

Customer Satisfaction Across Organizational Units

by

Edward C. Malthouse
James L. Oakley
Bobby J. Calder
Dawn Iacobucci

July 2003

Authors' Note:

Edward C. Malthouse is an Associate Professor, Integrated Marketing Communications, Medill School of Journalism, Northwestern University. James L. Oakley is an Assistant Professor of Management, Krannert School of Management, Purdue University. Bobby J. Calder is the Charles H. Kellstadt Distinguished Professor of Marketing, Kellogg School of Management, Northwestern University. Dawn Iacobucci is Professor of Marketing, Kellogg School of Management, Northwestern University. The authors would like to thank the Media Management Center at Northwestern University for financial support and assistance and Solucient for allowing us to use their Healthplus survey data.

Direct all correspondence to Edward C. Malthouse, Integrated Marketing Communications, Northwestern University, 1845 Sheridan Road, Evanston, IL 60208-2175; phone 847-467-3376; fax 847-491-5925; email ecm@northwestern.edu.

Customer Satisfaction Across Organizational Units

Abstract

This paper examines customer satisfaction models for assessing the relationship of overall satisfaction with a product or service and satisfaction with specific aspects of the product or service for organizations having multiple units or subunits. These units could be stores, markets, dealers, divisions, etc. We suggest a methodology for studying whether the drivers of overall satisfaction vary across such units. For cases where the drivers do vary across subunits, we show how additional variables can be included in a model to account for the variation. We illustrate this approach by studying customer satisfaction in the newspaper and healthcare industries. We use Generalizability theory can be used to evaluate the reliability of scales from multi-stage cluster sample designs. It is argued that the approach has important implications for both theory and practice.

Introduction

Many studies have related overall satisfaction with some product or service to satisfaction with specific aspects of the product or service (Oliver 1980, 1993; Parsuraman, Berry, and Zeithaml 1988, 1991; Anderson and Sullivan 1993; Garbarino and Johnson 1999; DeWulf, Odekerken-Schröder, and Iacobucci 2001). Customers may explain their satisfaction with a product or service in terms of specific aspects such as the product attributes, price, customer service, or a combination of these various features. The objective of such studies is to understand how specific types of customer satisfaction affect overall satisfaction, usually by examining the slopes from a regression analysis. This paper extends this approach by allowing the slopes to vary over predefined “subunits” of customers. We hypothesize that different subunits within an organization or industry may show different relationship between specific aspects of satisfaction and overall satisfaction, i.e., there may be different utilities for the specific aspects of satisfaction.

The problem of whether the relationship between specific aspects of satisfaction and overall satisfaction varies by subunits has both practical and theoretical importance. As a practical matter, such variation could be important for marketing decisions. For example, an automotive manufacturer may have multiple dealers (the subunit). A marketing manager would want to know if all dealers should focus on the same aspects of satisfaction or whether the customers of one dealer may have different priorities than another. If there is variation in the utilities across subunits, can the variation be “explained” by, for example, the geographic location of the dealership? A second example is a national retailer with multiple stores (the subunit). It would not be surprising for consumers in densely populated urban areas to place a high utility on

dimensions such as location and convenience while these same dimensions might be less important in sparsely populated rural areas. A third example is a media organization with multiple properties (subunits). Newspaper owners often own several newspapers (subunits) in different markets. Should all of the owner's newspapers focus on the same customer satisfaction dimensions? Banks have multiple branches. Perhaps the drivers of satisfaction for large branches are different than for those of small branches?

Variation in the specific-general satisfaction relationship across organizational subunits also has important theoretical implications for satisfaction research. The goal of theoretical research is to test universal hypotheses that apply across observational units (Calder, Phillips, and Tybout 1981, Calder and Tybout 1999). Research attempts to expose these hypotheses to rejection by the empirical test. A study of the specific-general satisfaction relationship in a single organization provides such a test. However, testing the relationship across several organizational units provides an even stronger test in that the theoretical relationship is exposed to additional opportunities for empirical rejection. And, beyond this, if the hypothesized relationship is not found for some units, this offers the possibility of developing richer theoretical hypotheses that take into account the effects of other variables.

Much of academic services marketing research is of the single organization sort. It often posits certain effects and evaluates the extent to which the effects hold using a random sample of customers from a single company (Schlesinger and Zornitsky 1991; Hallowell 1996; Loveman 1998; Garbarino and Johnson 1999; Bolton 1998). Occasionally, the effect will be evaluated on a small convenience sample of companies (Parasuraman, Berry, and Zeithaml 1991; Zeithaml,

Berry, and Parasuraman 1996). While such studies are certainly important, they are not strong tests in the above sense. The ideal study would be one with a *random* sample of organizational units *and* a *random* sample of consumers from each selected unit.

Thus, for both practical and theoretical reasons, this paper focuses on the extent to which specific-general satisfaction effects vary across units. If the effects are the same across units, a manager may be able to use one strategy for all units. To the extent that effects vary across units, the company would want to consider different strategies for different units. And, at the theoretical level, the multiple units provide a stronger test of a hypothesized general effect.

We also stress the importance of explaining the variation in effects across units or subunits. One way to approach this question is to partition the units or subunits into strata. For example, the locations of retail stores could be classified into rural, small city, suburban, and urban types. We want to quantify how much variation in effects there is both within and across strata. If the within-stratum variation is small and the between-stratum variation is great (e.g., rural stores all have the same needs but rural stores have different needs than urban ones), the manager might develop separate strategies for each stratum. The academic researcher likewise would postulate a richer theory incorporating the strata as variables.

In this paper we present methods for addressing these issues. The methods are applied to multiunit data from two different industries. We illustrate how these methods could be useful to a marketing manager of a particular company and how they can be used to study “general truths” in marketing.

Literature Review

Since we are proposing a method for analyzing the dependence of overall satisfaction with a product or service on specific aspects of customer satisfaction, our review of the relevant literature will begin with a brief discussion of the extant literature on customer satisfaction.

