EVALUATING THE IMPACT OF SERVICE QUALITY DISCONFIRMATIONS ON STORE CHOICE

James B. Wiley, University of Alberta, Edmonton
Paul D. Larson, University of Alberta, Edmonton

ABSTRACT

This paper describes an application of standard choice model methods to fast food restaurant choice leading to confirmation of the role of perceived disconfirmation in these choices. The research was done in conjunction with an on-going project to develop an instrument that measures retail service quality. A distinguishing characteristic of the instrument is that it will focus on attributes that are actionable and of managerial significance. Actionable attributes can be manipulated by the firm. For example, waiting lines and product variety are actionable attributes. Managerially significant attributes are sufficiently valued by customers so that they will change their behavior in response to the actions of firms. This paper shows how choice model methods can be used to evaluate the managerial significance of actionable attributes.

INTRODUCTION

Firms can increase consumer satisfaction and improve their competitive position by reducing price, which increases consumers' ability to buy, or by increasing quality, which increases customers' willingness to buy. A key notion of the marketing concept and the emerging concept of market orientation (Kohli and Jaworski, 1990; Narver and Slater, 1990) is that firms can increase willingness to buy by augmenting the product, for example by improving service quality. Erevelles and Leavitt (1992) suggest that consumer satisfaction is the central element of the marketing concept.

The objective of the on-going research is to develop measures of quality specifically tailored to the retail sector. The long-term objectives of the research are to develop measurement procedures that have the following characteristics:

- Capable of *intra* and *inter* firm comparisons,

- Capable of cross national (cross border) comparisons,

- Capable of cross linguistic (cultural) comparisons,

- Capable of cross time comparisons.

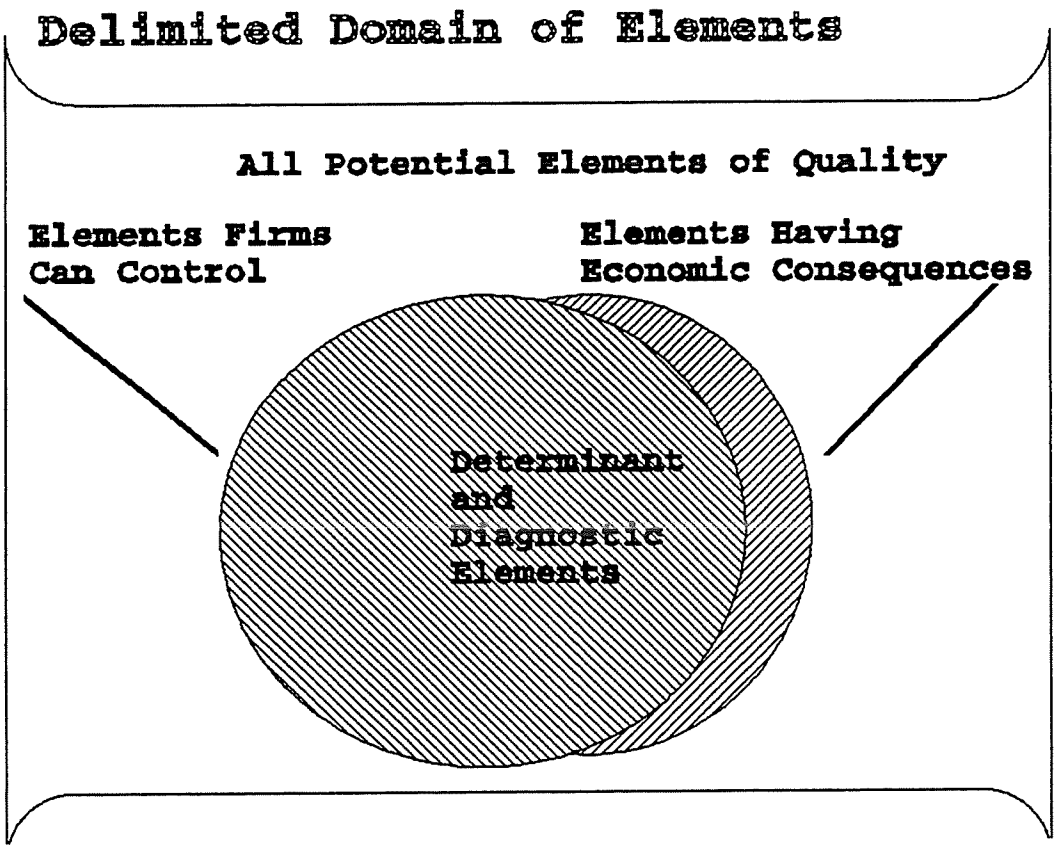
- Preferably could be administered by mail or telephone.

