Published In: Drew, James H. and Ruth N. Bolton, "The Structure of Customer Satisfaction: Effects of Survey Measurement,"Journal of Customer Satisfaction, Dissatisfaction and Complaining Behavior, 4, 1991, 21-30.

THE STRUCTURE OF CUSTOMER SATISFACTION: EFFECTS OF SURVEY MEASUREMENT

James H. Drew, GTE Laboratories Incorporated
Ruth N. Bolton, GTE Laboratories Incorporated

INTRODUCTION

Measurement errors and questionnaire effects in survey data can have substantial impact on estimates of statistical relationships. Although the customer satisfaction literature contains some discussion of scaling issues (Westbrook and Oliver 1981), statistical models of customer satisfaction have not explicitly recognized measurement issues. This paper enhances the traditional model of customer satisfaction and attitude formation by explicitly incorporating these effects, thus linking the three separate research streams whose bases are in customer satisfaction, structural equation modeling and in survey research. A structural equations/measurement error model is developed that distinguishes between observed variables from a satisfaction survey and underlying constructs. The model explicitly incorporates measurement error and questionnaire effects where response correlations exist among proximate survey items or those with similar wordings or response categories. The model is estimated with a telephone service survey in which identical overall quality items are included at both the beginning and end of the questionnaire, with a series of attribute rating questions in the middle.

BACKGROUND

Customer Satisfaction and Attitude Models

Customer satisfaction or dissatisfaction (CS/D) is considered to be a function of the disconfirmation arising from discrepancies between prior expectations and actual performance (Cardozo 1965; Oliver 1980; Olshavsky and Miller 1972; Olson and Dover 1976). For some products and services, expectations and perceptions of performance levels affect customer satisfaction directly, as well as indirectly via disconfirmation (Churchill and Surprenant 1982; Tse and Wilton 1988). In addition, recent studies suggests that CS/D depends on customers' causal attributions (Oliver and DeSarbo 1988) and their perceptions of equity (Oliver and Swan 1989). In summary, CS/D is considered to depend on the following constructs.

$$CS/D = g\ (EXP_{t0},\ PERF_{t1},\ DISC_{t1},\ EQUITY_{t1},\ ATTRIB_{t1}) \quad (1)$$

where:

EXP = a vector representing customers' expectations at time t0 about k product/service attributes

PERF = a vector representing the customer's perceptions at time t1 of actual product/service attribute performance levels;

DISC = the discrepancy between the customer's expectations and perceptions at time t1;

EQUITY = the customer's perceptions of the equity of the transaction; and,

ATTRIB = the customer's causal attributions at time t1 about the outcome of the transaction.

Based on adaptation level theory (Helson 1964), Oliver (1980, 1981) postulated that changes in attitude are dependent on the consumer's post-use satisfaction/dissatisfaction level. In symbols,

$$ATTITUDE_{t1} = f(ATTITUDE_{t0}, CS/D_{t1}). \quad (2)$$

In other words, satisfaction/dissatisfaction is an affect which acts as an intervening variable to modify a prior attitude into the current attitude.

Statistical Relationships

The CS/D literature has typically represented theoretical constructs by a single measure or the sum of several measures. However, the following studies provide evidence to suggest that a distinction should be made between the survey question, which is an observed variable that reflects measurement error, and the latent, theoretical construct.

Bitner (1990) estimated a model of the relationship between customer attitudes and dissatisfaction with data from a laboratory experiment about travel agency service. In her study, she operationalized each CS/D and attitude construct with a single measure or the sum of several measures, reporting coefficient alphas varying between 0.52 and 0.92.

Bolton and Drew (1991a, 1991b) estimated equation (2) for local telephone service. Following Oliver's (1989) notion that expectations for continuous services are not actively processed, they postulated that customer's attitudes about local telephone service were a function of disconfirmation and perceptions about performance levels only. This model was fit to both cross sectional and longitudinal panel data with good results. They operationalized each construct with a single questionnaire item. However, in their longitudinal study, Bolton and Drew (1991a) reported that each interview wave showed about half of all respondents changing their overall quality rating from the preceding interview -- even in serving areas in which no equipment or services were altered during that time period. Furthermore, correlations over successive interviewing waves ranged from 0.48 to 0.55. In the authors' experience with other panel studies, interview waves spaced from one to three months give correlation coefficients of about the same size.

