

Order Effects in Customer Satisfaction Modelling

Seigyoung Auh

University of Melbourne

Linda Court Salisbury

University of Michigan Business School

Michael D. Johnson

University of Michigan Business School

This research examines the effects of question order on the output of a customer satisfaction model. Theory suggests that locating product attribute evaluations prior to overall evaluations of satisfaction and loyalty should increase the impact of performance drivers in the model, explain more variation in the overall evaluations, and make positive satisfaction and loyalty evaluations more extreme. Our results show that, although customers' overall evaluations are more extreme and better explained when provided after attribute evaluations, the impact of satisfaction drivers is relatively unaffected. Consistent with expectations, question order does affect the explained variation in satisfaction and the levels of satisfaction and loyalty. Implications for satisfaction modelling are discussed.

Keywords: order effect, customer satisfaction, loyalty, polarization, temporarily accessible, PLS, critical incident technique (CIT)

Introduction

Customer satisfaction modelling has become an important tool for setting quality improvement priorities and improving marketing program effectiveness. Satisfaction models provide information regarding how companies or products perform on various benefits and attributes as well as the importance, or impact, the benefits and attributes have on satisfaction and subsequent intentions and behaviours. Within the customer satisfaction literature, the effect that the order in which quality, satisfaction, and loyalty are measured has on the output of a satisfaction model remains relatively unexplored. Satisfaction survey questions are typically asked in the order in which they are subsequently modelled, where attribute performance measures are obtained prior to overall satisfaction and loyalty measures. Yet this practice may serve to effectively create a model in customers' minds.

Consider the following scenario. A marketing manager at a major retail store is considering conducting a customer satisfaction and loyalty study. The manager constructs a questionnaire that includes items measuring customer satisfaction and loyalty, along with the drivers of customer satisfaction and loyalty identified in qualitative studies such as focus groups. Without expecting any foreseeable consequences, the manager orders the questions in the questionnaire such that the drivers of satisfaction are asked first followed by the overall satisfaction and loyalty questions.

Six months later, a new manager replaces the previous manager and uses the same questionnaire items as the previous manager. However, the new manager reverses the order of the questions by moving the overall evaluations to the beginning, thinking that the attribute evaluations may bias or otherwise affect the overall evaluations. The items that measure satisfaction and loyalty are asked first, followed by the drivers of satisfaction. The new manager is surprised by the results when he/she finds that the drivers of satisfaction explain significantly less variance in satisfaction, and the overall level of satisfaction and loyalty expressed by the respondents has significantly decreased. During this 6 month period, no turbulence in the industry or any other industrial organizational factors have been observed that would explain this discrepancy in findings. What could possibly be the rationale as to why we would observe such a difference? The goals of this research are to provide a theoretical explanation of this phenomenon and conduct an empirical test of the theoretical framework.

The objective of this study is to explore how the order in which satisfaction survey data is collected can affect satisfaction model results. The effect of eliciting specific beliefs before or after overall evaluations has been explored in the general attitude literature (Tourangeau and

Rasinski 1988; Tourangeau *et al.* 1989), multi-attribute modelling literature (Bickert 1993), life satisfaction literature {McClendon and O' Brien 1988), and in voting intentions {Simmons, Bickart, and Lynch 1993). Our research attempts to build on the findings from these studies and extend work into a customer satisfaction and loyalty context. Schul and Schiff {1993) explored order effects for measuring general satisfaction, satisfaction with specific domains, and frequency of negative experiences. An important difference in our study is that we explore how order effects affect the drivers of satisfaction and loyalty, which requires different theoretical arguments and methodology. We test for order effects using an empirical study of hair care services. The results show that when attribute performance measures are collected prior to satisfaction and loyalty measures {versus after), the performance measures explain greater variation in satisfaction and result in higher overall satisfaction and loyalty scores. Yet the impact that the attribute and benefit evaluations have on satisfaction and loyalty were relatively independent of question order. This invariance of impact scores on satisfaction and loyalty across different measurement orders is discussed in greater length in the discussion section later in the paper. Implications for data collection and modelling are discussed as well.

Satisfaction Modelling

Customer satisfaction is typically modeled within a system of causal relationships. The system runs from the concrete attributes that describe a product or service, to the benefits or consequences these attributes provide customers, to customers' overall evaluation of their purchase and consumption experience (customer satisfaction), and subsequently to the behavioural intentions or behaviours that result (Johnson and Gustafsson 2000). These may include repurchase intentions, actual repurchase, product recommendations or word-of-mouth. Satisfaction models rest heavily on expectancy-value model formulations (Fishbein and Ajzen 1975; Johnson *et al.* 2001), where beliefs about the consumption experience (quality dimensions and price) affect customer satisfaction as a type of overall evaluation or attitude. Satisfaction, in turn, affects customers' behavioural intentions and behaviours. The benefit, satisfaction, and loyalty constructs in these models are inherently abstract, or latent, variables measured using multiple concrete proxies or measurement variables. Benefits are measured using product or service attribute evaluations, satisfaction is measured using different overall evaluation standards (such as overall satisfaction, overall performance versus expectations, overall performance versus an ideal), and loyalty is often measured using behavioural intentions (such as the likelihood of repurchase or recommendation to others).

A popular method for estimating and operationalising a satisfaction model is partial least squares, or PLS (Gustafsson and Johnson 1997; Steenkamp and van Trijp 1996). PLS is essentially an iterative estimation procedure that integrates principal-components analysis with multiple regression (see Fornell and Cha 1994 or Wold 1982 for a description of PLS). When estimating a PLS model using reflective measures, the procedure essentially extracts the first component from the measures for each latent variable and uses these principal components within a system of regression models. The PLS algorithm then adjusts the principal-component weights to maximize the variance explained among the dependent variables of interest. For practical purposes, this means that PLS attempts to predict outcomes of significance to managers, such as satisfaction, loyalty, and profitability. Because PLS is conceptually similar to principal components, the latent variables in PLS are also easily operationalised as principal components, or weighted indices, of the measurement variables. A product's performance indicators, such as satisfaction and loyalty, are operationalised as latent variables (through a weighted index of multiple survey measures) to provide managers with explicit benchmarks for evaluating their performance. When this performance information is combined with the impact scores from the regression estimates, managers have the diagnostic information they need to make key resource allocation decisions. The PLS satisfaction model helps managers identify product improvement opportunities by revealing where the attribute impact is high but performance scores are relatively low. This information, evaluated together with improvement cost data, enables managers to make key resource allocation decisions. For example, a branch manager of a bank may find out that online banking has a high impact on customer satisfaction, but his branch is performing poorly in this particular area. This type of information guides the manager to effectively allocate resources to rectify the problem at hand.