We focus on service quality attributes that are actionable (can be influenced by management), determinant (there is variation between firms), and of managerial significance (valued by consumers). This focus derives from the following premises: There are at least two "constituencies" for service quality research, 1) The firm's customers and 2) decision-makers within the firm that use the information.

If service quality measures are to have impact on retailer performance, retailers must make use of them. For this to occur, there must be motivation for their use; that is, there must be economic consequences to the retailer (benefits gained or costs avoided). The *antecedents* of *economic consequences* are *actions taken by customers*. Decision makers will be motivated to use measures to the extent their use influences consumer actions. As a result of these premises, attention is restricted to a subset of all attributes which potentially may be related to quality. This is the subset illustrated in Figure 1. In other words, of all attributes that conceivably might have something to do with quality, we restrict ourselves to the subset consisting of the intersection of those which a) can be manipulated by retailers and b) will have economic consequences as a result of customers' actions.

A methodology for evaluating the impact of disconfirmations on store choice is described and tested in the present paper. Store choice is operationalized in the form of "choice experiments." The relevance of choice experiments to the present project is evident because an objective of the project is to develop an instrument which focuses on those aspects of service quality which can impact the behavior of customers, and store choice is preeminent among behaviors of

Figure 1



interest.

Since the late 1970s marketers have modeled choice experiments using logit regression. This function has the form:

$$P(i|k) = \frac{e^{V_{i0} + \sum_{w=1}^{natt} v_w x_{iw}}}{\sum_{j=1}^{m} e^{V_{j0} + \sum_{w=1}^{natt} v_w x_{jw}}} \quad (1)$$

where k is the choice set, v_{i0} is "utility" of brand i based on aggregate choice data, v_w is weight attached to attribute w, x_{iw} is the perceptual judgment associated with the attribute, and *natt* is the number of attributes.

A key issue in the use of Equation 1 concerns the way the "x's" are defined and measured. Two approaches have been suggested: (1) the so-called "gaps" formulation -- used in the SERVQUAL instrument (Parasuraman et. al. 1988) for example -- and (2) the perceived disconfirmations approach. For reasons discussed in the following section, the "x's" are operationalized in the present paper as perceived disconfirmations, i.e., *direct judgements of whether the retail store exceeds, or falls short of, expectations on specific attributes.*

In this preliminary study, three issues are evaluated. First, does knowledge of individuals' perceived disconfirmations of the service quality provided by a firm improve the ability to predict behavior relative to the ability to predict based on aggregate behavior, i.e., market shares? Richard and Allaway (1993) address this question in the

context of the home delivery pizza market and provide some evidence that service quality does impact choice. Second, does weighting of perceived disconfirmations by stated importance of the attribute improve prediction over what can be achieved based on perceived disconfirmations alone? Third, are choices in choice tasks consistent in the sense that different choice tasks provide equivalent parameter values, i.e., are they reliable.

BACKGROUND

There is considerable agreement among scholars regarding many aspects of service quality. There is consensus that it should 1) be defined from the perspective of the consumer, it is their *perceptions* of performance that are important, 2) these perceptions are on multiple attributes, i.e., they are not unidimensional, and 3) the state of satisfaction or dissatisfaction with service quality is a reaction to a comparison process (Oliver 1979). The potential advantage of a multi-attribute conceptualization of "service quality" over a simpler "overall" judgment is that knowledge of the structure of the evaluation can lead to diagnosis of the strengths and weaknesses of the firm's performance.

