QUALITY AND RELIABILITY ENGINEERING INTERNATIONAL
Qual. Reliab. Engng. Int. 2006; **22**:581–589
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/qre.763

Special Issue

Customer Satisfaction in the Airline Industry: the Case of British Airways

Giovanna Nicolini and Silvia Salini*,†
Department of Economics, Business and Statistics, University of Milan, Italy

In this paper two different non-classic methods, based on the analysis of qualitative data, are applied to evaluate customer satisfaction. The airline industry is considered with British Airways used as a case study. First, a classification algorithm based on the decision tree theory is performed. By conserving the original ordinal measuring scale for the items, it is possible to rank the dimensions in order to obtain a map of preference, select the groups of subjects more or less satisfied and delineate the peculiar characteristics of each group. Second, a probabilistic model, the Rasch model, is applied with the aim of distinguishing the two components (satisfaction and quality) that influenced the answer to each item and obtaining a numerical interval measure for both components. Copyright © 2006 John Wiley & Sons, Ltd.

Received 20 June 2005; Revised 21 November 2005

KEY WORDS: customer satisfaction; classification tree; Rasch model

1. INTRODUCTION

The analysis of quality and satisfaction is based on two different approaches involving objective and subjective aspects. This is always the case if quality and satisfaction are to be determined for a product, but it is not always true if the analysis is related to a service. The gathering of objective aspects is usually made with quantitative variables, whilst the gathering of subjective aspects, which are expressed with latent variables, is carried out using manifest variables. Such variables are mainly expressed in an ordinal scale, and, therefore, they cannot be processed as quantitative variables and they are not adaptable to the formation of summary indicators. The fuzziness between quality and satisfaction in analysing 'services' (which will be treated in Section 2), and the necessity to also adopt latent variables for quality evaluation leads to seek analysis methods more adequate to this context. In fact, in this paper we propose to apply two methodologies particularly suitable for qualitative data; in particular a probabilistic model, *the Rasch model* (covered in Section 3), originally adopted in psychometric studies and recently applied, with some changes in interpretation, to the quality-satisfaction context[1], and a *data driven* method that originates from Statistical Learning Theory[2], and is known by the name of *decision trees* (covered in Section 4). The latter is not included in quality and satisfaction measure techniques but nevertheless, if applied to manifest variables used to evaluate quality and satisfaction, provides an interesting segmentation of the reference sample, capable of indirectly providing the motivations at the various levels of quality and satisfaction. Therefore, the decision trees technique can be adopted as a

*Correspondence to: Silvia Salini, Department of Economics, Business and Statistics, Via Conservatorio 7 (III floor, Room 31), University of Milan, 20122 Milan, Italy.
†E-mail: silvia.salini@unimi.it

complementary instrument for the specific quality-satisfaction measure techniques. In Section 5 we present a joint application of the Rasch model[3] and of decision trees to measure quality satisfaction of the service offered by one of the largest worldwide airline companies.

2. ASPECTS OF SERVICE QUALITY AND SATISFACTION

Evaluations of service quality are not exclusively linked to the final result, as for products, but, and to a greater extent, to the whole *process* through which the service is provided. Consumer evaluation of the service received is shaped by factors linked to the psychological interaction that is established during the exchange transaction, and by factors connected to technical-specific characteristics of the service. The former factors concern behaviour, sensations and psychological benefits, which are difficult to measure, whilst the latter factors can be evaluated by objective indicators similar to those utilized for product quality.

To define the quality of a service Lehtinen and Lehtinen[4] identify two quality aspects, namely *process quality* and *results quality*. The former corresponds to the customer personal and subjective evaluation with respect to their participation to a service process; whilst the latter is related to customer evaluation with respect to process output. Such service output is also linked strictly to customer personal evaluations, since it is often connected with the customer's own subjective experience and presents very few tangible elements.