Customer Satisfaction

Customer satisfaction is a key and valued outcome of good marketing practice. According to Drucker (1954), the principle purpose of a business is to create satisfied customers. Increasing customer satisfaction has been found to lead to higher future profitability (Anderson, Fornell, and Lehmann 1994), lower costs related to defective goods and services (Anderson, Fornell, and Rust 1997), increased buyer willingness to pay price premiums, provide referrals, and use more of the product (Reichheld 1996; Anderson and Mittal 2000), and higher levels of customer retention and loyalty (Fornell 1992; Anderson and Sullivan 1993; Bolton 1998). Increasing loyalty, in turn, has been found to lead to increases in future revenue (Fornell 1992; Anderson, Fornell, and Lehmann 1994) and reductions in the cost of future transactions (Reichheld 1996; Srivastava, Shervani, and Fahey 1998). All of this empirical evidence suggests that customer satisfaction is valuable from both a customer goodwill perspective and an organization's financial perspective.

A firm's future profitability depends on satisfying customers in the present – retained customers should be viewed as revenue producing assets for the firm (Anderson and Sullivan 1993; Reichheld 1996; Anderson and Mittal 2000). Empirical studies have found evidence that

improved customer satisfaction need not entail higher costs, in fact, improved customer satisfaction may lower costs due to a reduction in defective goods, product re-work, etc. (Fornell 1992; Anderson, Fornell, and Rust 1997). However, the key to building long-term customer satisfaction and retention and reaping the benefits these efforts can offer is to focus on the development of high quality products and services. Customer satisfaction and retention that are bought through price promotions, rebates, switching barriers, and other such means are unlikely to have the same long-run impact on profitability as when such attitudes and behaviors are won through superior products and services (Anderson and Mittal 2000). Thus, squeezing additional reliability out of a manufacturing or service delivery process may not increase perceived quality and customer satisfaction as much as tailoring goods and services to meet customer needs (Fornell, Johnson, Anderson, Cha, and Everitt 1996).

Measuring Customer Satisfaction

While it seems clear that increasing customer satisfaction is beneficial to a marketing manager, how to measure it is less clear. Customer satisfaction has been studied from the perspective of the individual customer and what drives their satisfaction (Oliver and Swan 1989; Oliver 1993; Fournier and Mick 1999) as well as from an industry-wide perspective to compare customer satisfaction scores across firms and industries (Fornell 1992; Anderson, Fornell, and Lehmann 1994; Fornell et al. 1996; Mittal and Kamakura 2001), while other research has examined customer satisfaction in a single organization (Schlesinger and Zornitsky 1991; Hallowell 1996; Loveman 1998) or across several organizations (DeWulf, Odekerken-Schröder, and Iacobucci 2001). In addition, specific tools for measuring customer satisfaction have been developed in the

past, including SERVQUAL (Parasuraman, Berry, and Zeithaml 1988, 1991). Thus, there exists an ample literature on which to draw when attempting to measure customer satisfaction.

In attempting to measure customer satisfaction, it is possible that attributes can have different satisfaction implications for different consumer and market segments – the usage context, segment population, and market environment can influence satisfaction and product use (Anderson and Mittal 2000). Failure to take into account segment-specific variation may lead a firm to focus on the wrong aspect for a given set of consumers (Anderson and Mittal 2000). Furthermore, consumers with similar satisfaction ratings, yet different characteristics, may exhibit different levels of repurchase behavior (Mittal and Kamakura 2001). It is clear, then, that market and consumer segments should be important factors to consider when measuring customer satisfaction and its implications.

Garbarino and Johnson (1999) did consider segments in the customer base in their study of satisfaction where they analyzed the different role played by satisfaction between low relational and high relational customers. Their study, however, involved customers from only a single organization. Our approach extends this work by studying customers from multiple organizations, and shares some similarities with Anderson and Sullivan (1993) with respect to the type of analysis and sampling methods. The goals of their research, however, were to study the antecedents and consequences of customer satisfaction rather than investigate how different types of satisfaction may influence the overall measure of customer satisfaction. In addition, our theoretical approach shares some similarities to Hutchison, Kamakura, and Lynch (2000) who posited that unobserved heterogeneity is a problem for interpreting results from behavioral

experiments. The basic point of their argument is that aggregation may create effects that do not exist in any segments, or may mask effects that do exist. The present study makes a similar point and provides an analytical method for overcoming such a problem.

Kekre, Krishnan, and Srinivasan (1995) examine heterogeneity of effects across individual customers of a single company using a random effect ordered probit model. These models are similar to the hierarchical linear models considered here, and a single customer could be considered a subunit. Our study extends this previous work by allowing for multiple levels of randomization. For example, we have random samples of organizations and random samples of subunits within the organizations. An additional extension is that we attempt to explain the variation across subunits.

Subsegments vs. Subunits

Other authors have examined the heterogeneity of customer satisfaction effects. Danaher (1998) shows how latent class regression can be used to segment customers and estimate regression effects by segment simultaneously. Our work is different in that we assume pre-defined subunits – our concern is not to *define* segments that have different effects. For the problems examined here, the subunits already exist. Danaher (1998) identifies segments of customers (end users) who place different emphasis on different service attributes. Malthouse (2002) defines such a process as *subsegmentation*. A firm has targeted a *market segment* and acquired customers/end users. It then *subsegments* these customers/end users from a *market segment* into smaller, more homogeneous groups based on some criteria such as utility for aspects of the product in the case of Danher (1998).

An important conceptual question concerns when one approach should be preferred over the other. We make two points in response to this question. First, the pre-defined subunit approach to studying heterogeneity is more appropriate when the resulting managerial actions will be implemented at the subunit level. Second, managerial actions implemented at the subunit level are most reasonable when there is homogeneity within a subunit and heterogeneity across subunits; when this is not the case the organization should seek actions that can be implemented for subsegments of customers within a subunit. We give several examples to illustrate these points.

