

Determining Attribute Importance in a Service Satisfaction Model

Anders Gustafsson
Karlstad University

Michael D. Johnson
University of Michigan

Determining the importance that customers place on the product and service attributes that drive their satisfaction with, and loyalty to, service providers is an essential part of a firm's resource allocation process. An unsettled issue is whether importance measures should come directly from customers or be derived statistically and, if so, how. The authors compare direct importance ratings with a variety of methods for statistically deriving attribute importance in a customer satisfaction model. Using three data sets, the methods are compared on criteria that include their ability to explain variation in satisfaction, to identify customers' more important attributes, and to be interpretable. The findings suggest that because each of the tested methods has its strengths and weaknesses, it is essential to choose a method that is compatible with the research goals and context.

Keywords: attribute importance; customer satisfaction modeling

The importance that customers place on service quality attributes as drivers of satisfaction and loyalty is a critical input to a firm's resource allocation strategy and quality improvement efforts. From a strategy standpoint, logical areas for improvement are those that are both important to customers and on which a firm is doing poorly (Martilla and James 1977). From a quality improvement standpoint, attribute importance and performance measures are critical inputs to internal change processes and tools such as quality function deployment (Mazur 1998). However, there remains a paucity of research that compares alternative methods for determining attribute importance measures within the context of a customer satisfaction model (Johnson and Gustafsson 1997). A major unsettled issue is whether importance measures should come directly from customers or be derived statistically from satisfaction and loyalty evaluations. There are pervasive arguments for deriving importance measures statistically from overall evaluations (Dillon *et al.* 1997; Gustafsson and Johnson 1997). Yet the available empirical research that compares direct and derived importance measures suggests that direct measures predict future behavior well (Griffin and Hauser 1993). Direct measures of importance have been particularly popular in a service setting in the context of service quality measures such as SERVQUAL (Parasuraman, Zeithaml, and Berry 1988).

The goal of this research is to provide insight into this debate through a systematic study of alternative methods for determining the drivers of satisfaction and loyalty. We argue and show that there is relatively wide variation in statistically derived importance measures depending on the approach used and that direct measures fall somewhere between the best and the worst of the statistical methods. The performance of each method and resulting importance measures is examined using several criteria. These include a method's ability to explain variation in customer satisfaction and loyalty, to provide diagnostic importance measures, and to be interpretable.

We begin by describing customer satisfaction and loyalty modeling and the methods that we compare. In addition to direct importance ratings, we examine different statistical methods for deriving importance measures including multiple regression (MR), normalized pairwise estimation (NPE), two variations on partial least squares (PLS), and a type of principal components regression (PCR). We then develop our criteria for evaluating importance measures. An empirical analysis of three data sets highlights several interesting results. Generally, the MR and NPE approaches explain more variation in satisfaction. NPE has the additional advantage of avoiding negative measures that are difficult to interpret. The measures derived using PLS and PCR are more diagnostic in terms of being able to identify customers' most important attributes. There is also evidence that direct importance ratings are "forward-looking" as their relative ability to explain customer loyalty is greater than their relative ability to explain satisfaction *per se*.

CUSTOMER SATISFACTION MODELING

Identifying product and service attributes of import to customers has been a major focus of consumer and market researchers for decades. More traditional approaches to identifying importance include conjoint measurement, which focuses on evaluations of product and service concepts within controlled settings (Green and Srinivasan 1990), and choice modeling (as through analysis of diary or scanner-panel data; Guadagni and Little 1983). More recent attention has turned to understanding the attributes and benefits that drive evaluations of the consumption experience, or customer satisfaction, and subsequent loyalty (Dillon *et al.* 1997; Gustafsson and Johnson 1997; Ryan, Rayner, and Morrison 1999; Steenkamp and van Trijp 1996).

In this study, we use company-specific survey instruments that have been developed over time for the purpose of understanding customers' consumption experiences. The surveys are designed with a customer satisfaction model in mind. We want to determine what attributes and benefits have the most impact on, and explain most variation in, customer satisfaction. There are other alternatives to

identifying service qualities in need of improvement, such as the SERVQUAL instrument (Parasuraman, Zeithaml, and Berry 1988). This instrument identifies the service quality gaps between rated expectations and performance to identify and develop improvement priorities. Although the SERVQUAL questions are derived from a five-factor structure, the approach does not estimate a satisfaction model. Moreover, research has shown that the SERVQUAL scale does not always exhibit the five-factor structure it was designed for (Carman 1990; Johnson *et al.* 2001). Thus, the approach is less appropriate when modeling customer satisfaction or loyalty *per se*.

We focus specifically on cumulative customer satisfaction, defined as an overall evaluation of a customer's purchase and consumption experience to date (Fornell 1992; Fornell *et al.* 1996). Other research focuses on more transaction-specific satisfaction (Boulding *et al.* 1993; Oliver 1997). Because customer loyalty and repurchase decisions are based on a broader consumption history and explain more variation in loyalty (Lervik Olsen and Johnson 2003), cumulative satisfaction is simply more useful for our purposes. We define customer loyalty as a customer's predisposition to repurchase from a product or service provider (a behavioral intention), which serves as a proxy for actual retention and profit (Fornell *et al.* 1996).

In contrast to traditional conjoint measurement or choice modeling, cumulative satisfaction models emphasize customers' perceptions of product performance, the overall evaluations that result, and the behavioral intentions they create. Cumulative satisfaction models rest heavily on multidimensional expectancy-value model formulations (Bagozzi 1992; Fishbein and Ajzen 1975) that use latent variables (Gustafsson and Johnson 1997; Johnson *et al.* 2001). Accordingly, customers have distinguishable beliefs or perceptions regarding their consumption experience that we label customer benefits. Each of these benefits is described using one or more concrete product attributes. These benefits are the primary antecedents of customer satisfaction as a type of overall evaluation of the consumption experience. This satisfaction, in turn, influences customers' behavioral intentions in the form of a predisposition to repurchase the product or service again (loyalty).

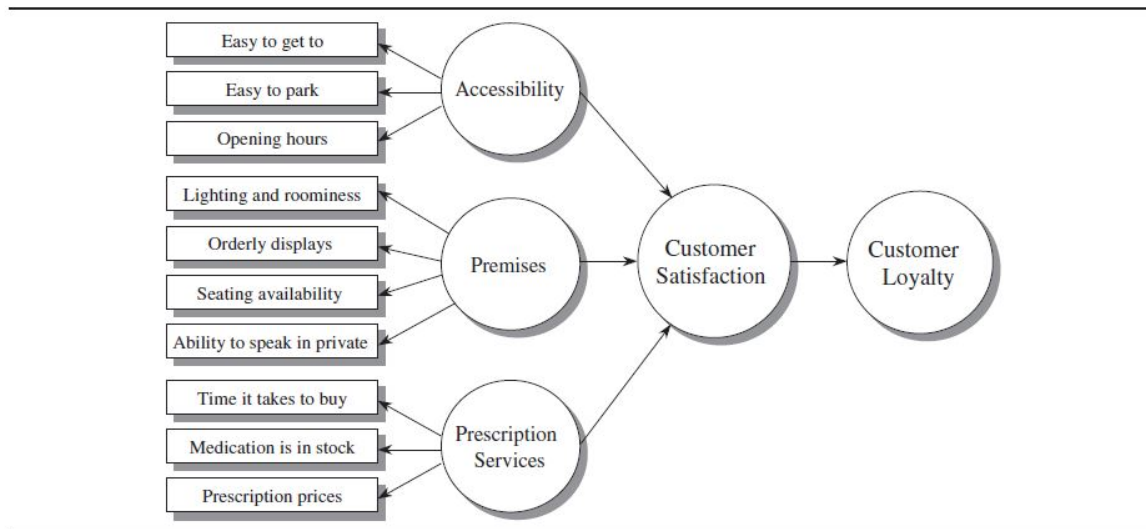
Figure 1 illustrates a satisfaction model using a subset of attributes and benefits from our empirical study of pharmacy services. The benefits include the accessibility of the pharmacies, the quality of the physical environment or premises, and the quality of prescription services. Each benefit is itself a function of multiple underlying attributes. Whereas the attributes describe more concrete and descriptive aspects of a product or service offering, the benefits describe the more general or abstract qualities that customers derive from the attributes (Gustafsson and Johnson 1997). The perceived accessibility or convenience of the pharmacy is, for example, a function of how easy the pharmacy is to get to, how easy it is to park, and the opening hours.