As a result, service quality is an elusive and abstract construction and, therefore, any attempt to measure it proves difficult. Parasuraman *et al*.[5] show that difficulties arise from the structure through which the service is provided, which the authors have characterized by intangibility, heterogeneity and inseparability: services are *intangible*, since they cannot be easily measured, counted, inventoried, tested and verified in order to ensure their quality before they are marketed; services are *heterogeneous* because it is difficult to guarantee their uniformity, as they are the result of inter-personal interactions (customer and staff), rather than of machines producing objects in series with the same characteristics. Service to customers can differ from company standards, and can easily change from supplier to supplier, from customer to customer or from context to context. Consumer behaviour, which is difficult to standardize, represents a fundamental component of the service process and, therefore, it is capable of influencing the result of service process; services are *inseparable* especially if they imply a strong interaction between the customer and the person providing the service, where customer input is decisive.

As a result of the objective impossibility of defining quality of service, we are led to study the psychological processes that underlie the *perceived quality* of the service. Perceived quality is linked, but not equivalent, to the concept of satisfaction; it results from comparing expectations and perceptions of service performance. The *perceived service quality model* proposed by Gronroos[6] is articulated in three quality dimensions.

(i) *Technical quality*. It is the result of *know-how* available to the company and relates to the technical output of the service process[‡]. Evaluations of this dimension are often objective since it is often possible to identify technical parameters for appraisal.
(ii) *Functional quality*. This represents the way the service is provided. High functional quality guarantees high satisfaction. Investing in functional quality is essential, especially for companies operating in the service sector where technical quality is very similar among competitors and any effort to make a substantial differentiation appears very hard[§].
(iii) *Corporate image*. This is the dominant dimension influencing *expected service*[¶]. It is dependent on the previous two dimensions, but also on exterior factors typical of the political-social-cultural context where

[‡]For instance, in the case of an airline company, a technical quality may be the transportation of a passenger from one airport to another.
[§]For instance, in the air transport industry, airline companies use almost identical fleets, have similar types of seats and services on board, often have stopovers in the same airports where they share the same facilities and staff. For companies operating in sectors characterized by lower differentiation, the research of distinctive elements becomes a primary objective.
[¶]Expected service differs from *perceived service*. The second results from customer evaluation of technical and functional qualities. Perceived service quality is obtained by comparing perceived service with expected service. A fundamental goal of a company, which aims at attaining good perceived service quality, is to have perceived service and expected service coincide.

DOI: 10.1002/qre

the company operates, on so-called 'word-of-mouth' (an effective communication instrument not directly controlled by the company) and on traditional marketing activities (advertising campaigns, pricing and public relations). The latter factors directly influence consumer expectations and, consequently, play a subtle role in the overall process of perception of quality by the consumer.

These three dimensions of quality are inter-connected: an acceptable technical quality is a requirement for a satisfactory functional quality, a good functional quality is capable of softening the negative effects due to temporary lack of technical quality, a company image properly studied can exercise a strong influence on customer perceptions related to technical–functional aspects of the service.

The perceived service quality model anticipates the modern formulation of the gap paradigm that is mainly based on *expectations, perceived performance, discrepancy and satisfaction*. The most recent versions of the paradigm consider perceived quality and satisfaction as two distinct constructions influenced by different variables; in fact, quality changes do not necessarily correspond to satisfaction changes of the same sign and intensity.

