 1983; Parasuraman, Zeithaml, and Berry, 1988). Various studies (e.g., Babakus and Boller, 1992; Babakus and Mangold, 1992; Carman, 1990; Finn and Lamb, 1991; Spreng and Singh, 1993) have measured service quality using the disconfirmation model, that is, measuring both expectations and perceptions and equating service quality evaluation to the difference scores derived from the two measures. Most of these studies, however, have found a poor fit for the

disconfirmation model. Researchers have raised several issues about problems with the disconfirmation model. Teas (1993) writes that the disconfirmation model has conceptual, theoretical, and measurement problems and suggests that alternative perceived quality models be used. Spreng and Olshavsky (1992) believe that the disconfirmation paradigm suffers due to problems with measuring expectations. In addition, all of these studies have been cross-sectional and it is yet to be determined whether the traditional disconfirmation model (based on difference scores) would be better supported in a longitudinal study.

It may be noted that the two exceptions to the cross-sectional approach have found inconsistent results. In a simulated longitudinal study, Boulding et al. (1993) found that expectations influence perceptions, which in turn influence overall service quality. Bolton and Drew (1991a) found both perceptions and disconfirmation to have a direct effect on overall service quality. This study was longitudinal in that the same measures were taken repeatedly over time. Expectations were not measured separately and both perceptions and disconfirmation were measured at the same time.

Increasingly, researchers (Andaleeb and Basu, 1994; Mittal and Lassar, 1996) are simply measuring perceptions as indicators of service quality (ignoring expectations completely) and are finding good predictive power in their studies. Relative to the data collection process, measuring only perceptions is attractive because it requires half the number of items that the traditional disconfirmation approach requires. Some researchers (Babakus and Boller, 1992; Cronin and Taylor, 1992) have compared *computed* difference scores with perceptions to conclude that perceptions are a better predictor of service quality than disconfirmation. However, this comparison of perceptions and disconfirmation has been conducted only in cross-sectional studies to date, and only with single-item measures of overall service quality. Moreover, it has ignored measured disconfirmation, or direct measures of disconfirmation.

What is needed is a longitudinal field study that measures expectations *before* the service and both perceptions and measured disconfirmation *after* the service. In addition, the dependent variables, whether they be overall service quality or behavioral intentions, need to be multi-item constructs whose reliability can be ascertained. Such a design will allow a more rigorous comparison of perceptions to *computed* disconfirmation (i.e., difference scores) over the longitudinal setting, and also allow a comparison of perceptions to *measured* disconfirmation (i.e., direct scores) after the service. If perceptions can be shown to be more reliable measures than either type of disconfirmation using a more rigorous research design, this would offer additional support to the increasing notion that perceptions are the preferred measures for service quality.

Proposition

3a: Perception measures are superior to measured disconfirmation.

Proposition

3b: Perception measures are superior to computed disconfirmation.

Difference Scores versus Measured Disconfirmation

Another debate in the service quality literature centers on the use of (computed) difference scores versus the measured disconfirmation approach (e.g., Brown, Churchill, and Peter, 1993; Parasuraman, Berry, and Zeithaml, 1993). This discussion addresses the issues related to using a difference score (a mathematical calculation of perceptions-minus-expectations) versus a measured disconfirmation score (a direct mental estimation of perceptions compared to expectations) to measure service quality.

Peter, Churchill, and Brown (1993) have cautioned against using difference scores in consumer research. They report that difference scores (i.e., the subtraction of one measure from another to create a measure of a distinct construct) possess recognized psychometric problems. Namely, difference scores typically are less reliable than other measures, they may appear to demonstrate discriminant validity when this conclusion is not warranted, they may be only spuriously correlated to other measures (because they do not discriminate from at least one of their components), and they may exhibit variance restriction. On the other hand, Parasuraman, Berry, and Zeithaml (1993) have reported evidence that counters these claims. Based on the empirical results of their research, these authors indicate that there is no serious threat to reliability, the problem of inflated discriminant validity is unlikely, and variance restriction is only a problem when difference scores are used as the dependent variable in multivariate analysis.

Because of the potential problems with difference scores, many researchers strongly recommend the use of direct comparison measures of service quality (Babakus and Boller, 1992; Bolton and Drew, 1991a; Brown, Churchill, and Peter, 1993; Carman, 1990). As Peter, Churchill, and Brown (1993) explain, the direct comparison measurement approach requires subjects to mentally consider differences rather than have the researcher calculate an arithmetic difference for them.

If both types of disconfirmation are similar in their predictive ability, explicit expectations data are not needed. Also, if measured disconfirmation has the same diagnostic power as computed disconfirmation, then psychometric problems associated with the latter can be avoided. It is clear that there is greater support for measured disconfirmation, but this support needs to be tested rigorously. Interestingly, all of the comparison studies of these two types of measures have been conducted in cross-sectional settings. What is needed is a longitudinal research design, where computed disconfirmation is measured as originally conceived in the literature; that is, with expectations measured *before* the service is delivered and perceptions measured *after* it is delivered.

Proposition

- 4:** Measured disconfirmation is superior to computed disconfirmation

Cross-Sectional versus Longitudinal Design

The majority of empirical studies conducted to measure service quality have been cross-sectional (e.g., Babakus and Boller, 1992; Carman, 1990; Finn and Lamb, 1991;

Parasuraman, Zeithaml, and Berry, 1988) where both expectations and perceptions were measured after the service had been delivered. This approach assumes that expectations before the service are identical to expectations after the service and does not account for the fact that expectations may change over time or after the delivery of the service. A study by Bolton and Drew (1991a,b) did use a longitudinal design but measured perceptions and disconfirmation together at three different times and did not measure expectations separately. Another study (Boulding et al., 1993) measured expectations before and perceptions after the service and found that consumers do alter their expectations and perceptions over time. However, as mentioned earlier, this study was based on a scenario-based laboratory experiment rather than a true longitudinal design, suggesting the need for further research.

Research has not clearly determined whether customers use their original expectations to evaluate what they receive or whether they use their modified expectations (changed as a result of the service delivery or because of the passage of time) as their standard for comparison. A longitudinal design in a field setting can help resolve this issue by comparing both ways of measuring disconfirmation. By measuring expectations *before* service delivery and both perceptions and measured disconfirmation *after* service delivery in a real service setting, researchers can make various comparisons among measures to evaluate whether longitudinal studies are superior to cross-sectional studies in the measurement of service quality. In addition, because longitudinal studies are cumbersome, costly, and, by definition, time consuming, as researchers we need to establish whether the effort and cost involved in conducting a longitudinal study is justified by an improved understanding of service quality and greater predictive power. Based on our propositions P3b and P4, it is expected that both perceptions and measured disconfirmation will be superior to computed disconfirmation. In this case, a cross-sectional study would be preferred to a longitudinal one.

Proposition

- 5: *A cross-sectional research design is superior to a longitudinal design for measuring service quality.*

The three propositions related to measurement (P3, P4, and P5) are depicted in Figure 1 (presented earlier). Empirical support for all five propositions is also indicated in this figure.

METHODOLOGY

To examine the measurement issues, a longitudinal study was conducted in which expectations of relevant factors were measured *before* the service. *After* the service was provided, half the respondents were asked about their *perceptions* of the service and the other half were asked about their *perceptions compared to their expectations, or measured disconfirmation*. Expectations were subtracted from perceptions to calculate *computed disconfirmation*. Thus, this research design allowed: 1) a comparison of perceptions versus

(both types of) disconfirmation, 2) a comparison of *computed* disconfirmation (over a longitudinal setting) versus *measured* disconfirmation, and 3) a comparison of cross-sectional versus longitudinal design. In addition, measurements of service quality, customer satisfaction, and behavioral intentions were made *after* the service, which (together with the perception and disconfirmation measures) allowed the testing of the conceptual issues: 1) is the antecedents model of service quality superior to the components model, and 2) does customer satisfaction mediate the effect of service quality on behavioral intentions?

The Context

The data used in this study were collected from institutional customers of the pictorial directory division of a national photographic company. The institutions participating in the study were churches.³ Churches produce pictorial membership directories approximately every three years, to facilitate interaction among church members and to provide a visual history of the church.⁴

At the national level, the pictorial directory industry consists of five major competitors. Although some minor variations exist, these five companies tend to produce a directory that is quite similar in quality, features, and price range. The price to the churches is generally low and the vendors hope to profit from subsequent sales of portraits to church members. Given the commonality of the pictorial directory product and the lack of variation in price, *the choice between directory vendors is largely determined by the quality of the service delivered by the directory vendors.*

The directory vendor's salespeople play a critical role in the delivery of the service to the church members. After the contract is signed, the salesperson/service provider typically remains the main contact with the church. S/he will assist the minister, or a church directory committee, in the selection of pictorial directory features, such as the cover and layout. Often the service provider will assist the church in organizing the membership for their photographic sessions. Also, the service provider is responsible for responding to requests from the church, answering questions, resolving problems, and generally ensuring that the service proceeds as planned throughout the entire process.