One of the primary functions of such a model is to provide input to a firm's priority setting and subsequent resource allocation decisions. The priority setting process requires two key inputs. One is the relative importance of the various attributes and benefits toward improving customer satisfaction. The other is performance data on the attributes and benefits. Performance benchmarks are obtained from the survey-based attributes and benefits that drive satisfaction, often relative to a set of direct competitors within a market segment.

Importance-performance analysis then determines where a firm should concentrate its resources to improve performance (Martilla and James 1977). The essential aspects to improve are those where importance is high and performance is low. This effectively focuses resources where they have the greatest impact on satisfaction and subsequent loyalty. Those aspects where performance and impact are both high illuminate a firm's competitive advantage. It is essential to at least maintain, if not improve, performance on these drivers. When both importance and performance are low, customers are telling us not to waste resources improving these areas. More interesting is the low importance/high performance category. This may be an area where resources have been wasted in the

past be- cause the improvements were not important to customers. Alternatively, these may be drivers of satisfaction that customers

FIGURE 1
A Partial Satisfaction Model for the Pharmacy



consider basic and necessary. Such benefits and attributes may be consistently provided to customers and, as a result, have little to no impact on satisfaction. In this research, we focus specifically on the determination of importance.

METHODS FOR DETERMINING IMPORTANCE

A key feature of satisfaction models is that the benefit, satisfaction, and loyalty constructs in the models are inherently abstract or latent variables. The most common way to empirically measure these latent variables is through the use of multiple concrete proxies or measurement variables. These measures should cover different unique dimensions in the latent variables; they should not be the same questions repeated in slightly different ways (Drolet and Morrison 2001; Garpentine 2001). Benefits are measured using their attributes, satisfaction is measured using different overall evaluation standards (such as overall satisfaction, overall performance versus expectations, overall performance versus an ideal), and loyalty is measured using behavioral intentions (such as the likelihood of re-purchase or recommendation to others).

There are two general methods for determining the importance of attributes and benefits in a satisfaction model. The researcher can ask customers directly how important attributes are using scale ratings, point allocation, or paired comparison rating. The alternative is to statistically estimate the satisfaction model to derive attribute and benefit importance. We compare direct scale ratings of attribute importance to five common approaches to statistical estimation: (a) MR, (b) NPE, (c) PLS with reflective attribute specifications, (d) PLS with formative attribute specifications, and (e) a type of PCR.

Multiple Regression (MR)

In MR, the researcher simply regresses an entire set of product attribute ratings against some dependent variable of interest, such as satisfaction (Griffin and Hauser 1993). This approach is the

easiest to implement statistically. It is also the most problematic. The primary problem with this approach is that it does not take into account the distinction between measurement and latent variables in a satisfaction model. As a result, multicollinearity among the independent variables can be severe. Consider that most of the multicollinearity within a customer satisfaction model exists among multiple measures of the same underlying constructs (such as the attributes of a given benefit or the multiple measures of satisfaction). As described later, reflective PLS and PCR remove much of this multicollinearity using the measurement variables to operationalize the latent variables as indices prior to running the regressions. The primary advantage of the MR approach is that it uses all available information in the attribute performance measures to explain satisfaction. However, the severe multicollinearity in the case of MR may cause some positively valenced attributes to have negative coefficients that are difficult to interpret (Ryan, Rayner, and Morrison 1999).

Normalized Pairwise Estimation (NPE)

NPE is a very simple algorithm for dealing with multicollinearity that has become popular among practitioners. The procedure is described in Rust and Donthu (2003) as follows. First, correlations are obtained between each of the predictor variables and the dependent variable. An ordinary least squares (OLS) multiple regression is run, and the R^2 obtained. If the predictors are uncorrelated, then the sum of the squared correlations equals the R^2 from the multiple regression. If the predictors are correlated, however, the sum of the squared correlations will be larger than the R^2 . Let us call the sum of the squared correlations S^2 , and let r^2 be the square of the correlation between predictor (attribute) i and the dependent variable (e.g., satisfaction). Then the estimated importance measure for predictor i is equal to $(r^2 R^2 / S^2)$. Conceptually, NPE adjusts individual correlations based on the total correlation in the model.

Like MR, NPE uses all available information in the independent variables to explain the dependent variable. The particular advantage of the NPE approach is that it addresses the multicollinearity problem that occurs with multiple regression (where overfitting can result in negative importance measures). As long as the correlation between an attribute and satisfaction are positive, the importance measure is positive. A disadvantage with both MR and NPE is that these methods do not use the type of model structure in Figure 1. No distinction is made between measurement variables (such as attributes) and more latent variables (such as customer benefits, satisfaction, and loyalty). Both MR and NPE are more data driven than theory driven. As a result, a company may not take advantage of the lens of the customer (Johnson and Gustafsson 2000). Furthermore, the information generated by NPE is correlation based, and correlations do not have the desired interpretation as impact scores. The attributes never compete with each other to explain variation in satisfaction. This results in two potential limitations. First, attribute importance measures using NPE may be less diagnostic than other methods. Second, NPE may overstate the importance of a particular attribute.

Partial Least Squares (PLS)

Using Reflective or Formative Measures

In contrast, statistical estimation of a satisfaction model through PLS accommodates the fact that the model is a network of cause-and-effect relationships (as from benefits, to satisfaction, to loyalty) that contains latent variables (Gustafsson and Johnson 1997; Johnson and Gustafsson 2000; Ryan, Rayner, and Morrison 1999; Steenkamp and van Trijp 1996). We focus for the moment on PLS using reflective measures. PLS is essentially an iterative estimation procedure that integrates principal components analysis with multiple regression (Fornell and Cha 1994; Wold 1966). The objective of PLS is to explain variance in the endogenous variables in a satisfaction model (such as satisfaction, loyalty, or

profit). Because PLS is based on principal components, the latent variables are operationalized as weighted indices of their measurement variables. And because these constructs are used as input to regression models, the weights and path coefficients relating attributes to benefits to satisfaction are akin to beta coefficients or impact scores.

Measurement variables in a PLS model can be operationalized as either reflective or formative. The reflective PLS procedure extracts the first principal component from each subset of measures for the various latent variables in a model and then uses these principal components within a system of regression models. Going back to Figure 1, one equation would explain satisfaction with pharmacy services using the three benefits as independent variables, whereas a second equation would explain loyalty using satisfaction as an independent variable. The algorithm then goes through a series of iterations in an attempt to improve the ability to explain variation in the dependent measures in the regression models by adjusting the principal component weights. Performance measures for the latent variables in a PLS model are operationalized as principal components, or simple weighted averages of the measurement variables. Attribute-level impacts are determined by multiplying the unstandardized weight that an attribute contributes toward a benefit times the impact the benefit has on satisfaction in the regressions (Gustafsson and Johnson 1997).

An alternative within PLS is to specify formative indicators for the attribute to benefit relationships. The reflective versus formative distinction is illustrated using the simple causal model in Figure 2 in which two benefits affect satisfaction, which in turn influences loyalty. In each case, the latent variables are described using multiple measurement variables (attributes A1 through A6 for the benefits, measures S1 through S3 for satisfaction, and measures L1 through L3 for loyalty).

One of the benefits in Figure 2 specifies a reflective relationship between the latent and measurement variables. The latent variable is reflected in the measurement variables as indicated by the direction of the arrows for attributes A1 through A3. This approach takes the theoretical or latent variable as the starting point and proposes or implies specific observable events or measures; the observations are reflective of the underlying constructs. Contrast this with the other benefit in Figure 2 where attributes A4 through A6 are specified as formative. The benefit in this case is assumed to be made up of, or defined by, a collection of measurement variables. In essence, the reflective case presumes that the observable measures are dependent on the latent or abstract construct, whereas the formative case presumes that the latent variable is defined by, or dependent on, what is observed (Fornell and Cha 1994).