Order Effects in Data Collection

Survey or questionnaire items measuring the antecedents of satisfaction are typically asked before the satisfaction items, and the items measuring proposed consequences are asked after the satisfaction items (Johnson and Gustafsson 2000). This can result in a question order effect whereby the researchers essentially simulate, or create, the proposed model relationships for the consumer via the sequencing of items (Feldman and Lynch 1988). For example, attribute ratings are followed by overall satisfaction and loyalty evaluations. Previous research has shown that after answering an attitude question, the respondent's attitude toward an object becomes more accessible (Fazio, Powell and Williams 1989). As a result of ordering items in a sequence mirroring the proposed causal relationships, evaluations of the attributes

and benefits measured in the survey become highly accessible in memory for the respondent as he or she answers the satisfaction and loyalty items.

From another theoretical perspective, the survey process can be viewed as an ongoing conversation between the researcher and the survey respondent (Sudman, Bradburn, and Schwarz 1996). Grice (1975) articulated several assumptions, or maxims, governing everyday conversation. Grice's "maxim of relation" requires that speakers make their contribution to a conversation relevant. Thus, in a survey, if several specific attribute ratings precede more summary evaluations, a conversational perspective suggests that the ratings are important and should be considered in the overall evaluations (Schwarz, Strack and Mai 1991).

McClendon and O'Brien (1988) reported question order effects when measuring overall subjective well-being vis-a-vis well-being in specific life domains (e.g., health, family, work, marriage, and school). They found that whether the determinants of overall subjective well-being were asked before or after overall subjective well-being had a significant effect on the magnitude of the regression coefficients. Moreover, their study revealed that the closer the determinants were relative to the overall subjective well-being question, the greater the size of the regression coefficients. They suggested that a recency effect, along with an accessibility explanation, explains these results. Managerially, this means that resource allocation decisions are overly affected by where an area of product or service performance happens to be evaluated in a survey.

In contrast, Tourangeau *et al.* (1989) did not find question order effects for so-called context items and target issues. Target issues are analogous to an overall or summary evaluation. Context items are groups of items that are related to target issues but are not determinants or subsets of the target issues. For example, Tourangeau *et al.* (1989) used questions on government responsibility to help the poor or economic individualism as context questions and favourability or opposition towards welfare programs as the target issue. The results suggest that the valence of the context items has a significant impact on how people respond to the target item. In essence, they argue for carry over effects between context and target questions. Their empirical results, however, are inconsistent with McClendon and O'Brien (1988) in that no question order effect was found between context and target questions. Whether the context question was asked before or after the target question did not have any significant effect on correlation between context questions and target questions (McClendon and O'Brien 1988; Table 5, pp.513). This may be because a target question is not a summary evaluation of context questions but merely a set of substantively related questions.

Schul and Schiff (1993) found order effects when measuring satisfaction with customers of a national telecommunications company. They argue that the overall satisfaction evaluation is a linear combination of positive and negative evaluations of domain specific satisfaction ratings. When the domain specific satisfaction questions are asked prior to the overall satisfaction question, positive and negative evaluations of domain specific satisfaction questions are more or less equally represented in the overall satisfaction rating. However, when overall satisfaction questions are asked prior to the domain specific satisfaction question, negative evaluations of domain specific satisfaction questions are more heavily represented in the overall satisfaction evaluation because of their greater accessibility. They argue that negative experiences are more easily accessible than positive experiences because negative information is processed more intensively and elaborately.

It is important to note that the average general satisfaction measured by Schul and Schiff (1993) was very low. This is atypical for most satisfaction studies, which tend to have relatively high average satisfaction ratings (Fornell 1995). Thus, an essential question that needs to be addressed is what if we find little or no negative experience with a product or service provider? Would we find similar order effects if most of the survey respondents reported satisfaction that was relatively high?

The mixed empirical results motivate additional research in a marketing context, especially where customers are relatively satisfied with the product or service. Our research addresses order effects in a service experience context with satisfaction as the overall summary evaluation, and we further include loyalty in the model as the ultimate dependent variable. Loyalty is conceptualized as a predisposition to choose one product or service provider over other competitors and is more closely related to actual behaviour than satisfaction. We test to see if the order effects found in the existing literature extend to the marketing arena. Further, we go beyond merely modelling a summary evaluation and include a proxy for actual choice behaviour, loyalty.

Hypotheses

There are three implications of these theoretical arguments for satisfaction modelling. One is that the diagnostic value of the satisfaction drivers should be enhanced when overall satisfaction questions are collected after attribute ratings. The model coefficients, or impact scores, associated with satisfaction drivers in our model should be relatively greater when overall satisfaction is elicited after attribute ratings because the overall ratings should be

reflections of the attribute ratings. In contrast, when overall satisfaction and loyalty questions are evaluated prior to attribute performance, respondents will rely more on stored overall evaluations in memory that are less affected by the attribute ratings per se. A second implication is that attribute ratings, when combined to create benefit indices, should explain more variation in satisfaction and loyalty when the attribute ratings are collected first. We state these predictions formally as hypotheses one and two.

H1: The impact scores of performance drivers on satisfaction and loyalty are greater when the performance drivers (attributes and benefits) are measured prior to satisfaction and loyalty rather than after.

H2: Performance drivers (attributes and benefits) explain more variation in satisfaction and loyalty when the drivers are measured prior to satisfaction and loyalty rather than after.