The Sample

The churches participating in this study had recently contracted with a national photographic company to produce a pictorial membership directory. Located throughout the United States, the churches contacted were selected in a systematic random fashion from a list provided by the participating directory company. The study participants, usually ministers, were directors of their church's directory program. Of the 500 churches contacted in this study, 432 agreed to participate. However, due to the longitudinal nature of this study, the sample was reduced to 397 usable pairs of questionnaires. This nevertheless yielded a response rate of 79.4%.

Qualitative Research

A series of focus groups was conducted to form the basis for later instrument development. The first focus group consisted of ten church ministers with previous directory experience. A modified critical incident technique was used to draw out concrete examples of "critical" service situations, whether exceptionally good or exceptionally bad. Next, the participants were shown the SERVQUAL scale (Parasuraman, Zeithaml, and Berry, 1988) and asked to assess the instrument's ability to describe factors relevant for service quality in the church directory situation. The group used the critical incident examples to note any factors missing from the SERVQUAL scale. Similarly, any items from the existing scale that seemed irrelevant to the discussion were noted. A second focus group was conducted with a group of six directory salespeople and a similar procedure was followed to determine factors relevant for service quality in this context and whether the existing scale captured these factors. A third focus group was conducted with four church directory sales managers, who integrated the discussions from the previous two focus groups.

The results of this three-step qualitative research verified that modifications were needed at both the individual item and the factor levels of SERVQUAL. Four factors related to service quality emerged from the discussion: Reliability, Personal Attention, Comfort, and Features. Specifically, the participants believed that service providers needed to be reliable (i.e., consistent) and give them personal attention. They also indicated that pressure-selling tactics were common in the industry. Hence, being comfortable with the service provider was considered critical. Items capturing lack of sales pressure and feeling safe were added to fully capture this factor. Finally, the participants suggested that the features provided in the directory were important for service evaluation. Because customers do not visit the directory producer's offices or physical facilities, items addressing these issues in SERVQUAL were not applicable. Hence, items relating to directory features were added, but those relating to physical facilities were dropped.

Scale Development for Exogenous Variables

Specific items were developed for each factor (Reliability, Personal Attention, Comfort, and Features) either by modifying SERVQUAL or from our qualitative analysis (see Appendix). Reliability and Comfort were each measured using three items from SERVQUAL and two developed from focus group comments (see Appendix). Personal Attention was created using three items from Responsiveness, one from Assurance, one from Reliability, and two items from focus group comments (see Appendix). (The reason for this selection was to match the items closely to comments from our focus groups that suggested that courteousness, helpfulness, and sincerity were all part of personal attention.) Specific topics mentioned by the focus group respondents, such as portrait setting, quality of electronic video image proofing, quality of photographs, and quality of directory were captured in the items developed to measure Features (see Appendix).

In phrasing the items, expectations (measured before the service) were prefaced with "Based on your experience with church directory services in general (or Based on what

"Company X" has told you about their service) do you expect that _____.⁵ For perceptions (measured after the service), respondents were asked, "To what extent did "Company X" do the following _____. To capture measured disconfirmation (after the service), respondents were asked, "Compared to what you expected, to what extent did "Company X" do the following _____. All 21 items from the Appendix were used to fill in the blanks for each of these measures. (Computed disconfirmation was calculated by subtracting expectations measured prior to the service from perceptions measured after the service.)

Measures of Endogenous Variables

Although service quality and customer satisfaction are treated as exogenous variables in testing P2, they are viewed as endogenous variables in the remaining tests and in the comprehensive framework. Hence they are discussed here along with measures of behavioral intentions. All three constructs were measured after the service was delivered.

Overall service quality was measured using four Likert scale items. The items captured either *perceived service quality* or *service quality as measured disconfirmation* (depending on the assigned group) using similar prefacing phrases as for the factors. The four five-point items (with endpoints strongly agree/strongly disagree) referred to "excellent overall service," "service of a very high quality," "a high standard of service," and "superior service in every way" (Dabholkar, 1995a; Spreng and Mackoy, 1996). Customer satisfaction was also captured as *perceptions* or *disconfirmation* (depending on the assigned group) using similar prefacing phrases as before. Three five-point items were used with endpoints completely satisfied/completely dissatisfied, very pleased/very displeased, and absolutely delighted/absolutely terrible (Westbrook, 1980; and Westbrook and Oliver, 1981). Intentions to use the service in the future and intentions to recommend the service to others were captured with two five-point scales each, with endpoints very unlikely/very likely and definitely would/definitely would not (Fishbein and Ajzen, 1975).

Data Collection Procedure

The data were collected through telephone interviews and each respondent was contacted twice. The first contact occurred within one month of the church's commitment to order a pictorial directory and the purpose was to collect expectations data. The second contact occurred shortly (usually within one month) after service delivery and the purpose was to collect perceptions and measured disconfirmation data. In all cases, the same individual was interviewed in both contacts. Each telephone interview required approximately twenty-five minutes to complete. At the start of the project there was some concern that subjects would tire of the lengthy survey and tend to terminate prior to completion. However, the problem did not develop. Instead, in general, respondents were very willing to devote the time necessary to complete the survey. The telephone interviewers were undergraduate and graduate students with previous experience in telephone interviewing.

Each interviewer was required to attend a one-hour training session. Additionally, each interviewer observed a supervisor conducting two or three telephone interviews before attempting to complete a call themselves. As a final step in the training, interviewers made two or three calls with a supervisor listening to and critiquing their procedure. Supervisors continued to monitor interviewers throughout the study to assure the quality of the interview process.

Confirmatory Factor Analysis

Preliminary confirmatory factor analysis was conducted with the 21 items used to capture factors relevant for service quality. Four factors were specified (Reliability, Personal Attention, Comfort, and Features) and were tested for three sets of measures (perceptions, measured disconfirmation, and computed disconfirmation). The CFIs for these preliminary analyses were 0.83 for perceptions, 0.91 for measured disconfirmation, and 0.90 for computed disconfirmation. Whereas, two of these CFIs were acceptable, one was somewhat low. Moreover, it was seen that five items (# 5, 7, 11, 12, and 21, see Appendix for item description) had high modification indices irrespective of the measure (i.e., modification indices for these items were high for all three sets).

Based on a re-examination of the wording of these five items and their lack of fit with their respective factors, these items were dropped. The first was an item on the Reliability factor (item 5, Appendix). This item cross-loaded heavily on the Personal Attention factor probably due to the phrase “interactions with you” and, therefore, it was dropped. The second item was from the Personal Attention factor (item 7, Appendix) and cross-loaded on the Reliability factor, probably due to the phrase “tell you exactly when” and, therefore, it was also dropped. The next two items, also from the Personal Attention factor, (items 11 and 12, Appendix) had high cross-loadings on the Comfort factor, most likely due to the “problem solving” and “prompt service” references. These were also dropped due to their ambiguity. The fifth and last item to be dropped was from the Features factor (item 21, Appendix) and referred to the overall directory rather than its features or associated process and as a result did not fit well with the factor.

Confirmatory factor analysis was repeated with the remaining sixteen items for the three sets of measures and all the results showed a good fit for the 4-factor structure, with no high modification indices. The CFIs for these analyses were 0.91 for perceptions, 0.95 for measured disconfirmation, and 0.95 for computed disconfirmation. (The results of the full confirmatory analysis on *all* exogenous and endogenous variables are shown in Table 1A.) Reliabilities based on Cronbach's alpha were computed for each factor, using the three sets of measures. The values ranged from 0.76 to 0.91 (see Table 1B), more than acceptable for scale development (Nunnally and Bernstein, 1994).