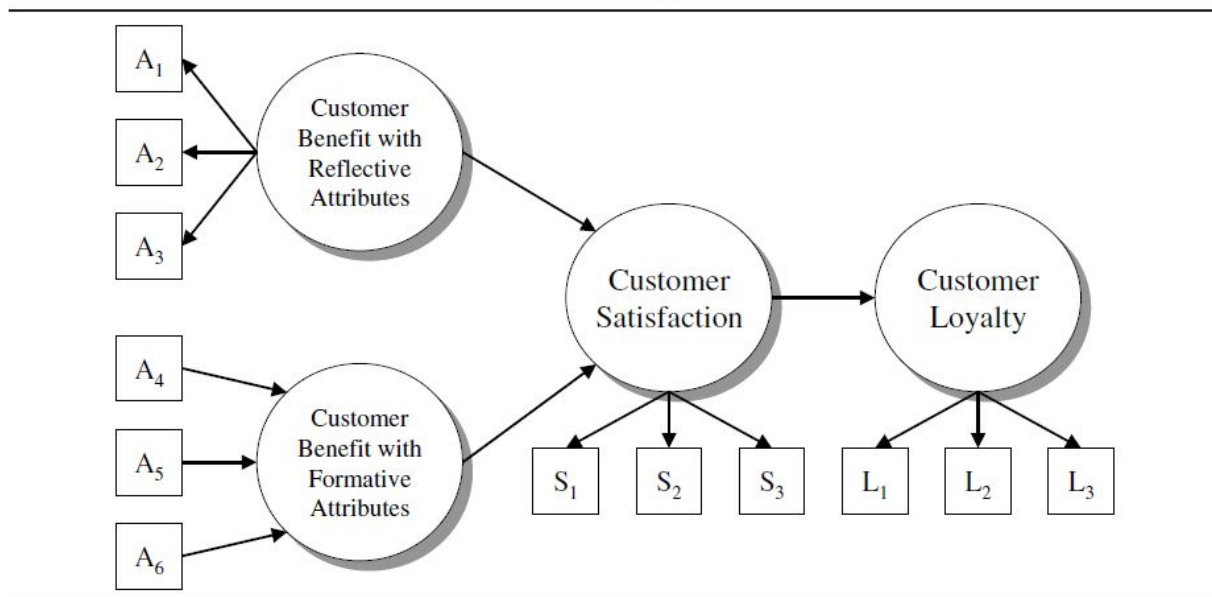
Computationally, the weights in PLS are determined in different ways depending on the formative or reflective specification. When estimating a PLS model with reflective indicators, the latent variable indices are akin to principal components. The weights are simply the attribute-level loadings after rescaling. They are a function of the correlation among a construct's measurement variables. In the formative specification mode, the multiple measures of a latent variable are regressed directly against another latent variable (e.g., the attributes of "service" are regressed directly against satisfaction), and the unstandardized multiple regression coefficients are the measurement variable weights.

A disadvantage of PLS is that it ignores some information in the independent variables (attributes). The benefit indices do not retain all of the information in the original attributes. Thus, PLS should explain less variation in satisfaction than MR or NPE. And although PLS reduces the amount of multicollinearity in the model (because the highest correlations are among the attributes of a given benefit), it does not eliminate the problem. As formative PLS includes some use of multiple regression at the measurement variable level (regressing subsets of attributes against satisfaction), it shares some of the same problem inherent in the multiple regression approach.

The choice of reflective or formative depends on several factors. The measurement variable weights under a formative specification are most directly interpretable as impact scores (as is the case for the path coefficients) because they compete with each other to explain satisfaction. The implication is that formative weights may be more diagnostic or vary from attribute to attribute within a given benefit. At the same time, a reflective specification is more defensible in most applications. For constructs such as satisfaction and loyalty, which are likely manifested in a wide variety of measures, Johnson and Gustafsson (1997) argued that only a reflective specification is justified. Finding a finite number of subdimensions or components that make up or define a construct such as overall satisfaction is problematic. If, however, there are a specified and reasonable number of attributes that define a particular customer benefit, a formative specification is appropriate. Put simply, a formative specification is much more demanding in that the latent variable should be more completely specified or defined. For example, Johnson and Gustafsson (1997) contrasted reflective and formative attribute specifications within a PLS-based satisfaction model for a large furniture retailer and found that the formative weights were better at distinguishing important from unimportant attributes. But small negative weights for some of the attributes were difficult to interpret. We test between PLS using reflective attribute specifications and PLS using formative attribute specifications. However, in both cases, satisfaction is a reflective construct.

An alternative to PLS for estimating structural equation models with latent variables is a maximum likelihood-based procedure such as covariance structure analysis (CSA) using, for example, LISREL (Jöreskog 1970). Whereas PLS is prediction oriented, CSA focuses on explaining covariance or the strength of relationships. CSA is a very appropriate method when testing among alternative model specifications based on strong theory and data (Lervik Olsen and Johnson 2003). However, several considerations make CSA less appropriate when *operationalizing* an existing quality or satisfaction model. Foremost, the latent variables in CSA are based on true-score theory. One implication is that CSA

FIGURE 2
Reflective and Formative Attribute Specifications



is restricted to reflective indicators. Another implication is that, when CSA models are used to develop attribute weights in an applied setting, the researcher typically reverts to OLS-based regression. It is simply not possible to interpret the weights at the measurement variable level for CSA in the same

manner as when PLS is used. Because the main focus in this research is on making comparisons of measurement variables, CSA is less appropriate. In addition, CSA requires larger sample sizes that are not always available in practice (Fornell and Bookstein 1982). Compared to PLS, for example, CSA is more susceptible to improper or nonconvergent results when estimating a complex model with many measurement variables (Bagozzi and Yi 1994).

Principal Components Regression (PCR)

An alternative to both reflective and formative PLS is a hybrid approach that combines the use of principal components analysis and MR. This approach is termed PCR (Frank and Friedman 1993; Massy 1965). In traditional PCR, all the attribute ratings for a product are factor-analyzed to produce a set of independent components or factors (Ryan, Rayner, and Morrison 1999). The problem is that the approach is completely data driven and atheoretical. We use a variation on PCR that is designed to estimate an existing satisfaction model, such as a model that is based on theory, which has evolved in a company over time or is developed through qualitative research (the lens of the customer). Using the approach, the benefit categories (attribute clusters) in the model are used to structure the analysis. The researcher extracts the first principle component from each subset of measures for each benefit, the satisfaction measures as a group, the loyalty measures as a group, and so on. The principal components are then used as input to a series of regression models. As with PLS, the attribute-level weights are calculated by multiplying the weight of an attribute on a benefit times the beta coefficient or impact score for that benefit on satisfaction.

This approach to PCR is actually a special case of PLS in which all observable variables or measures are reflective and there is no iteration or adjustment in the measurement variable weights. An important advantage of PCR is that it is relatively easy to implement using a variety of existing statistical packages, whereas PLS requires special software. It is conceptually very close to PLS in that it takes into account the difference between latent and measurement variables and provides regression-based impact scores for the path coefficients. Research in other domains suggests that PCR provides model fits that are quite close to PLS (Frank and Friedman 1993). The primary disadvantage of PCR compared to PLS is that it ignores residual variation among the measurement variables of different constructs; the variable weights are not adjusted to explain more variation in the dependent variables in a model. And like PLS, PCR does not use all of the information in the attribute measures and may not eliminate overfitting due to multicollinearity.

Direct Importance Measures

A categorically different approach is to ask customers directly for importance information. The two most common of the direct approaches are direct rating and point allocation methods (Bottomley, Doyle, and Green 2000; Griffin and Hauser 1993; Doyle, Green, and Bottomley 1997). Using direct ratings, respondents rate the importance of individual benefits or attributes on a scale ranging, for example, from *not at all important* to *very important* (Jaccard, Brinberg, and Ackerman 1986). These direct ratings are similar to the ratings of what customers “should expect” or desire in the SERVQUAL model (Parasuraman, Zeithaml, and Berry 1988). Using point allocation methods, respondents allocate a given number (say 100) importance points among a set of attributes. A third approach, used primarily in the quality area, involves the use of paired comparison ratings. Here respondents rate the relative importance of attribute pairs.

On the basis of prior research, we use direct ratings of attribute importance as a basis of comparison to the statistical methods. The paired comparison approach, as typified by the analytic hierarchy process (Saaty 1980), is relatively difficult for respondents to provide. It has also been

criticized for producing arbitrary measures of importance (Dyer 1990). Between the direct rating and point allocation methods, Griffin and Hauser (1993) found that they yield similar results. However, other research shows systematic differences between direct ratings and point allocation (Bottomley, Doyle, and Green 2000; Doyle, Green, and Bottomley 1997). Specifically, Bottomley, Doyle, and Green (2000) showed that direct ratings are preferred for two important reasons. First, respondents prefer direct ratings to point allocation. Second, direct ratings provide more stable weights. In our experience, direct importance ratings are also more common in practice.