Consider the case of a newspaper owner, discussed in more detail below. An owner in the U.S. has multiple newspapers and wants to know whether to invest in improving either the service or the content of its individual papers. Investing in content could involve hiring additional reporters so that local news can be covered more thoroughly, adding pages to existing sections, adding special-interest sections, etc. For most newspapers in the U.S. these actions would have to be taken at the subunit level. One might object by suggesting, for example, that large metropolitan newspapers (which represent only a small percentage of U.S. newspapers) could improve content for specific suburban communities by hiring reporters and adding customized local sections. We would argue that the suburban “zone” would be a subunit. A second example can be when actions primarily involve reach media. If a company is communicating a single message with, for example, television, newspapers, billboards, etc., the message must be tailored to the subunit reached by the media. A third example is managerial actions that are most naturally applied at the subunit level of retail stores, car dealerships, supermarkets, and bank branches, as discussed

previously. A corporation could send employees of certain subunits, but not all, for specialized customer service training programs. Corporations often choose where to locate subunits, and might opt for more expensive locations in regions where “convenience” is more important. In addition, pricing strategies often must be executed at the subunit level (Singh, Chintagunta, and Dube 2002).

Of course, there are numerous examples of situations where customer subsegmentations are more appropriate. See Danaher (1998) or Malthouse (2002) for further discussion and examples.

The present research represents the first study of which we are aware to measure customer satisfaction from a representative sample of customers who are in turn from a representative sample of organizations in a single industry. The analysis was replicated in a second industry to confirm that the findings are not unique to a single industry.

Customer Satisfaction And Heterogeneity

Answering the two key questions we have posed, 1) the extent to which effects vary across subunits and 2) explaining the variation in effects across subunits, requires a special sampling design. This section discusses the designs and models that are required to answer these questions.

The first question asks how much variation there is across subunits. Answering this requires a two-stage cluster sample.

1. Draw a random sample of subunits. We demonstrate this with samples of organizations from two different industries, daily U.S. newspapers and HMOs.

2. For each sampled organization, draw a random sample of consumers who are familiar with the organization's product or service.

The primary sampling units (PSUs) are organizations and the secondary sampling units (SSUs) are consumers. For example, we study a newspaper owner below by first drawing a sample of newspapers owned by the company and then consumers in the newspaper markets.

Data from such designs can be easily analyzed in readily available commercial software packages that estimate *mixed* linear models such as SAS (Littell, et al. 1996) or S-PLUS (Pinheiro and Bates 2000); we estimate all models in this paper using proc mixed in SAS under the default restricted maximum likelihood (REML). The specific mixed model that we will use is called a *random coefficient* or *hierarchical linear* model (Bryk and Raudenbush 1992; Kreft and deLeeuw 1998). To illustrate these models, suppose that we have specific types of satisfaction; in the newspaper example below, these will be satisfaction with the newspaper content and satisfaction with service. Let y_{ij} , x_{ij1} , x_{ij2} denote the measures of overall, content, and service satisfaction, respectively, of customer j in the market of newspaper i . We assume that

$$(1) \quad y_{ij} = (\beta_0 + b_{i0}) + (\beta_1 + b_{i1})x_{ij1} + (\beta_2 + b_{i2})x_{ij2} + e_{ij},$$

where (b_{i0}, b_{i1}, b_{i2}) is a normal random vector of regression coefficients¹ with mean $(0,0,0)$ and e_{ij} is normal with mean 0 and variance σ^2 . The values of $(\beta_0, \beta_1, \beta_2)$ are estimates of the coefficients for the entire population. This model looks very similar to a multiple regression of y_{ij} on the two predictor variables. One difference is that the intercepts and slopes are divided into two

components. For the first predictor variable x_1 , the industry average is β_1 , and b_{i1} indicates how subunit i differs from the industry average. The variance of b_{i1} is usually of great interest. If the variance is 0 then all of the subunits have the same slope suggesting that the manager does not need vary the emphasis on this dimension across subunits; if the variance is great then different subunits have different utilities for this dimension.

Another difference between the random coefficient model proposed in (1) and a multiple regression is the way that the parameters are estimated. One could fit a multiple regression for every subunit, which would involve estimating four parameters, (α, b_{i1}, b_{i2}) and the variance of e_{ij} , for every subunit. Thus, in the newspaper study described below with 101 newspapers, we would estimate 404 parameters. Computing the average of the 101 estimates (a_i, b_{i1}, b_{i2}) would be analogous to the values of $(\alpha, \beta_1, \beta_2)$ from the random coefficient model. A problem with this approach is that it treats every newspaper separately and does not exploit the possibility that newspapers may have aspects in common. Notice that the random coefficient model in (1) involves 7 free parameters (the intercept, 2 slopes, 3 variances for the random intercepts and slopes, and the error variance) rather than 404 as described above. This count is the same regardless of the number of subunits (newspapers) sampled. Just as with random effect ANOVA models, inference on $(\alpha, \beta_1, \beta_2)$ and the variances of the random coefficients applies to the population from which the subunits were sampled (assuming we have a random sample of organizations).²

¹ Hierarchical linear models also usually assume that the random coefficients (b_{i0}, b_{i1}, b_{i2}) are uncorrelated with the error term e_{ij} and that the random coefficients are uncorrelated with each other. This is called the *variance component* model in SAS.

² We know from basic introductions to the linear model that we should not generalize to the population when the model is comprised of fixed effects. Given that we are randomly sampling organizations, we will be in a stronger

The second question concerns *explaining* the variation in effects. The approach we follow here is to group the subunits into types. For example, retail stores could be classified into three types, urban, suburban, or rural. If there were considerable variation across types but little variation within type, the manager would need different strategies for different types. Alternatively, if there were little variation across types but substantial variation within types, the proposed typology would be of little use to the manager.

Before developing a model for this situation, it is necessary to decide whether the types are fixed or random. This decision will imply slightly different models, analogous to fixed- and random-effect ANOVA models. We begin with the random case, which is more relevant to the academician. It is appropriate to treat the types as random if the types can be regarded as a “random” sample from some larger universe of types. In sampling vernacular, this is a three-stage cluster sample. There is first a random sample of types, second a random sample of subunits nested within the types, and third a random sample of customers within each subunit. An academic researcher studying the newspaper industry could draw a random sample of owners (types), then a random sample of newspapers (subunits) owned by each of the selected companies, and finally a random sample of consumers within each selected newspaper market. A fixed effects model is appropriate when the types are not to be regarded as a random sample. The urban-suburban-rural typology above would be best modeled as a fixed effect.

position to make statements of external validity. To do so, the data must be analyzed properly as random effects, e.g., via random coefficient models, rather than as fixed effects, e.g., as separate regressions.

