Transportation Planning and Technology

ISSN: 0308-1060 (Print) 1029-0354 (Online) Journal homepage: http://www.tandfonline.com/loi/gtpt20

A Stated Preference Experiment for Measuring Service Quality in Public Transport

Laura Eboli & Gabriella Mazzulla

To cite this article: Laura Eboli & Gabriella Mazzulla (2008) A Stated Preference Experiment for Measuring Service Quality in Public Transport, Transportation Planning and Technology, 31:5, 509-523, DOI: 10.1080/03081060802364471

To link to this article: http://dx.doi.org/10.1080/03081060802364471

Transportation Planning and Technology, October 2008
Vol. 31, No. 5, pp. 509–523

ARTICLE

A Stated Preference Experiment for Measuring Service Quality in Public Transport

LAURA EBOLI & GABRIELLA MAZZULLA

Department of Land Use Planning, University of Calabria, Rende (CS), Italy

(Received 18 January 2007; Revised 14 February 2008; In final form 16 July 2008)

ABSTRACT This paper develops a Stated Preference (SP) experiment that provides a way to measure service quality in public transport. The paper introduces an empirical procedure for optimising the SP experiment. This procedure permits the identification of the choice alternatives defining the experiment by simulating the choices of a user sample. By using the data collected from an experimental survey, a Multinomial Logit model was calibrated. This model is a way of identifying the importance of service quality attributes on global customer satisfaction and calculating a Service Quality Index, which provides an operationally appealing measure of current or potential service effectiveness.

KEY WORDS: Service quality; public transport; stated preference; experiment; simulation

Theoretical Framework

Service quality is a subject that has aroused considerable interest both in academic research and in public and private service sectors, where managers are inclined towards customer-focused service and continuous performance improvement. Specifically in public transport, service quality is a matter of the greatest importance because an improvement of quality levels can attract further users. An increase in

Correspondence Address: Laura Eboli, Department of Land Use Planning, University of Calabria, 87036 Rende (CS), Italy. Email: laura.eboli@unical.it

ISSN 0308-1060 print: ISSN 1029-0354 online
DOI: 10.1080/03081060802364471

public transport use, with a concurrent reduction in the use of the private car, could help to reduce many problems like traffic congestion, air and noise pollution, and energy consumption.

For these reasons, many techniques for measuring service quality and defining the importance of service quality attributes on global satisfaction have been proposed in the literature. These techniques are more often not based on customer evaluation. Service quality and customer satisfaction can be evaluated according to different methods: by asking customers their perception/satisfaction on service quality, expectation/importance, or both perception and expectation; in addition, perception can be compared with the zone of tolerance of expectations (the range defined by the maximum desired level and minimum acceptable level of expectations) (Figini, 2003). Customers can be asked to rate or rank some service attributes. Furthermore, a rating on overall satisfaction can be asked for.

Methods of measuring service quality and customer satisfaction can be identified in two different categories. The first category includes techniques of statistical analysis, such as quadrant and gap analysis, factor analysis, scatter graphs, bivariate correlation, cluster analysis, and conjoint analysis. Some of these techniques provide an evaluation of the service attributes; others provide the relationship of the service attributes with overall satisfaction. Various authors have introduced some indices for measuring overall satisfaction or service quality. For a more detailed discussion of these techniques see Kano *et al.* (1984), Zeithaml *et al.* (1986), Berger (1993), Akan (1995), Cuomo (2000), Hill (2000, 2003), Bhave (2002) and Hartikainen *et al.* (2004).

The second category of methods consists in the estimation of coefficients by modelling. Here models are used that relate global service quality (dependent variable) to some attributes (independent variables). Some of these models are regression models, structural equation models (SEM) (Bollen, 1989) and Logit models. Examples of SEM are reported in Vilares and Coelho (2003) and Grønholdt and Martensen (2005); an ordinal regression technique was proposed by Siskos *et al.* (1998).

As far as the authors' are aware, few methods for measuring customer satisfaction and service quality in public transport have been reported in the literature. In the *Handbook for Measuring Customer Satisfaction and Service Quality*, published by the US Transportation Research Board, the 'impact score' technique is described (TRB, 1999), while a linear regression model was proposed by Jones (Swanson *et al.*, 1997).