All of the direct methods assume that customers both understand what the researcher means by “important” and what attributes are important to them. Even if customers know what is important to them, they must be willing to tell you. As a result, direct importance measures may result in socially acceptable or status quo answers and poor discrimination (as when customers rate all attributes as relatively important). The amount of information in self-reported importance measures also drops off as the number of attributes increases (Scott and Wright 1976).

Because statistical estimation of attribute and benefit importance is more objective and unbiased, it is arguably superior to direct customer ratings (Gustafsson and Johnsson 1997; Hayes 1998). Yet the only empirical study in which direct and derived importance measures are compared shows the opposite result. Griffin and Hauser (1993) compared direct importance measures with those obtained using multiple regression. They report on data obtained from a consumer products firm where importance was measured using three direct methods (a 9-point direct rating scale and two forms of point allocation—a constant sum scale of 100 points and an anchored scale in which 10 points are allocated to the most important attribute and up to 10 points are allocated to other attributes). Their results reveal high reliability among the three direct measures. Moreover, these scales each correlate highly with customer preference for seven product concepts from a product development team, which varied on the different performance dimensions.

The authors then regressed attribute performance ratings for the consumer product against a rating of satisfaction. They found that the revealed importance measures did not correlate with preferences for the hypothetical product concepts. They also report briefly on another study in which revealed importance measures were obtained for a high-cost durable product. In both studies, they found several negative coefficients (where some positively scaled attributes have a negative effect on satisfaction) and poor face validity for the revealed importance measures. However, these results should be interpreted with caution. The direct measures were only compared to the multiple regression approach, which is arguably the weakest of the statistical approaches examined here. Moreover, customer satisfaction is meant to describe a customer’s accumulated experience with an existing product or service, not his or her interest in, or preferences for, hypothetical product concepts.

EVALUATION CRITERIA

We compare the different methods for determining importance on a series of criteria to evaluate their performance in a satisfaction-modeling context. All comparisons are made on the measurement variable level because three of the methods (MR, NPE, and direct rating) only produce results at this level. Our criteria are a method’s ability to (a) explain variation in customer satisfaction; (b) identify customers’ most important attributes; (c) avoid negative, uninterpretable importance measures; and (d) explain variation in loyalty. Although any one method may perform well on any one of the criteria, our goal is to understand which method or methods perform well across all the criteria and just how great the differences are.

Variation Explained

As the fitting objective of the OLS estimation methods is to explain variation in the dependent variables, percent- age of variance explained (R^2) is a natural criterion on which to evaluate them (Fornell and Cha 1994). This criterion is also used to evaluate the direct ratings of importance using a simple multiattribute model formulation that is analogous to the multiple regression approach. The attribute ratings, importance measures, satisfaction measures, and loyalty measures in our empirical studies are all rated on 1- to 10-point scales. For each individual, we multiply the rated importance measures times the rated performance measures for each attribute, sum the products, and divide by the sum of the weights (to normalize the function) as follows:

$$\hat{Y}_j = \frac{\sum x_{ij} p_{ij}}{\sum x_{ij}}, \quad (1)$$

Where Y_j is the predicted satisfaction (or loyalty) for individual j , x_{ij} is rated importance of attribute i for individual j , and p_{ij} is rated performance for attribute i for individual j . Note that this function uses individual-level weights (importance ratings) to predict overall satisfaction or loyalty. In contrast, the statistical approaches produce aggregate-level weights estimated across respondents. This suggests an alternative to the weights in equation (1) where we calculate a second predicted satisfaction (or loyalty) for individual j , denoted \hat{Y}_j , as follows:

$$\hat{Y}_j = \frac{\sum \bar{x}_i p_{ij}}{\sum \bar{x}_i}, \quad (2)$$

is the average importance rating on attribute i across a similar population of customers (respondents). It is unclear at this point as to which of the two versions will be a better predictor of satisfaction and/or loyalty. Equation (1) has the advantage of using individually customized versus aggregate-level weights. Equation (2) has the advantage of aggregation over potentially error-laden individual responses. The R^2 s are calculated directly.

Diagnosticity

Diagnosticity is the ability of the method to identify just which attributes and benefits are most important, or most diagnostic, in affecting customer satisfaction. When setting priorities for quality improvement, the emphasis is on identifying just which area a company should invest in to ensure or increase satisfaction. Doyle, Green, and Bottomley (1997) showed how different direct methods vary with respect to their diagnosticity. Their benchmark for comparison is a linear relationship between the importance measures and their rank order. Deviations from linearity directly affect the measures' diagnosticity. For example, the authors show that direct ratings (from *not at all important* to *very important*) deviate from linearity in a concave fashion. That is, respondents are less able to distinguish between relatively important attributes and less important attributes.

Following Doyle, Green, and Bottomley's (1997) approach, we test for diagnosticity by regressing both the attribute ranking and a quadratic term for the ranking against the importance measures for a particular method. We look for two results. First is a large and significant linear relationship between the ranks and the importance measures. Second is a convex relationship (negative

quadratic) between rank and importance. The latter suggests that the method is particularly good at distinguishing among the more important attributes in a set. In contrast, a concave relationship (positive quadratic) between rank and importance indicates a distinct lack of diagnosticity among the more important attributes in a set.

Negative Measures

Using a direct rating scale, all of the attribute importance measures are positive as long as the scale values are positive. As described earlier, the multicollinearity across attribute ratings for a given offering can create problems in the form of negative weights (beta coefficients) for the regression-based methods. Based on our earlier discussion, these problems should be the most severe for the multiple regression approach because it completely ignores the relatively high collinearity among multiple measures of the same latent variable. Although NPE uses an MR to determine the total amount of correlation in a model, its estimates are based on correlations, which will only be negative if the correlation itself is negative. The PLS and PCR approaches use the satisfaction model (where attributes are used to create benefit-level indices) to reduce the multicollinearity to a level that produces meaningful benefit impacts. Attribute weights or impacts are then calculated by multiplying the weight or contribution of an attribute to its benefit times the impact of the benefit on satisfaction. For both PCR and PLS with reflective attributes, there should be relatively few negative importance weights. Because PLS with formative attributes is more similar to multiple regression, it should fall somewhere in between. We examine the incidence of negative measures across the methods.

Predictions

In sum, an apparent inconsistency in the satisfaction literature prompted our study. While statistical methods should produce more objective attribute and benefit importance measures, the one empirical study to date supports direct measures as superior. Our proposed explanation is that the performance of direct importance measures depends critically on the statistical benchmark. MR and NPE should explain the most variation in satisfaction and loyalty. Whereas MR should be very prone to negative, uninterpretable importance measures in the face of multicollinearity, NPE should not. But as NPE relies on pairwise correlations, where attributes do not compete to explain a dependent variable, importance measures obtained from this method should be less diagnostic and may be inflated. The advantage of PLS and PCR is that these methods build on a firm's satisfaction model and recognize latent variables. These models constitute a firm's operational theory that relates attributes to benefits, benefits to satisfaction, and satisfaction to loyalty. We expect these methods to avoid much of the multicollinearity problem that plagues MR yet provide diagnostic importance measures. Finally, we expect direct importance measures to perform somewhere between the best and the worst of the statistical methods. Beyond such general predictions, our study is exploratory. Prior research has simply not compared the range of methods studied here.

EMPIRICAL STUDY

Our empirical study consists of making comparisons across three data sets collected specifically for the purpose of the research. All of the satisfaction models for the data sets follow the basic structure described in Figure 1 where groups of attributes provide customers with particular benefits and these benefits drive satisfaction. We place the following boundaries on our analyses and presentation for brevity. For comparison reasons, only attribute-level importance measures are examined. Recall that benefit-level impacts are used to calculate attribute-level importance in those methods where the

satisfaction model is used to create latent variables (reflective PLS, formative PLS and PCR). For the multiple regression, NPE and direct ratings, attribute-level importance measures are the only output. We focus on a limited model in which satisfaction or loyalty is the endogenous or dependent variable. Finally, we use a principal component of satisfaction and loyalty measures to provide the multiple regression, NPE, and direct measures with dependent variables that are comparable (in abstraction and sensitivity) with that used in the structural modeling methods.