Measurement Error and Questionnaire Effects

The statistics literature provides substantial evidence of the importance of specifying models that account for measurement error. (See Fuller 1987). When the intent is to calculate the effect of one variable on another through a

regression equation, ignored measurement error will tend to dilute the true strength of the effect. Simple regression coefficients will be attenuated and appear to be closer to zero than they really are. The more complex statistical relationships among CS/D constructs may be much stronger or weaker than the estimates obtained in empirical studies.

The survey research literature also suggests that CS/D models should account for questionnaire effects, as well as measurement error. Andrews (1984) found that a "typical" survey item, yielded 50-83% valid variance, 0-7% methods variance, and 14-18% residual variance. Over two thirds of the variation in measurement quality could be explained by survey design characteristics. For example, data quality increased with the number of items in the scale (at least up to 5-7) and and increased when questions appeared later in the questionnaire (between positions 26-100). Similarly, Churchill and Peter (1984) found that measure characteristics had a major influence on obtained reliability estimates. For example, reliability increased with the number of scale points per scale.

Oliver (1980) has suggested that CS/D researchers have ignored measurement error issues because they have assumed (incorrectly) that CS/D is an attitude and that research on attitude measurement is applicable. A notable exception is a study by Westbrook and Oliver (1981) that showed that Likert and semantic differential scales had higher reliabilities and convergent and discriminant validities than other satisfaction scales.

THEORETICAL AND PRACTICAL CONSIDERATIONS

Measurement Error

The classic formulation of statistical measurement error in an observed variable (Y) is modeled as a sum of y, whose variability is due only to sampling error, and u, which is measurement error:

$$Y = y + u.$$

It follows that u is the difference between the observed Y and the latent factor y for a given subject, and the relative size of its variance can be assessed by calculating the correlation coefficient between two successive observations Y_1 and Y_2 on the same subject. The farther the coefficient is from 1.0, the greater the relative size of the measurement error, and the greater the discrepancy between a variable we observe and an underlying factor.

Example. The consequences of ignoring measurement error can be illustrated by the following study. In 1988, service rating data were collected from 370 GTE customers in one geographic area for the month of February. In this survey, customers were questioned about their telephone service during the preceding thirty days and they were asked whether they had reported any service problems to the telephone company. Telephone repair records for the preceding 18 months were compiled for those interviewed customers as a forward record check on those customers citing a repair report. There was found to be a problem with both under- and over-reporting on this survey, with the preponderance of

reporting discrepancies occurring among those customers that did not acknowledge actual reports. Nevertheless, GTE was very interested in assessing the effect the making of a trouble report may have on the customer's ratings of overall quality. Let the overall quality rating for the ith customer be denoted by OVQ_i, the trouble report indicator from the survey be TR_i, and the trouble report indicator from company records by tr_i. Then the regression of OVQ on objective tr is estimated, by ordinary least squares, as

$$\underset{(0.0469)}{OVQ = 3.8287} \; - \; \underset{(0.1420)}{0.3537\ tr}$$

and the regression of OVQ on the errorful TR is estimated as

$$\underset{(0.0456)}{OVQ = 3.800} \; - \; \underset{(0.2121)}{0.2118\ TR}$$

The coefficient of tr in the first equation is significantly different from 0 (p=0.0132), while the coefficient of TR in the second equation is not. The manager who relies on the errorful second equation will mistakenly downplay the importance of trouble reports.

This calculation shows the effect of confusing the difference between an objective variable (actual trouble reports) and an indicator (reported trouble reports). With variables which are inherently subjective, such as attitudes and beliefs, a corresponding adjustment is to use structural equation/measurement error models exemplified by LISREL, or instrumental variable techniques. Drew and Bolton (1991b) use the latter approach to estimate a model for attitude change due to service improvements, where they found moderate differences in coefficients when instruments were unused and indicator/construct differences were ignored.

Questionnaire Effects

As discussed above, the survey research literature provides considerable evidence that statistical relationships among measures can result from survey design factors. A very important survey design factor for CS/D research is question order. In theory, relationships among measures (i.e., measurement models) and relationships among constructs (i.e., structural models) need not correspond. Indeed, causal direction for constructs and their corresponding measures may actually be opposite. For example, a prior attitude theoretically precedes a current attitude, but a questionnaire may elicit measures of these attitudes in chronologically backward order. (Many customer surveys do, in fact, measure the current attitude first, followed by an item comparing current to prior attitudes.)