The most commonly encountered instruments which operationalize the above three areas of consensus are based on the notion of gaps between the perceived performance on relevant attributes and the recipient's expectations for what the performance should be. This view of service quality matches the "disconfirmation of expectations" model of consumer satisfaction (Rogers, Peyton and Berl, 1992). In its most general form, a gaps formulation may be conceptualized in terms of the following compensatory model:

$$S/D_{jk} = \Sigma W_{ik} [P_{ijk} - St_{ik}], \text{ where} \quad (2)$$

i = the attribute or dimension of performance, j = the firm, and k = the respondent. Then, S/D_{jk} = respondent k's satisfaction score with firm j, P_{ijk} = respondent k's perception of the performance of firm j on attribute i, and St_{ik} is the respondent's standard (or norm) for performance on attribute i.

The weight component has not been widely studied in satisfaction research and has proven problematic in multi-attribute model research (Wilkie and Pessimier 1973; Wiley, 1977). However, there is revived interest in the component in conjunction with "importance-performance analysis" (Martilla and James 1977; and Burns 1986). The weight component also potentially provides for individual differences in the compensatory rate of substitution between attributes. Since it can be anticipated that there may be interest in the importance/weighting component in service quality research, the component is included in the present study.

The SERVQUAL instrument currently is the most widely one used based on a gaps formulation (Parasuraman et. al. 1988). Finn and Lamb (1991) tested SERVQUAL in a retail setting. Using LISREL they did not find a structure consistent with those originally proposed by its developers. In a health care setting, Babakus and Mangold (1989) were not able to reproduce the Parasuraman et. al. results. Similar results by Koelemeijer (1991) lead her to conclude that SERVQUAL does not possess construct validity in a retail store setting. However, SERVQUAL was not specifically developed for retail applications.

In summary, while there is wide consensus on the general form a measure of retail quality should take, there currently does not appear to be available a valid and reliable measure of quality for the retail sector. This may be because retail establishments draw from at least three domains in their efforts to provides value to their customers; goods and merchandise, services, and utilities specifically created by institutions in the retail sector. None of the SERVQUAL items, for example, pertain to goods or merchandise issues.

Direct Judgments versus Subjective Disconfirmations

The two most controversial components of the model are the nature of the standards and the way the gap between perceptions and standards should be operationalized. A widely used approach for operationalizing the gap between perceptions and standards is a self-explicated approach where respondents give direct judgments of the two constructs and the gap is computed by subtracting

one judgment from the other. Several groups of authors have identified problems with this approach. Wall and Payne (1973) provide results which indicate that calculating difference scores masks the true relationship between variables, even when the true relationship has a form such as (2). Peterson and Wilson (1992) provide related arguments specifically targeted to consumer satisfaction research. Koelemeijer (1991) summarizes four major methodological shortcomings of the approach:

- The expectations likely are derived from previous experience. Hence, differences between expectations and perceptions in many cases may be small;

- Ceiling or floor effects might occur causing difficulty in capturing the difference between expectations and perceptions. Since the expectations likely are based on experiences, there is little chance that perceptions can be above expectations and some likelihood that perceptions will be close to expectations;

- Lack of reliability inherent in using any difference scores. Reliability decreases as the variance of either measure increases, and as the correlation between the two measures increases;

- Perceived service quality, measured as inferred disconfirmation (P-E) cannot be used as a predictor of behavior together with expectations and perceptions, as the model will be overspecified.

Numerous alternative conceptualizations of the "standards" have been proposed. For example, Spreng and Olshavsky (1992) develop a "desires-as-standard" model of consumer satisfaction. In a recent article, Woodruff, et. al. (1991) identify several conceptualizations for standards that have been proposed in this journal.