Other Tests to Validate Constructs and Measures

One of the issues under consideration (corollary to proposition P2) was the overlap or distinction between service quality and customer satisfaction. The fact that the full

TABLE 1

A. Confirmatory Factor Analysis for All Exogenous and Endogenous Variables

Based on:	df	χ^2	SD RMR	NNFI	CFI
Perceptions	254	551.95 ($p = .00$)	.05	0.90	0.91
Measured disconfirmation	254	484.05 ($p = .00$)	.04	0.94	0.95
Computed disconfirmation	254	450.66 ($p = .00$)	.05	0.93	0.94

B. Reliabilities for Factors Related to Service Quality

Based on:	Cronbach's Alpha			
	Reliability	Personal Attention	Comfort	Features
Perceptions	0.79	0.88	0.79	0.76
Measured disconfirmation	0.87	0.91	0.88	0.82
Computed disconfirmation	0.79	0.88	0.84	0.78

confirmatory factor analysis was well-supported (see Table 1A) provides evidence of discrimination among *all* the constructs. Nevertheless, the discriminant validity of these two constructs was ascertained using a χ^2 difference test. The χ^2 difference was significant (at $p < .001$) for both perceptions and measured disconfirmation⁶. Consequently, the constructs can be seen as distinct. Cronbach's alphas for service quality were 0.92 for perceptions and 0.94 for measured disconfirmation. Cronbach's alphas for customer satisfaction were 0.92 for both perceptions and measured disconfirmation.

A similar test was run for intentions to use the service and intentions to recommend the service in the future. Although discriminant validity was supported (χ^2 difference significant at $p < .001$), the modification indices and standardized residuals for these two variables were unacceptably high. Consequently, it was decided to treat behavioral intentions as a single construct. During the structural equations analysis that follows, the second intention to use item and the second intention to recommend item were found to have high modification indices, this time on service quality and customer satisfaction. Given that these items had been reverse coded and may have been problematic, it was decided to drop them. The remaining two items measuring behavioral intentions had a correlation of 0.86 for both sets of data.

The correlations among all the constructs for the three measures are presented in Table 2. Additional χ^2 difference tests were conducted for pairs of constructs with higher correlations. In each case, the χ^2 difference was significant (at $p < .001$) indicating that all the constructs in the study were independent. This was a further confirmation of discriminant validity among the constructs in addition to the results of the confirmatory factor analysis (see Table 1A). Additionally, collinearity diagnostics were used for the exogenous variables. The variance inflation factor test showed lack of collinearity among these variables for all three measures (i.e., VIFs < 5). The perceptions measures had the lowest VIFs (ranging from 1.49 to 2.76) and measured disconfirmation had the highest VIFs (1.92 to 4.97). The VIFs for computed disconfirmation fell in the middle (ranging from 1.82 to 3.53).

TABLE 2

Correlation Matrices for All Exogenous and Endogenous Variables

A. Perceptions Data							
	REL	PA	COMF	FEAT	SQ	CS	BI
Reliability	1.000						
Personal attention	0.542	1.000					
Comfort	0.516	0.734	1.000				
Features	0.309	0.437	0.572	1.000			
Service quality	0.673	0.772	0.822	0.593	1.000		
Customer satisfaction	0.666	0.657	0.715	0.438	<u>0.848</u>	1.000	
Behavioral intentions	0.399	0.449	0.474	0.265	<u>0.548</u>	<u>0.795</u>	1.000
B. Measured Disconfirmation Data							
	REL	PA	COMF	FEAT	SQ	CS	BI
Reliability	1.000						
Personal attention	0.810	1.000					
Comfort	<u>0.807</u>	<u>0.843</u>	1.000				
Features	<u>0.611</u>	0.625	0.682	1.000			
Service quality	0.735	0.772	<u>0.816</u>	0.648	1.000		
Customer satisfaction	0.627	0.575	<u>0.691</u>	0.526	<u>0.768</u>	1.000	
Behavioral intentions	0.367	0.367	0.451	0.336	<u>0.535</u>	<u>0.772</u>	1.000
C. Computed Disconfirmation Data							
	REL	PA	COMF	FEAT	SQ	CS	BI
Reliability	1.000						
Personal attention	0.619	1.000					
Comfort	0.625	0.765	1.000				
Features	0.463	0.538	0.692	1.000			
Service quality	0.547	0.633	0.648	0.423	1.000		
Customer satisfaction	0.521	0.531	0.554	0.301	<u>0.848</u>	1.000	
Behavioral intentions	0.326	0.345	0.350	0.172	<u>0.548</u>	<u>0.795</u>	1.000

Variance Restriction and Other Patterns

Perceptions for each item were compared with measured disconfirmation and with computed disconfirmation for possible variance restriction and any other patterns. Perceptions ranged from 3.63 to 4.57, measured disconfirmation ranged from 3.25 to 3.93, and computed disconfirmation ranged from -0.06 to -0.69. Standard deviations were similar for all three measures and ranged from 0.78 to 1.59. Thus, computed disconfirmation did not seem to suffer from variance restriction. Nevertheless, the negative numbers suggested a kind of double counting because the perceptions measures likely incorporated *implicit* standards of comparison (see Dabholkar, 1993). Finally, the gaps for all the items followed a similar pattern, irrespective of how they were measured (i.e., computed or measured disconfirmation).

ANALYTICAL RESULTS

Each model (proposed and alternative) is tested in terms of model fit using “Regression in SEM,” a special form of structural equations modeling (SEM) with LISREL (Jöreskog and Sörbom, 1993). The analysis is based on covariance matrices and all the indicators for each construct are summed to provide a single indicator in each case. One reason for using this special technique is that the exogenous variables are highly correlated, and although collinearity diagnostics and discriminant validity tests show no problems in the data itself, the traditional SEM approach in correcting for measurement error is likely to create collinearity and mask some of the effects of these variables. Jöreskog and Sörbom (1993) explain that with this technique, the focus should be on the relative importance of the predictor variables, as long as the fit is acceptable.

A second reason to use this method is that measurement error is low in the data and does not need to be corrected through SEM. Low measurement error is evidenced both by high reliabilities of the constructs and by the fact that regression analysis allows more effects to be significant than does traditional SEM, a somewhat unusual case. Some researchers suggest that in such instances regression offers more reliable results than traditional SEM (Jaccard and Wan, 1996). However, by using Regression in SEM, the advantages in using regression are maintained, but in addition, it is possible to include more than one dependent variable, allow predictor variables to be correlated, and estimate model fits. Compared to traditional SEM, this technique has the added advantage of parsimony and is more robust against violations of assumptions about multivariate normality.

The criteria for model comparison are as follows. First, model fits are ascertained to ensure that the models can be included in further comparisons. Two of the alternative models (Components and Independent Effects) are saturated (based on the theory underlying these models). Hence model fits cannot be ascertained through LISREL as degrees of freedom = 0. Instead, F-values from regression analysis are provided for these two models to show that these models *are* supported and can be compared to other models. For the rest of the models, fit indices can be computed in LISREL. RMSEA is not an appropriate fit index given the extremely low degrees of freedom with such models, but three highly recommended fit indices, standardized RMR, NNFI, and CFI (Jaccard and Wan, 1996; Kline, 1998) are provided.

Next, the models are compared for the number of significant gammas and the absence of negative gammas. The reason for including this criterion is that nonsignificant and negative gammas often indicate misspecified models (Maddala, 1977). Structural equations modeling textbooks (Jaccard and Wan, 1996; Kline, 1998) increasingly point out that a good fit alone does not imply a strong relationship, and that it is possible to have a perfect fit with little correlation. In other words, fit indices rule out bad models but do not necessarily indicate good models; it is important to study the number and signs of significant gammas, as well as effect sizes and R² values to assess the strength of models.

The next criteria therefore include comparisons of R² values and effect sizes. The nature of the study (testing different measures for the same constructs) precludes statistical comparisons of effect sizes and R² values, and these comparisons are therefore qualitative.

Two other criteria are used for further assurance in model selection. At a basic level, the underlying simple correlations are examined to study patterns and differences across models and measures. At a more sophisticated level, proposed models are checked for omitted paths, that is, direct paths to the endogenous variable are added to the proposed models and tested (Kline, 1998). If models with these added paths (that are omitted from the proposed model) were supported, this would detract from the support for the proposed model. Finally, although model fit is not a sufficient criterion to compare models, high model fits lend greater credence to the rest of the comparison criteria.

Testing the Conceptual Framework

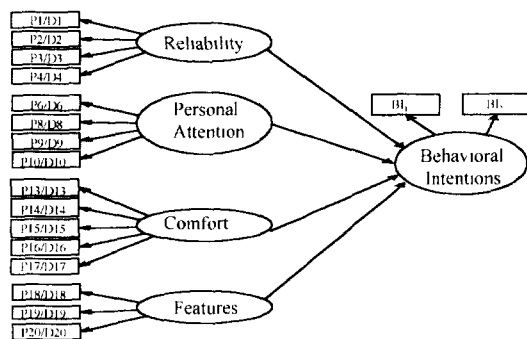
First proposition P1 is tested against an alternative model and then proposition P2 is tested against two alternative models (see Figure 1). The comprehensive conceptual framework is to be tested only after these proposed submodels have been tested and supported. This two-step process is an essential part of theory building for service quality and its related constructs given all of the unresolved issues in the literature.