Question order can also affect the extremity of overall evaluations and behavioural intentions in a satisfaction survey creating polarization effects. Collecting attribute performance ratings prior to satisfaction and loyalty evaluations effectively creates a schema for the respondent in which the articulated attributes and benefit categories play a central role. This should lead to a polarization of the overall evaluation or attitude (Tesser 1978), in this case the satisfaction and loyalty evaluations. Because performance data in a satisfaction survey is generally negatively skewed (relatively positive), we expect satisfaction and loyalty evaluations to be greater when they are collected after the attribute ratings. The positive attribute evaluations, being more accessible, increase subsequent overall evaluations and intentions resulting in polarization effects for satisfaction and loyalty. Intuitively, if you remind a customer of all the reasons why they are generally happy with a product, their overall evaluations and intentions should increase.

Additional support for a polarization effect comes from cognitive dissonance theory (Festinger 1975), whereby people try to seek consistency in their beliefs about the world around them. Survey respondents may utilize a cognitive dissonance reduction strategy. When respondents answer positively to attribute items and are subsequently asked about their overall evaluation of their hair care provider, the previously answered positive responses work as a reinforcement mechanism in bolstering their overall satisfaction beliefs. This results in a polarization effect. Stated formally:

H3: Satisfaction and loyalty evaluations are greater when the performance drivers {attributes and benefits} are measured prior to satisfaction and loyalty rather than after.

Method

The hypotheses were tested using a survey of satisfaction with hair care services. Hair care service was selected for the following reasons. First, hair care service was a category that was sufficiently familiar for our survey respondents (college students) to provide reliable and quality data. Second, it was a service category that both males and females used on a relatively consistent basis such that our empirical results would not be biased toward either gender. Third, it possessed several relevant attribute/benefit categories to allow us to detect differences across conditions. Order effects were tested by changing the relative order of survey items between two groups of respondents to see if different impacts and values emerged. Attribute ratings, organized by benefit category, were presented before satisfaction and loyalty ratings in one group and after satisfaction and loyalty ratings in a second group.

The attributes and benefit categories used in the survey were obtained using CIT, the critical incident technique {Bitner, Booms, and Tetreault 1990; Grove and Fisk 1997; Hayes 1998; Meuter *et al.* 2000). The CIT approach focuses on input from the customer to develop survey items. A convenience sample of ten adults was asked to list five to ten positive experiences (likes) as well as five to ten negative experiences {dislikes} regarding their past or current hair care service. The interviews lasted approximately 15-20 minutes and were audiotaped for subsequent analysis. A total of 100 different critical incidents were obtained from these ten interviewees, ranging from 8 to 24 per interview. Two of the authors then served as judges to independently categorize the critical incidents into 39 distinct performance attributes and then categorize these attributes into five benefit-level performance categories. The classifications were highly reliable (over 80% agreement for both attributes and benefits) and disagreements were resolved through discussion. The benefits (hair cut quality, provider interaction, atmosphere, timeliness, and scheduling) served as drivers of satisfaction in the satisfaction model, which is presented in Figure 1. Table 1 shows the 15 attribute performance measures used to operationalise the benefits as latent variables. The benefit items were constructed using 10-point disagree to agree scales. Satisfaction and loyalty were also operationalised as latent variables using the measures listed in Table 1. They were collected using 10-point semantic differential and likelihood scales.

Preliminary data modelling (using the entire data set) revealed that two benefit categories, timeliness and scheduling, had direct effects on loyalty in addition to their indirect effects on loyalty via satisfaction. These links are intuitive as both timeliness and scheduling are factors that allow customers to access the hair care service when they need it, thus affecting loyalty independent of the service experience per se. The model in Figure 1 was subsequently used in our analyses.

The self-administered survey was completed by 191 undergraduate business students who earned course credit for participating. We acknowledge the recent work by Peterson (2001) on the nature of using college students as subjects in social science research. Results from his meta analysis suggest that college student samples provide greater homogeneity than do non-student samples. However, his research points to the startling fact that this greater homogeneity does not automatically translate into greater effect sizes and more powerful hypothesis testing. Peterson's (2001) work even reveals that the magnitude and directionality between constructs are reversed 19% of the time when using student samples. Therefore, we are cautious in generalizing our conclusions too broadly from this particular research and acknowledge the need for replication on a more representative sample.

Subjects were randomly assigned to one of two conditions. The attributes first group received a survey that asked them to evaluate the attributes of their hair care provider's service, organized by benefit categories, before being asked to evaluate their satisfaction with and loyalty toward the hair care provider. The overall evaluations first group received a survey that asked them to evaluate their satisfaction with and loyalty toward the hair care provider before being asked to evaluate the attributes. Twelve subjects were deleted from the analysis due to an excessive number of missing values. An additional three subjects were removed because they indicated that a family member or friend acted as their hair care provider. Consequently, 176 subjects were included in the final analysis ($n = 87$ and 91 respectively for the attributes first and overall evaluations first conditions). Males and females were about equally represented in our sample, with 51% male and 49% female subjects.

Results

The models were estimated using PLS. We first report the quality of the measurement model and then discuss the latent variable model results pertinent to our hypotheses. The measurement model is assessed on two criteria; reliability and discriminant validity. The model

results demonstrate high reliability for the measurement variables and high discriminant validity among the latent variables. Reliability in a PLS model is measured through the communality of the measurement variables (the squared correlation between a variable and the index, or latent variable, of which it is a part; Fornell and Cha 1994), which should exceed 0.5 or 50%. This implies that the loadings should exceed 0.707 to ensure that at least half of the variance in the observed measurement variable is shared with the construct. (The squared correlation equals the variance explained, where $0.707^2 = 50\%$.) Table 2 reports the average communality for each latent construct in the model. The average communality of the measurement variables was high for each construct in each model, ranging from 0.702 to 0.884 for the attributes first model and from 0.655 to 0.882 for the overall evaluations first model. Average communality was greater than 0.5 in 14 of 14 cases (100%). Another criterion used to evaluate the validity of the measurement model is examining the discriminant validity of the measurement model. The discriminant validity of a measurement model is evaluated by determining whether each latent variable, or construct, shares more variance with its measurement variables than it does with other constructs in the model. (Fornell and Cha 1994). This can be performed by looking at the percent of measurement variable loadings that exceed the latent variable correlations. For example, we have a total of five comparisons to make when comparing customer satisfaction and loyalty because they are operationalised using three and two measurement variables, respectively. We examined the loadings for each of these five measurement variables to see if any of these are smaller than the correlation between customer satisfaction and loyalty. The satisfaction model had a total of only two out of 240, or 0.83%, comparisons across the two conditions where a latent variable correlation exceeded a measurement variable loading. The two cases that violated the discriminant validity criteria were in the overall evaluations first condition and pertained to the relationship between hair quality and customer satisfaction and between customer satisfaction and customer loyalty. Overall, however, these results provide strong support for the discriminant validity of the constructs.