The most widespread application of the gap model is the SERVQUAL method proposed by Parasuraman *et al*.[7] in 1988 on the basis of the theoretical model (*gap model*) proposed by the same authors in 1985. This model overcomes difficulties in finding objective evaluations by considering consumer's subjective judgements in relation to their expectations and perceptions. It results in a quantitative instrument measuring quality indirectly, since it provides information on consumer perceived quality through the indirect comparison between perceived and expected services, rather than through the direct consumer evaluation process. Among the numerous critics to the SERVQUAL model, the most relevant criticism regards the concept of service quality adopted by the model: doubts arise on the validity of the hypothesis that quality is sought in the shifting of expectations from perceptions. Experimental studies conducted in the psychometric field by Babakus and Boller[8] reveal that evaluations on perceptions already include differences between perceptions and expectations. The introduction of these differences in the model would tend, according to this critic, to create redundancy in the model itself. The arguments put forward against the principles of the gap model are clarified in the model proposed by Cronin and Taylor[9], known as SREVPERF. However, this technique is also not entirely satisfactory, in the same way as all techniques that measure quality and satisfaction using quantitative methods on categorical data. In general, manifest variables utilized to record latent aspects of perceived quality and of satisfaction are measured with ordinal scales that are not sufficiently adaptable for measuring models of a quantitative nature, such as SERVQUAL, SERVPERF or, for instance, multiple regression. This is why some techniques have been proposed that allow the transformation of values from an ordinal scale into values expressed in a metric scale by formulating appropriate hypotheses (nonlinear regression model with latent variables, monotone regression model[10]) or techniques that, although complying with the characteristics of the ordinal scale, concentrate on its probability distribution (logistic regression model). Also other techniques allow the identification of a set of quantitative measures that are invariant from and independent of both subjective and objective aspects (Rasch model).

However, beyond the methods adopted to measure service quality and satisfaction, widespread agreement exists in considering service quality as preceding satisfaction and in understanding perceived quality as a global evaluation, while some confusion arises when consumer perceptions are considered with respect to service quality and when customer satisfaction is conceptually defined.

3. THE RASCH MODEL

The Rasch model, which was proposed in the psychometric field by Danish mathematician Georg Rasch in 1960[3], and item response models in general, are based on the hypothesis that a person's propensity to choose a specific value in the measuring scale is influenced by two factors:

- an *ability* factor or propensity, a person's specific aptitude towards a latent aspect;
- a *difficulty* factor specific to the item.

The first factor expresses the quantity of latent aspect possessed by the ith person and is function of the complex conglomerate of subjective aspects typical of the respondent (preferences, opinions, tastes and

DOI: 10.1002/qre

personal habits) onto which are superimposed expectations, frames of mind and subjective interpretation of questions and methods of the measuring scale in the questionnaire.

The second factor, attribute difficulty, expresses the quantity of latent aspect represented by the jth attribute and is linked to the more objective aspects typical of the item under investigation. The Rasch analysis is more frequently applied to data-sets resulting from an ensemble of aptitude tests, physical and psycho-motor tests and introspective questionnaires: answers are expressed in a dichotomous or ordinal scale for each of the items considered.

In the examination of such measures of ordinal and subjective nature given in the first instance, the Rasch analysis allows one to obtain quantitative estimates of objective measures of item difficulty (irrespective of the ability of the components of the sample considered) and of person propensity/ability (irrespective of item difficulty).

The objective of the Rasch model is to obtain, from dichotomous or ordinal methods, through which respondent evaluation is expressed, a set of *invariant* quantitative measures, capable of identifying a measuring scale independent from the subjectivity of the persons analysed and also independent from the attributes considered[11].

Translating what has been said so far into operating terms, given a sample of n persons asked to express their own opinion in relation to k attributes on a scale of $m+1$ modalities, we use θ_i to represent the ability of the subject, i.e. the quantity of latent feature present in the ith person, $(i = 1, 2, \ldots, n)$, β_j the item difficulty, i.e. the quantity of latent feature represented by the jth attribute, $(j = 1, 2, \ldots, k)$ and δ_h the threshold-measure between the two adjacent modalities h and $h-1$, i.e. the measure at which the two modalities have the same probability.

Before the construction of the model, we consider the modalities scale into which each of the n members of the sample has to express an evaluation for each of the k attributes considered. This scale, dichotomous or ordinal, is changed with the introduction of numbers in place of original modalities.