Let y_{ijk} , x_{ijk1} , x_{ijk2} denote the measures of overall, content, and service satisfaction, respectively, of customer k in the market of newspaper j of type i . For fixed effects models, we assume that

$$(2) \quad y_{ijk} = (\alpha + \alpha_i + a_{ij}) + (\beta_1 + \beta_{i1} + b_{ij1})x_{ij1} + (\beta_2 + \beta_{i2} + b_{ij2})x_{ij2} + e_{ijk},$$

where $(a_{ij}, b_{ij1}, b_{ij2})$ is a normal random vector of regression coefficients with mean $\mathbf{0}$ and e_{ij} is normal with variance σ^2 . The variables $(\alpha, \alpha_i, \beta_1, \beta_{i1}, \beta_2, \beta_{i2})$ are fixed effects. The slopes and intercepts have been divided into a sum of three components. Consider the slope for x_{ij1} . The overall average slope (for the entire company or industry) is β_1 . The effect for type i is β_{i1} ; for example if the first type is rural, the second type is urban, and the slope for urban stores is systematically greater than for rural stores on this dimension, then $\beta_{21} > \beta_{11}$. The effect for a specific subunit j is b_{ij1} . Our interest will primarily be with the variance of b_{ij1} ; if the variance is 0 then this variable has the same effect for all subunits within the type and the manager would not need to vary this dimension. If the variance is large, the manager would want to consider varying this dimension across subunits even within type.

The random effects model is very similar:

$$(3) \quad y_{ijk} = (\alpha + a_i + a_{j(i)}) + (\beta_1 + b_{i1} + b_{j(i)1})x_{ij1} + (\beta_2 + b_{i2} + b_{j(i)2})x_{ij2} + e_{ijk},$$

where $(a_i, b_{i1}, b_{i2}, a_{j(i)}, b_{j(i)1}, b_{j(i)2})$ is a normal random vector of regression coefficients with mean $\mathbf{0}$, and e_{ij} is normal with variance σ^2 . Again, the intercepts and slopes are partitioned into three components. Suppose we are academic researchers with a random sample of owners, a random

sample of newspapers from the sampled owners, and a random sample of consumers from the sampled newspapers. For the first variable x_{ijl} , the overall average slope (for the industry) is β_1 . The effect for owner i is b_{i1} . For example, if the slope for this dimension for owner 1 is greater than the slope for owner 2, then $b_{11} > b_{21}$. We will be primarily interested in the variance of b_{i1} . If the variance is 0 then this dimension has the same effect for all owners. If the variance is large then the effect is different across owners. The effect for newspaper j of owner i is $b_{j(i)1}$. If the variance of this term is 0 then the effect is the same for all newspapers within an owner.

It is possible to estimate the slopes for individual organizations using empirical best linear unbiased predictors (EBLUPs, see Laird and Ware 1982; SAS Institute 1997, pp. 640-1; Pinheiro and Bates 2000, section 2.5). If the sample sizes in subunits are sufficiently large, one could also fit separate OLS regressions or a single ANCOVA model (interactions between subunit and predictor variables with a common error variance).

Empirical Evaluation

We evaluate our hypotheses using data from the newspaper and health care industries. For each of the industries, we develop extensive case studies by developing measures of overall satisfaction and satisfaction with specific attributes, estimating the random coefficients model described above, and discussing the implications of the fitted parameters.

Newspaper Satisfaction

The objective of this analysis is to understand how specific types of customer satisfaction affect overall satisfaction with the newspaper. In particular, we will examine how satisfaction with the content of a newspaper and the customer service offered by the newspaper affect overall satisfaction.

Data. Our sampling plan can be summarized briefly as follows. We first compiled a sampling frame of 864 daily U.S. newspapers using lists of newspapers from the Newspaper Association of America, the Audit Bureau of Circulation³ (ABC), and *Editor and Publisher*. We drew a stratified random sample of 101 U.S. daily newspapers, stratifying on market and newspaper characteristics such as circulation, urbanicity, competition, market penetration, and the geographical extent of distribution. We mailed 110,000 surveys to consumers in the 101 newspaper "markets," where a market is the set of zip codes that account for at least 80% of circulation; markets were defined using data from the Audit Bureau of Circulation. The number of surveys mailed to each market was selected to produce approximately the same number respondents in an effort to provide a balanced sample of consumers. We included a \$3 incentive with each survey. In total, 37,036 responded, giving a response rate of 37% after dropping undeliverable surveys. The distribution of the number of responses in each market had a mean of 337, standard deviation of 46, minimum of 271, and a maximum of 472. We then did a telephone survey of 2000 non-responders. The telephone survey was used in forming weights for readers and nonreaders, along with U.S. Census data on age and gender.

Scale Development. The questionnaire included items measuring satisfaction with content and customer service of a particular newspaper. The exact question wording of all items is provided in the appendix. We conducted a factor analysis estimated under maximum likelihood on the 38 items, followed by a varimax rotation. A scree test suggested two meaningful factors. In interpreting the rotated factor pattern, we considered an item to load on a given factor if the factor loading was .5 or greater. Using this criterion, we found that one question from the service scale did not load highly on either factor: Telemarketing calls. This item was dropped from further analyses. The three items “classified ads,” “inserts for food,” and “ads for clothing, health, and stores” did not have loadings greater than .5 and were dropped. Coefficient alpha for the content and service scales was .96 and .89, respectively. We used the simple averages of the items as estimates of the scales in subsequent analyses. The Pearson correlation between the two factors is .46.

Our method of measuring customer satisfaction follows the disconfirmation paradigm in the marketing literature (Oliver 1980, 1993). We measure overall satisfaction with the newspaper with the following question measured on a 5-point scale: "Overall, how would you rate the ___ newspaper? Even if you are not very familiar with the paper, rate how good you think it would be." Drolet and Morrison (2001) suggest that single-item scales may be used in this context.