Some Logit models have been proposed by Hensher (Prioni & Hensher, 2000; Hensher, 2001; Hensher & Prioni, 2002; Hensher *et al.*, 2003). Logit models are discrete choice models based on random utility theory that a choice probability is calculated for each alternative; the utility of each choice alternative is the sum of a systematic component and a random one; all random components are distributed according to a Gumbel random variable (Ben-Akiva & Lerman, 1985; Cascetta, 2001). By the estimation of coefficients of the models the importance of service quality attributes on global customer satisfaction can be evaluated. A Service Quality Index (SQI) is then calculated by using estimated coefficients. This index, as the utility related to each alternative of choice, is calculated like a linear combination of attributes, each one weighted as to its importance. In one study of the bus industry (Hensher *et al.*, 2003), 13 service quality attributes were selected; each attribute varied on three levels producing different alternatives (bus packages).

A Multinomial Logit (MNL) model was proposed in Prioni and Hensher (2000) and Hensher and Prioni (2002). This model provides the set of indicators required to represent a user-based measure of service quality. A survey was addressed to bus users of 25 private bus operators in New South Wales (Australia). In Hensher *et al.* (2003), a Nested-Logit model was proposed for comparing the service quality levels within and between the bus operators. A Mixed Logit model was proposed by Hensher (2001) in order to explore observed and unobserved heterogeneity among users. Each of these models was calibrated using Stated Preference (SP) data. The major advantage of SP data compared with Revealed Preference (RP) data is given by the possibility of considering a more extensive attributes space (Pearmain *et al.*, 1991; de Dios Ortùzar & Willumsen, 1994).

SP data applications, related to stated choices in a specific hypothetical context, have assumed a growing importance in the few last decades. Various authors have proposed a variety of methods for using this kind of data, and some models derived from them (Pearmain *et al.*, 1991; de Dios Ortùzar, 1992); however, many authors assert that a direct application of these models in order to forecast the choices made by users is not appropriate (Bradley & Daly, 1992; Fujiwara & Sugie, 1992). For this reason, some authors have proposed joint calibration models using RP and SP data (cf. Ben-Akiva & Morikawa, 1990). In the literature three methods for the joint calibration have been reported: the first, proposed by Ben-Akiva and Morikawa (1990), is based on a sequential estimation procedure; the second, proposed by Bradley and Daly (1991), is based on the simultaneous estimation of the RP/SP model parameters and the parameter θ; the third, proposed by Swait and Louviere (1993), is based on an iterative estimation procedure.

This paper introduces an SP experiment that provides a service quality measure. A Logit model for calculating a SQI is proposed. In the following section, the experimental context is described and the descriptive statistic results of the survey realised to support the research are summarised. In this section a simulation procedure for optimising the SP experiment is also introduced. In the third section the proposed model is described and the calibration results are shown. Finally, a brief concluding section is reported.

Experimental Context

Survey Design and Statistical-Descriptive Analysis of Data

A sample survey of University of Calabria students was conducted. The campus is situated in the urban area of Cosenza (in the south of Italy) and is attended by approximately 32,000 students and 2000 members of staff (2006).

In a typical working day, about 8000 students travel by urban bus. The survey, realised in the winter of 2006, involved a sample of 470 students who live in the urban area and habitually use the bus to reach the campus. Therefore, the sampling rate was approximately 5.8%. Respondents were asked to provide information about their trip habits getting to and from the university and, in addition, about public transport service quality.

The interview was divided into three sections: in the first and second section some information about socioeconomic characteristics (gender, age, income and car availability) and travel habits was elicited. The last section included an SP experiment submitted to users, in which they were asked to make a choice between the current bus service, which was the alternative representing user habitual service, and two hypothetical bus services, which were the SP alternatives. The current alternative was defined by the user, who assigned a value to some service quality attributes on a scale with more levels, according to the bus service used at the time of the interview. Each SP alternative represented a combination of the attribute levels. The choice alternatives were defined by nine attributes varying on two levels; some levels were not available for the current service. Table 1 reports the attribute levels.