Proposition P1

The alternative Components model suggests that the factors (Reliability, Personal Attention, Comfort, and Features) operate as components of service quality, and that they have direct effects on behavioral intentions (see Figure 2A). A Regression in SEM is run to test this model and the results are presented in Table 3A. The proposed Antecedents model suggests that overall service quality mediates between the four factors and behavioral intentions (see Figure 2B). A Regression in SEM is run to test this model as well and the results are presented in Table 3B. Both models are tested with all three measures—perceptions, measured disconfirmation, and computed disconfirmation.

The Components model is saturated, and hence fit indices cannot be estimated in LISREL. However, F values show that the fit is acceptable. The Antecedents model has excellent fits in terms of standardized RMR, NNFI, and CFI. The Components model has negative gammas and fewer significant gammas than the Antecedents model (see Table 3), suggesting poor model specification'. Misspecified models often have nonsignificant or negative gammas, which change with the addition of an appropriate mediating variable (Maddala, 1977). As soon as a mediator variable (overall service quality) is introduced (i.e., the Antecedents model), the negative gammas disappear and there are fewer non-significant gammas. The substantial improvement in the model when service quality acts as a mediating variable (over the direct effects model) provides strong support for proposition P1.

The Components model also has lower R^2 values for behavioral intentions than for the Antecedents model, but this difference is <0.1 . This difference may be statistically significant, but it is not possible to test it as explained earlier. At the same time, the R^2

A. Components of Service Quality

Where P_i and D_i are indicators for perceptions and disconfirmations for $i = 1-20$ (see Appendix). BI₁ are indicators for behavioral intentions, and SQ_i are indicators for service quality

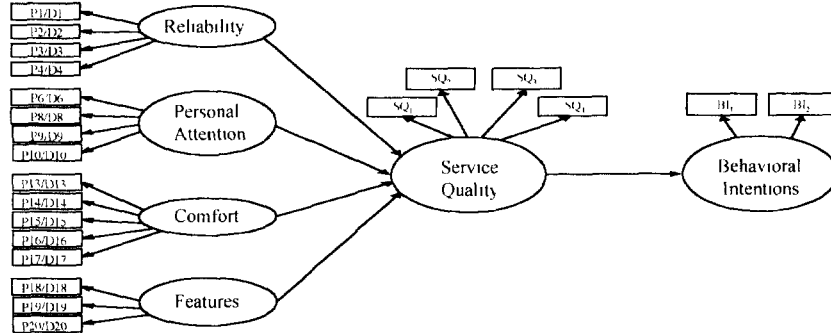
B. Antecedents of Service Quality

FIGURE 2

Testing Factors as Components versus Antecedents and the Behavioral Consequences of Service Quality

values for service quality (in Table 3B) are substantial showing that the factors explain a much larger variance in service quality (Antecedents model) than they do in behavioral intentions (Components model). Viewing the underlying simple correlations (see Table 2) shows further that the factors have a higher correlation with service quality than with behavioral intentions. Finally, an alternative antecedents model with additional direct paths from the factors to behavioral intentions shows that these paths are not supported when the indirect paths through service quality are present. In other words, the proposed Antecedents model has no omitted paths. All of these results as well as the excellent fits for the Antecedents model offer strong support for proposition P1 (i.e., the antecedents model of service quality is superior to the components model).

Proposition P2

Earlier we had seen through confirmatory factor analysis and a test of discriminant validity that service quality and customer satisfaction are distinct constructs, thus

TABLE 3
Factors as Components vs Antecedents of Service Quality (SQ)

A. Components of SQ: Direct Effect on Behavioral Intentions (BI) (Alternate Model)									
γ ($\rightarrow$ BI)									
Based on	df	χ^2	F*	<p*	Personal				
					Reliability	Attention	Comfort	Features	R ² _{BI}
Perceptions	0	0	18.03	0.001	0.35 ^a	n.s.	0.54 ^a	n.s.	0.47
Measured disconfirmation	0	0	12.62	0.001	n.s.	(-0.33) ^c	0.59 ^a	n.s.	0.29
Computed disconfirmation	0	0	9.37	0.001	0.22 ^b	n.s.	0.48 ^a	(-0.28) ^a	0.34

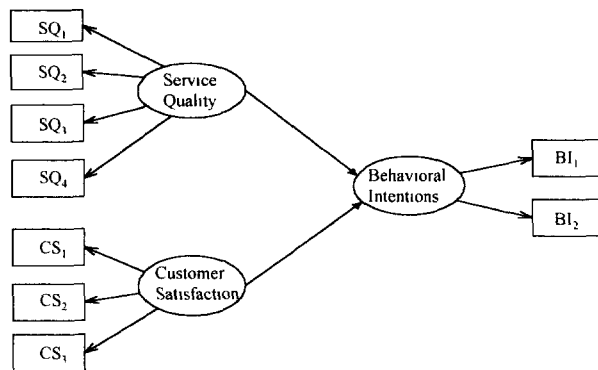
B. Antecedents of SQ: Mediating Effect of Service Quality (Proposed Model)										
γ ($\rightarrow$ SQ)										
Based on:	df	χ^2	p	SD	RMR	NNFI	CFI	Personal		
								Reliability	Attention	Comfort
Perceptions	4	23.10	.0001	0.04	0.04	0.90	0.97	0.30 ^a	0.27 ^a	0.42 ^a
Measured disconfirmation	4	14.35	.006	0.02	0.02	0.95	0.99	n.s.	0.22 ^c	0.51 ^a
Computed disconfirmation	4	29.26	7E-06	0.04	0.04	0.86	0.96	0.15 ^c	0.23 ^b	0.31 ^b

β				
Features	SQ $\rightarrow$ BI	R ² _{SQ}	R ² _{BI}	
0.21 ^a	0.73 ^a	0.81	0.51	
0.16 ^c	0.54 ^a	0.70	0.31	
n.s.	0.79 ^a	0.48	0.39	

*OLS fit provided. (Not possible to test SEM model fit for saturated model, with $df = 0$.)

Notes: ^a $p < .001$ ^b $p < .01$ ^c $p < .05$.

A. Independent Effects of Service Quality and Customer Satisfaction on Behavioral Intentions



Where SQ_i are indicators measuring service quality, CS_i are indicators measuring customer satisfaction, and BI_i are indicators measuring behavioral intentions

B. Customer Satisfaction as a Mediator

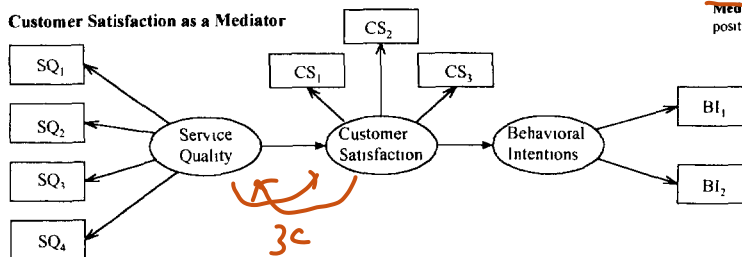


Figure 3C (Service Quality as a Mediator) simply exchanges the positions of SQ and CS in Figure 3B

FIGURE 3

Exploring the Link Between Service Quality and Customer Satisfaction and Testing the Mediating Role of Customer Satisfaction in Predicting Behavioral Intentions

supporting the corollary to proposition P2. To test proposition P2, three separate models are run. The first alternate model is the Independent Effects model where service quality and customer satisfaction have direct effects on behavioral intentions (see Figure 3A). The proposed model is where customer satisfaction mediates the effect of service quality on behavioral intentions (see Figure 3B). The second alternate model is where service quality mediates the effect of customer satisfaction on behavioral intentions (see note for Figure 3C). All three models are tested with Regression in SEM and the results are presented in Table 4. Further, all three models are tested both with perceptions and measured disconfirmation data⁸.