Example. Drew (1987) analyzed a survey of residential telephone subscribers in which an initial overall quality item was followed by questions rating such telephone service attributes as local dial service, long distance service and billing service. In addition to the predictable result that overall quality was a function of these attributes, it was found that models for each of the attributes was significantly enhanced by the inclusion of the overall quality

indicator. For instance, local dial quality was found to be a function of the perceived frequencies of line static, dial tone problems, call blockages, cut-offs and overall quality. In this situation, there is evidence that the positional priority of overall quality prior to the attributes in the questionnaire produces a statistical effect despite the theoretical priority of the attributes over overall quality.

AN ENHANCED MODEL OF CUSTOMER ATTITUDES AND CS/D

Prior research has developed models of the relationships between customer attitudes, customer satisfaction/dissatisfaction, and their antecedents. These models have utilized a single measure (i.e., survey question) for each construct and have ignored measurement error and questionnaire design (i.e., methods) effects. This study extends these models to incorporate measurement error and questionnaire design effects. A unique feature of this study is the exploration of the effect of question order.

In the remainder of this section we investigate the following two issues.

- Is there is a substantive difference in customers' responses to questions that differ only in their position in a survey questionnaire?

- What is the magnitude of question order effects (i.e., methods effects) compared with the magnitude of structural (i.e., construct) effects?

The first question is investigated using simple contingency tables. The second question is investigated by developing a statistical model of the measurement relationships as well as the structural relationships among questionnaire items.

Form of the Study

To manipulate question order, GTE conducted a survey in which the same question appeared twice in the questionnaire: at the beginning of the questionnaire and at the end. This survey was administered to residential telephone customers during August 1990. A question about overall quality appeared at both the beginning and the end of the survey. After the first overall quality question, ratings of various services were obtained in the following order: local dial, inter-LATA long distance, intra-LATA long distance, billing, and operator service. All attribute questions were measured on the same four point scale, while the overall quality questions were measured on identical five-point scales.

The Effect of Position on Overall Quality

Ratings changes. The following contingency table shows the number of customers giving each of the sixteen possible combinations of responses to the question positioned at the beginning (OVQ_b) versus the question positioned at the end (OVQ_e). The decimal numbers below each of the counts in the margins are percentages for the marginal number.

OVQ_e / OVQ_b	POOR, BELOW AVG.	AVG.	GOOD	EXC.	
POOR, BELOW AVG.	5	7	1	1	14 (4.3)
AVG.	4	24	35	7	70 (21.3)
GOOD	1	12	109	41	163 (49.7)
EXC.	1	4	12	64	81 (24.7)
	11 (3.4)	47 (14.3)	157 47.9)	113 (34.4)	328

The ratings for the end-placed item, OVQ_e are generally more favorable than those for OVQ_b, since the "Excellent" rating is nearly ten percentage points higher for the former. A simple McNemar-like test would compare the number of those improving their rating over the survey, 92, to the total number changing ratings, 126, multiplied by 0.5, the proportion expected to increase under the null hypothesis that increases equal decreases. The p-value of this statistic is smaller than 0.001 and confirms intuition about differences in responses to the two items, in spite of their identical wording.

Potential Explanations The favorable increase may be associated with increased recall of information over time or it may occur with no changes in service memories. Respondents might recall the frequency or recency of a problem event or the variety of services offered, and this new information might modify their response. In the absence of changes in service memories, a change in the respondent's rating standards, or increased yea-saying (due to rapport with the interviewer), could be responsible for the increase.

Specific explanations for the ratings increase include:

1) An increased desire to please the interviewer, who is identified as calling for the telephone company.

2) The diminished importance of a service problem, in view of the many other services the company competently provides.

3) The realization, over the course of the interview, that an initially remembered service problem actually occured somewhat infrequently.

4) The realization, over the course of the interview, that an initially remembered service problem actually fell outside the 30-day reference period.

The phenomenon of general yea-saying does not seem to be operating here because there is no overall increase in ratings from the beginning to the end of the questionnaire. Indeed, "Poor" ratings for the individual services occur in the same proportions regardless of whether the respondent changed his overall quality rating or not. The motivation for withholding a second unfavorable overall rating from the interviewer as a company representative is also unclear. Hence, the following paragraphs consider how customers' recall of new information may modify their responses.