Data Sets

Three data sets were collected from convenience samples of university students who completed written surveys as part of their course work. Three services for which the students had significant experience were surveyed: postal services, a supermarket chain, and pharmacy services (usable observations of 99, 91, and 70 respectively). Each company's or agency's own satisfaction survey was used as a base, into which direct importance ratings were incorporated. Each attribute was rated on a direct-rating scale from 1 (*not at all important*) to 10 (*very important*). The survey measures used to operationalize the benefit and satisfaction indices for each data set are shown in Appendix A.

We take the companies' models and associated surveys as given because our interest is in operationalizing an existing model (versus developing a better model). The models and surveys are all currently being used by the companies or organizations from which they are taken and have evolved over time based on both qualitative data and quantitative analysis. Before describing our results, we use diagnostic information from the reflective PLS estimation results to evaluate the quality of the structural models, which we describe prior to presenting our main results.

The data were first screened to make sure that respondents had the proper experience with all the constructs. The only data set that posed a problem was the pharmacy survey, where 24 out of the original 94 respondents (25.5%) were removed based on a lack of experience with prescription services. The criterion used was simply that these respondents had not filled out that section of the questionnaire. The next step was to eliminate all attributes (variables) that had too many missing values. Replacing missing values using methods such as average value substitution leads to an underestimation of the beta coefficients due to reduced variation. Consequently, the variables need to be screened to make sure that they contain a sufficient percentage of responses. Downey and King (1998) argued and showed that average value substitution is appropriate when there are 20% or fewer values missing. This rule was applied to our data sets. Overall missing values were not a large problem. All variables with more than 20% missing values were removed. Most variables (80%) had less than 5% missing values, and there were three variables that had between 10% and 20% missing values. A small number of latent variables with only a single measurement variable were also eliminated from the analyses to maximize any observable differences among the estimation methods.

Model Quality

As noted, the quality of the measurement models may affect how well the different methods perform. In the case of a poor measurement model, as when the latent variables in the model are more highly correlated with each other than with their measurement variables, the models may not be measuring what they purport to measure. If the models are poor, even PLS may have significant problems with multicollinearity. Model quality, in this case, is a reflection of the company's or research agency's prior qualitative research and model development.

There are guidelines to apply when determining the quality of models based on PLS output (Fornell and Cha 1994). One is that the measurement variable loadings in PLS should exceed .707 to ensure that at least half of the variance in the observed variables is shared with the construct (the

squared correlation equals the variance explained, where $.7072 = 50\%$). This criterion is referred to as communality. The second criterion used to evaluate the discriminant validity of the model is to explore whether each latent variable shares more variance with its measures than it does with other constructs in the model. This may be examined by looking at the percentage of measurement loadings that exceed the latent variable correlations.

According to these criteria, the postal services model with 15 measurement variables is the best model; there are no communality problems, and the measurement loadings always exceed associated latent variable correlations. The supermarket model is the largest with 33 indicators. In this case, two out of eight benefit-level latent variables have an average communality just below 50%, whereas 92% of the measurement loadings exceed the latent variable correlations. The model with the most problems among the three models is the pharmacy model (with 29 measures). Here three out of nine latent variables have an average communality below 50%, and 88% of the indicators' loadings exceed the latent variable correlations. Although all of the models are relatively good from a measurement standpoint and representative in our view of what one sees in practice, some have more weaknesses

TABLE 1
Satisfaction Variance Explained by Method and Model (in Percentages)

	<i>Postal Services</i>	<i>Supermarket</i>	<i>Pharmacy</i>	<i>Average</i>
Reflective partial least squares	59	67	66	64
Formative partial least squares	74	71	76	74
Principal components regression	55	66	67	63
Normalized pairwise regression	78	82	85	82
Multiple regression	78	82	85	82
Direct ratings ^a	55	40	44	46
Direct ratings ^b	53	62	42	52

NOTE: Variance explained for multiple regression and normalized pairwise regression are the same.

a. Using individual ratings.

b. Using aggregate ratings.

than others. It is natural for estimation problems to grow as the quality of the measurement models decline. From a quasi-experimental stand- point, the variation in model quality provides a natural basis for examining how the differences among the estimation techniques vary from model to model.

Results

Variance explained. Table 1 shows the variance in satisfaction explained (R^2) for the methods and models. As MR and NPE use all of the information in the attribute measures, they naturally explain the largest amount of variance (78%-85%) where the average is 82%. The formative PLS models are next in explained variance with an average of 74% followed by the reflective PLS and PCR methods with 64% and 63%, respectively. Across the models, reflective PLS and PCR are quite similar. One of the more interesting results is the ability of the direct measures to predict customer satisfaction, at least in an absolute sense. When building a model at the individual level, we reach an average R^2 of 46%. When aggregated importance measures are used, the average R^2 increases to 52%. Although lower than the R^2 averages for the statistical methods, the direct measures are not that much worse than, for example, the reflective PLS and PCR methods (especially for the postal service and supermarket data sets). It is also clear that the direct measures benefit from using more aggregate-level measures.

Diagnosticity. As described earlier, we follow Doyle, Green, and Bottomley's (1997) approach to evaluate the distribution of importance weights and their diagnosticity. Specifically, the importance measures obtained from each method and model are used to estimate the following regression equation:

$$\text{Importance} = \text{constant} + \beta_1 \text{Rank} + \beta_2 \text{Rank}^2 \text{Residual}, \quad (3)$$

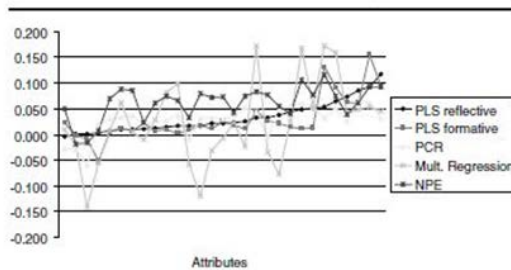
where *importance* is the estimated (or directly rated) importance for each attribute and *rank* is the rank order importance of the attribute. We rank the importances from 1 (*most important*) to *n* (*least important*) for each individual method. The *rank2* residual is a term obtained by first regressing the square of the rankings against the ranking to remove the linear component. This reduces the collinearity between the linear and quadratic terms to provide more stable estimates. Recall that we look for a significant linear relationship between the ranks and importance measures as revealed by a significant negative coefficient for *rank*. That is, importance should decrease as an attribute's rank increases from 1 to *n*. For a method to provide diagnosticity among the more important attributes, we also look for a significant negative quadratic term for

TABLE 2
Regression Coefficients for Rank and Rank²Residual by Method and Model

	Postal Services		Supermarket		Pharmacy	
Reflective partial least squares						
Rank (<i>t</i>)	-0.94	(-18.344)	-0.96	(-70.969)	-0.84	(-21.955)
Rank ² residual (<i>t</i>)	-0.14	(-2.644)	-0.12	(-8.943)	-0.34	(-8.912)
Formative partial least squares						
Rank (<i>t</i>)	-0.89	(-17.995)	-0.88	(-11.493)	-0.76	(-9.252)
Rank ² residual (<i>t</i>)	-0.27	(-5.578)	-0.12	(-1.639*)	-0.36	(-4.451)
Principal components regression						
Rank (<i>t</i>)	-0.92	(-13.836)	-0.97	(-33.652)	-0.96	(-13.322)
Rank ² residual (<i>t</i>)	-0.17	(-2.500)	-0.07	(-2.348)	0.29	(4.024)
Normalized pairwise estimation						
Rank (<i>t</i>)	-0.966	(-12.307)	-0.993	(-20.062)	-0.987	(-17.028)
Rank ² residual (<i>t</i>)	0.006	(0.077*)	0.296	(5.982)	0.220	(3.802)
Multiple regression						
Rank (<i>t</i>)	-0.99	(-25.342)	-0.99	(-31.326)	-0.95	(-18.015)
Rank ² residual (<i>t</i>)	-0.02	(-0.528*)	0.00	(0.019*)	-0.06	(-1.068*)
Direct ratings						
Rank (<i>t</i>)	-0.97	(-9.928)	-0.98	(-18.077)	-1.00	(-22.631)
Rank ² residual (<i>t</i>)	0.28	(2.861)	0.35	(6.506)	0.32	(7.211)

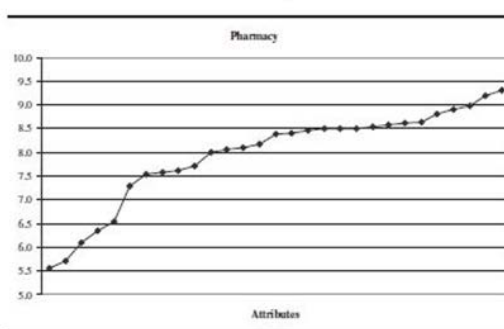
*Indicates not significant at $p < .05$.