In the case of a dichotomous original scale, a scale of values 0 and 1 is introduced meaning respectively failure or success of the test $(h = 0; 1)$.

In the case of an ordinal scale the transformation is made recurring to the numerical values $h = 0, 1, \ldots, m$ (with $m \geq 2$) reflecting the hierarchical order of the same modalities.

The Rasch model is based on the fundamental hypothesis that the interaction of such factors is expressed by the formula

$$\theta_i - \beta_j - \delta_h \tag{1}$$

and that, therefore, the estimates provided by the model are a direct function of person ability θ_i and an inverse function of item difficulty β_j and of threshold difficulty in separating modalities δ_h (*additivity* property).

According to another fundamental assumption of the model, the so-called *Rasch theorem*, estimated parameters of an ability factor are identified irrespective of attribute difficulties and separation thresholds among modalities (*separability* property).

The probability that a subject responds with a certain modality is expressed in probabilistic units (*logits*) of the type

$$\ln\left[\frac{P(success)}{P(failure)}\right] \tag{2}$$

that vary on an interval scale.

The simplest version of the Rasch model is to be used when considering a dichotomous modalities scale with alternative responses, exhaustive or mutually exclusive of the n subjects interviewed.

In this case, the probability that the ith person overcomes the test associated jth item $(x_{ij} = 1)$ is only a function of attribute and subject parameters:

$$P(x_{ij} = 1) = \frac{\exp(\theta_i - \beta_j)}{1 + \exp(\theta_i - \beta_j)} \tag{3}$$

DOI: 10.1002/qre

The polytomic model is employed when, for each attribute, responses are required on a measuring scale of $m + 1$ modalities (with $m \geq 2$).

Some noteworthy polytomic models include:

- the *rating scale* model, the measuring scale that presents modalities having the same distance from one another (the separation threshold, δ_h, among modalities will therefore be constant for each item considered);
- the *partial credit* model, in which there are no equidistant modalities and thus parameters δ_h differ;
- the *rank* model, in which observations are ranks.

Of these models, the partial credit model will be considered here. Subject ability is, in this context, measured on a scale with modalities corresponding to numbers.

In the polytomic model, the linear expression (1) becomes

$$h\theta_i - h\beta_j - \delta_h \tag{4}$$

Consequently, the probability that the ith subject of the sample shows his own opinion with respect to the jth attribute with the hth modality becomes

$$P(x_{ij} = h) = \frac{\exp[h\theta_i - (h\beta_j + \delta_h)]}{\sum_{h=0}^{m} [h\theta_i - (h\beta_j + \delta_h)]} \tag{5}$$

In a recent paper[1], the Rasch model was extended to measuring quality and satisfaction. The response expressed through a questionnaire on the quality and satisfaction of a service contains a subjective component, linked to the personality, to the frame of mind and to the temperament of the respondent (perceived quality) and an objective component (objective quality) within the service itself and independent from any personal evaluation. The size of the gap between the two components usually depends on the competence of the person who is about to express an opinion on the service.

In standard applications of the Rasch model, difficulty is associated to item and ability is associated to person whereas, in this paper, quality is associated to item and satisfaction is associated to person. The interpretation of satisfaction and quality parameters changes compared to the interpretation of ability and difficulty parameters. In particular, high values for item parameters, which originally indicated high difficulties, now indicate low quality. Instead, the reading of subject parameters remains direct: originally high values indicated very skilful persons, now indicate very satisfied persons.

In conclusion, the Rasch model generates two sets of numerical parameters, measured on an interval scale:

(1) a set of numerical coefficients (β_j) associated to k items;
(2) a set of numerical coefficients (θ_i) associated to n persons.

Through calculation, a ranking of the items with regard to their quality can be obtained from the coefficients β_j. The model estimates a second series of parameters, θ_i, that express each person's satisfaction; this parameter can be considered as a customer satisfaction index.