Diminished importance. A new variable (MODIFY) is created to record whether a customer's overall quality rating increased, decreased or remained constant over the interview. A simple systematic way of exploring possible disconfirmers is to calculate Pearson chi-square statistics for contingency tables of this variable versus candidate questions that may have diminished/increased importance. These candidates include the questions that are asked after the first overall question, namely questions about local dial, inter-LATA long distance, intra-LATA long distance, billing, and operator service.

It was found that ratings changes from local calling to inter-LATA long distance were significantly associated with MODIFY ($p=0.045$) in the sense that ratings increases from the former to the latter service tended to produce increases from OVQ_b to OVQ_e. This finding suggests that the opinion change occurs early in the interview, but does not explain why. It does cast doubt on the notion that the salience of a service problem is diluted by recalling the extent and competence of other offered services. The long distance service question occurs too early in the probing of service experiences to allow much dilution, and there is no statistical evidence that similar rating changes among local service and those mentioned even later in the instrument have any effect on overall rating changes.

Recalled frequency. To investigate whether an initially-remembered service incident set actually occurred less often than the respondent originally thought, we constructed contingency tables of perceived problem frequency by MODIFY. For each network service (local dial, inter-LATA, and intra-LATA service), the customer is asked whether he/she has experienced any of several common problems (static, call blockages, dial tone absence, cut-offs) in the last 30 days, and if so, how frequently they were perceived to have occurred. If "frequency effects" occur, one might expect a greater tendency to increase the overall quality rating if these frequencies were low, than if they were high. This tendency would probably be greater if the network problem were not experienced at all, for in that case there would be no service event whose memory was modified by the survey. However, this hypothesis is not supported by the data. The contingency tables show a much greater increase in the overall quality rating when perceived frequencies were large, not when they were small. This effect can perhaps be attributed to the likelihood that with frequent problems, the initial overall quality rating is already low, and most changes in the overall quality rating must be an increase.

Reference period effects. Neither of the two overall quality items specifies a

period of service that the customer should rate. However, nearly all the intervening questions--about specific services and their attributes--do specify a 30-day reference period. For instance, the customer is asked "How would you rate the quality of local calls you have dialed from this number *in the past 30 days?*" Since the second overall quality item occurs after numerous questions with a 30 day reference period, customers' responses may have been influenced by the reference period.

The absence of service problems in the reference period may be a sufficient reason for an overall quality ratings increase, so that the initial misplacement of such problems within that period could explain differences in the two rating questions. (In fact, past analyses and customer debriefings have suggested that problem absence actually induces "Excellent" ratings.) In this study, "Excellent" ratings on the ending overall quality item come from 22.6% (47/208) of the sample experiencing one or more transmission problems, while 41.1% (92/224) of those with no such problems give "Excellent" ratings. This difference is significant at the 0.001 level (as calculated by Pearson's chi-square statistic). Since problem absence tends to increase "Excellent" ratings in overall quality, this reasoning may apply to other items as well, and one might anticipate a rise in "Excellent" ratings from the initial overall quality item (with no reference period) and subsequent items with the 30-day period.

The initial overall quality responses which have no set reference period, can be compared with the (immediately following) local dial responses, which have a 30-day reference period. Despite the different subject matter of the items, it is interesting to note that the percentage of "Excellent" responses increases from 24% to 38% from one item to another, with most of the increase coming from the 29% of respondents who change their "Good" ratings to "Excellent." In contrast, there is virtually no net change in ratings between the local dial quality item and the proximate long distance quality item, and so the net migration from one response to another is nearly symmetric.

Summary. There is a large observed difference in rating level between two overall quality questions which differ only in their position in a survey questionnaire. Although no definite conclusions can be reached with the type of evidence available to us, the most supportable explanation for this difference is the questionnaire-associated effect that the intervening items induce a shorter reference period for the latter overall quality item.

A STRUCTURAL MODEL WITH MEASUREMENT AND QUESTIONNAIRE DESIGN EFFECTS

In this section, we develop a structural model based on equation (2) and incorporate measurement error and methods (i.e., question order) effects. The structural model incorporates the question order effects identified in the preceding paragraphs. Estimated statistical relationships between constructs may be different than those observed in prior research because we have accounted for these methods effects.