Each SP experiment was defined by the current bus service and two SP alternatives. Only three alternatives were submitted to users because it was considered that they could have some difficulties in making a choice between more than three alternatives when several attributes define the alternatives (cf. Prioni & Hensher, 2000; Hensher & Prioni, 2002). The SP alternatives were coupled and joined to the current alternative producing several types of experiment, each of which was submitted to a group of users. The selection and the coupling of the SP treatments were the major difficult phases in the survey design. For this

Table 1. Service quality attributes and levels

Service quality attributes	Levels
Walking distance to the bus stop	Same as now (1); 10 minutes more (0)
Frequency	Every 15 minutes (1); same as now (0)
Reliability	On time (1); late (0)
Bus stop facilities	Bus shelter, seats and lighting (1) no shelter, no seats, no lighting (0)
Bus crowding	No overcrowded (1); overcrowded (0)
Cleanliness	Clean enough (1); not clean enough (0)
Fare	25% more than the current fare (1); same as now (0)
Information	Timetable, map, announcement of delays (1) no timetable, no map, no announcement of delays (0)
Transit personnel attitude	Very friendly (1); very unfriendly (0)

reason an empirical procedure was proposed. An example of an experiment is shown in Table 2. To some users, two experiments were submitted. In these cases only an SP alternative was replaced in the second experiment in order to reduce the fatigue effect on the respondent.

The sample was spread over 46% male and 54% female respondents. Eighty-nine percent of the student sample was between 18 and 24 years old. The sample was divided, also, between 'in course' and 'out course' students; in Italy, the 'out course' condition relates to a university

Table 2. Example of an SP experiment submitted to the interviewee

Attributes	Actual service	Service bus A	Service bus B
Walking distance to the bus stop	Same as now	10 minutes more	Same as now
Frequency	Same as now	Same as now	Every 15 minutes
Reliability	On time	Late	Late
Bus stop facilities	No shelter, no seats, no lighting	Bus shelter, seats and lighting	No shelter, no seats, no lighting
Bus crowding	Overcrowded	Overcrowded	No Overcrowded
Cleanliness	Clean enough	Clean enough	Not clean enough
Fare	Same as now	Same as now	25% more than the current fare
Information	No timetable, no map, no announcement of delays	Timetable, map, announcement of delays	No timetable, no map, no announcement of delays
Transit personnel attitude	Very friendly	Very friendly	Very unfriendly

student who has not completed their studies in the prescribed time. The 'in course' students represented 78% of the total. About 50% of students belonged to a middle class of family income and about 35% to a lower-middle class. Almost all the students did not have the possibility of using a car to reach the campus (92%).

Simulation Procedure

The empirical procedure permits the simulation of the choices of a sample of users, according to which a model is calibrated. The choice scenarios are defined when the statistical tests of goodness-of-fit are verified (Figure 1).

The simulation of the choices is made by considering a sample of users selected in a preliminary survey, and by using a basic model specified as a function of the attributes defined on *a priori* basis. This sample of users should have the same socioeconomic characteristics as the sample in the real survey. Each interviewee describes their current context by taking into account the attributes and levels available at the time of the survey. Specifically, the data of a sample of 513 interviewees were used; these data were collected by a survey effected in December 2004 on the University of Calabria students. Some information about the tours to reach the campus and about some service quality attributes were asked to the users. These data and information permitted construction of the current contexts.

The SP treatments were selected from the full factorial design, which included all the possible combinations among the attribute levels. Each of nine attributes has two levels of variation and then the complete factorial involves 2^9 or 512 combinations. Usually, the selection of the treatments can be effected by some partialisation techniques of full factorial design. Generally, these techniques are used in order to select treatments that retain the orthogonality of the comparisons, i.e. the independence between attribute combinations. By using the partialisation techniques the possibility of estimating some interaction effects is lost. Two partialisation techniques of full factorial design are usually adopted. The first is the *block decomposition* of the full factorial design, which is based on the principle of subdividing the set of scenarios into groups (blocks) to submit to different decision makers. The second technique, known as *fractional factorial design*, completely eliminates some scenarios while retaining orthogonal comparisons which allow estimation of the main effects. A fractional factorial design can be obtained from a full design through a 'defining relationship', in which the level of a given factor is obtained from those of other