The Independent Effects model is saturated, and hence fit indices cannot be estimated in LISREL. However, F values show that the fit is acceptable. The Customer Satisfaction as a Mediator model has excellent fits in terms of standardized RMR, NNFI, and CFI. The Service Quality as a Mediator model has an unacceptable NNFI (despite acceptable CFI and standardized RMR values) in one case (i.e., perception measures), and all completely unacceptable fit indices for the other case (i.e., measured disconfirmation). This is clearly a poor model with omitted direct paths (from

TABLE 4
Independent Effects of Service Quality (SQ) and Customer Satisfaction (CS) versus Mediating Role of CS and SQ on Behavioral Intentions (BI)

A. Independent Effects of SQ & CS on BI (Alternate Model 1)									
Based on	df	χ^2	F*	<p*	γ			R^2_{BI}	
					SQ→BI	CS→BI	CS→BI		
Perceptions	0	0	163.07	0.001	0.29 ^b	0.65 ^a	0.65 ^a	0.58	
Measured disconfirmation	0	0	129.34	0.001	n.s.	0.92 ^a	0.92 ^a	0.54	
B. Mediating Role of CS on BI (Proposed Model)									
Based on	df	χ^2	p	SD RMR	NNFI	CFI	β		R^2_{BI}
							SQ→CS	CS→BI	
Perceptions	1	9.21	0.002	0.03	0.94	0.98	0.68 ^a	0.96 ^a	0.56
Measured disconfirmation	1	0.04	0.851	0.003	1.00	1.00	0.61 ^a	0.91 ^a	0.54
C. Mediating Role of SQ on BI (Alternate Model 2)									
Based on	df	χ^2	p	SD RMR	NNFI	CFI	γ		R^2_{BI}
							CS→SQ	SQ→BI	
Perceptions	1	28.48	0.0	0.06	0.78	0.93	1.06 ^a	0.73 ^a	0.51
Measured disconfirmation	1	71.26	0.0	0.13	0.28	0.76	0.97 ^a	0.54 ^a	0.31

*OLS fit provided. (Not possible to test SEM model fit for saturated model, with df = 0.)

Notes: ^ap < .001 ^bp < .01 ^cp < .05.

CS to BI) causing the deplorable fits. Hence, it is discarded from further consideration in this test.

The Independent Effects model has fewer significant gammas than the Customer Satisfaction as a Mediator model (see Table 4). The fact that service quality and customer satisfaction are highly correlated across all three measures (see Table 2), may provide a possible explanation for the nonsignificant gamma for service quality in the Independent Effects model. Yet, the nonsignificant gamma is only present for measured disconfirmation where the correlation between service quality and customer satisfaction is in fact *lower* than it is for the perceptions data. Hence, the nonsignificant gamma for service quality is more likely to be the result of poor model specification (Maddala, 1977). With the correctly specified model, the nonsignificant gamma disappears. This improvement in the model in which customer satisfaction acts as a mediating variable (over the independent effects model) supports proposition P2.

The Independent Effects model also has much smaller effect sizes than the Customer Satisfaction as a Mediator model in three out of four cases (see Table 4). Although the difference in effect sizes cannot be compared statistically, the smaller gammas in the former model are again likely to be the result of poor model specification (Maddala, 1977). With the correctly specified model, effect sizes increase, thus supporting P2. The R^2 values for behavioral intentions are not very different across the two models and in fact are slightly higher for the Independent Effects model in one case. However, R^2 values for behavioral intentions with only service quality as a determinant (see Table 3B) are smaller, suggesting that customer satisfaction has a stronger effect on behavioral intentions than does service quality. The results support the idea that although service quality has an impact on behavioral intentions, customer satisfaction acts as a strong mediator.

The same conclusion may be drawn from the underlying correlations as well, where the factors are more highly correlated with service quality, and customer satisfaction is more highly correlated with behavioral intentions (see Table 2). In addition, a test of an alternate proposed model with an indirect path from service quality to behavioral intentions shows no support for such a path. In other words, the proposed model has no omitted path. Finally, the proposed model has excellent fits, and along with all the other results strongly supports proposition P2 (i.e., the mediating role of customer satisfaction on behavioral intentions is superior to both alternative models).

Overall Framework

Having tested and supported each submodel against alternative models, the next step is to construct and test the full model to verify that the overall framework is supported (see Figure 4). The full model is tested for both perceptions data and measured disconfirmation and the results are presented in Table 5. (Again, this model is not tested with computed disconfirmation data because part of that dataset is identical to the perceptions data, and would give misleading results.) Both models have good fits and high variance explained for the endogenous variables. Thus, the broad conceptual framework proposed for service

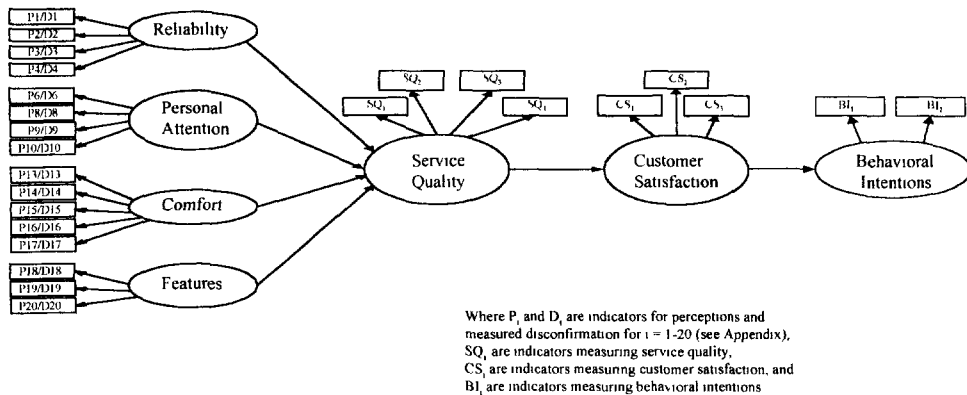


FIGURE 4

Testing the Comprehensive Framework for the Antecedents and Consequences of Service Quality, with Customer Satisfaction as a Mediator

quality is strongly supported. Factors act as antecedents rather than as components of service quality (proposition P1) and customer satisfaction plays a mediating role on the effect of service quality on behavioral intentions (proposition P2).

Testing the Measures

To compare the three measures of service quality, the results are re-examined within (rather than across) the different models tested. First, computed disconfirmation is evaluated against the other two measures for a concurrent test of propositions P3b, P4, and P5. Then, measured disconfirmation is evaluated against perceptions for a test of proposition P3a. The same criteria that were used for comparing models are used now for comparing measures, with two exceptions. The criterion of omitted paths is only relevant for comparing alternate conceptual models and therefore cannot be applied for comparing measures. However, another criterion, parsimony in data collection, is relevant for comparing measures whereas it was not applicable for model comparison.

Computed Disconfirmation versus Perceptions and Measured Disconfirmation

In terms of model fit, computed disconfirmation has the lowest F value of the three measures in the Components model (see Table 3A). This measure also has an unacceptable NNFI in the Antecedents model whereas the other two do not (see Table 3B). Unacceptable model fit may be enough reason to drop this measure. However, other criteria are examined for a fuller comparison.

TABLE 5

Comprehensive Framework for the antecedents and Consequences of Service Quality																						
Based on:	df	χ^2	p	SD	RMR	NNFI	CFI	Reliability	γ ($\rightarrow$ SQ)					β								
									Personal					Attention	Comfort	Features	SQ $\rightarrow$ CS	CS $\rightarrow$ BI	R^2_{SQ}	R^2_{CS}	R^2_{BI}	
									Perceptions	Measured disconfirmation	Perceptions	Measured disconfirmation	Perceptions									Measured disconfirmation
Perceptions	9	40.65	6E-06	0.05	0.05	0.92	0.97	0.30 ^a	0.27 ^a	0.42 ^a	0.21 ^a	0.68 ^a	0.96 ^a	0.81	0.72	0.56						
Measured disconfirmation	9	21.54	0.01	0.03	0.03	0.97	0.99	n.s.	0.22 ^c	0.51 ^a	0.16 ^c	0.61 ^c	0.91 ^a	0.70	0.59	0.54						
Notes:									^a $p < .001$	^b $p < .01$	^c $p < .05$.											

Both computed and measured disconfirmation show a negative gamma in the Components model (see Table 3A) and one nonsignificant gamma in the Antecedents model (see Table 3B). In contrast, perception measures have no negative gammas in the Components model (see Table 3A) and no nonsignificant gammas in the Antecedents model (see Table 3B). These results suggest that perception measures are superior to both disconfirmation measures using the criterion of counting gammas. However, a question may be raised whether possible collinearity in the disconfirmation measures may lead to negative or nonsignificant gammas. Hence, further criteria are examined.

The R^2 value for service quality is far lower for computed disconfirmation (0.48) than the R^2 values for perception (0.81) and measured disconfirmation (0.70). Although R^2 for behavioral intentions is higher for computed disconfirmation than for measured disconfirmation for both models, the former data overlap with perception measures for behavioral intentions and hence this comparison is misleading. Further, R^2 for behavioral intentions is lower for computed disconfirmation than for perceptions *despite* the overlap of the data—0.34 versus 0.47 for the Components model, and 0.39 versus 0.51 for the Antecedents model.