We also evaluated reliability using generalizability theory (Rentz 1987; Finn and Kayandé 1997). We assume the role of a manager working for the owner of multiple newspapers. We take newspapers as the facet of differentiation and use the relative error variance component in

³ See <http://www.accessabc.com>.

computing reliability estimates. The sources of variation are the person, item, newspaper, and item×newspaper. All variance components were estimated in proc mixed using SAS version 8.2. The reliabilities of the content scale are .9587, .9637, and .9452 for owners A, B, and C, respectively. The reliabilities for the service scale are .7398, .9912, and .9183.

Results. We first show how these methods are relevant to the owner of a specific chain of newspapers. The three largest newspaper owners all had 6 or more newspapers in the sample of 101. We estimate the model in equation (1) separately for each of these three owners, labeled Owner A, B, and C. The estimates of this model are presented in Table 1. First consider Owner A. The overall slope for content is .44 and the overall slope for service is .18 with *t*-statistics 16.79 and 6.06, respectively. Bring (1994) recommends using *t*-statistics in comparing the “importance” of predictor variables; with this criterion content is a stronger predictor of satisfaction than service. As always, managerial judgment and expertise should be used comparing regression coefficients, and the owner must consider the costs of moving readers one unit along the content and service satisfaction scales. For example, satisfaction with content could conceivably be improved with tactics such as running in-paper promotion (e.g., telling readers that the content is excellent), adding additional national and international wire stories, hiring additional reporters so that more local stories can be covered, having existing staff write in a style more pleasing to the audience, etc. Some of these tactics are inexpensive while others are very expensive. These costs could vary across newspaper and should be taken into account when deciding which actions to take.

The $V(b)$ column gives the variance of the newspaper-specific effects for this owner. Note that the variances for content and service are both close to zero. In fact, we cannot reject (at the .05 level) the null hypothesis that the variances are 0 using the Wald Z tests provided by SAS. We thus conclude that the slopes are equal for all newspapers owned by this owner. One strategy works for all the newspapers owned by Owner A.

Insert Table 1 Here

The conclusions for Owners B and C are the same. For Owners B and C content has a steeper slope than service and this is true for all newspapers owned by the companies (since $V(b)$ is not significantly different from 0 for content or service for either owner). The approach suggested in this paper is thus useful to the manager.

As academic researchers, we note that the effects are strikingly similar across owners as well. The “grand” effects for content are .44, .51, and .52 and the grand effects for service are .18, .17, and .16. We can reject the null hypothesis that the slope for content equals the slope for service in all three cases with $P < .01$. The t -statistics for content are 16.79, 9.03, and 25.44 and the t -statistics for service are 6.06, 4.12, and 8.02. This suggests a general truth for the newspaper industry: to increase overall satisfaction with any newspaper, satisfaction with content has a greater effect than satisfaction with service. Thus far, we have exhibited strong evidence in support of this statement for the three largest owners, but is this true for all owners and newspapers? To answer this question, consider that we have a random sample of owners and a

random sample of newspapers owned by these companies. So that we can make within-owner variance estimates, we take only the owners with two or more newspapers in the sample. With this restriction, we have 17 owners and 57 newspapers. We estimate the model specified by equation (3) with the estimates in Table 2.

Insert Table 2 Here

The overall slope for content is .5081 and for service .1478 and we can reject the null hypothesis that the slopes are equal with $P < .0001$. These values are similar to the slopes for the three largest owners. The inter-newspaper (within-owner) variance of the content slope is not significantly different from zero and the variance of the service slopes is small, although $P = .024$. This suggests that the newspapers owned by each of these respective owners are pretty similar to each other and that these owners can use a single strategy for all newspapers in the family. The variances across owners are small and barely significant with this large sample.

Health Insurance Satisfaction

The objective of this example is similar to the newspaper example. We want to see how *overall* satisfaction with a health plan depends on two specific types of satisfaction: satisfaction with the *costs* and satisfaction with the *medical care*. There are several managerial questions that are important to large health insurers. Health insurers offer several distinct types of plans, including Health Maintenance Organizations (HMOs), Preferred Provider Organizations (PPOs), and Point of Service (POS) plans. HMOs tend to cost less than PPOs, but PPOs offer greater flexibility. In general, POS plans attempt to combine the cost savings of an HMO with the flexibility of PPOs. In managing the overall satisfaction with various plans, can insurers use the same strategies for all three types of plans? A second question concerns inter-market variation. Different markets

within the US are in different stages of evolution and are governed by different state-specific legislation. Does an insurer need to use different strategies in different markets? Will a strategy that works in California also work in Tennessee? Answering these questions will require a model of the form in equation (2).

Data. We use data from the 1997 HealthPlus survey conducted by Solucient (see <http://www.solucient.com>). This survey used a combination of phone and mail to collect information about healthcare attitudes and opinions from a random sample of 108,007 people in 35 Scarborough⁴ markets. Respondents were recruited with a phone survey during which basic demographic information was recorded. Those who agreed to participate were mailed an extended questionnaire. We shall restrict our attention to those who indicated that their primary healthcare plan was HMO, PPO, or POS, dropping observations indicating a Medicare HMO. The question wording used to determine the type of healthcare plan is provided in the appendix.

We take the perspective of a manager at a large insurance company offering HMO, PPO, and POS plans in multiple markets. In particular, we evaluate the effects of specific types of satisfaction for three large insurance companies, labeled A, B, and C. Company A has 13 markets in this study and a total sample size of 2,347. Company B has 11 markets in the study and a sample size of 1,732. Company C has 11 markets and 1,309 observations.

Scale Development. The survey included questions measuring overall satisfaction and satisfaction with the plan administration and medical care. We performed a factor analysis

⁴ Scarborough Research is a market research firm that regularly collects data in 75 U.S. markets. Additional information on Scarborough and their products can be obtained at <http://www.scarborough.com/>.

estimated with maximum likelihood, followed by a varimax rotation. The scree test indicated that the 2-factor solution was reasonable. In interpreting the factors, again, we considered a loading .5 or greater as indicating an item loaded on a particular factor. All of the items under medical care loaded on the same factor, labeled *medical care*. The remaining items loaded on the second factor, labeled *cost*. The values of alpha for medical care and cost were .94 and .87, respectively. We estimated factor scores with the simple average of the items. We also evaluated reliability using generalizability theory. We take the perspective of a researcher at one of the three companies and use market as the facet of differentiation. The sources of variation are the person, item, market, and item \times market. The reliabilities for the medical care scale are .9803, .9753, and .9338 for companies A, B, and C, respectively. The reliabilities for the cost scale are .9664, .7393, and .8813.