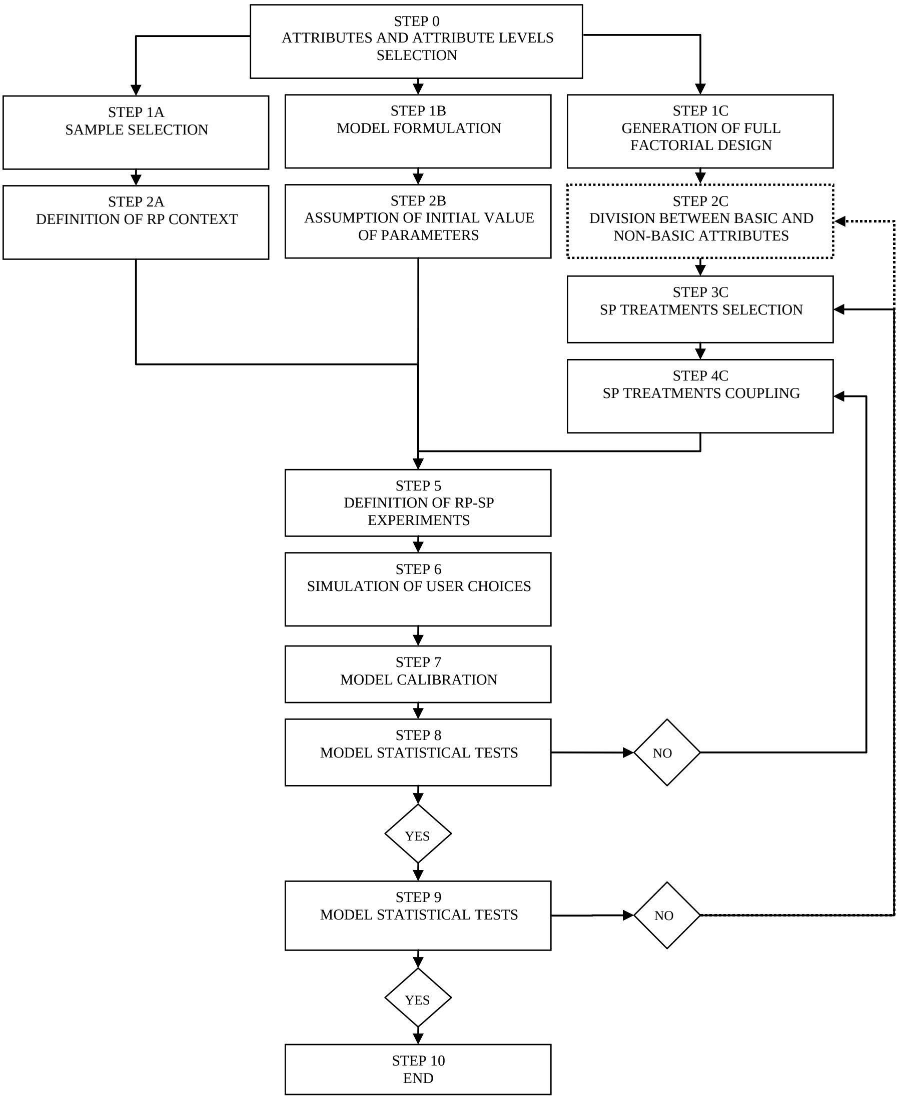

Figure 1. Simulation procedure flowchart

factors. In many cases the two techniques are used contemporaneously (Cascetta, 2001).

In this experimental context, the treatments that consider some main effects and some low-level interaction effects were selected. In order to select the treatments, the service quality attributes were subdivided into 'basic' and 'non-basic' attributes. The basic attributes represent the service performances that can compromise the service quality when their level is low. The non-basic attributes are secondary characteristics that have a direct influence on service quality if present, while they do not compromise it if absent. Specifically, bus crowding, service frequency, service reliability, fare and walking distance to the bus stop are considered as basic attributes. Bus stop facilities, cleanliness of vehicles, information and transit personnel attitude are considered as non-basic attributes. The distinction between the two categories of service quality attributes was made by considering the preferences expressed by the sample of users in the preliminary survey. For the selection of the treatments, the relationships between the attributes of the same category and between the attributes belonging to different categories were taken into account. As an example, some treatments in which only a basic attribute had a higher level of variation or only a basic and a non-basic attribute had a higher level of variation were selected.