4. DECISION TREES

Decision trees represent one of the simplest and often most effective tools used in classification and regression problems. It is advantageous to use a classification or regression tree in the context of service quality evaluations because, given a response variable on global satisfaction in the questionnaire, it is possible to select only the items, or predictors, deemed important in the determination of customer satisfaction. Moreover, the method allows us to identify what levels of single items discriminate between satisfied and dissatisfied customers. Basically, the final result obtained is a certain number of consistent groups of customers, easily identifiable by the rules generated by the tree characterized by the same level of satisfaction, expressed by a score which

DOI: 10.1002/qre

is a conditional frequency when the response variable is a category, and a conditional mean when the response variable is a numerical variable. Classification with trees originates from supervised learning techniques and is, therefore, a *data driven* approach does not need *a priori* parametric hypotheses on data distributions and particular measuring scales for variables and it is also capable of classifying spaces that are nonlinearly separable. Unlike the techniques typical of *statistical learning theory*[12], such as *support vector machines* and *neural networks*, the decision tree is a useful and intuitive interpretative model as well as an efficient forecasting model.

Segmentation analysis outputs are visualized through a *tree diagram*, which starts with a node (*root*) where all n statistical units of the sample are placed and expands through ramifications of nodes including subsets of smaller and homogenous units, resulting from recursive segmentation. The node generating the segmentation is called *parent*, while descending nodes are called *children*; the terminal nodes are named *leaves*.

On the basis of the number of children nodes generated at each step, segmentation techniques can be distinguished as follows:

- *binary*, for two-way repartitions;
- *ternary*, for three-way repartitions;
- *multiple*, for k-way repartitions.

The tree algorithm is an iterative procedure through which the sample of statistical units considered is split into several groups, according to the following sequence:

(1) selection of explanatory and dependent variables;
(2) choice of the segmentation criterion;
(3) choice of the stop criterion;
(4) assignment of one of the J modalities of the dependent variable to each leaf, definition of the classification rule $\mathrm{d}(x)$ and estimate of misclassification rates.

5. QUALITY AND SATISFACTION OF THE BRITISH AIRWAYS SERVICE

The data obtained from the questionnaires filled in by 19 653 passengers in March 2003 on board British Airways flights have been analysed[‖].

5.1. Classification tree

Having indicated global satisfaction and overall experience as dependent variables, and the 10 dimensions of the service (*overall booking, check-in, transfer, lounge, departure, cabin environment, meal, in-flight entertainment, goods for sale and cabin crew*) as explanatory variables, the CHAID tree[13] has been applied. The cluster of people, who have expressed satisfaction in relation to overall experience, is split in two nodes with respect to the dimension to which is associated the highest value of the χ^2 statistic and which, therefore, makes the conditional frequencies differ the most. The dimension chosen at the first level of segmentation is *cabin crew* (with $\chi^2 = 8\,810\,766$), which splits the root in nodes 1 and 2, containing respectively the persons satisfied and dissatisfied with the service offered by cabin crews. Any new node is submitted to further segmentation procedures as far as assigned stop rules allow (maximum number of levels lower than the root, minimum number of persons contained in each parent and child nodes, χ^2 statistic close to zero with p-value higher than 0.05). In addition to *cabin crew*, we identify *cabin environment, departure* and *meal and refreshment* as the more influential dimensions of the overall satisfaction of respondents. Furthermore, the segmentation technique CHAID has enabled the synthesis of the classification rules of respondents. Terminal nodes (leaves) of the tree

[‖]Intercontinental flights have been ignored.

DOI: 10.1002/qre

Table I. β parameters of attributes of service

ID	Attributes	β
6	Crew	−1.376
1	Booking	−0.282
2	Check-in	−0.222
3	Departure	0.185 417
4	Cabin	0.440 972
5	Meal	0.679 861

diagram have identified clusters of persons with a similar level of satisfaction. The analysis of the composition of such clusters of people on the basis of demographic, socio-economic, geographic and behavioural characteristics recorded in the first section of the questionnaire are particularly interesting. In particular, the group of least satisfied people is mainly composed of males, aged under 24 and over 50 years who travel in 'economy' class (*Euro Traveller, U.K. Domestic* and *World Traveller*) and do not possess fidelity cards (*Executive, Gold, Silver*).