Results. We estimate the model in equation (2), assuming a random sample of markets (subunits) and a random sample of consumers within the selected markets. The model is estimated for three large insurers with the results summarized in Table 3.

Insert Table 3 Here

We begin our discussion of the results in Table 3 by examining variation in effects across markets. The variation in slopes across markets is indicated in the $V(b)$ column. Company B has significant variation in the slopes for both medical care and cost across markets, suggesting that this company may have to tailor its strategies for different markets. The other companies have

little variation in slopes across markets, suggesting that there is less need for individual-market strategies.

For Company C, the overall slope for medical care is significantly greater than that for cost – the null hypothesis $\beta_{\text{Med Care}} = \beta_{\text{Cost}}$ has $P=.0009$. We cannot reject this null hypothesis for Company A. Company B should evaluate this at the market level, since it has significant variation in slopes across markets. The costs of implementing tactics to affect customer satisfaction on these dimensions should be considered.

Does the importance of satisfaction with cost and medical care vary by plan type? For company A, satisfaction with medical care has larger slopes for HMOs than for PPOs or POS products and satisfaction with cost has a larger slope for PPOs and POS products; PPOs and POS products are not significantly different on either dimension. These directional results are not surprising in view of the nature of these products. This suggests similarity between PPO and POS products offered by A. For Company B, the cost slopes for HMOs and POS products are not significantly different, but all other differences are significant. No differences are significant for Company C.

Discussion

Customer satisfaction studies that examine the dependence of overall satisfaction with a product or service on various specific features of the product or service are common. This paper gives empirical results from the newspaper and healthcare industries that show that the nature of the dependence can vary substantially across subunits (stores, markets, etc.) of an organization. For one subunit, some specific type of satisfaction may be a strong predictor of overall satisfaction while for another subunit the same specific type of satisfaction may have little or no relationship

to overall satisfaction. In such cases the organization may need different strategies for different subunits. Moreover, these results indicate the need for richer theoretical hypotheses including more variables.

This paper also indicates the utility of the methodology used for studying the variation in effects across subunits. An organization draws a random sample of subunits (many firms in practice regularly measure satisfaction for all subunits) and a random sample of customers within the subunits. Hierarchical linear models (HLM) are used to evaluate (1) how strongly each specific type of satisfaction is related to overall satisfaction and (2) whether the strength of these relationships varies across subunits. Because the subunits were selected randomly, the inference from the HLMs can be extended to the population from which the subunits were sampled. Thus, a firm may be able to reduce costs of satisfaction studies by not sampling every subunit. Of course, if a firm is using satisfaction surveys to monitor satisfaction levels of individual subunits, for example to be used in determining compensation of managers in that subunit, the firm will have to draw sufficiently large samples in every subunit. Academic researchers evaluating theory with a random sample of subunits (companies/organizations) can assess to what extent the theory applies across companies.

In cases where the drivers of satisfaction vary across subunits, this paper shows how to include additional variables in the model to account for such variation. For example, customers of a health insurance provider have different utilities for medical quality and cost depending on whether the customer has an HMO, PPO, or POS plan.

One would expect that predictors of customer satisfaction would vary across organizational units in some instances, but not in others. We provide illustrations of situations where there is no significant variation across subunits as well as situations where such variation exists. In the case where this variation is present, the results obtained in this research provide evidence for such variation and a means by which to analyze data from a set of such organizational units.

Table 1

Parameter estimates of hierarchical model for newspaper satisfaction for the three largest owners

	Owner A			Owner B			Owner C		
	β	<i>t</i> -Stat	<i>V</i> (<i>b</i>)	β	<i>t</i> -Stat	<i>V</i> (<i>b</i>)	β	<i>t</i> -Stat	<i>V</i> (<i>b</i>)
Intercept	1.13**	12.83	0.0000	0.67**	4.39	0.0916	0.78**	9.25	0.0152*
Content	0.44**	16.79	0.0000	0.51**	9.03	0.0151	0.52**	25.44	0.0000
Service	0.18**	6.06	0.0018	0.17**	4.12	0.0066	0.16**	8.02	0.0004
Residual			.7061**			.1623**			.5961**

Note: * indicates $P < .05$ and ** indicates $P < .01$

Table 2

Parameter estimates of the three-level (owners, newspapers, consumers) model for the newspaper industry

	β	t -Stat β	Inter-Owner Variance $V(b)$	Inter-Newspaper (Within Owner) Variance $V(b_{i(i)})$
Intercept	0.8445**	16.77	.0141*	.0222**
Content	.5081**	38.38	.0014*	0.0000
Service	.1476**	11.57	.0010*	0.0008*
Residual			0.5569**	

Note: * indicates $P < .05$ and ** indicates $P < .01$

Table 3
Parameter estimates from three health-care insurance companies

	Company A			Company B			Company C		
	β	t -Stat	$V(b)$	β	t -Stat	$V(b)$	β	t -Stat	$V(b)$
Intercept	-0.23	-1.56	.0019	-0.25	-1.08	0.3181*	-0.04**	-0.22	.0213*
HMO	-0.28	-1.59		0.03	0.15		-0.48*	-2.22	
POS	0.57	2.54		-0.31	-1.14		-0.22	-0.70	
PPO	0	-		0	-		0	-	
Med Care	0.37**	0.0338	.0000	0.42**	7.16	0.0210*	.54**	10.00	0.0000
HMO	0.24**	4.96		0.22**	4.27		0.01	0.18	
POS	-0.10	-1.44		0.36**	4.93		0.00	0.09	
PPO	0	-		0	-		0	-	
Cost	0.48**	13.60	.0000	0.45**	7.78	0.0190*	.28**	6.65	.0000
HMO	-0.18**	-4.42		-0.23**	-5.12		0.07	1.34	
POS	-.03	-0.57		-0.30**	-5.27		0.05	0.77	
PPO	0	-		0	-		0	-	
Residual			.3601**			.3158**			.3536**

Note: * indicates $P < .05$ and ** indicates $P < .01$

References

Anderson, Eugene W. and Vikas Mittal (2000), "Strengthening the Satisfaction-Profit Chain," *Journal of Service Research*, 3 (2), 107-120.