In a typical SP experiment all the selected treatments are submitted simultaneously to the interviewee; however, in this case there are too many selected treatments and the number of attributes defining the treatments is too high. Therefore, the generation of some couplings of SP treatments was necessary; these couplings were joined to the current alternative described by the user. In addition, whereas the SP alternatives are defined *a priori* by the analyst, the current alternatives are described at the time of interview; this fact permits a forecast of the effects owing to the SP alternatives. For this reason, the simulation procedure was necessary. In this context a random utility model, i.e. an MNL model, was used. The systematic utility functions of the alternatives are linear combinations of the service quality attributes, as shown in the following expressions:

$$V_j = \beta_{WTime} WTime_j + \beta_{Freq} Freq_j + \beta_{Rel} Rel_j + \beta_{Stop} Stop_j + \beta_{Crow} Crow_j + \beta_{Clean} Clean_j + \beta_{Fare} Fare_j + \beta_{Inf} Inf_j + \beta_{Per} Per_j \quad (1)$$

with j varying from 1 to n, in which n is the number of selected treatments.

The initial value of the parameters of each attribute were supposed according to the information gained from the preliminary survey and by considering similar models reported in the literature (cf. Prioni & Hensher, 2000; Hensher & Prioni, 2002). Usually, the initial value of the parameters is estimated from the model calibrated by using RP data collected before the survey. This is possible if there are more alternatives, e.g. different transport modes, in the RP context. But, in this specific case, only one alternative is available in the RP context, i.e. the current bus service.

The utility function of each alternative is:

$$U_j = V_j + \varepsilon_j \tag{2}$$

A random residual distributed according to a normal function, with mean equal to 0 and variance equal to 0.5, has been assigned to each user. The values of the residuals relate to choice probabilities generated as random numbers between 0 and 1. For each iteration of the procedure, the utility of each alternative is calculated and the choice of each user is simulated; the Logit model is calibrated according to simulated choices. By the statistical tests and the comparisons with the results of a previous iteration, the most convenient couplings of the SP treatments are chosen and then the SP experiments to submit to users in the real survey are defined. When the SP treatments are selected, all the possible treatment couplings are tested in a cycle of simulations and then the best model is chosen. Other cycles can be effected by selecting other treatments.

Simulation Results

By applying the simulation procedure to the experimental context analysed, 50 treatments were selected and 32 couplings were composed. This goal was achieved by several simulations. Some simulations provided models with incongruent results in which some parameters did not have a correct sign and were not statistically significant. In other models some parameters were significant at a level lower than 90% and some goodness-of-fit statistics did not have satisfactory values (as likelihood ratio (LR) and Rho-squared statistics, final Log-Likelihood, etc), although all the parameters had a correct sign. The models with the highest values of Log-Likelihood were selected among the statistically significant models. The parameters and the statistical tests are reported in Table 3.

The differences between the models were not significant, according to the LR statistic. Specifically, the LR value comparing the first simulation with the second is equal to 6.77, while the value comparing the first simulation with the third is 1.32; finally, the LR value comparing the second simulation with the third is 5.46 (against a critical value equal to 16.92, with 9 d.o.f. at 95% level of significance). In addition, the other goodness-of-fit statistics had comparable values. Therefore, the model relating to the third simulation was considered for the selection of treatments and the choice of couplings, according to the best parameter *t*-student values.