Hypothesis one predicts that when attribute evaluations are collected before rather than after overall evaluations, impact scores {path coefficients} are higher. Figure 2 shows the impact scores for the model under both the attributes first and overall evaluations first conditions. Overall, the impact scores were relatively consistent between the two order conditions. This consistency was demonstrated in several ways. The total effect that all five benefit areas have on satisfaction {the sum of the impact scores} was relatively equal between the two conditions {1.075 and 1.088, respectively, for the attributes first and overall evaluations first conditions}. The total effect of the benefits on loyalty {both directly and indirectly via satisfaction} was also

almost identical {0.864 versus 0.863, respectively). Finally, a jack-knife analysis was performed to determine the significance of the impacts. The significant ($p < .05$) paths for the attributes first model included hair quality and provider interaction on satisfaction (0.628 and 0.200, respectively) and the direct effect of scheduling on loyalty (0.314). The significant paths for the overall evaluations first model also included hair quality and provider interaction on satisfaction (0.620 and 0.235, respectively) and the direct effect of scheduling on loyalty (0.192). Also significant in the latter model was the direct effect of timeliness on loyalty (0.207). Overall, therefore, while there was some variance in the impacts, there was no evidence to support greater overall impacts in the attributes first condition-. Thus, hypothesis one was not supported. We provide detailed explanations about this finding in the discussion section below.

Hypothesis two predicts that the attributes, operationalised as benefit indices, should explain more variation in the overall evaluations when the attributes are collected prior to the overall evaluations. This prediction was supported by the satisfaction results, where the variance explained for the attributes first group ($R^2=0.746$) was higher than that for the overall evaluations first group ($R^2=0.618$).

This suggests that placing benefit items before satisfaction items increases the accessibility of the attributes and benefits when evaluating hair care provider satisfaction. Interestingly, however, the result did not extend to the loyalty evaluations where the variance explained was almost equal ($R^2 =0.578$ and 0.587 for the attributes first and overall evaluations first conditions, respectively). Table 3 summarizes the variance explained for both satisfaction and loyalty across the two conditions.

Thus, H2 is partially supported. We suggest two possible explanations why the hypothesis was supported for satisfaction but not for loyalty. One is that the effect is short lived, and does not extend past the initial satisfaction evaluations. The other is that loyalty evaluations are less susceptible to the effect than is satisfaction. Loyalty evaluations are closer to preferences and choice, whereas satisfaction evaluations are closer to attitudes.

The polarization hypothesis (H3) was tested by conducting mean difference tests (one-tailed t-tests) for satisfaction and loyalty across the two experimental groups. The results supported our hypothesis that measuring attribute performance before satisfaction and loyalty leads to more extreme evaluations in both satisfaction and loyalty. The attributes first group had a mean satisfaction evaluation (7.56) that was significantly greater (more extreme) than the mean satisfaction evaluation for the overall evaluations first group (7.14; $t(174) = 1.78$, $p < .05$). In contrast to H2, we found that polarization extended to the loyalty evaluations, where the attributes first group had a higher mean loyalty evaluation than the overall evaluations first

group (8.72 versus 8.12). This difference was also significant ($t(174) = 2.52, p < .01$). This suggests that collecting the attribute performance measures immediately prior to the satisfaction and loyalty items leads to a polarization of both the satisfaction and loyalty evaluations. Therefore H3 was strongly supported. Table 4 summarizes the results of our hypothesis tests.

Discussion

Theoretical Implications

Customer satisfaction surveys typically collect measures in the order in which they are subsequently analysed in a satisfaction model. We argue that this practice creates a model in customers' minds. Psychological theory suggests that presenting attribute ratings prior to overall evaluations of satisfaction and loyalty should make the attribute information more accessible when making the overall evaluations. This should increase the impact of performance drivers in the model, explain more variation in the endogenous variables, and make already positive satisfaction and loyalty evaluations even more extreme. We tested these predictions by varying the order in which attribute and overall evaluation questions were presented in a survey of hair care services.

Our results revealed that the order of data collection did not systematically affect the ability of the satisfaction model to identify the drivers of satisfaction and loyalty, as measured by the impact that benefit categories have on satisfaction and loyalty. This finding was inconsistent with what we predicted in hypothesis one. We suggest two reasons why question order did not affect the ability of the satisfaction model to identify the drivers of satisfaction and loyalty in our study. First, the attributes in our survey are ones whose evaluations were permanently accessible to survey respondents, and therefore less subject to question order effects. The attributes were identified by customers themselves rather than the researchers. Second, we conceptualized satisfaction as derived from an accumulation of experiences over time rather than an evaluation of a particular transaction or episode. Thus satisfaction is more a stored evaluation in memory rather than an evaluation constructed on the spot, making it less sensitive to question order effects. The managerial significance of this finding is that cumulative evaluations of satisfaction provide a more stable dependent variable for deriving impact scores.