5.2. *Rasch analysis*

Given that segmentation is valuable in customer classification and in ordering categorical variables, it nevertheless cannot identify indices of quality of items and satisfaction of people. The Rasch model, only recently applied to the analysis of quality/customer satisfaction, proves particularly useful because it is capable of simultaneously analysing both *perceived quality* for each subject, and *objective quality* for each attribute of the service provided. The Rasch model has been applied to the same group of people that underwent segmentation. The parameters have been estimated only considering the 12 710 non-*extreme* subjects**. The Cronbach index Alpha is equal to 0.731 and the attributes are therefore able to measure a common feature.

According to the interpretation of the item parameter given in Section 3, the highest quality is associated to attribute *cabin crew*, whilst the lowest quality to attribute *meal*, as shown in Table I.

Figures 1 and 2 show attribute curves. On the horizontal axis the subject parameter θ_i is shown, whilst the vertical axis shows probabilities related to each response category (satisfied = 1/dissatisfied = 0). In particular, Figure 1 shows the case of the attribute to which the highest quality is associated (*cabin crew*). It is observed that, regardless of the satisfaction level shown on horizontal axis (subjective factor), there is a very high probability that the subject is satisfied, thus implying the presence of real product quality. Instead, in Figure 2 the attribute considered is *meal*, i.e. the one with the lowest quality associated to it. The probability that the subject expresses satisfaction is low, in this case, irrespective of subjective factor.

5.3. *Final remarks*

The decision tree has highlighted what are the important items in the evaluation of global satisfaction. The Rasch analysis allowed to rank the items on the basis of their objective quality. At the operating level, British Airways should invest in low-quality items, for instance *meal and refreshments*, which are important for subject global satisfaction.

With regard to customer satisfaction, the coefficients θ_i can be utilized as a *customer satisfaction index*. This index can be used to segment people, both through the application of regression trees and through algorithms of cluster analysis. Furthermore, the index so obtained, unlike the score produced with the classification tree of Section 5.1, which is the same for any final node and thus for any subject in the same

**The extreme person are subjects have manifested their own degree of satisfaction with the same score in all the items.

DOI: 10.1002/qre

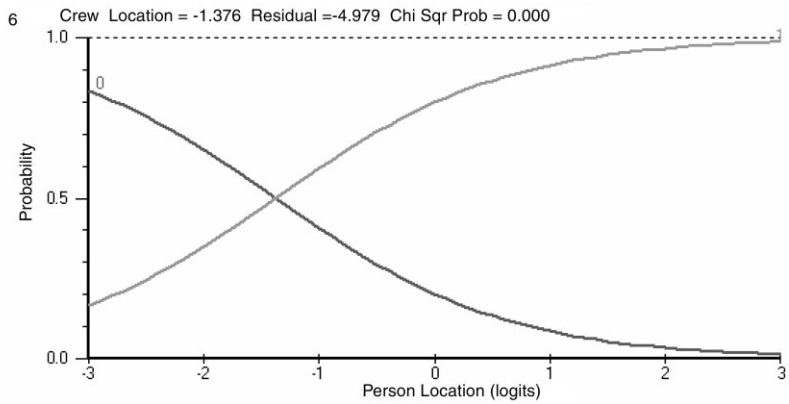

Figure 1. Probability category curve related to the attribute with the best quality (*cabin crew*)