Anderson, Eugene W. and Mary W. Sullivan (1993), "The Antecedents and Consequences of Customer Satisfaction for Firms," *Marketing Science*, 16 (2), 129-45.

Anderson, Eugene W., Claes Fornell, and Donald R. Lehmann (1994), "Customer Satisfaction, Market Share, and Profitability: Findings from Sweden," *Journal of Marketing*, 58 (3), 53-66.

Anderson, Eugene W., Claes Fornell, and Roland T. Rust (1997), "Customer Satisfaction, Productivity, and Profitability: Differences Between Goods and Services," *Marketing Science*, 16 (2), 129-145.

Bolton, Ruth N. (1998), "A Dynamic Model Of The Duration Of Customer's Relationship With A Continuous Service Provider: The Role Of Satisfaction," *Marketing Science*, 17 (1), 45-65.

Bring, John (1994), "How to Standardize Regression Coefficients," *The American Statistician*, 48 (3), 209-213.

Bryk, Anthony S. and Stephen W. Raudenbush (1992), *Hierarchical Linear Models*, Newbury Park, CA: Sage.

Calder, Bobby J., Lynn W. Phillips, and Alice M. Tybout (1981), "Designing Research for Application," *Journal of Consumer Research*, 8 (2), 197-207.

Calder, Bobby J. and Alice M. Tybout (1999), "A Vision of Theory, Research, and the Future of Business Schools," *Journal of the Academy of Marketing Science*, 27 (3), 359-366.

Calder, Bobby J. and Edward C. Malthouse (2002), "The Behavioral Score Approach to Dependent Variables," *Journal of Consumer Psychology*, forthcoming.

Danaher, Peter J. (1998), "Customer Heterogeneity in Service Management," *Journal of Service Research*, 1 (2), 129-139.

DeWulf, Kristof, Gaby Odekerken-Schröder, and Dawn Iacobucci (2001), "Investments in Consumer Relationships: A Cross-Country and Cross-Industry Exploration," *Journal of Marketing*, 65 (4), 33-50.

Drolet, Aimee L. and Donald G. Morrison (2001), "Do We Really Need Multiple-Item Measures in Service Research?" *Journal of Service Research*, 3(3), 196-204.

Drucker, Peter F. (1954), *The Practice of Management*, New York: Harper & Row.

Finn, Adam and Ujwal Kayandé (1997), "Reliability Assessment and Optimization of Marketing Measurement," *Journal of Marketing Research*, 34 (2), 262-275.

Fornell, Claes (1992), "A National Customer Satisfaction Barometer: The Swedish Experience," *Journal of Marketing*, 56 (1), 6-21.

Fornell, Claes, Michael D. Johnson, Eugene W. Anderson, Jaesung Cha, and Barbara Everitt Bryant (1996), "The American Customer Satisfaction Index: Nature, Purpose, and Findings," *Journal of Marketing*, 60 (4), 7-18.

Fournier, Susan and David Glen Mick (1999), "Rediscovering Satisfaction," *Journal of Marketing*, 63 (4), 5-23.

Garbarino, Ellen and Mark S. Johnson (1999), "The Different Roles of Satisfaction, Trust, and Commitment in Customer Relationships," *Journal of Marketing*, 63 (2), 70-87.

Hallowell, Roger (1996), "The Relationships of Customer Satisfaction, Customer Loyalty, and Profitability: An Empirical Study," *International Journal of Service Industry Management*, 7 (4), 27-42.

Hutchison, J. Wesley, Wagner A. Kamakura, and John G. Lynch, Jr. (2000), "Unobserved Heterogeneity as an Alternative Explanation for 'Reversal' Effects in Behavioral Research," *Journal of Consumer Research*, 27 (3), 324-344.

Kekre, Sunder, Mayuram S. Krishnan, and Kannan Srinivasan (1995), "Drivers of Customer Satisfaction for Software Products: Implications for Design and Service Support," *Management Science*, 41 (9), 1456-1470.

Kreft, Ita and Jan deLeeuw (1998), *Introducing Multilevel Modeling*, London: Sage.

Kreft, Ita, Jan DeLeeuw, and L. Aiken (1995), "The effect of different forms of centering in hierarchical linear models," *Multivariate Behavioral Research*, 30, 1-22.

Laird, Nan and James Ware, "Random-Effects Models for Longitudinal Data," *Biometrics*, 38, pp. 963-974.

Littell, Ramon, George Milliken, Walter Stroup, and Russell Wolfinger (1996), *SAS System for Mixed Models*, Cary, NC: SAS Institute.

Loveman, Gary W. (1998), "Employee Satisfaction, Customer Loyalty, and Financial Performance: An Empirical Examination of the Service Profit Chain in Retail Banking," *Journal of Service Research*, 1 (1), 18-31.

Malthouse, Edward C. (2002), "Database Subsegmentation," in Dawn Iacobucci and Bobby J. Calder (eds.) *Kellogg on Integrated Marketing*, New York: John Wiley & Sons, Inc.

Mittal, Vikas and Wagner A. Kamakura (2001), "Satisfaction, Repurchase Intent, and Repurchase Behavior: Investigating the Moderating Effect of Customer Characteristics," *Journal of Marketing Research*, 38 (1), 131-142.

Oliver, Richard L. (1980), "A Cognitive Model of the Antecedents and Consequences of Satisfaction Decisions," *Journal of Marketing Research*, 17 (3), 460-469.

Oliver, Richard L. (1993), "Cognitive, Affective, and Attribute Bases of the Satisfaction Response," *Journal of Consumer Research*, 20 (3), 418-430.