This second explanation reinforces the notion that evaluations fall into one of two general categories: more permanently or temporarily accessible. The accessibility of information generally depends on the frequency and recency of its use (Higgins 1996). More permanently accessible information is frequently used and less subject to question order effects. Information

that has just been used, such as a previous survey question, is temporarily accessible information. Temporarily accessible information underlies most question order effects in surveys (Sudman, Bradburn, and Schwarz 1996; Tourangeau and Rasinski 1988). Service or product attributes that are important to customers and influence their satisfaction judgments are likely to have more permanently accessible attribute evaluations stored in memory. When other less-important attributes are included as survey items, their evaluations are stored as temporarily accessible information in memory that can impact subsequent survey questions. Consequently, they will have different impacts on satisfaction depending on whether they are asked before or after satisfaction. More permanently accessible attribute evaluations, on the other hand, should be less prone to question order effects in satisfaction surveys. The hair care service attributes included in our satisfaction survey were determined by eliciting input from customers using the critical incidence technique. This allowed us to determine which attributes were most important to customers, suggesting that the attributes included in our survey were more likely to have been more permanently accessible attribute evaluations. The results of hypothesis one support the notion that the drivers of satisfaction were likely to have been more permanently accessible attribute evaluations.

Our conceptualization of customer satisfaction as a cumulative evaluation may also explain why we found no systematic effect of question order on benefit impact scores. The marketing and satisfaction literature discuss two types of satisfaction: transaction-specific and cumulative satisfaction (Fornell *et al.* 1996; Rust, Zahorik, and Keiningham 1995). Cumulative satisfaction is an overall evaluation of a customer's total purchase and consumption experiences with a particular product or service to date (e.g., Fornell 1992, Johnson and Fornell 1991). Transaction-specific customer satisfaction is an evaluation of a specific product purchase and consumption experience or service encounter. We measured cumulative satisfaction in our survey.

The attitude literature distinguishes between a constructed attitude or evaluation versus a retrieved attitude or evaluation (Bettman and Zins 1976; Bettman, Luce, and Payne 1998; Hastie and Park 1986; Payne, Bettman, and Johnson 1992). The constructivist view posits that attitudes are constructed, or made up on the spot, depending on how a question is framed or elicited.

Findings such as preference reversals and sensitivity to context effects are typical examples under this school of thought. On the other hand, arguments also exist for a stored and retrieved

view of attitude. The proponents here contend that attitudes, once formed, are stored in memory and do not change easily. As a result, they are retrieved rather than constructed.

If overall satisfaction is constructed, the impact of performance drivers on satisfaction will vary depending on their location in a survey. When attribute items are located prior to overall satisfaction items, attribute evaluations will contribute more heavily in constructing overall satisfaction than if they were located after overall satisfaction items. However, if satisfaction is retrieved based on some stored evaluation, the impact of attribute evaluations on the overall satisfaction evaluation should be consistent whether attribute items are located before or after satisfaction items. Retrieval of a stored evaluation (satisfaction) is expected to be less affected by the accessibility of attribute evaluations than a satisfaction evaluation that is constructed on the spot. We conceptualized and measured satisfaction as a cumulative construct similar to a stored and retrieved evaluation. For example, the survey instructed respondents to consider all of their experiences to date with their hair care provider rather than on a particular salient or most recent experience. Therefore, respondents may have been more likely to retrieve a stored overall satisfaction evaluation in memory rather than construct an overall satisfaction evaluation on the spot. This would make the attribute (benefit) impact scores less sensitive to question order.

Our finding that the impacts of the benefits on satisfaction are invariant across different order conditions is contrary to what McClendon and O'Brien (1988) found. They found a significant difference in regression coefficients on subjective well-being when they altered the order of questions. The dependent variable in the McClendon and O'Brien's {1988} study was subjective well-being whereas in our study it was cumulative satisfaction. Thus the effect of question order on model parameters may depend on the nature of the dependent variable under investigation. Our invariance finding extended to loyalty as well. Loyalty is closer to actual choice behaviour than is satisfaction when the two constructs are ordered in a cause and effect relationship. As noted earlier, our results suggest that the impact of benefits on satisfaction and loyalty are quite robust to order effects in a product/service context.

Consistent with hypothesis two, the attributes first model explained more variation in satisfaction than the overall evaluations first group. This is consistent with McClendon and O'Brien's (1988) results for overall subjective well-being. We found no difference between the two order conditions in variance explained for loyalty. We believe that the attributes first model would have explained more variance in loyalty had we used a stronger order manipulation.

Hypothesis three was confirmed when we found that both the satisfaction and loyalty evaluations were more positive for the attributes first group than for the overall evaluations first

group. This finding is consistent with Schul and Schiff's (1993) finding that general satisfaction with an organization was higher when domain-specific satisfaction items were located before general satisfaction items. However, their theoretical explanation of the finding is different from ours. They suggest that general satisfaction was more heavily influenced by positive (versus negative) domain-specific evaluations in the specific-general question order than in the general-specific question order.

We did not predict any asymmetric effects of attribute evaluations. We attribute our finding to polarization effects in the attributes first group. We have discussed what our results imply from a theoretical perspective. However, this does not directly answer the question, "what does all of this mean for a manager trying to satisfy her customers and build loyalty?" The next section attempts to address this important issue.

Managerial Implications

Our results have important implications for practical applications of satisfaction modelling. First, it is critical to use consistent survey question order when evaluating satisfaction and loyalty benchmarks, either for a single firm over time or for multiple firms at the same time. Recall our example at the beginning of the paper where two retail managers used identical survey questions but obtained dissimilar results due to different question order. A recent trend in several industries today is to track the level of customer satisfaction over time using tracking studies. These tracking studies are implemented to show that a firm is a leader compared to its competitors in providing customer satisfaction in a particular industry. However, if several companies use different survey question orders for their attribute and satisfaction evaluations, the tracking results will not be directly comparable because question order effects may bias the results. Therefore, it is critical for multiple firms engaged in tracking studies to have consistent ordering of the attribute and satisfaction questions in a survey. Similarly, objective third parties who provide customer satisfaction ratings for consumers by compiling satisfaction information across firms need to be aware of the survey instruments used to collect the satisfaction data. Unless attribute and satisfaction question order is consistent across competitors, such information compiled from several firms may not give an accurate comparison of satisfaction ratings.