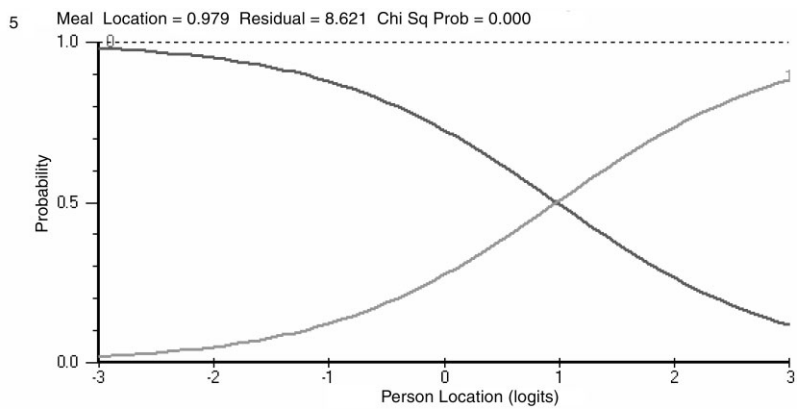

Figure 2. Probability category curve related to the attribute with the worst quality (*meal and refreshment*)

group, is an index measured on an interval scale. However, it appears that there is coherence between the two measures: the subjects giving greatest satisfaction according the CHAID segmentation are those with the highest values of the parameter θ_i in the Rasch model.

Acknowledgements

The authors are grateful to Marco Battaglia of British Airways and NOP Research for giving the dataset.

REFERENCES

1. De Battisti F, Nicolini G, Salini S. The Rasch model to measure the service quality. *The ICFAI Journal of Service Marketing* 2005; in press.
2. Breiman L, Friedman JH, Olshen RA, Stone CJ. *Classification and Regression Trees*. Chapman and Hall: New York, 1984.

DOI: 10.1002/qre

3. Rasch G. *Probabilistic Models for Some Intelligence and Attainment Tests*. Danish Institute for Educational Research: Copenhagen, 1960.
4. Lehtinen U, Lehtinen J. Two approaches to service quality dimensions. *The Service Industries Journal* 1991; **11**(3):287–303.
5. Parasuraman A, Zeithaml V, Berry L. A conceptual model of service quality and its implications for future research. *Journal of Marketing* 1985; **49**:41–50.
6. Gronroos C. A service quality model and its marketing implications. *European Journal of Marketing* 1984; **2**:244–249.
7. Parasuraman A, Zeithaml V, Berry L. SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing* 1988; **64**:40.
8. Babakus E, Boller G. An empirical assessment of the SERVQUAL scale. *Journal of Business Research* 1992: **24**: 253–268.
9. Cronin J, Taylor SA. Measuring service quality: A re-examination and extension. *Journal of Marketing* 1992: **56**:55–68.
10. Zanella A. Valutazione e modelli interpretativi di customer satisfaction: Una presentazione d'insieme. *Working Paper Serie E.P. n. 105*, Istituto di statistica, Università Cattolica del Sacro Cuore, Milano, 2001.
11. Andrich D. A rating formulation for ordered response categories. *Psychometrika* 1978; **43**(4):561–573.
12. Hastie T, Tibshirani R, Friedman JH. *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer: Berlin, 2001.
13. Kass GV. An exploratory technique for investigating large quantities of categorical data. *Applied Statistics* 1980; **29**:119–127.

Authors' biographies

Silvia Salini holds a degree in Statistics from the Catholic University of Milan and a PhD degree in Statistics from the University of Milan 'Bicocca', Italy. Currently she is Assistant Professor of Statistics at the Department of Economics, Business and Statistics of the University of Milan. Her main research interests are multivariate statistics analysis, data mining and statistics for social science.

Giovanna Nicolini holds a degree in Statistics from University of Rome 'La Sapienza', Italy. Currently she is a Full Professor of Statistics at the Department of Economics, Business and Statistics of the University of Milan. She is also the coordinator of the PhD in Statistics of University of Milan 'Bicocca'. Her main research interests are sampling techniques and customer satisfaction.

DOI: 10.1002/qre