Model Specification. The following models explicitly account for possible differences between survey items, such as OVQ_b and OVQ_e, and the theoretical

concepts they putatively measure. Since we have seen that the two items differ dramatically in their responses, it is prudent to initially model each question as an indicator of its own corresponding construct, labeled *initial quality* and *final quality*. Using the arguments of an earlier section, the final quality construct is a function of initial quality, performance and disconfirmation, where disconfirmation is expanded to mean the discrepency between successive attitudes. The attribute measures of local dial, long distance and billing (LOCQ, LDQ and BILLQ, respectively) are indicators of both current performance and (because they too are attitudes) disconfirmation, and therefore may be products of both overall quality constructs. Since attitude modification appeared to occur during the transmission problems portion of the interview and reference period effects were detected, the existence of problems in the reference period (PROBEX) is a disconfirmation measure and is thus a product of the *final quality* construct.

Since responses to an item depend on the survey context in which it is asked, we postulate that each measure is a function of the major attribute measure just preceding it in the questionnaire. In our survey, OVQ_b influences LOCQ, which in turn influences LDQ, and BILLQ influences OVQ_e. Furthermore, similarity of items, or of response categories may generate a relationship, while dissimilarity dampens one. Thus OVQ_b has an effect on OVQ_e, while LOCQ and PROBEX are not directly related despite their proximity in the questionnaire. Numbering the items OVQ_b, LOCQ, PROBEX, LDQ, BILLQ, and OVQ_e according to their relative position in the questionnaire and letting $ß_{ij}$ represent the regression coefficient for the effect of item i on item j, the null model we consider can be diagrammed as in Figure I.

Figure 1 here

Transformation of the Data. To fit this model, some special handling of the variables and their correlation coefficients is needed. Since each survey response is ordinal, polychoric correlations should be used instead of the usual product-moment variety, and some merging of categories is necessary so that these correlations are consistent with the postulate of underlying multivariate normality. To ensure the identification of the labeled coefficients, the coefficients of relations between the two quality latent variables and their indicators are both assumed to be 1.0, and the variance of the initial quality latent variable is estimated to be 0.40, based on previous repeated measurement studies. This model will be loosely called the "unconstrained model."

Results. The fit of this model seems to be acceptable. The chi-square value indicating the closeness of its predicted correlation matrix with the observed polychoric matrix is an insignificant 9.12 with six degrees of freedom. More important is the fact that the residuals from the fit of the polychoric correlation matrix show no systematic patterns, have elements with values unimodally distributed about zero, and that the largest value of the standardized residuals is only 1.72.

We wish to formally test two hypotheses:

1) that each $ß_{ij}$ labeled in the diagram is non-zero, and

2) whether the two latent quality variables are distinct.

The chi-square values associated with the unconstrained model, with the four models with one of the $ß_{ij}$'s set equal to 0, and with the model equating the latent variables, are given below.

Model	d.f.	Chi-Square
Unconstrained	6	9.12
$ß_{12}=0$	7	24.22
$ß_{24}=0$	7	44.03
$ß_{56}=0$	7	41.83
$ß_{16}=0$	7	93.66
initial quality=final quality	7	14.84

It follows from these values that each $ß_{ij}$ is non-zero. For instance, the test that $ß_{12}$ = 0 consists of comparing 24.22 - 9.12 = 15.10 to a chi-square random variable with 7 - 6 = 1 degree of freedom. This value is significant at the 0.0001 level. Likewise, the three other tests yield highly significant test statistics. We conclude that proximate item relationships exist apart from their being common indicators of latent variables.

The test statistic for the model with only one latent quality variable is 14.84 - 9.12 = 5.72, which is well above the .975 quantile, so this model can be rejected at the 2.5% level. In this model both OVQ_b and OVQ_e are linked to latent quality by a coefficient of 1.0, so that the test can be performed by subtracting chi-square values. With the decrease in latent variables, it is possible to allow one of these coefficients to be unconstrained. In this case, the chi-square value is 13.63 with seven degrees of freedom. Given the rather smaller chi-square for the "unconstrained" two latent variable model above, this is informal evidence that the latent variables are indeed different.

Some aspects of the model help describe these two types of quality. It can also be demonstrated that the attribute variables are functions of both *initial quality* and *final quality*, so it appears that at least the three attributes of local dial, long distance and billing services enter into the evaluation at the beginning of the interview. The coefficients joining each of these attribute variables to *final quality* are, however, larger than the *initial quality* coefficients, so there may be a more explicit role for functional service attributes in the production of final quality. It also appears that the variance of final quality is smaller than that for initial quality, for the

former is estimated at 0.10, compared to the externally estimated 0.40 for the latter. A picture thus emerges of final quality as a more focused predictable measure than initial quality.