Another area that question order effects can have a profound effect in is new product design, product development, and quality improvement programs. Techniques such as conjoint analysis use questionnaires to assess not only the part-worth utility but also the overall evaluation of a product comprised of multiple attributes (Green and Srinivasan 1990). Our

research suggests that the order in which questions are asked to respondents can have a significant impact on how each individual attribute is perceived. Question order effects are also relevant for managers implementing quality improvement programs such as better customer service, more accurate billing, or improved after sales service. A typical method used by companies is to conduct a survey asking general questions on the overall performance of each quality improvement area, followed by more specific questions related to each of these broader areas (or vice versa). For example, a marketing manager interested in improving his or her firm's performance on delivery may ask two types of questions in a survey. One is the overall rating of the delivery process, ranging from poor to excellent. The second are more specific questions that fall under the umbrella of the overall delivery process such as ratings of being on-time, being complete, arriving without any damage, and safe and proof free packaging. The more specific questions offer managers actionable guidelines for prioritizing which product or process attributes to allocate resources for improvement. The quality improvement priorities derived from the survey results could vary by question order.

In spite of these potential pitfalls, our results suggest that impact scores may be relatively invariant to order effects. However, this finding may be limited to attributes whose evaluations are more permanently accessible and satisfaction ratings that are cumulative in nature. We suggest that managers use techniques such as the critical incident technique to elicit the product/service quality dimensions most important to their customers. These managers are more likely to use attribute survey items that have more permanently accessible evaluations and are less vulnerable to question order effects. This is reassuring to managers who use the impacts to set priorities for quality improvement. Benefit impact scores and satisfaction ratings, together with cost information, allow managers to make well-informed quality improvement decisions. An interesting opportunity for future research is to compare question order effects for temporarily accessible attribute evaluations versus more permanently accessible attribute evaluations and for transaction-specific satisfaction versus cumulative satisfaction.

An important question for managers is, "which survey question order is more appropriate?" We argue that the order one chooses should depend primarily on the type of product or service under investigation. For low involvement products, such as soft drinks and household staple products, where the mental framework of information processing is more of a top-down approach, customers are more likely to rely solely on stored evaluations when purchasing. This warrants more of an overall evaluations first type questionnaire format. Conversely, for high involvement products, such as computers or expensive household appliances, a bottom-up approach to information processing may be more prevalent.

Consumers may use more explicit problem solving (attribute and benefit consideration) prior to any final evaluation or repurchase decision. This suggests that asking attribute performance first may be more natural. An interesting future research opportunity would be to test if the level of product involvement moderates question order effects.

Acknowledgement

We would like to thank the editor for her encouragement throughout the review process and also the constructive comments from the reviewers.

Tables

Table 1. Latent and Measurement Variables

Latent Variables	Measurement Variables
Hair Quality	My hair care provider cuts my hair evenly.
	I like the way my hair looks after it is cut or styled.
	My hair care provider is precise.
	My hair care provider is skilled.
Provider Interaction	My hair care provider is careful.
	My hair care provider remembers my name.
Atmosphere	My hair care provider remembers things about me.
	I enjoy the topics that my hair care provider and I talk about.
Timeliness	I have fun while I'm at the shop.
	I like the general atmosphere in the shop
Scheduling	I don't have to wait very long to be served.
	I can get into and out of the shop quickly.
Customer Satisfaction	I can get an appointment when I need it.
	I can change an appointment when I need to.
	I can always get an appointment with the hair care provider I want.
Customer Loyalty	Overall satisfaction
	Overall performance versus expectations
	Overall performance versus an ideal hair care provider
	Likelihood of revisiting
	Likelihood of revisiting given a 10% increase in prices

Table 2. Average Community by Latent Variable

Average Community	Condition	
	<i>Attribute First</i>	<i>Overall Evaluation First</i>
Hair Quality	0.804	0.738
Provider Interaction	0.830	0.825
Atmosphere	0.884	0.882
Timeliness	0.756	0.746
Scheduling	0.740	0.655
Customer Satisfaction	0.810	0.731
Customer Loyalty	0.702	0.841

Table 3. Variance Explained (R^2) in the Latent Variables by Condition

Variance explained (R^2)	Condition	
	Attribute first	Overall evaluation first
Customer Satisfaction	0.746	0.618
Customer Loyalty	0.547	0.587

Table 4. Summary of Hypotheses

Hypotheses	Description	Results
Hypothesis 1	Impact scores <i>Attributes first</i> > Impact scores <i>Overall evaluations first</i>	Not supported
Hypothesis 2	R^2 <i>Attributes first</i> > R^2 <i>Overall evaluations first</i>	Partially supported
Hypothesis 3	CS (CL) <i>Attributes first</i> > CS (CL) <i>Overall evaluations first</i>	Supported

Note: CS and CL refer to customer satisfaction and customer loyalty, respectively