Table 3. Results of simulations

		First simulation		Second simulation		Third simulation	
Acronym	Parameter	Estimation	*t*-Student	Estimation	*t*-Student	Estimation	*t*-Student
WTIME	β_{WTIME}	−0.326	−7.6	−0.335	−8.3	−0.324	−8.0
FREQ	β_{FREQ}	1.924	8.3	2.234	9.3	2.169	9.4
REL	β_{REL}	0.771	5.1	0.691	4.6	0.729	4.8
STOP	β_{STOP}	0.210	1.3	0.243	1.6	0.214	1.4
CROW	β_{CROW}	1.006	6.7	0.961	6.0	0.987	6.0
CLEAN	β_{CLEAN}	0.321	2.0	0.369	2.4	0.301	2.0
FARE	β_{FARE}	−1.703	−1.7	−1.102	−1.1	−1.390	−1.4
INF	β_{INF}	0.575	3.8	0.628	4.2	0.574	3.9
PER	β_{PER}	0.182	1.2	0.233	1.6	0.242	1.6
Final value of log-likelihood		−360.556		−357.168		−359.900	
Log-likelihood with zero coefficients		−563.588		−563.588		−563.588	
Rho-squared		0.360		0.366		0.361	
Rho-squared corrected		0.344		0.350		0.345	
Likelihood ratio		406.064 (χ^2 = 16.919)		412.840 (χ^2 = 16.919)		407.376 (χ^2 = 16.919)	
% right		81.60% (419/513)		83.04% (426/513)		82.26% (422/513)	

Estimation Results

The MNL choice model previously shown has been calibrated and validated by using the data collected in the real survey. A total of 640 observations was incorporated in the estimation of the model.

All the variables are dichotomous, except *WTime* and *Fare* which are continuous and measured in minutes and in Euros (€), respectively.

The results of the calibration are shown in Table 4. All parameters have a correct sign and assume a value statistically different from zero, at a 95% level of significance. The service frequency is a statistically strong attribute, with a t-value equal to 11.5. As expected, *WTime* and *Fare* assume a negative sign, in order to indicate that an increase of fare and distance from the bus stop involves a decrease of utility.

The model verifies the statistical tests on the goodness-of-fit. Rho-squared is equal to 0.331. This value may not appear very good, but it relates to a non-linear model. As reported in Louviere *et al.* (2000), the corresponding r-squared in a linear model would be equal to 0.8 approximately. The LR statistic is much higher than the critical value with 9 d.o.f. and the% Right has a good value (almost 70%).

The parameter of the variable *Fare* has the highest value because the sample is composed of students with a middle or lower-middle family income. Nevertheless, by multiplying this parameter by the value of the corresponding attribute, the fare effect on disutility is comparable with the frequency effect.

The utility of each alternative is an SQI of each bus package and the parameter values are the attribute weights (Prioni & Hensher, 2000).

Table 4. Results of the model estimation

Variable	Acronym	Parameter	Estimation	t-Student
Walking distance to the bus stop	Wtime	β_{WTime}	−0.135	−8.5
Frequency	Freq	β_{Freq}	2.564	11.5
Reliability	Rel	β_{Rel}	1.309	8.8
Bus stop facilities	Stop	β_{Stop}	0.583	3.9
Bus crowding	Crow	β_{Crow}	0.554	3.2
Cleanliness	Clean	β_{Clean}	0.827	5.9
Fare	Fare	β_{Fare}	−7.619	−8.5
Information	Inf	β_{Inf}	0.558	3.7
Transit personnel attitude	Per	β_{Per}	0.491	3.5
Final value of log-likelihood	−470.376			
Log-likelihood with zero coefficients	−703.112			
Rho-squared	0.331			
Rho-squared corrected	0.318			
Likelihood ratio	465.484 ($\chi^2 = 16.919$)			
% right	68.59% (439/640)			

Short of the attributes with a negative parameter, service frequency is the attribute with the higher weight on global service quality. Indeed, an increase in bus frequency from one bus every hour to one bus every 15 minutes produces, *ceteris paribus*, an increase of about 2.6 on the SQI.

Other important attributes are reliability, cleanliness and bus stop facilities. A simultaneous improvement of the three attributes is comparable to an improvement in terms of frequency.

Other MNL choice models were calibrated in order to verify the strength of the simulation procedure. Some models, in which other treatments and couplings of treatments were used, gave results worse than the model chosen by the simulation. In these models the goodness-of-fit statistics had worse values and some parameters were not statistically significant. As an example, in a model also calibrated on 640 observations, the ρ^2 statistic was equal to 0.18 and the final value of Log-likelihood was equal to -576.844. These values are clearly lower than the values reported in Table 4. In addition, the estimated parameter corresponding to the personnel attitude did not have a correct sign. By using the simulation procedure for the calibration of this model we obtained similar results.