Similarly, effect sizes are lower for computed disconfirmation measures as compared with the other two measures, irrespective of the model (see Table 3). In examining underlying correlations, computed disconfirmation measures do not show the unequivocal support for the conceptual framework (i.e., for P1 and P2) that the other two measures do. The last criterion, parsimony in data collection, suggests that perceptions and measured disconfirmation would be easier to collect than computed disconfirmation, which requires twice as many items irrespective of whether the data are collected longitudinally or cross-sectionally. All of these results/attributes, along with poor model fit for computed disconfirmation, clearly support propositions P3b and P4, namely, that computed disconfirmation is inferior to perceptions (P3b) and to measured disconfirmation (P4).

Proposition P5 simply follows from these results taken together. Given that both perceptions and measured disconfirmation can be measured in a cross-sectional design, these results indicate that a study of service quality should not necessitate a longitudinal design. Thus, proposition P5 (i.e., preference for cross-sectional over longitudinal design) is supported.

Measured Disconfirmation versus Perceptions

First, the results of the *alternate* conceptual models (Tables 3A, 4A, and 4C) are examined to compare the two measures. In Tables 3A and 4A, perception measures have higher F values and higher R^2 than measured disconfirmation (although in Table 4A, the R^2 difference is <0.1). Perception measures also do not yield a negative gamma as does measured disconfirmation in Table 3A. Similarly, perception measures do not yield a nonsignificant gamma as does measured disconfirmation in Table 4A. In Table 4C, perception measures have far better fit indices (although the NNFI is unacceptable) than measured disconfirmation (which has completely unacceptable fit

indices). Finally, perception measures have higher effect sizes for service quality and customer satisfaction than measured disconfirmation in Tables 4A and 4C.

Next, looking at Tables 3B, 4B, and 5 (which show results of the *proposed* conceptual models), it is seen that both measures have excellent fits. Although the fits are higher (almost perfect) for measured disconfirmation, this does not necessarily imply a better model. The strength of SEM models should be assessed by examining significant gammas, effect sizes, and R^2 values (Jaccard and Wan, 1996; Jöreskog and Sörbom, 1993; Kline, 1998). Perceptions have more significant gammas for the factors than measured disconfirmation (Table 3B). Perceptions also have higher effect sizes for service quality and/or customer satisfaction as predictor variables (in all three tables). R^2 values appear to be clearly superior for perceptions versus measured disconfirmation in Table 3B for both service quality and behavioral intentions. In Tables 4B and 5, R^2 values appear to be superior for perceptions versus measured disconfirmation for both customer satisfaction and service quality, but are very close for behavioral intentions.

The conclusion is that there is strong (but partial) support for proposition P3a. Measured disconfirmation has better fits in most cases but perception measures have excellent fits as well and better fits in a few cases. Perception measures also yield a greater number of significant gammas and show an absence of negative gammas as compared to measured disconfirmation (see Tables 3A, 3B, and 4A). For the results in Tables 3A and 3B, the issue of possible collinearity in the factors with measured disconfirmation reduces support for perception measures on this particular criterion. However, for the results in Table 4A, the factors are not included and the issue of possible collinearity in the measured disconfirmation is not relevant. In fact service quality and customer satisfaction show a lower intercorrelation for measured disconfirmation than they do for perceptions (see Table 2). Hence, in this case, the nonsignificant gamma suggests that measured disconfirmation may be inferior to perceptions. As discussed earlier, perception measures also have greater effect sizes over measured disconfirmation (in all the proposed models, i.e., Tables 3B, 4B, and 5). Finally, perception measures consistently have higher R^2 values over measured disconfirmation for service quality and customer satisfaction (with differences >0.1 across the board) and also for behavioral intentions (although this latter difference may not be statistically significant).

DISCUSSION

This study examines several critical issues related to the conceptualization and measurement of service quality. Based on various debates in the literature, a comprehensive framework is proposed related to issues about antecedents, consequences, mediators, and measurement of service quality. This framework is tested with a series of alternate models in a longitudinal, empirical study.

The first conceptual issue examined was whether factors related to service quality should be viewed as its components or its antecedents. The results clearly show support

for the antecedents model. In other words, consumers evaluate different factors related to the service but also form a separate overall evaluation of the service quality (which is not a straightforward sum of the components). The factors serve as antecedents to this overall evaluation that in turn influences behavioral intentions. The antecedents model provides a more complete understanding of service quality and how these evaluations are formed. As seen in the literature, the components-to-antecedents transition is a natural progression in the understanding of critical constructs.

This view of service quality has benefits for practitioners as well. For predictive purposes, marketers could simply measure this overall evaluation and will find it easier to do so regularly than to measure all the items related to the factors. Another advantage of using the antecedents model is that measures of overall service quality can provide better feedback to managers regarding overall impressions of their service. In the components model framework, there is no separate, reliable construct of overall service quality, and marketers would have to use *all* the items related to the factors just to measure customer evaluations of service quality. For diagnostic purposes, the factors should certainly be measured and evaluated.

The second conceptual issue investigated was the mediating role of customer satisfaction on the effect of service quality on behavioral intentions. A strong mediating role was found, confirming that it is important to measure customer satisfaction separately from service quality when trying to determine customer evaluations of service. It was also determined that customer satisfaction is a much better predictor of behavioral intentions, whereas service quality is more closely related to specific factor evaluations about the service. These findings together support a conceptual framework in which the relevant factors act as antecedents to service quality and where customer satisfaction strongly mediates the effect of service quality on behavioral intentions.

Although the study was conducted with nonprofit institutional customers who can be viewed as "retail buyers," the results should apply equally well to final consumers and profit organizations. It is recommended that future studies investigate the validity of these findings in a variety of contexts. Broad support of the chronological order of service evaluations (see Figure 1) would serve as a guide to practitioners in studying different aspects of their service delivery. For prediction purposes, they could focus on customer satisfaction, whereas for investigative purposes they could focus on service quality. In other words, practitioners can measure either construct of service evaluation, or both, depending on their objective(s).

A related issue is the distinction (as opposed to overlap) between service quality and customer satisfaction. In our context, we found the constructs to be distinct, although highly correlated (see Table 2). For situations with minimum emotional content, or where performance always falls *within* the zone of indifference, or over time, there might be even greater overlap between the two constructs (Dabholkar, 1995b). In contrast, for situations with high emotional content, or where performance always falls *outside* the zone of indifference, the causal sequence between the two constructs may be reversed (Dabholkar, 1995b). These atypical situations should be explored in future studies to continue to deepen our understanding of service evaluations.

Another contribution of this study is that both qualitative and quantitative research

techniques were used in determining factors relevant for service quality. Further, the findings from both approaches were consistent. Overall, the factors that respondents in focus groups suggested were relevant for service quality were found to be important predictors of service quality in the quantitative study. Additionally, “comfort” appeared to be the most important factor in both approaches. In the quantitative study, “comfort” appeared to have the highest effect sizes regardless of model or measure; in the qualitative studies, respondents stressed how important being comfortable was to them. It is clear that feeling safe and comfortable with the service provider is critical for service quality evaluations in this context. Future research should continue to use both qualitative and quantitative techniques to determine which factors are important to customers in evaluating service quality in different contexts.

We also investigated three major measurement issues in this study. The first measurement issue related to perceptions versus disconfirmation as measures of service quality. The perceptions measures were clearly superior to computed disconfirmation on all criteria. Perception measures also performed better than measured disconfirmation on most criteria and were comparable on a few criteria. The findings provide strong support for using perception measures over disconfirmation, especially when the objective is to increase prediction and/or explanation. From the practitioners’ perspective, perception measures provide more links between the endogenous variable (service quality) and its exogenous determinants, thus offering a means to understand and test the impact of a greater number of antecedent factors.

The second issue is related to a preference between disconfirmation measures. This is relevant because perception measures do not identify gaps (i.e., discrepancies between customers’ expectations and perceptions) that provide valuable process-improvement information to practitioners. Although there was no significant difference between computed and measured disconfirmation in terms of variance restriction or gap analysis, measured disconfirmation was found to be superior to computed disconfirmation on the majority of criteria. The findings suggest the use of measured over computed disconfirmation, when the objective is gap analysis. In other words, direct measures of disconfirmation are preferred over difference scores computed from expectations and perceptions measured separately.

It is possible that for studies with a shorter lag time between measuring expectations and perceptions than in this study, computed disconfirmation measures might fare somewhat better, although these measures do have some inherent disadvantages. One disadvantage is the misleading negative service evaluations obtained by subtracting expectations from perceptions, which suggest that customers were dissatisfied, whereas the perception measures clearly show that they were not. Further, it appears that expectations may change over time and can be captured both in the perception measures (where implicit standards were evidently being used) and in the measured disconfirmation (where explicit comparisons were made after the service). For computed disconfirmation, however, it is somewhat artificial to ask respondents for their evaluations (i.e., perceptions), and then to go back to what they had said they expected before the service and artificially subtract these numbers.