FIGURE 3
Attribute-Level Importance Measures From the Pharmacy Model



NOTE: PLS = partial least squares; PCR = principal components regression; MR = multiple regression; NPE = normalized pairwise estimation.

FIGURE 4
Attribute-Level Direct Importance Ratings From the Pharmacy Model



rank2residual. That is, importance should decrease at a decreasing rate as an attribute's rank increases. The results are presented in Table 2.

The various methods, including the direct ratings, are similar with respect to the linear effect of rank. The differences across methods are primarily with respect to the quadratic term. The formative and reflective PLS methods are most likely to highlight those attributes that are most important to customers based on a greater incidence of significantly negative quadratic terms. For example, all three of the data sets have a significant negative effect for *rank2residual* under reflective PLS. In contrast, the MR estimates show only a linear relationship between importance and rank, and NPE shows two significant positive quadratic terms. To illustrate, Figure 3 shows the distribution of importance weights for each statistical method using the pharmacy data set. The figure uses the rank order from the reflective PLS model as a basis for comparison. The convex nature of the reflective PLS results are visible in the figure. The figure also reveals some similarity between the formative PLS and multiple regression results.

Another important finding from Table 2 is that the direct ratings show a significant positive quadratic effect in all three data sets. This suggests that the direct ratings are more diagnostic among the lower ranked attributes than among the higher ranked attributes. Figure 4 illustrates the finding using the distribution of direct importance ratings for the pharmacy data. As reported above, and consistent with Doyle, Green, and Bottomley's (1997) results, there is both a significant linear effect for rank and a significantly positive quadratic effect for rank in all three data sets for the direct ratings. Thus, the direct ratings are poor at distinguishing among the more important attributes in each model. The concave nature of the relationships illustrates how the method is more likely to identify what is least important to customers than what is most important to customers. It is worth noting that the most important attributes according to the direct ratings are distributed across the different latent variables in each model. The results are not just a case of all the attributes of a particular benefit being equally important. Interestingly, the NPE approach is similarly undiagnostic for the supermarket and pharmacy data sets. Recall that this is likely due to the fact that the NPE approach is based on pairwise correlations. The attribute performance measures do not compete with each other to explain variation (except when computing the total correlation).

The size of the positive quadratic terms for the direct measures is generally larger than the size of the negative quadratic terms for the statistically derived importance measures (see Table 2). This suggests that the concavity observed for the direct ratings is greater than the convexity observed using, for example, the PLS methods. Another interesting observation is that the PLS and PCR approaches are

TABLE 3
Average Attribute Importance by Method and Model

	<i>Postal Services</i>	<i>Supermarket</i>	<i>Pharmacy</i>	<i>Average</i>
Reflective partial least squares	0.044	0.021	0.032	0.032
Formative partial least squares	0.040	0.021	0.029	0.030
Principal components regression	0.043	0.021	0.028	0.031
Normalized pairwise estimation	0.052	0.025	0.061	0.046
Multiple regression	0.040	0.023	0.023	0.029

TABLE 4
Incidence of Negative Importance Measures by Method and Model

	<i>Postal Services</i>		<i>Supermarket</i>		<i>Pharmacy</i>		<i>Total</i>	
Reflective partial least squares	0	(0)	0	(0)	1	(0)	1	(0)
Formative partial least squares	1	(0)	4	(1)	3	(0)	8	(1)
Principal components regression	0	(0)	1	(0)	5	(0)	6	(0)
Normalized pairwise estimation	0	(0)	0	(0)	2	(0)	2	(0)
Multiple regression	5	(0)	14	(2)	11	(1)	30	(3)
Number of attributes	15		33		29		87	

NOTE: Number of significant negative coefficients is shown in parentheses. Not included are the direct ratings that are positive by definition. In the case of normalized pairwise estimation, the number of negative correlations is shown.

most similar on this criterion when the quality of the measurement model is higher, as for the postal services and supermarket data sets.

Figure 3 illustrates another difference between the NPE estimates and the other statistical methods. The NPE importance measures are higher. Table 3 shows the average level of importance by statistical method and model (data set). The differences are greatest for the pharmacy data set and smallest for the postal service data set. This is consistent with the earlier argument that, because the method is correlation based where importance measures do not compete with each other to explain satisfaction, the NPE approach may overstate importance. Again, however, the differences are greatest when the measurement model quality is lower (the pharmacy data). When the model quality is higher, as for the postal service and supermarket data, the NPE estimates are only marginally higher than the other approaches.

Negative measures.

Another evaluation criterion is each method's ability to avoid negative weights or measures that, for the statistical methods, are difficult to interpret as importance. Recall that the direct ratings are positive by definition. Table 4 shows the total number of negative measures in each case as well as the number that are negative and significant (in parentheses). Our expectation was that the NPE approach should perform quite well on this criterion, followed by the PLS and PCR methods, whereas MR should perform the worst. As can be seen from Table 4, these predictions were generally confirmed. In the case of MR, 30 out of 87 attributes show negative values, 3 of which are significant. Although one might argue that these 3 attribute coefficients are significantly negative by chance, they represent 33% of all significant attributes for the MR results. The majority of the negative measures for the PLS and PCR methods come from the pharmacy and supermarket models. The negative coefficients in these cases are quite small, most of which are only negative at the fourth decimal point. Because NPE avoids overfitting by relying on pairwise correlations, negative measures should only occur when the correlations themselves are negative. Two negative but nonsignificant correlations occurred for the pharmacy data set.

TABLE 5
Loyalty Variance Explained by Method and Model (in Percentages)

	<i>Postal Services</i>	<i>Supermarket</i>	<i>Pharmacy</i>	<i>Average</i>
Reflective partial least squares	19	36	48	34
Formative partial least squares	27	39	50	38
Principal components regression	10	27	44	27
Normalized pairwise estimation	29	31	52	37
Multiple regression	29	31	52	37
Direct ratings ^a	35	20	34	30
Direct ratings ^b	35	20	43	33

a. Using individual ratings.

b. Using aggregate ratings.

Performance of the direct measures.

Although the direct importance measures show no negative values by definition, they explain less variation in satisfaction and are less diagnostic than the other methods. One explanation is that the direct measures focus more on what is salient or important going forward, whereas the statistically derived measures reflect what is more diagnostic among customers' more recent experiences

(satisfaction). If this is the case, then the direct measures should perform better when used to explain a more forward-looking dependent variable. In the case of a satisfaction model, a natural forward-looking variable to examine is loyalty as a measure of customers' predisposition to behave favorably toward the product or company. This loyalty is operationalized in our data sets using either a measure of future repurchase likelihood or recommendation likelihood.

We tested this explanation by looking at the gap between satisfaction variance explained and loyalty variance explained for each method. If direct ratings are tapping into a more long-run importance or salience, their relative ability to explain loyalty should improve. From Table 5, we find that the average variance explained for loyalty using the statistical methods is 27% for PCR, 34% for reflective PLS, 37% for MR and NPE, and 38% for formative PLS. This is equivalent to what is found in the American customer satisfaction index where on average 36% of the loyalty latent variable is explained (Fornell *et al.* 1996). For the direct ratings, the average variance explained for loyalty is 33% for the aggregate-level measures and 30% for the individual measures. The observed gaps in variation explained between the statistical methods and the aggregate direct measures are much smaller than the earlier reported gaps in satisfaction variance explained (see Table 1). This lends credence to the notion that direct measures are tapping into a more forward-looking importance or salience. As a result, their performance improves when explaining loyalty rather than satisfaction.

SUMMARY AND DISCUSSION

Academics and practitioners alike use various methods to determine the importance of attributes and benefits in a service quality and satisfaction model. These include direct ratings of importance and statistically derived impact scores. Despite the attention given to customer satisfaction and loyalty in recent years, very little research has explored which of these methods provides the better measures of importance. Although there are good arguments for deriving importance measures statistically from over-all evaluations, prior empirical evidence suggests that direct measures are superior to at least one statistical method—MR.