In the present work we seek to avoid the methodological problems of separate perception and standard judgments by asking the respondent to make a direct disconfirmation judgement of the difference within the squared brackets of (2). Rogers, Peyton and Berl (1992) refer to this approach as a "perceived disconfirmation" (as opposed to a derived disconfirmation) measure. The perceived disconfirmation approach offers the potential benefit of avoiding some of the controversy surrounding the way standards should be defined by enabling the respondent to use his or her own standard in arriving at their judgment. Tse and Wilton (1988) and others (Anderson, 1973; Churchill and Surprenant, 1982) have shown the approach gives good results.

OUTLINE OF THE STUDY

One hundred and three students participated in an omnibus study in which they responded to a variety of questions pertaining to service quality issues. Restaurants located in a mall within the university were the outlets selected for study, since virtually all of the respondents were familiar with them. Fifty students received a version of the questionnaire which included two choice tasks. Forty-seven of them provided useable data for this illustrative study.

The choice tasks asked respondents to imagine that they were hungry and wanted to get something to eat at the mall. The scenario continued:

> You go to the mall and it turns out not all of the outlets are open. In fact, only the ones indicated in each set are open.
>
> For each set, circle the outlet where you would eat. If you would not eat at any of the outlets, circle "none."
>
> "None" means that you would rather go hungry than eat at one of the indicated available outlets.

One of the choice tasks contained eight choice sets, the other contained seven choice sets. For example, the first three choice sets of the first task were:

Set 1	Taco Time	Academy Pizza	A&W	none
Set 2	A&W	none		
Set 3	Louie's	Academy Pizza	A&W	none

Table 1
Confirmation/Disconfirmation Question

The following set of statements relate to your feelings about four fast food outlets in the HUB Mall:

A&W	Louie's Submarine
Academy Pizza	Taco Time

For each statement, please show the extent to which you believe the outlet exceeds, equals, or falls short of meeting your expectations on each attribute.

Once again, circling a 7 means that you believe that the outlet exceeds your expectations for fastfood outlets in general, and circling a 1 means that it falls short of your expectations for fastfood outlets. Circle 4 if the outlet exactly meets your expectations for fastfood outlets on that attribute.

There are no right or wrong answers -- all we are interested in is a number that best shows your perceptions of the four outlets.

What are your perceptions about A&W on each of these dimensions?

The visual appeal of A&W's physical facilities.
Greatly exceeds my expectations ... Greatly falls short
7 6 5 4 3 2 1

The variety of items that A&W offers.
Greatly exceeds my expectations ... Greatly falls short
7 6 5 4 3 2 1

The time A&W employees spend chatting with the customers.
Greatly exceeds my expectations ... Greatly falls short
7 6 5 4 3 2 1

A&W's range of prices within each food category.
Greatly exceeds my expectations ... Greatly falls short
7 6 5 4 3 2 1

The length of A&W's waiting lines.
Greatly exceeds my expectations ... Greatly falls short
7 6 5 4 3 2 1

The sit down facilities at A&W.
Greatly exceeds my expectations ... Greatly falls short
7 6 5 4 3 2 1

The ordering of the outlets were randomized in the sets. The first task occurred at the beginning of the questionnaire, the second occurred at the end of the questionnaire. In addition, the respondents indicated perceptions on six attributes using the confirmation/disconfirmation format illustrated in Table 1 and indicated the importance they attach to the attributes using a 7-point "very important" (7) to "very unimportant" (1) rating scale.

A set of six retail service quality attributes were selected for study (see Table 1). These were chosen to represent the domain of relevant attributes and to include elements of product quality (e.g., food price as a proxy), service quality (e.g., waiting lines), and merchandising (e.g., food item variety). Though all relevant, these six were also expected to vary across respondents in terms of importance.

DESCRIPTION OF THE METHODOLOGY

The criteria for estimating the parameters of (1) maximizes the log likelihood ratio LL(b), where b is the number of parameters in the model. A test of the hypothesis that adding (or deleting) direct judgements of disconfirmations to a model that includes only the v_{i0} parameters can be formulated as:

test statistic = -2(LL(model with disconfirmations) - LL(model without disconfirmations)) (3)

which asymptotically is distributed as chi-square with degrees of freedom equal to the number of parameters in the less restricted model less the number of parameters in the more restricted model. This statistic is used to evaluate three issues in the following study: a) whether addition of individual perceptions adds to the ability to predict choices based solely on the v_{i0} parameters, b) whether weighting perceptions by importance adds to the ability to predict choices over what can be accomplished with perception data alone, and c) whether choices are reliable in the sense that v_{io}'s are consistent across choice tasks.