A third disadvantage is that whereas perceptions and measured disconfirmation can

be measured in a cross-sectional study, computed disconfirmation is ideally measured in a costly, longitudinal study (where expectations are measured before the service and perceptions after). Further, irrespective of research design, computed disconfirmation measures take twice the space as the other two given that expectations and perceptions need to be captured separately. Given the superiority of perception measures and measured disconfirmation (over computed disconfirmation), as well as the parsimony in data collection associated with these two measures, a cross-sectional design is recommended over a longitudinal design for understanding and testing service quality.

Thus, with regard to measurement issues, this study suggests the following. Measure perceptions rather than disconfirmation when the objective is to predict service quality or to gauge its determinants. If gap analysis is the objective, measured disconfirmation is recommended over computed disconfirmation. A longitudinal design is unnecessary in either case. However, on-going measures of overall service quality (as a separate, multi-item construct) are recommended to keep in touch with customers' evaluations. For specific, in-depth studies with the dual objectives of prediction and gap analysis, service quality should be measured in a cross-sectional design after the service is delivered. In addition, ideally perceptions should be measured for half the sample and measured disconfirmation for the other half. Both scales would be of equal length and together would offer high predictive power as well as the opportunity for gap analysis.

In conclusion, we proposed and found strong support for a conceptual framework for understanding service quality, its antecedents, consequences, and mediators. We also addressed major issues related to service quality measurement and found fairly clear answers to these issues. Based on the two sets of findings, we have made specific recommendations to improve future research (and practice) involving service quality conceptualization and measurement. We have also raised several related research issues and suggested avenues for further investigation of these issues.

Three caveats about the study need to be noted. The issue of possible collinearity in the measured disconfirmation may have affected the counting of gammas as a criterion. In this study, this issue was only relevant for one proposition. In addition, we used a large number of comparison criteria to ensure that we did not base model or measure selection on a single criterion. Future research should continue to use a variety of criteria to compare alternative models or measures. At the same time, the issue of collinearity and its effect on evaluation of model specification bears further investigation, whether for service quality measures or measures of other constructs.

Second, we did not explore antecedents of customer satisfaction in our qualitative research because our focus was on service quality. However, future studies should explore the antecedents of both constructs to further indicate how they may be similar or different in order to increase our understanding of both types of service evaluations. Finally, we measured behavioral intentions rather than actual behavior, and thus may have overestimated predictive power. Future studies could track actual repurchase and word-of-mouth behavior to more accurately determine the predictive power of service quality and customer satisfaction evaluations.

APPENDIX

Service Quality: Factor Structure and Concomitant Items

<i>SERVQUAL Dimension</i>	<i>Service Quality Dimensions in this study</i>	<i>Item</i>
Reliability	Reliability	1 When Company X's employees promise to do something by a certain time, they will do so.
*	Reliability	2 When Company X's employees promise to do something in a certain way, they will do so.
Reliability	Reliability	3. Company X will maintain error-free records.
Reliability	Reliability	4. Company X will perform the service right the first time.
* +	Reliability	5. Company X's interactions with you will be consistent.
*	Personal Attention	6. Company X's employees will give you personal attention.
Responsiveness +	Personal Attention	7. Company X will tell you exactly when services will be performed.
Responsiveness	Personal Attention	8 Company X's employees will never be too busy to respond to your requests.
Assurance	Personal Attention	9. Employees of Company X will be courteous with you.
*	Personal Attention	10. Employees of Company X will be willing to help you
Reliability +	Personal Attention	11. When you have a problem, Company X will show a sincere interest in solving it.
Responsiveness +	Personal Attention	12. Employees of Company X will give prompt service
Assurance	Comfort	13. Employees of Company X will have the knowledge to answer your questions.
*	Comfort	14. Company X's employees will have the ability to solve your problems.
*	Comfort	15. Company X's salespeople will not pressure church members to buy portraits.
Assurance	Comfort	16. The behavior of Company X's employees will instill confidence in you.
Assurance	Comfort	17 You will feel safe in your transactions with Company X.
Tangible	Features	18. Company X will provide a portrait setting in the church that is visually appealing.
*	Features	19. Company X will provide high quality electronic video image proofing.
*	Features	20. Company X will provide high quality photographs
* +	Features	21 Company X will produce a high quality church directory
* Items not in SERVQUAL + Items dropped after preliminary confirmatory factor analysis		

Notes

1. The word "dimensions" is similar in meaning to "components" which is a different conceptualization from that of "antecedents." Therefore, the word "factors" is used in this study to refer to those attributes that are relevant for service quality evaluations.

2. A practical reason for using overall measures of service quality (i.e., without reference to

any specific factors) is to capture customer evaluations of overall service quality directly. Such measures provide better feedback to managers about how customers view overall service and better prediction of behavioral intentions than any *computed* value of service quality based on several "dimensions."

3. As the research issues of interest were relevant to all types of services, the driving reasons for the selection of a particular industry were to ensure: (1) a large enough sample to conduct structural equations modeling on different subsets to allow comparisons between perceptions, computed disconfirmation, and measured disconfirmation, and (2) respondents who would be accessible before and after the service, and who would be willing to participate fully in the data collection process.

4. We viewed the churches as a form of nonprofit retail organization and hence viewed their concern with service quality as that of any other retail buyer and therefore relevant to researchers of retailing. In agreeing to be part of a longitudinal study, the motivation of these respondents was to share information to help improve the service to their ultimate "retail" customers, the church members.

5. Analysis comparing the two types of expectations measured in the study (i.e., based on experience or promises) is excluded for space considerations.

6. There is no difference between the computed disconfirmation data and the perceptions data in examining service quality, customer satisfaction, and behavioral intentions, as these variables are all measured *after* the service. The difference only arises when the factors related to service quality are included in the analysis, and expectations of the factors *before* the service are subtracted from perceptions of the factors *after* the service to determine *computed* disconfirmation.

7. As the same exact measures are used in the components model as in the antecedents model, the gammas are not affected by possible collinearity in any of the measures in this particular comparison.

8. See Note 6.

REFERENCES

- Andaleeb, Syed S. and Amiya K. Basu (1994). "Technical Complexity and Consumer Knowledge as Moderators of Service Quality Evaluation in the Automobile Service Industry," *Journal of Retailing*, **70** (4): 367–381.
- Anderson, Eugene W. and Mary W. Sullivan (1993). "The Antecedents and Consequences of Customer Satisfaction for Firms," *Marketing Science*, **12** (2): 125–143.
- Babakus, Emin and Gregory W. Boller (1992). "Empirical Assessment of SERVQUAL Scale," *Journal of Business Research*, **24**: 253–268.
- Babakus, Emin and W. Glynn Mangold (1992). "Adapting the SERVQUAL Scale to Hospital Services: An Empirical Investigation," *Health Service Research*, **26** (6): 767–780.
- Bansal, Harvir S. and Shirley Taylor (1997). "Investigating the Relationship Between Service Quality, Satisfaction and Switching Intentions." Pp. 304–313 in Elizabeth J. Wilson and Joseph C. Hair (Eds.), *Developments in Marketing Science*. Coral Gables, FL: Academy of Marketing Science.
- Bitner, Mary J. (1990). "Evaluating Service Encounters: The Effects of Physical Surroundings and Employee Responses," *Journal of Marketing*, **54** (April): 69–82.
- Bolton, Ruth and James H. Drew (1991a). "A Multistage Model of Customers' Assessments of Service Quality and Value," *Journal of Consumer Research*, **17** (March): 375–384.