Conclusions

The main purpose of this research has been to explore the optimal design of an SP experiment for measuring service quality in public transport. For this aim an empirical procedure for simulating user choices has been proposed. The procedure, based on the use of real data collected by a preliminary survey, allows the most convenient couplings of the SP treatments to be chosen and the experiments to be defined.

The procedure was necessary because the experimental design was very complex owing to the numerousness of the selected treatments and the attributes defining the treatments. Generally, in a typical SP experiment, the number of treatments and attributes is so small that all the selected treatments can be submitted simultaneously to the interviewees, or blocks of treatments can be submitted to groups of interviewees; but, in the experimental context considered in this research, there were too many selected treatments defined by a very high number of attributes. Therefore, the simulation procedure was indispensable in order to generate the couplings of SP treatments to submit to the interviewees with the alternative representing the current service.

The procedure was also necessary in order to reduce the sample numerousness and to submit the most efficient experiments (more information) to users, in order to obtain economic benefits.

By means of statistical tests and comparisons with the results of several simulations, the most convenient experiments have been defined. These experiments were submitted to a sample of users in a real survey in order to calibrate a Logit model. The model calibrated by the real data verified the statistical tests on the goodness-of-fit. All parameters have a correct sign and assume a value statistically different from zero, at a 95% level of significance. Thanks to these good results we can say that the simulation procedure is appropriate.

The utility of each alternative represents an SQI of each bus package and the values of parameters are the attribute weights. This is a way for quantifying the improvement of service quality as a consequence of an improvement of the service quality attributes. SQI is useful to planners and transit operators for measuring the importance of service quality attributes.

The proposed model might not seem very useful for measuring the service quality of public transport in an urban area, because it relates to a specific category of users, the students of a university campus. Nevertheless, one should consider that the experimental context is represented by the urban area of Cosenza in which the campus is the major centre of attraction. Therefore, students are a relevant part of total public transport users because the other user categories scarcely travel by public transport.

Further developments of this research may be identified in the use of more complex Logit models, like Hierarchical-Logit or Mixed Logit models. By considering Hierarchical-Logit structures, the behavioural differences between market segments defined by user socioeconomic characteristics could be taken into account. These segments could be represented by students belonging to different age ranges or family income classes, or attending different courses of study.

By considering Mixed Logit models the heterogeneity among users could be investigated and differences in user perceptions and responses could be considered. In fact, the dimensions of quality, viewed from a customer's perspective, are complex, and perceptions about qualitative characteristics of service are very different among users; for these reasons, it could be very useful to consider this heterogeneity explicitly in the model formulation in order to obtain more realistic results.

References

Akan, P. (1995) Dimensions of service quality: A study in Istanbul, *Managing Service Quality*, 5(6), pp. 39–43.

Ben-Akiva, M. E. & Lerman, S. R. (1985) *Dicrete Choice Analysis: Theory and Application to Travel Demand* (Cambridge, MA: The MIT Press).

Ben-Akiva, M. & Morikawa, T. (1990) Estimation of travel demand models from multiple data sources, in: *11th International Symposium on Transportation and Traffic Theory*, Yokohama, Japan.

Berger, C., Blauth, R., Boger, D., Bolster, C., Burchill, G., DuMouchel, W., Pouliot, F., Richter, R., Rubinoff, A., Shen, D., Timko M. & Walden, D. (1993) Kano's methods for understanding customer-defined quality. *The Center for Quality Management Journal*, 2(4), pp. 3–36.

Bhave, A. (2002) Customer Satisfaction Measurement, *Quality & Productivity Journal*, February.

Bollen, K. A. (1989) *Structural Equations with Latent Variables* (New York: John Wiley & Sons Inc).

Bradley, M. & Daly, A. (1991) Estimation of logit choice models using mixed stated preference and revealed preference information, in: *6th International Conference on Travel Behaviour*, Quebec.

Bradley, M. & Daly, A. (1992) Uses of the logit scaling approach in stated preference analysis, in: *6th World Conference on Transportation Research*, Lyon.