There are some other aspects of these models which are noteworthy. We postulated earlier that PROBEX is a disconfirmation measure, so that it should not be directly linked to the initial quality latent variable. When such a link is tested the decrease in chi-square value, from 9.12 to 7.32, is insignificant at any level less than 0.15, as the theory suggests. There also is no evidence of a link between LDQ and BILLQ. This result is probably due not only to major differences in the services these attributes represent, but to the relatively large number of intervening items between these two questions in the interview. Finally, the variances of the expressions of measurement error in all of the questionnaire items are all significantly larger than zero. It follows that there is non-negligible measurement error in these data whose presence must be modeled.

DISCUSSION

Our models and supporting discussion hold a number of fundamental lessons for telephone company managers and for others who seek customer reaction to their services via a survey program. First, our findings stress the importance of disconfirmation in both determining and changing attitudes, as evidenced its role in differentiating between our two types of quality. *Indeed, disconfirmation can be affected by a perceived change in service, or by a change in the perspective of the customer. In this study, disconfirmation operates over the brief duration of the survey interview*. The evaluated service does not change, but the customer's recall of service performance attributes does. Managers of improving services should be particularly careful to appreciate the attitude-enhancing effect of a customer interview.

Customer satisfaction is an elusive thing, both in its concept and its manifestations. Far from being the "meter on a machine" that some have thought attitude surveys to be, a customer's response to a survey item will depend on its position and neighboring items in the questionnaire. It is therefore perilous to make inferences about customer satisfaction from an isolated Good/Excellent (or other response subset) percentage. Comparable statistics, such as satisfaction rates in a single area over time, are far more interpretable as indicators of perceived service performance. Furthermore, satisfaction is elusive as an underlying concept, as shown by our model's distinction between *initial quality* and *final quality*. It follows that even within a single service and a single measurement program, satisfaction is not a monolithic concept, and a complete customer satisfaction program will obtain data with several indicators of satisfaction and of its related concepts (including various sub-attributes) so that the variability, reliability and characteristics of the constructs can be described.

In addition to the aforementioned areas of program design, this work sheds some light on survey analysis and problem diagnosis. For example, it is tempting to measure the magnitude of a problem by the number of variables it affects. However, since attribute items can be correlated by their questionnaire

proximity as well as by the commonality of their underlying factors, a problem does not necessarily increase in seriousness according to its survey extent. A problem with the long distance survey item, say, may simply reflect a local dial problem through the correlation of the items. More generally, a temporal decline in one or more ratings should ideally be "cleansed" of the effects of the other model variables by examining residuals for that variable rather than the raw values. Finally, changes in the two quality constructs should be calculated since a drop in one but not the other is likely to aid in the diagnosis. For example, a drop in initial but not final quality perhaps indicates a service problem in an area not explicitly measured by the survey.

CONCLUDING REMARKS

We have attempted to tie three themes in customer satisfaction research together in this paper. Classic models of attitude formation are invoked to construct a basic structure for the analysis of telephone service survey data. Distinctions are drawn between observed variables from a satisfaction survey and underlying constructs. These items and constructs are incorporated in a structural equations/measurement error model whose basic relationships are generated by the attitude models. The third research stream entering the model specifically recognizes questionnaire effects where response correlations exist among proximate survey items or those with similar wordings or response categories. A brief synopsis of previous work is presented where analysis suggests the existence of these effects.

These ideas are tested with a telephone service survey in which identical overall quality items are included at both the beginning and end of the questionnaire, with a series of attribute rating questions in the middle. Dramatic differences in rating levels between the two overall quality items are examined with some simple analyses and subsequent hypotheses. The higher ratings for the end-of-the-questionnaire item are ultimately attributed to an implicit change in the reference period of the rating. This change appears to be centered in a location of the questionnaire whose items are coded to appear as a disconfirmation measure in the final model.

The model provides comforting evidence of the primacy of the disconfirmation measure, and of the indispensability of modeling relationships among proximate and otherwise similar survey items. Statistical evidence is also produced for the distinction between two latent constructs of overall quality, the first reflecting a top-of-mind evaluation and the latter being a more considered attitude incorporating the attributes in the preceding parts of the interview.