- Bolton, Ruth and James H. Drew (1991b). "A Longitudinal Analysis of the Impact of Service Changes on Customer Attitudes," *Journal of Marketing*, **55** (January): 1–9.
- Boulding, William, Ajay Kalra, Richard Staelin, and Valarie A. Zeithaml (1993). "A Dynamic Process Model of Service Quality: From Expectations to Behavioral Intentions," *Journal of Marketing Research*, **30** (February): 7–27.
- Brown, Tom J., Gilbert A. Churchill, Jr, and J. Paul Peter (1993). "Improving the Measurement of Service Quality," *Journal of Retailing*, **69** (Spring): 127–139.
- Carman, James M. (1990). "Consumer Perceptions of Service Quality: An Assessment of the SERVQUAL Dimensions," *Journal of Retailing*, **66** (Spring): 33–55.
- Cronin, J. Joseph and Steven A. Taylor (1992). "Measuring Service Quality: A Reexamination and Extension," *Journal of Marketing*, **56** (July): 55–68.
- Cronin, J. Joseph and Steven A. Taylor (1994). "SERVPERF Versus SERVQUAL: Reconciling Performance-Based and Perceptions-Minus-Expectations Measurement of Service Quality," *Journal of Marketing*, **58** (January): 125–131.
- Dabholkar, Pratibha A. (1993). "Customer Satisfaction and Service Quality: Two Constructs or One?" Pp. 10–18 in David W. Cravens and Peter R. Dickson (Eds.), *Enhancing Knowledge Development in Marketing*. Chicago, IL: American Marketing Association.
- Dabholkar, Pratibha A. (1995a). "The Convergence of Customer Satisfaction and Service Quality Evaluations with Increasing Customer Patronage," *Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior*, **8**: 32–43.
- Dabholkar, Pratibha A. (1995b). "A Contingency Framework for Predicting Causality between Customer Satisfaction and Service Quality." Pp. 101–108 in Frank Kardes and Mita Suján (Eds.), *Advances in Consumer Research*, Vol. 22. Provo, UT: Association for Consumer Research.
- Dabholkar, Pratibha A., Dayle I. Thorpe, and Joseph O. Rentz (1996). "A Measure of Service Quality for Retail Stores: Scale Development and Validation," *Journal of the Academy of Marketing Science*, **24** (1): 3–16.
- Finn, David W., and Charles W. Lamb (1991). "An Evaluation of the SERVQUAL Scales in a Retailing Setting." Pp. 480–493 in Rebecca Holman and Michael R. Solomon (Eds.), *Advances in Consumer Research*, Vol. 18. Provo, UT: Association for Consumer Research.
- Fishbein, Martin and Icek Ajzen (1975). *Belief, Attitude, Intention, Behavior: An Introduction to Theory and Research*. Reading, MA: Addison-Wesley Publishing Company.
- Gotlieb, Jerry B., Dhruv Grewal, and Stephen W. Brown (1994). "Consumer Satisfaction and Perceived Quality: Complementary or Divergent Constructs?," *Journal of Applied Psychology*, **79** (6): 875–885.
- Grönroos, Christian (1978). "A Service-Oriented Approach to Marketing of Services," *European Journal of Marketing*, **12** (8): 588–601.
- Grönroos, Christian (1984). "A Service Quality Model and Its Marketing Implications," *European Journal of Marketing*, **18** (4): 36–44.
- Iacobucci, Dawn, Kent A. Grayson, and Amy L. Ostrom (1994). "The Calculus of Service Quality and Customer Satisfaction: Theoretical and Empirical Differentiation and Integration." Pp. 1–67 in Teresa A. Swartz, David H. Bowen, and Stephen W. Brown (Eds.), *Advances in Services Marketing and Management*, Vol. 3. Greenwich, CT: JAI Press.
- Iacobucci, Dawn, Amy L. Ostrom, and Kent A. Grayson (1995). "Distinguishing Service Quality and Customer Satisfaction: The Voice of the Consumer," *Journal of Consumer Psychology*, (3), 277–303.
- Jaccard, James and Choi K. Wan (1996). *LISREL Approaches to Interaction Effects in Multiple Regression*. Thousand Oaks, CA: Sage Publications.
- Jöreskog, Karl G. and Sörbom, Dag (1993). *LISREL8 User's Reference Guide*. Chicago, IL: Scientific Software.

- Kline, Rex B. (1998). *Principles and Practice of Structural Equations Modeling*. New York: Guilford Press.
- Lewis, Robert C. and Bernard H. Booms (1983). "The Marketing Aspects of Service Quality." Pp. 99–107 in Leonard L. Berry, G. Lynn Shostack, and G. Upah (Eds.), *Emerging Perspectives on Services Marketing*. Chicago, IL: American Marketing Association.
- Maddala, G. S. (1977). *Econometrics*. New York: McGraw-Hill.
- Miller, John A. (1976). "Exploring Some Alternative Measures of Consumer Satisfaction." Pp. 661–664 in Kenneth L. Bernhardt (Ed.), *Marketing: 1776–1976 and Beyond*. Chicago, IL: American Marketing Association.
- Mittal, Banwari and Walfried M. Lassar (1996). "The Role of Personalization in Service Encounters," *Journal of Retailing*, **72** (1): 95–109.
- Nunnally, Jum C. and Ira H. Bernstein (1994). *Psychometric Theory*, 3rd edition. New York: McGraw-Hill.
- Oliver, Richard L. (1981). "Measurement and Evaluation of Satisfaction Processes in Retail Settings," *Journal of Retailing*, **57** (Fall): 25–48.
- Oliver, Richard L. (1993). "A Conceptual Model of Service Quality and Service Satisfaction: Compatible Goals, Different Concepts." Pp. 65–85 in Teresa A. Swartz, David E. Bowen, and Stephen W. Brown (Eds.), *Advances in Marketing and Management*. Greenwich, CT: JAI Press, Inc.
- Parasuraman, A., Leonard L. Berry, and Valarie A. Zeithaml (1991). "Refinement and Reassessment of the SERVQUAL Scale," *Journal of Retailing*, **67** (Winter): 420–450.
- Parasuraman, A., Leonard L. Berry, and Valarie A. Zeithaml (1993). "More on Improving Service Quality Measurement," *Journal of Retailing*, **69** (Spring): 140–147.
- Parasuraman, A., Valarie A. Zeithaml, and Leonard L. Berry (1988). "SERVQUAL: A Multiple-Item Scale for Measuring Consumer Perceptions of Service Quality," *Journal of Retailing*, **64** (Spring): 12–40.
- Parasuraman, A., Valarie A. Zeithaml, and Leonard L. Berry (1994). "Reassessment of Expectations as a Comparison standard for Measuring Service Quality: Implications for Future Research," *Journal of Marketing*, **58** (January): 111–124.
- Peter, J. Paul, Gilbert A. Churchill, and Tom J. Brown (1993). "Caution in the Use of Difference Scores in Consumer Research," *Journal of Consumer Research*, **19** (March): 655–662.
- Spreng, Richard A. and Robert D. Mackoy (1996). "An Empirical Examination of a Model of Perceived Service Quality and Satisfaction," *Journal of Retailing*, **72** (2): 201–214.
- Spreng, Richard A. and Richard W. Olshavsky (1992). "A Desires-as-Standard Model of Consumer Satisfaction: Implications for Measuring Satisfaction," *Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior*, Vol. 5. 45–54.
- Spreng, Richard A. and A. K. Singh (1993). "An Empirical Assessment of the SERVQUAL Scale and the Relationship Between Service Quality and Satisfaction." Pp. 1–6 in David W. Cravens and Peter R. Dickson (Eds.), *Enhancing Knowledge Development in Marketing*. Chicago, IL: American Marketing Association.
- Taylor, Steven A. and Thomas L. Baker (1994). "An Assessment of the Relationship Between Service Quality and Customer Satisfaction in the Formation of Consumers' Purchase Intentions," *Journal of Retailing*, **70** (2): 163–178.
- Teas, R. Kenneth. (1993). "Expectations, Performance Evaluation, and Consumers' Perceptions of Quality," *Journal of Marketing*, **57** (October): 18–34.
- Teas, R. Kenneth. (1994). "Expectations as a Comparison Standard in Measuring Service Quality: An Assessment of a Reassessment." *Journal of Marketing*, **58** (January): 132–139.
- Westbrook, Robert A. (1980). "A Rating Scale for Measuring Product/Service Satisfaction," *Journal of Marketing*, **44** (Fall): 68–72.

- Westbrook, Robert A. (1983). "Consumer Satisfaction and the Phenomenology of Emotions During Automobile Ownership Experiences." Pp. 2-9 in Ralph L. Day and H. Keith Hunt (Eds.), *International Fare in Consumer Satisfaction and Complaining Behavior*. Bloomington, IN: Indiana University.
- Westbrook, Robert A. and Richard L. Oliver (1981). "Developing Better Measures of Consumer Satisfaction: Some Preliminary Results." Pp. 94-99 in Kent B. Monroe (Ed.), *Advances in Consumer Research*, Vol. 8. Arlington, VA: Association for Consumer Research.
- Westbrook, Robert A. and Richard L. Oliver (1991). "The Dimensionality of Consumption Emotion Patterns and Consumer Satisfaction," *Journal of Consumer Research*, **18** (June): 84-91.
- Zeithaml, Valarie A., Leonard L. Berry, and A. Parasuraman (1996). "The Behavioral Consequences of Service Quality," *Journal of Marketing*, **60** (April): 31-46.