Oliver, Richard L. and John E. Swan (1989), "Consumer Perceptions of Interpersonal Equity and Satisfaction in Transactions: A Field Survey Approach," *Journal of Marketing*, 53 (2), 21-35.

Parasuraman, A., Leonard L. Berry, and Valarie A. Zeithaml (1988), "SERVQUAL: A Multiple-Item Scale for Measuring Consumer Perceptions of Service Quality," *Journal of Retailing*, 64 (Spring), 12-40.

Parasuraman, A., Leonard L. Berry, and Valarie A. Zeithaml (1991), "Refinement and Reassessment of the SERVQUAL Scale," *Journal of Retailing*, 64 (Spring), 12-40.

Pinheiro, Jose and Douglas Bates (2000), *Mixed-Effects Models in S and S-PLUS*, New York: Springer.

Reichheld, Frederick (1996), *The Loyalty Effect*, Cambridge, MA: Harvard Business School Press.

Rentz, Joseph O. (1987), "Generalizability Theory: A comprehensive Method of Assessing and Improving the Dependability of Marketing Measures," *Journal of Marketing Research*, 24 (1), 19-28.

SAS Institute Inc. (1997), *SAS/STAT Software: Changes and Enhancements through Release 6.12*, Cary, NC: SAS Institute Inc.

Schlesinger, Leonard A. and Jeffrey Zornitsky (1991), "Job Satisfaction, Service Capability, and Customer Satisfaction: An Examination of Linkages and Management Implications," *Human Resource Planning*, 14 (2), 141-149.

Singh, Vishal, Pradeep Chintagunta, and Jean-Pierre Dube (2002), "Balancing Profitability and Customer Welfare: An Application to Zone Pricing by a Supermarket Chain," *Working Paper*.

Srivastava, Rajendra K., Tasadduq A. Shervani, and Liam Fahey (1998), "Market-Based Assets and Shareholder Value: A Framework for Analysis," *Journal of Marketing*, 62 (1), 2-18.

Zeithaml, Valarie A., Leonard L. Berry, and A. Parasuraman (1996), "The Behavioral Consequences of Service Quality," *Journal of Marketing*, 60 (2), 31-46.

Appendix

Newspaper Content

The question wording for the items measuring satisfaction with content was as follows: "Please rate the ___ newspaper on each of the following kinds of content. To answer use a 5-point rating scale for which a 1 means that it is poor and a 5 means it is excellent. Pick a single number between 1 and 5 to rate how good a job the newspaper does with each type of content. If you are unsure, just indicate your expectations. Then please indicate how important each is to you personally by checking one box in the importance section for each type of content." The importance question used the scale 1="Little/None", 2="Some", and 3="A lot". The survey included 30 content areas:

- Arts (e.g., dance, classical music, museums, etc.)
- Automotive
- Business, Economics and Personal Finance
- Community announcements (including weddings, events, etc.)
- Education
- Environment
- Fashion and Beauty
- Food
- Health, Fitness and Medicine
- Home, Garden and Real estate
- Jobs and Career
- Movies
- Natural disasters / Accidents
- Obituaries
- Ordinary People
- Parenting and Relationships
- Police / Crime / Courts / Legal
- Politics / Government
- Popular Music
- Religion / Spirituality
- Science and Technology
- Sports
- Television
- Travel
- War / International conflict
- Weather
- Ads and inserts for food and groceries
- Ads for clothing, health & beauty aids, and stores other than supermarkets
- Classified ads for jobs and employment opportunities / automobiles / real estate
- Ads for entertainment events and sporting events.

Newspaper Service Satisfaction

The wording for measuring satisfaction with service was identical to the content question, except that the word "content" was replaced with "service" and satisfaction was measured on a 10-point scale. Eight aspects of service were measured:

- When and how the paper is delivered
- The cost of home delivery
- The accuracy of my bill
- Customer service
- Easy to buy at the store or vending machines
- The condition/completeness of the paper when I get it
- Telemarketing calls
- Quality of the newspaper's paper, ink, and type size

Type of Health Insurance

Respondents were asked "What type of health insurance do you have (CHECK ONE BOX ONLY):"

- Medicare only
- Medicare plus supplemental coverage
- Medicare HMO
- Medicaid HMO
- Other Medicaid
- Military health insurance
- Preferred provider organization (PPO)
- Health maintenance organization (HMO)
- Point of service (POS)
- Other
- Have health insurance but don't know the type
- No health insurance

Respondents were asked "Which insurance company currently provides *most of your coverage*, that is, your *primary* plan?" Respondents were given a list of all the companies in a given market and the option of writing in another company. We attempted to match all write-ins with a listed company.

Medical Care Satisfaction

The survey included two sections of questions relating to satisfaction. "*Medical Care: How satisfied are you with each of the following? (Please check one box for each item: very satisfied, somewhat satisfied, neither satisfied nor dissatisfied, somewhat dissatisfied, very dissatisfied, and no experience with this).*" Respondents rated the following items:

- Thoroughness of treatment
- Attention given to what you have to say
- Number of doctors you have to choose from
- Ease of choosing a personal physician

- Quality of doctors in the health plan
- Location of hospitals in the health plan
- Location of doctor's offices in the health plan
- Choice of pharmacies available to you
- Overall quality of care and services

Additional items were included on the survey, but were dropped here because few respondents had "experience" with the area, e.g., "Care given for emergencies."

Plan Administration Satisfaction

The second satisfaction section asked "*Plan Administration*: Thinking about your current health plan, how satisfied are you with each of the following?" The items with which a substantial number of respondents had experience include:

- Your insurance coverage overall, taking everything into account
- The part of the premium you pay for the covered services
- Amount you pay out-of-pocket (for example: co-payments, deductibles, payments for services not covered)
- The value of your health care coverage considering what you have to pay

Overall Satisfaction with Health Plan

Respondents were also asked the following question: "Overall, how satisfied with your health plan? (completely satisfied, couldn't be better; very satisfied; somewhat satisfied; neither satisfied nor dissatisfied; somewhat dissatisfied; very dissatisfied; completely dissatisfied, couldn't be worse)"