RESULTS

There are three questions evaluated in this paper. From a methodological standpoint, the logical first question is whether choices in choice tasks are consistent in the sense that different choice tasks provide equivalent parameter values. Reliability on choices is a necessary, but not sufficient, condition for using choice experiments to answer the second question. The second and substantive question is whether knowledge of individuals' perceptions of service quality improve the ability to predict behavior, relative to the

ability to predict based solely on market shares. Finally, the third question asks whether weighting of service quality perceptions by importance improves prediction over what can be achieved based on perceptions alone.

Are Choices Reliable?

Are choices in choice experiments reliable in the sense that parameter estimates based on different choice tasks are equivalent. A straightforward way of answering this question is to compare the likelihood ratio of a model which has separate estimates for each set with the likelihood ratio for a model which constrains the two sets of parameters to be equal. The results of these two analyses are presented in Table 2. The likelihood ratio of the unconstrained model is -653.87 with 8 parameters. The likelihood ratio for the restricted model is -656.01 with 4 parameters. The resulting test statistic is $\chi^2 = 2.24$ with 4 degrees-of-freedom is not significant, implying that the parameters estimated from the two choice tasks are equal and that choices are reliable. The constrained and unconstrained estimates are presented in Table 2.

Do Disconfirmations Help?

Does adding confirmation/disconfirmation perceptions improve the ability to predict choices? While an affirmative answer to this question is not a sufficient condition for establishing that a *change* in the disconfirmations would lead to a *change* in behavior, it likely is a necessary condition. That is, if knowledge of disconfirmations cannot improve prediction of choice, it is unlikely that knowledge of *changes* in disconfirmations will predicts *change* in choice. The statistical results relevant to this issue are presented in Table 3. The likelihood ratio value for a model with the v_{i0} parameters alone is - 653.87 with 8 parameters (separate v_{i0}'s for the four outlets based on each choice task). The likelihood ratio value for a model with the v_{i0} parameters and disconfirmation judgments is -532.17 with 15 parameters. The resulting test statistic is $\chi^2 = 101.70$ with 7 degrees-of-freedom. The improvement is significant beyond the .005 level. Using only the aggregate choice data 45% correct prediction of choices is achieved (against a naive base of 1/5, or 20%). Including knowledge of the disconfirmations increased correct predictions to 56%.

Table 2
Log Likelihood and Parameter Estimates For Aggregate v_{i0}'s

Unrestricted Model

LL(B) = -653.87372

	Parameter Estimate	SE of Parameter	Asymptotic t-Stat	Pr(Z>|t|)
A&W-1	1.1029460	.20038010	5.504	.0000
AP-1	1.5645210	.19619674	7.974	.0000
LS-1	1.4219216	.19723002	7.209	.0000
TT-1	1.3794199	.19759112	6.981	.0000
A&W-2	1.6549310	.25772392	6.421	.0000
AP-2	2.1214111	.25158935	8.432	.0000
LS-2	1.9841911	.25271257	7.852	.0000
TT-2	1.8780249	.25394843	7.395	.0000

Restricted Model

LL(B) = -656.01475

	Parameter Estimate	SE of Parameter	Asymptotic t-Stat	Pr(Z>|t|)
A&W-1	1.2983098	<- Constrained to equal A&W-2 !		
AP-1	1.7684219	<- Constrained to equal AP-2 !		
LS-1	1.6275906	<- Constrained to equal LS-2 !		
TT-1	1.5494158	<- Constrained to equal TT-2 !		
A&W-2	1.2983098	.15364659	8.450	.0000
AP-2	1.7684219	.15005420	11.785	.0000
LS-2	1.6275906	.15077724	10.795	.0000
TT-2	1.5494158	.15130727	10.240	.0000

Does Weighting of Disconfirmations Help?