This work has ramifications for the design and interpretation of customer satisfaction results. By demonstrating an increase in satisfaction over the course of an interview, the customer survey may claim an active role for itself in educating and satisfying respondents. By confirming the distinction between concepts and indicators, arguments can be made for the use of multiple quality indicators and indices constructed therefrom, and for the de-emphasis of single statistic, and single response category measures. Finally, the production of a viable quality model supports the diagnosis of poor ratings through modeled results rather than the collective use of single survey items,

since each attribute rating has been seen to contain extraneous information which must be filtered out.

REFERENCES

Andrews, Frank M. (1984), "Construct Validity and Error Components of Survey Measures: A Structural Modeling Approach," *Public Opinion Quarterly*, 48, 409-442.

Bitner, Mary Jo (1990), "Evaluating Service Encounters: The Effects of Physical Surrounding and Employee Responses," *Journal of Marketing*, 53 (2), 69-82.

Bolton, Ruth N. and James H. Drew (1991a), "A Longitudinal Analysis of the Impact of Service Changes on Customer Attitudes," *Journal of Marketing*, 55 (1), 1-9.

_____ and James H. Drew (1991b), "A Multi-Stage Model of Service Quality and Value," *Journal of Consumer Research*, Vol. 17 (March), 375-384.

Cardozo, Richard (1965), "An Experimental Study of Customer Effort, Expectation and Satisfaction," *Journal of Marketing Research*, 2 (3), 244-9.

Churchill, Jr., Gilbert A. and J. Paul Peter (1984), "Research Design Effects on the Reliability of Rating Scales: A Meta Analysis," *Journal of Marketing Research*, 21 (4), 360-75.

_____ and Carol Surprenant (1982), "An Investigation into the Determinants of Customer Satisfaction," *Journal of Marketing Research*, 19 (4), 491-504.

Drew, James H. (1987), "Modeling Halo Effects in Surveys for Low-Saliance Services," *Proceedings of the Social Statistics Section, American Statistical Association*. Washington, Amer. Stat. Assoc.

_____ and Ruth N. Bolton (1990), "The Customer Satisfaction Feedback Loop for Local Telephone Service," *Journal of Customer Satisfaction, Dissatisfaction and Complaining Behavior*, Vol. 3, 29-35.

Fuller, Wayne A. (1987), *Measurement Error Models*. New York: John Wiley and Sons.

Helson, Harry (1964), *Adaptation-Level Theory*. New York: Harper and Row Publishers.

Oliver, Richard L. (1980a), "A Cognitive Model of the Antecedents and Consequences of Satisfaction Decisions," *Journal of Marketing Research*, XVII (November), 460-9.

_____ (1980b), "Theoretical Bases of Consumer Satisfaction Research: Review, Critique and Future Direction," in *Theoretical Developments in Marketing*, Charles W. Lamb, Jr., and Patrick M. Dunne (eds.) Chicago: American Marketing Association, 206-210.

_____ (1981), Measurement and Evaluation of Satisfaction Processes in Retail

Settings," *Journal of Retailing*, 57 (Fall), 25-48.

_____ and Wayne S. DeSarbo (1988), "Response Determinants in Satisfaction Judgments," *Journal of Consumer Research*, 14 (March) 495-507.

._____ and John E. Swan (1989), "Equity and Disconfirmation Perceptions as Influences on Merchant and Product Satisfaction," *Journal of Consumer Research*, 16, 372-383.

Olshavsky, Richard W. and John A. Miller (1972), "Consumer Expectations, Product Performance, and Perceived Product Quality," *Journal of Marketing Research*, 9 (1), 19-21.

Olson, Jerry C. and Philip Dover (1976), "Effects of Expectations, Product Performance, and Disconfirmation on Belief Elements of Cognitive Structures," in *Advances in Consumer Research*, Provo, UT: Association for Consumer Research.

Tse, David K. and Peter C. Wilton (1988), "Models of Consumer Satisfaction Formation: An Extension," *Journal of Marketing Research*, 25 (2), 204-12.

Westbrook, Robert A. and Richard L. Oliver (1981), "Developing Better Measures of Consumer Satisfaction: Some Preliminary Results," in *Advances in Consumer Research*, 8, Kent B. Monroe, ed., Ann Arbor, Michigan: Association for Consumer Research, 94-99.