TABLE 6
Summary of Results Across Methods and Criteria

	<i>Reflective PLS</i>	<i>Formative PLS</i>	<i>PCR</i>	<i>NPE</i>	<i>MR</i>	<i>Direct Ratings</i>
Average satisfaction variance explained	64%	74%	63%	82%	82%	52% ^a
Data sets (out of 3) meeting diagnosticity criteria	3	2	2	0	0	0
Average attribute importance	0.032	0.030	0.031	0.046	0.029	—
Incidence of negative importance (out of 87):						
Total	1	8	6	2	30	—
(significant at $p < .05$)	(0)	(1)	(0)	(0)	(3)	—
Loyalty variance explained ^a	34%	38%	27%	37%	37%	33%

NOTE: PLS = partial least squares; PCR = principal components regression; NPE = normalized pairwise estimation; MR = multiple regression.

a. Based on aggregate importance ratings.

In this article, we examined a wide variety of methods for determining importance: MR, NPE, PLS with reflective attribute specifications, PLS with formative attribute specifications, a variation on PCR, and direct importance ratings. We compared the various methods on several criteria including: (a) satisfaction variance explained, (b) diagnosticity (the ability to identify what attributes are most important to customers), (c) the incidence of negative and uninterpretable measures, and (d) loyalty variance explained. The results are summarized in Table 6.

Based on our summary findings in Table 6, it is clear that no one method outperforms all the others—all methods have strengths and weaknesses. While the MR approach explains the largest amount of variance in satisfaction, it suffers the most from multicollinearity based on the incidence of

negative importance estimates. NPE explains the same amount of variance as does multiple regression. It avoids the problem of overfitting that can affect the regression-based methods because it relies on correlations that are adjusted based on the total correlation in a model. However, our results reveal that because the attributes do not compete with each other to explain variation in satisfaction, NPE importance measures are less diagnostic (less able to identify customers' most important attributes). NPE measures may also overstate importance somewhat depending on the quality of the measurement model.

The advantage of the PLS and PCR methods is that they recognize the existence of latent variables (such as customer benefits) and leverage a company's satisfaction model or implicit theory that relates attributes to benefits to satisfaction and loyalty. By using the structure in a satisfaction model to limit the problems of multicollinearity while allowing benefits (and attributes in the case of formative PLS) to compete with each other, these regression-based approaches provide more diagnostic importance measures. Our approach to PCR is most similar to the PLS methods when the quality of the measurement model is high. The major limitation of these approaches is that they do not use all of the information in the attribute measures and thus explain less variation in satisfaction.

The primary implication of these results is that the choice of method should suit the company's and researcher's purpose. If, for example, the goal is simply to explain variation and provide attribute importance measures that completely avoid the problems of multicollinearity, then NPE would be a suitable approach. If, in contrast, the goal is to identify those benefits and attributes that are most diagnostic in affecting satisfaction, our results suggest that the reflective PLS method is the method of choice, and formative PLS and PCR are close substitutes. Recall, however, that the choice of reflective and formative PLS depends on just how comprehensive the attribute specifications are.

We also contrasted the statistical methods with direct measures of importance based on Griffin and Hauser's (1993) findings that direct measures outperform the MR approach with respect to, for example, the incidence of negative estimates that are difficult to interpret. Our results confirm these findings. However, a statistical approach that models product benefits and satisfaction as latent variables within a system of cause-and-effect relationships outperforms the direct ratings. As noted, the direct ratings fail to explain nearly as much variation in satisfaction as do the other methods. Direct ratings also suffer from a severe "concavity bias" (Doyle, Green, and Bottomley 1997) in that they distinguish more among the less important attributes in a set than among the more important attributes in a set. A particularly interesting finding with respect to the direct ratings is that they explain nearly as much variation in loyalty as do the statistical methods. Our conclusion is that statistical estimates of importance identify those attributes that have had the greatest impact on a customer's more recent consumption experiences, whereas direct ratings capture what is more globally salient to customers and thus important over time. As direct and derived ratings contain somewhat different and complementary information, an implication of our results is that researchers might gainfully employ both measures to operationalize importance as a more latent construct to explain loyalty.

Overall, our study provides a more broad-based comparison of satisfaction research methods than previously available and yields some important conclusions. First, the NPE approach avoids the primary problem that plagues MR (multicollinearity) but is not as diagnostic as the other methods. Second, PLS using reflective indicators generally perform the best in terms of identifying attributes that are most important to customers. This is consistent with the arguments that this method is particularly well suited to operationalizing a customer satisfaction model for the purposes of driving quality improvement efforts (Gustafsson and Johnson 1997; Johnson and Gustafsson 2000; Ryan, Rayner and Morrison 1999). Finally, we find evidence that direct importance ratings are forward-looking as their relative ability to explain customer loyalty is greater than their relative ability to explain satisfaction *per se*.

APPENDIX A
Latent and Measurement Variables by Dataset

TABLE A1
Latent and Measurement
Variables for Supermarket

<i>Latent Variable</i>	<i>Measurement Variables</i>
Atmosphere	That it is a nice environment How fresh an impression the store gives The joy of food and inspiration the store gives Tidiness in the store
Location	The store's location in relation to the place where you normally start your purchase round How easy it is to get to the store Availability to other stores in or service establishments nearby How easy it is to park
Service	Queue time at the cash point The store's opening hours How the product groups are placed in relation to each other How easy it is to find your way in the store How easy it is to get through in the store
Communication	Advertisements you get home or in newspapers Display inside the store Price information
Personnel	The personnel's interest in you as a customer The personnel's competence if you ask something The personnel's availability if you need help The personnel's helpfulness The personnel's attention
Prices	The price level in your store The price level on basic products Price in relation to quality Price in relation to quality of fruit and vegetables Price in relation to quality of meat and cured meats and provisions Price in relation to other food stores Special offers
Assortment	The number of alternatives in the assortment How well the assortment corresponds to your needs How well-filled the shelves are (i.e., the product you wanted to buy is out of stock) The quality of fruit and vegetables The quality of meat and cured meats and provisions
Customer satisfaction	Overall satisfaction Overall performance versus expectations Overall performance versus an ideal convenience store
Loyalty	If you get a neighbor who asks you for advice, how likely is it that you would recommend ICA?

One potential limitation of our study is that we assume linear relationships between the

TABLE A2
Latent and Measurement
Variables for Postal Services

<i>Latent Variable</i>	<i>Measurement Variables</i>
Location	Overall location Availability to other stores service establishments nearby
Opening hours	Opening hours Monday to Friday Opening hours Saturday Opening hours overall
Facility	Feels inviting Impression of tidiness
Rapid service	Wait until attended Speed and efficiency of the office
Personnel	The personnel's friendliness The personnel's availability if you need help The personnel's interest in you as a customer
Competence	The personnel's helpfulness The personnel's competence of handling errands correctly The personnel's competence of giving information concerning the different services
Customer satisfaction	Overall satisfaction (first question) Overall satisfaction Overall performance versus expectations Overall performance versus an ideal convenience store
Loyalty	If there was an alternative to Posten, how likely would you be to still use Posten as your service provider?

TABLE A3
Latent and Measurement
Variables for the Pharmacy

<i>Latent Variable</i>	<i>Measurement Variables</i>
Accessibility	Easy to get to How easy it is to park Opening hours
Premises	Feels inviting Impression of tidiness Impression of lighting and roominess Seating availability Ability to speak in private with personnel concerning nonprescription drugs
Prescription services	The time it takes to get a prescription expedited Medication is in stock The price paid for the prescription

(continued)

TABLE A3 (continued)

<i>Latent Variable</i>	<i>Measurement Variables</i>
Nonprescription services	How well the assortment corresponds to your needs The product you wanted to buy is in stock Time it takes to shop Quality of the products Prices of the nonprescription drugs Prices of the body and hygiene products
Self-serve	Ease of finding (i.e., signs and placement of products) Ease of getting help of personnel
Empathy	The personnel understands the concerns of the customer Information given in a way that is easy to understand Making sure that the customer has understood the reply Understanding that the customer wants privacy
Willingness to serve	The personnel's willingness to take the time the customer needs The personnel's friendliness The personnel taking care of the customer when there are people standing in line The personnel's flexibility to solve problems
Written material	Availability of information in the areas of interest Easiness of understanding the information The information on the packages is readable
Customer satisfaction	Overall satisfaction (first question) Overall satisfaction Overall performance versus expectations Overall performance versus an ideal convenience store
Loyalty	If there was an alternative to Apoteksbolaget, how likely would you be to still use Apoteksbolaget as your service provider?