Figures

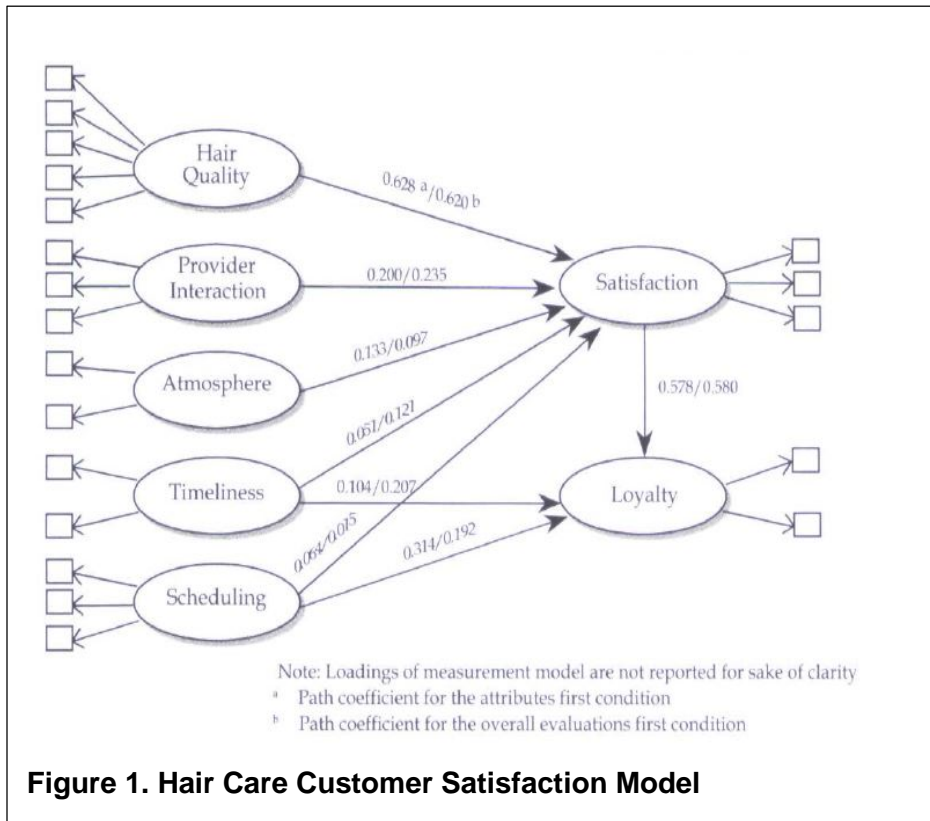
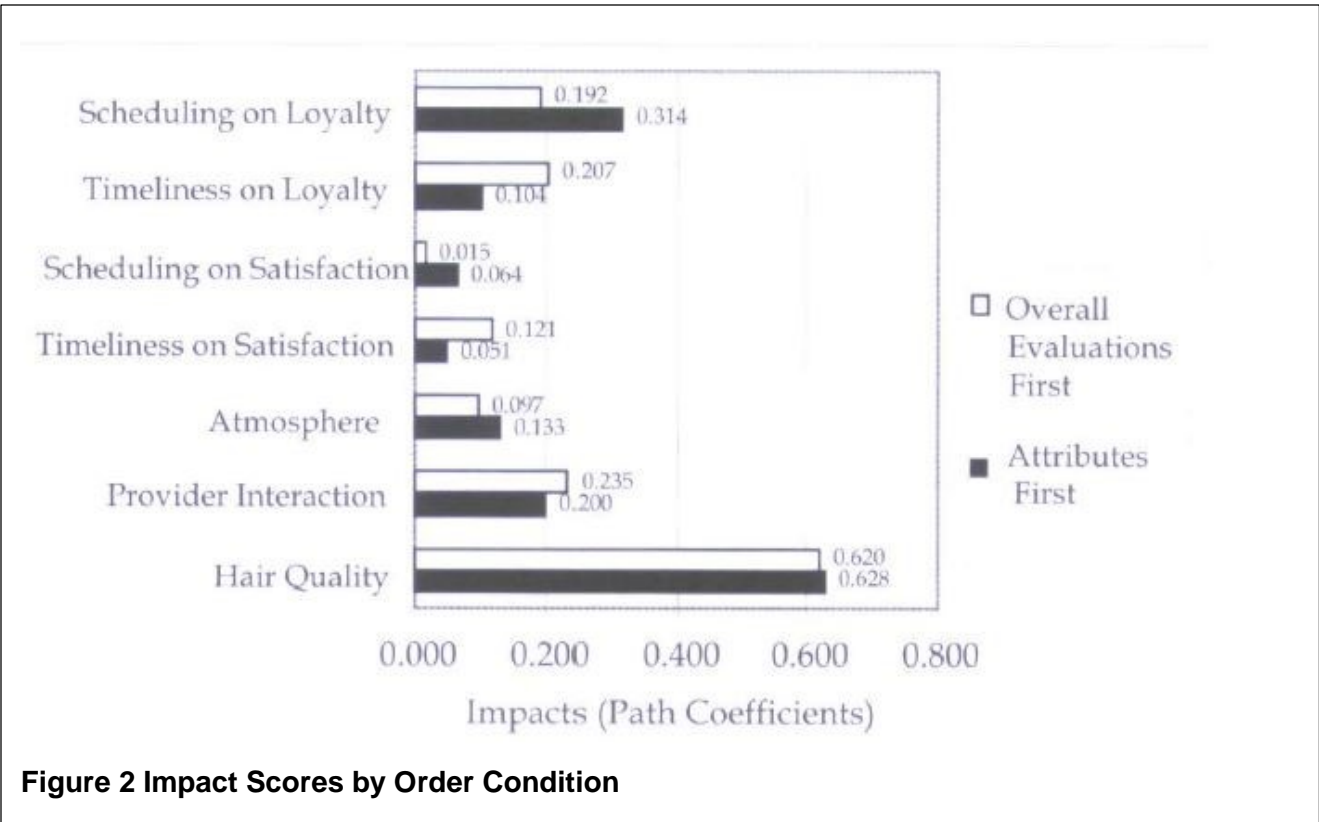