In order to evaluate this question, the disconfirmation judgments were multiplied by the individual's stated importance weights before including them in (1). Statistical results relevant to this issue are presented in Table 3. The likelihood ratio value for a model with the v_{i0} parameters alone is the same as above, -653.87 with 8 parameters. The likelihood ratio value for a model with the v_{i0} parameters and the weighted disconfirmation judgments is -517.07 with 15 parameters. The resulting test statistic is χ^2 =

136.80 with 7 degrees-of-freedom. Again, the improvement over the aggregate model is statistically significant at a level beyond the .005 level and there is improvement over the unweighted version. However, the substantive improvement is slight, 57.5 % correct prediction of choices is achieved against 56% correct without weighting.

Table 3
Model Comparisons

MODEL	LL	D.F.	χ^2	D.F	% CORRECT
Naive					20%
(1) Both Tasks	-653	8			45%
(1) Combined	-656	4	2.24	4	
(2) Unweighted	-532	15	101	7	56%
(2) Weighted	-517	15	136	7	57%

These findings are consistent with the findings in the multi-attribute modeling literature. For example, Lehmann (1971) found inclusion of weights contributed little to the ability of a multiattribute model to predict television show preference. One reason weights may contribute little to predictive power is that responses to the other model component -- beliefs in multi-attribute models or perceived disconfirmations in a expectations-performance model -- may implicitly incorporate the importance information. For example, respondents may give a wider range of responses or respond with greater reliability to important attributes. Work by Bass and Wilkie (1973) using a multi-attribute model indicates that normalizing belief measures before inclusion in the model increased the ability to predict a criterion. This result implies some correlation between the weight of the attribute and the variance of responses on it. Also, the relationship between level and value is conceived to be monotonic in virtually all multi-attribute formulations, including the relationship between performance and satisfaction in expectations-performance multiattribute models. It is well known that linear models are extremely robust when the functions linking independent and dependent variables are monotonic.

CONCLUSION

A procedure for evaluating the impact on disconfirmations on choice and other behaviors is illustrated. The results indicate that choice modeling has promise for selecting attributes that are sufficiently valued by customers to influence their behavior. Specifically, results of the study indicate that choices in choice experiments are reliable, that including information about subjective disconfirmations increases the ability to predict store choice beyond what can be achieved with aggregate information, such as market share, and weighting disconfirmations by importance offers improvement -- but very modest improvement -- in the ability to predict choices beyond the level achievable without the information.

Although choices from hypothetical choice sets are used in this study, the methodology can accommodate actual choice. Furthermore, other behaviors of interest also can be viewed as choices: customers' decisions to voice complaints or compliments, "exit" the store, retaliate against the store, and so forth. A more challenging test would involve using disconfirmations and actual choices. A more challenging test yet would involve experimental manipulation of disconfirmations and observing the choices within experimental and control groups.The methodology outlined can be used in each of these tests. The unresolved issues mainly are how to generate the appropriate data.

REFERENCES

Anderson, Rolf E. (1973), "Consumer Dissatisfaction: The Effect of Disconfirmed Expectancy on Perceived Product Performance," *Journal of Marketing Research*, 10, (February), 38-44.

Bass, F. M. and W. L. Wilkie (1973), "A Comparative Analysis of Attitudinal Predictions of Brand Preference," *Journal of Marketing Research*, 10, 262-69.

Babakus, E. and W. G. Mangold (1989), "Adapting the 'SERVQUAL' scale to Health Care Environment," in P. Bloom, R. Winer, H. Kassarjian, D. Scammon, B. Weitz, R. Speckman, V. Mahajan and M. Levy (eds.) *Enhancing Knowledge Development in Marketing*, Chicago: American Marketing Association, 195.

Burns, A. C. (1986), "Generating Marketing Strategy Priorities Based on Relative Competitive Position,"

Journal of Consumer Marketing, 3, 4, (Fall), 49-59.