Cascetta, E. (2001) *Transportation Systems Engineering: Theory and Methods* (Dordrecht: Kluwer Academic Publishers).

Cuomo, M. T. (2000) *La customer satisfaction. Vantaggio competitivo e creazione di valore* [Customer satisfaction: Competitive advantage and creation of value]. (Padua: CEDAM).

de Dios Ortùzar, J. (1992) Stated preference in travel demand modelling, in: *6th World Conference on Transportation Research*, Lyon.

de Dios Ortùzar, J. & Willumsen, L. G. (1994) *Modelling Transport* (London: John Wiley and Sons Ltd.)

Figini, M. (2003) *Dare valore alle esigenze dei clienti e dei dipendenti dell'azienda* [Giving value to customer and agency employee needs by means of customer satisfaction and groups of agency improvement]. (Milano: Franco Angeli).

Fujiwara, A. & Sugie, Y. (1992) Modification of stated preference mode choice models to improve the external validity, in: *6th World Conference on Transportation Research*, Lyon.

Grønholdt, L. & Martensen, A. (2005) Analysing customer satisfaction data: A comparison of regression and artificial neural networks, *International Journal of Market Research*, 47(2), pp. 121–130.

Hartikainen, M., Salonen, E. P. & Turunen, M. (2004) *Subjective Evaluation of Spoken Dialogue Systems Using SERVQUAL Method*. ICSLP, 2273–2276.

Hensher, D. A. (2001) Service quality as a package: What does it mean to heterogeneous consumers? in: *9th World Conference on Transport Research*, 22–27 July, Seoul.

Hensher, D. A. & Prioni, P. (2002) A service quality index for area-wide contract performance assessment regime, *Journal of Transport Economics and Policy*, 36(1), pp. 93–113.

Hensher, D. A., Stopper, P. & Bullock, P. (2003) Service quality-developing a service quality index in the provision of commercial bus contracts, *Transportation Research*, 37(A), pp. 499–517.

Hill, N. (2000) *Handbook of Customer Satisfaction and Loyalty Measurement* (Aldershot, UK: Gower Publishing Ltd).

Hill, N. (2003) *How to Measure Customer Satisfaction* (Aldershot, UK: Gower Publishing Ltd).

Kano, N., Seraku, N., Takahashi, F. & Tsjui, S. (1984) Attractive quality and must-be quality, *Hinshitsu*, 14(2), pp. 147–156.

Louviere, J., Hensher, D. A. & Swait, J. D. (2000) *Stated Choice Methods. Analysis and Application* (Cambridge, MA: Cambridge University Press).

Pearmain, D., Swanson, J., Bradley, M. & Kroes, E. (1991) *Stated Preference Techniques: A Guide to Practice* (2nd edn) (Rotterdam, Netherlands: Steer Davies Gleave and Hague Consulting Group).

Prioni, P. & Hensher, D. A. (2000) Measuring service quality in scheduled bus services, *Journal of Public Transportation*, 3(2), pp. 51–74.

Siskos, Y., Grigoroudis, E., Zopounidis, C. & Saurais, O. (1998) Measuring customer satisfaction using a collective preference disaggregation model, *Journal of Global Optimization*, 12, pp. 175–195.

Swait, J. & Louviere, J. (1993) The role of the scale parameter in the estimation and use of multinomial logit models, *Journal of Marketing Research*, 30, pp. 305–314.

Swanson, J., Ampt, L. & Jones, P. (1997) Measuring bus passenger preferences, *Traffic Engineering and Control*, 38(6), pp. 330–336.

Transportation Research Board (1999) *A Handbook for Measuring Customer Satisfaction and Service Quality*. Washington, DC: National Academy Press, Transit Cooperative Research Program, Report 47.

Vilares, M. J. & Coelho, P. S. (2003) The employee-customer satisfaction chain in the ECSI model, *European Journal of Marketing*, 37(11–12), pp. 1703–1722.

Zeithaml, V. A., Parasuraman, A. & Berry, L. L. (1986) *Servqual: A Multiple Item Scale for Measuring Perceptions of Service Quality* (Cambridge, MA: Marketing Science Institute).