Figure 1. Hair Care Customer Satisfaction Model



References

- Bettman, J. R. and Zins, M. A. (1976), "Constructive Processes in Consumer Choice", *Journal of Consumer Research*, Vol. 4, September, pp.75-85
- Bettman, J. R., Luce, M. F. and Payne, J. W. (1998), "Constructive Consumer Choice Processes", *Journal of Consumer Research*, Vol. 25, December, pp.187- 217
- Bickart, B. (1993), "Carryover and Backfire Effects in Marketing Research", *Journal of Marketing Research*, Vol. 30, February, pp.52-62
- Bitner, M. J., Booms, B. H. and Tetreault, M. S. (1990), "The Service Encounters: Diagnosing Favourable and Unfavourable Incidents", *Journal of Marketing*, Vol. 54, January, pp.71-84
- Fazio, R. H., Powell, M. C. and Williams, C. J. (1989), "The Role of Attitude Accessibility in the Attitude-to-Behaviour Process", *Journal of Consumer Research*, Vol. 16, December, pp.280-288
- Feldman, J. M. and Lynch, J. G. (1988), "Self-Generated Validity and Other Effects of Measurement on Belief, Attitude, Intention, and Behaviour", *Journal of Applied Psychology*, Vol. 73, No. 3, pp.421-435
- Festinger, L. (1975), *A Theory of Cognitive Dissonance*, Stanford University Press
- Fishbein, M. and Ajzen, I. (1975), *Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research*, Reading, MA, Addison-Wesley
- Fornell, C. (1992), "A National Customer Satisfaction Barometer: The Swedish Experience", *Journal of Marketing*, Vol. 56 January), pp.6-21
- Fornell, C. (1995), "The Quality of Economic Output: Empirical Generalizations About Its Distribution and Relationship to Market Share", *Marketing Science*, Vol. 14, No.3, pp.203-211
- Fornell, C. and Cha, J. (1994), "Partial Least Squares", In: Bagozzi, R. P. (Ed.), *Advanced Methods of Marketing Research*, Cambridge, MA, Basil Blackwell, pp.52-78
- Fornell, C., Johnson, M. D., Anderson, E. W., Cha, J. and Bryant, B. E. (1996), "The American Customer Satisfaction Index: Nature, Purpose and Findings", *Journal of Marketing*, Vol. 60, October, pp.7- 18
- Green, P.E. and Srinivasan, V. (1990), "Conjoint Analysis in Marketing: New Developments with Implications for Research and Practice", *Journal of Marketing*, Vol. 54, October, pp.3-19

- Grice, H. P. (1975), "Logic and Conversation", In: Cole, P. and Morgan, J. L. (Eds.), *Syntax and Semantics: Speech Acts*, New York, Academic Press, pp.41-S8
- Grove, S. J. and Fisk, R. (1997), " The Impact of Other Customers on Service Experiences: A Critical Incident Examination of Getting Along", *Journal of Retailing*, Vol. 73, No. 1, pp.217-224
- Hastie, R. and Park, B. (1986), "The Relationship between Memory and Judgment Depends on Whether the Judgment Task is Memory-Based or On-Line", *Psychological Review*, Vol. 93, pp.258-268
- Hayes, B. E. (1998), *Measuring Customer Satisfaction: Development and Use of Questionnaires*, Milwaukee, ASQC Quality Press
- Higgins, E.T. (1996), "Knowledge: Accessibility, Applicability, and Salience", In: Higgin, E. T. and Kruglanski, A. (Eds.), *Social Psychology: Handbook of Basic Principles*, New York, Guilford
- Johnson, M. D. and Fornell, C. (1991), "A Framework for Comparing Customer Satisfaction Across Individuals and Product Categories", *Journal of Economic Psychology*, Vol. 12, pp.267-286
- Johnson, M. D. and Gustafsson, A. (2000), *Improving Customer Satisfaction, Loyalty and Profit: An Integrated Measurement and Management System*, San Francisco, Jossey-Bass
- Johnson, M.D., Gustafsson, A., Andreassen, T.W., Lervik, L. and Cha, J. (2001), "The Evolution and Future of National Customer Satisfaction Index Models", *Journal of Economic Psychology*, Vol. 22, pp.217-245
- Joreskog, K.G. (1970), "A General Method for Analysis of Covariance Structures", *Biometrika*, Vol. 57, pp.239-251
- McClendon, M. J. and O'Brien, D. J (1988), "Question-Order Effects on the Determinants of Subjective Well-Being", *Public Opinion Quarterly*, Vol. 52, Fall, pp.151-364
- Meuter, L. M., Ostrom, A. L., Round tree, R. I. and Bitner, M. J. (2000), "Self- Service Technologies: Understanding Customer Satisfaction with Technology-Based Service Encounters", *Journal of Marketing*, Vol. 64, July, pp.50-64
- Payne, J. W., Bettman, J. R. and Johnson, E. J. (1992), "Behavior Decision Research: A Constructive Processing Perspective," *Annual Review of Psychology*, Vol. 43, pp.87-131
- Peterson, R. A. (2001), "On the Use of College Students in Social Science Research: Insights from a Second-Order Meta-Analysis", *Journal of Consumer Research*, Vol. 28, December, pp.450-461

- Rust, R. T., Zahorik, A. J. and Keiningham, T. L. (1995), "Return on Quality (ROQ): Making Service Quality Financially Accountable", *Journal of Marketing*, Vol. 59, April, pp.58-70
- Schul, Y. and Schiff, M. (1993), "Measuring Satisfaction with Organizations-Predictions from Information Accessibility", *Public Opinion Quarterly*, Vol. 57, pp.536-551
- Schwarz, N. Strack, F, and Mai, H. P. (1991), "Assimilation and Contrast Effects in Part-Whole Question Sequences: A Conversational Logic Analysis", *Public Opinion Quarterly*, Vol. 55, Spring, pp.3-23
- Simmons, C. J., Bickart, B. A. and Lynch, J. G. (1993), "Capturing and Creating Public Opinion in Survey Research", *Journal of Consumer Research*, Vol. 20, September, pp.316-330
- Steenkamp, J. E. M. and van Trijp, H. C. M. (1996), "Quality Guidance: A Consumer-Based Approach to Food Quality Improvement using Partial Least Squares", *European Review of Agricultural Economics*, Vol. 23, pp.195- 215
- Sudman, S., Bradburn, N. M. and Schwarz, N. (1996), *Thinking About Answers: The Application of Cognitive Processes to Survey Methodology*, San Francisco, Jossey-Bass
- Tesser, A. (1978), "Self-Generated Attitude Change", In: Berkowitz, L. (Ed.), *Advances in Experimental Social Psychology*, New York, Academic Press, pp.289-338
- Tourangeau, R. and Rasinski, K. A. (1988), "Cognitive Processes Underlying Context Effects in Attitude Measurement", *Psychological Bulletin*, Vol. 103, pp.299-314
- Tourangeau, R., Rasinski, K. A., Bradburn, N. and D'Andrade, R. (1989), "Carryover Effects in Attitude Surveys", *Public Opinion Quarterly*, Vol. 53, pp.495-524
- Wold, H. (1982), "Systems Under Indirect Observation Using PLS", In: Fornell, C. (Ed.), *A Second Generation of Multivariate Analysis: Methods*, pp. 325-347, New York, Praeger