Churchill, Gilbert A. and Carol Surprenant (1982), "An Investigation Into the Determinants of Satisfaction Research," *Journal of Marketing Research*, 19, (November), 491-504.

Erevelles, Sunil and Clark Leavitt (1992), "A Comparison of Current Models of Consumer Satisfaction/ Dissatisfaction," *Journal of Consumer Satisfaction/ Dissatisfaction & Complaining Behavior*, 5, 104-114.

Finn, D. W. and C. W. Lamb Jr. (1991), "An Evaluation of the SERVQUAL scales in a retail setting," *Advances in Consumer Research*, R. H. Holman and M. R. Solomon (eds.), 18, 483-490.

Koelemeijer, K. (1991), "Perceived Customer Service Quality: Issues on Theory and Measurement," proceedings *Sixth World Conference on Research in the Distributive Trades*, Sponsored by the Dutch Ministry of Economic Affairs, Services Directorate, 68-76.

Kohli, Ajay K. and Bernard J. Jaworski (1990), "Market Orientation: The Construct, Research Propositions, Managerial Implications," *Journal of Marketing*, 54, (April), 1-18.

Lehmann, D. R. (1971), "Television Show Preference: Application of a Choice Model," *Journal of Marketing Research*, 8, 47-55.

Martilla, J. A. and J. C. James (1977), "Importance-Performance Analysis," *Journal of Marketing*, 41, 1 (January), 77-79.

Narver John C. and Stanley F. Slater (1990), "The Effect of a Market Orientation on Business Profitability," *Journal of Marketing*, 54, (October), 20-35.

Oliver, R. L. (1979), "Product Dissatisfaction as a Function of Prior Expectation and Subsequent Disconfirmation: New Evidence," in *New Dimensions of Consumer Satisfaction and Complaining Behavior*, R. L. Day and H. K. Hunt, eds. Bloomington:Indiana University, 66-71.

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.

Peterson, R. A and W. R. Wilson (1992), "Measuring Customer Satisfaction: Fact and Artifact," *Journal of the Academy of Marketing Science*, 20, 61-71.

Richard, Michael D. and Arthur W. Allaway (1993) "Service Quality Attributes and Choice Behavior," *Journal of Services Marketing*, 7(1), 59-68.

Rogers, Hudson P., Reginald M. Peyton and Robert L. Berl (1992), "Measurement and Evaluation of Satisfaction Processes in a Dyadic Setting," *Journal of Consumer Satisfaction/Dissatisfaction & Complaining Behavior*, 5, 12-23.

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 & Complaining Behavior*, 5, 45-54.

Tse, David K. and Peter C. Wilton (1988), "Models of Consumer Satisfaction Formation: An Extension," *Journal of Marketing Research*, 25, (May), 204-212.

Wall, T. B. and R. Payne (1973), "Are Deficiency Scores Deficient?" *Journal of Applied Psychology*, 58, 322-326.

Wiley, James B. (1977), "Stability of Inferred 'Importance' and 'Ideal Points' in Additive Models", in H. C. Schneider, (ed.), *Proceedings American Institute for Decision Science*, San Francisco, IL, 192-94.

Wilkie, W. L. and E. A. Pessimier (1973), "Issues in Marketing's Use of Multiattribute Attitude Models," *Journal of Marketing Research*, 10, 428-41.

Woodruff, R. B., D. S. Clemons, D. W. Schumann, S. F. Gardial and M. J. Burns (1991), "The Standards Issue in CS/D Research: A Historical Perspective," *Journal of Consumer Satisfaction/Dissatisfaction & Complaining Behavior*, 4, 103-109.

ACKNOWLEDGMENTS

Research for this paper was supported by a grant from the Social Science and Humanities Research Council of Canada. The authors wish to acknowledge the suggestions of three anonymous reviewers.

Send correspondence regarding this article to:
James B. Wiley
Department of Marketing and Economic Analysis
University of Alberta, Edmonton
Alberta, T6G 2R6 CANADA