REFERENCES

- Anderson, Eugene W. and Vikas Mittal (2000), "Strengthening the Satisfaction-Profit Chain," *Journal of Service Research*, 3 (November), 107-20.
- Auh, Seigyoung and Michael D. Johnson (1997), "The Complex Relationship between Customer Satisfaction and Loyalty for Automobiles," in *Customer Retention in the Automotive Industry: Quality, Satisfaction and Loyalty*, M. D. Johnson, A. Herrmann, F. Huber, and A. Gustafsson, eds. Wiesbaden, Germany: Gabler, 141-66.
- Bagozzi, Richard P. (1992), "The Self-Regulation of Attitudes, Intentions, and Behavior," *Social Psychology Quarterly*, 55 (2), 178-204.
- _____ and Youjae Yi (1994), "Advanced Topics in Structural Equation Models," in *Advanced Methods of Marketing Research*, Richard P. Bagozzi, ed. Cambridge, MA: Blackwell, 1-52.
- Bottomley, Paul A., John R. Doyle, and Rodney H. Green (2000), "Testing the Reliability of Weight Elicitation Methods: Direct Rating versus Point Allocation," *Journal of Marketing Research*, 37 (November), 508-13.
- Boulding, William, Ajay Kalra, Richard Staelin, and Valarie A. Zeithaml (1993), "A Dynamic Process Model of Service Quality: From Expectations to Behavioral Intentions," *Journal of Marketing Research*, 30 (February), 7-27.
- Carman, James M. (1990), "Consumer Perception of Service Quality: An Assessment of the Service Quality Dimension," *Journal of Retailing*, 56 (3), 33-55.
- Dillon, William R., John B. White, Vithala R. Rao, and Doug Filak (1997), "Good Science: Use Structural Equation Models to Decipher Complex Customer Relationships," *Marketing Research*, 9 (Winter), 22-31.
- Downey, R. G. and Craig V. King (1998), "Missing Data in Likert Ratings: A Comparison of Replacement Methods," *Journal of General Psychology*, 125 (2), 175-92.
- Doyle, John R., Rodney H. Green, and Paul A. Bottomley (1997), "Judging Relative Importance: Direct Rating and Point Allocation Are Not Equivalent," *Organizational Behavior and Human Decision Processes*, 70 (April), 65-72.
- Drolet, Aimee L. and Donald G. Morrison (2001), "Do We Really Need Multiple-Item Measures in Service Research?" *Journal of Service Research*, 3 (3), 196-294.
- Dyer, James S. (1990), "Remarks on the Analytical Hierarchy Process," *Management Science*, 36 (March), 249- 58.
- Fishbein, M. and I. Ajzen (1975), *Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research*. Reading, MA: Addison-Wesley.
- Fornell, Claes (1992), "A National Customer Satisfaction Barometer: The Swedish Experience," *Journal of Marketing*, 56 (January), 6-21.
- _____ and Fred L. Bookstein (1982), "Two Structural Equation Models: LISREL and PLS Applied to Consumer Exit-Voice Theory," *Journal of Marketing Research*, 14 (November), 440-52.
- _____ and Jaesung Cha (1994), "Partial Least Squares," in *Advanced Methods of Marketing Research*, Richard P. Bagozzi, ed. Cambridge, MA: Blackwell, 52-78.
- _____ 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 (October), 7-18.
- Frank, Ildiko E. and Jerome H. Friedman (1993), "A Statistical View of Some Chemometrics Regression Tools," *Technometrics*, 35 (2), 109-35.
- Garpentine, Terry H. (2001), "A Practitioner's Comment on Aimee L. Drolet and Donald G. Morrison's 'Do We Really Need Multiple-Item Measures in Service Research?'" *Journal of Service Research*, 4 (2), 155-58.

- Green, Paul E. and V. Srinivasan (1990), "Conjoint Analysis in Marketing: New Developments and Directions," *Journal of Marketing*, 54 (October), 3-19.
- Griffin, Abbie and John R. Hauser (1993), "The Voice of the Customer," *Marketing Science*, 12 (1), 1-27.
- Guadagni, Peter M. and John D. C. Little (1983), "A Logit Model of Brand Choice Calibrated on Scanner Data," *Marketing Science*, 2 (Summer), 203-38.
- Gustafsson, Anders and Michael D. Johnson (1997), "Bridging the Quality- Satisfaction Gap," *Quality Management Journal*, 4 (3), 27-43.
- Hayes, Bob E. (1998), *Measuring Customer Satisfaction: Survey Design, Use, and Statistical Analysis Methods*, 2nd ed. Milwaukee, WI: ASQ Quality Press.
- Jaccard, James, David Brinberg, and Lee J. Ackerman (1986), "Assessing Attribute Importance: A Comparison of Six Methods," *Journal of Consumer Research*, 12 (March), 463-68.
- Johnson, Michael D. and Anders Gustafsson (1997), "Bridging the Gap II: Measuring and Prioritizing Customer Needs," in *Proceedings of the Third Annual International QFD Symposium: Volume 2*, Anders Gustafsson, Bo Bergman, and Fredrick Ekdahl, eds. Linköping, Sweden: Linköping University, 21-34.
- _____ and _____ (2000), *Improving Customer Satisfaction, Loyalty and Profit: An Integrated Measurement and Management System*. San Francisco: Jossey-Bass.
- _____, _____, Tor Wallin Andreassen, Line Lervik, and Jaesung Cha (2001), "The Evolution and Future of National Customer Satisfaction Index Models," *Journal of Economic Psychology*, 22 (April), 217-45.
- Jöreskog, Karl G. (1970), "A General Method for Analysis of Covariance Structures," *Biometrika*, 57, 239-51.
- Kumar, Piyush (2002), "The Impact of Performance, Cost, and Competitive Considerations on the Relationship between Satisfaction and Re-purchase Intent in Business Markets," *Journal of Service Research*, 5 (August), 55-68.
- Lervik Olsen, Line and Michael D. Johnson (2003), "Service Equity, Satisfaction, and Loyalty: From Transaction-Specific to Cumulative Evaluations," *Journal of Service Research*, 5 (3), 184-95.
- Martilla, John A. and John C. James (1977), "Importance-Performance Analysis," *Journal of Marketing*, 41 (January), 77-79.
- Massy, William F. (1965), "Principal Components Regression in Exploratory Statistical Research," *Journal of the American Statistical Association*, 60, 234-46.
- Mazur, Glenn (1998), "QFD for Service Industries," in *The QFD Hand- book*, Jack V. ReVell, John W. Moran, and Charles A. Cox, eds. New York: John Wiley, 139-62.
- Mittal, Vikas and Wagner Kamakura (2001), "Satisfaction, Repurchase Intent, and Repurchase Behavior: Investigating the Moderating Effects of Customer Characteristics," *Journal of Marketing Research*, 38 (1), 131-42.
- Oliver, Richard L. (1997), *Satisfaction: A Behavioral Perspective on the Consumer*. New York: McGraw-Hill.
- Parasuraman, A., Valarie A. Zeithaml, and Leonard L. Berry (1988), "SERVQUAL: A Multiple-Item Scale for Measuring Consumer Perceptions of Service Quality," *Journal of Retailing*, 64 (Spring), 12-40.
- Rust, Roland T. and Naveen Donthu (2003), "Addressing Multi-collinearity in Customer Satisfaction Measurement," working paper, Robert H. Smith School of Business, University of Maryland at College Park.
- Ryan, Michael J., Robert Rayner, and Andy Morrison (1999), "Diagnosing Customer Loyalty Drivers: Partial Least Squares vs. Regression," *Marketing Research*, 11 (Summer), 19-26.
- Saaty, T. L. (1980), *The Analytic Hierarchy Process*. New York: McGraw-Hill.

- Scott, James and Peter Wright (1976), "Modeling an Organization Buyer's Product Evaluation Strategy," *Journal of Marketing Research*, 13 (May), 211-24.
- Steenkamp, Jan-Benedict E. M. and Hans C. M. van Trijp (1996), "Quality Guidance: A Consumer-Based Approach to Food Quality Improvement Using Partial Least Squares," *European Review of Agricultural Economics*, 23, 195-215.
- Wold, Hermann (1966), "Estimation of Principal Components and Related Models by Iterative Least Squares," in *Multivariate Analysis: Proceedings of an International Symposium Held in Dayton, Ohio*, P.R. Krishnaiah, ed. New York: Academic Press, 391-420.