Capturing the differences in perceiving service quality of metro passengers of Madrid

Laura Eboli[1], Carmen Forciniti[1*], Gabriella Mazzulla[1]

[1]*Department of Civil Engineering, University of Calabria, Italy*

Abstract

Transit service quality can be measured by using passengers' perceptions about the services, through the well-known Customer Satisfaction Surveys. Defining and measuring transit service quality is still quite complex because service quality depends on several different aspects characterizing the service. The aim of this work is to verify the suitability of a particular and not common kind of survey performed for assessing the service quality levels of the metro system of Madrid (Spain), and characterized by the subdivision of a list of service attributes among sub-samples of users, in order to save time and fatigue in compiling the questionnaire. For achieving our objective, we calibrated Ordered Probit models from which we obtained the coefficients associated to each service attribute representing the importance of each aspect on the overall satisfaction; then, we made some comparisons among the models that were useful to establish the suitability of the survey.

Keywords: Madrid Metro system; Customer Satisfaction Survey (CSS); Ordered regression model.

1. Introduction

The achievement of a sustainable mobility is the challenge of our times. In recent years, the number of trips made using the private car increased significantly. Especially in large urban areas, this fact has led to problems of traffic congestion, pollution, and high levels of CO2 emissions. To reduce these problems, transport interventions aimed at encouraging non-motorized modes of transport and public transport systems (Eboli et al., 2018; Allen et al., 2018; 2019). In this perspective, the evaluation of transit service quality is an aspect of considerable importance. In fact, by improving the quality of the offered service, an ever-increasing number of users could be attracted. Even if the evaluation of service quality has become a common practice among transport companies, defining and measuring transit service quality is still quite complex, because service quality depends on the different aspects characterizing the service (de Oña et al., 2016; 2019). On the basis of the suggestions provided by the EN13816 (CEN/TC 320, 2002), several aspects can be included for assessing transit service quality, as service scheduling and reliability, service coverage, information, comfort, cleanliness, and

* Corresponding author: Carmen Forciniti (carmen.forciniti@unical.it)

safety and security. A comprehensive measure of service quality is represented by the perceptions of the users about the service, which can be detected through the well-known Customer Satisfaction Surveys (CSSs). Generally, users' perceptions are collected in terms of rates expressed for the various aspects describing the service, according to different evaluation scales. In carrying out a CSS, however, various difficulties have to be taken into account. First of all, the willingness of users to be interviewed. A serious problem is related to the fatigue effect, that might occur if the interview is too long. Questions appearing in the latter part of the interview might be answered in a more perfunctory manner (Rossi et al., 1983). Another drawback is the monetary cost to be incurred to carry out the survey due to the payment for the interviewers, which increases with the time necessary to complete the questionnaire (Transit Cooperative Research Program, 2005). Due to the period of economic recession in recent years, the resources required for the surveys have been limited (de Oña et al., 2018). For these reasons, transport companies are experimenting with different survey methods in order to offer respondents a shorter and easier way to complete questionnaire but still achieve satisfactory results.

In the present work, data collected through a particular structure of a CSS are analysed. Specifically, the survey includes four versions of the questionnaire, each containing a reduced number of service quality factors to be evaluated by the interviewed users. In other words, a complete list of service quality factors of a metro service was divided into four groups of service factors; each group was included in a version of the questionnaire. Consequently, the whole sample to which the survey is addressed can be divided into four sub-samples that answer different questions about the quality of the metro service. The service regards the metro of Madrid (Spain). The aim of this work is investigating on the benefits that can be drawn from carrying out CSSs by submitting different questionnaires to a certain number of sub-samples. For achieving our objective, Ordered Probit (OP) models were calibrated because customer satisfaction data are expressed according to an ordinal categorical scale. Specifically, one model was calibrated for each of the four groups of users, and a comparison between the results obtained from each model was made, with the final aim to discover the main differences and verify if the particular structure of the survey is suitable for the CSSs.

The paper is organized as follows. The next section regards the case study, with a description of the questionnaire and a characterization of the sample. Section 3 presents a brief theoretical framework about OP regression, the proposed models, and the model results with a discussion. Finally, in Section 4 some final remarks are presented.

2. Case study

The case study is represented by the metro system of Madrid (Spain), which has one of the most extended metropolitan networks in the world. The metro network reaches twelve municipalities in the hinterland of Madrid and about 80% of the inhabitants can reach a station at less than 600 meters from own home or own usual destinations (Calvo et al., 2019). In 2018, the Metro accommodated 657 million passengers. Compared with data recorded in 2017, the number of passengers increased by about 5%. Currently, on average more than 2 million of passengers daily choose to travel by Metro (Metro de Madrid, 2018). The data used for calibrating the proposed models were collected through the fourth campaign of surveys made by the company operating the metro service in 2015. The surveys aim to identify how the users evaluate the several attributes of transit service quality and the level of importance that users give to each element. The questionnaire contains four parts. The first one collects the socio-demographic characteristics of the interviewees and the main aspects of their trip. Part 2 regards the user's perception about the service quality of the metro system. The interviewee is called to declare the evaluation score about the overall service quality of the metro system, the evaluation score about the overall service quality referred to the line that the interviewee is using while answering the questionnaire, the evaluation score and the levels of importance that associate to specific aspects of the service. Part 3 regards the inconveniences caused by any possible service interruptions. In the last part, the passenger answers to questions regarding the intent to recommend the use of metro to other people. As mentioned in the introduction, the survey has a particular structure, because it includes four versions of the questionnaire. On the whole, the service quality factors investigated in part 2 cover all the aspects of the metro service, as operation, comfort, cleanliness, attention to client. However, in each version of the questionnaire, a small number (only six or seven) of factors were proposed to the interviewees. Consequently, the whole sample is divided into four sub-samples that evaluate different service quality factors.

Table 1: Samples characteristics.

		Sub-sample 1	*Sub-sample 2*	*Sub-sample 3*	*Sub-sample 4*
Gender	Males	40.3%	40.7 %	40.4 %	40.0 %
	Females	59.7%	59.3 %	59.6 %	60.0 %
Age	16-24	24.6 %	25.5 %	26.2 %	24.2 %
	25-34	24.9 %	25.3 %	24.3 %	25.8 %
	35-44	21.8 %	21.3 %	22.8 %	22.6 %
	45-54	16.7 %	16.7 %	16.3 %	17.0 %
	55-59	5.8 %	5.3 %	4.7 %	4.6 %
	>59	6.2 %	5.9 %	5.8 %	5.8 %
Nationality	National resident	84.6 %	83.9 %	83.7 %	83.3 %
	National tourist	1.9 %	2.1 %	1.7 %	1.6 %
	Foreigner resident	12.8 %	13.3 %	13.8 %	14.6 %
	Foreigner tourist	0.7 %	0.7 %	0.8 %	0.6 %
Ticket type	One-way ticket	6.8 %	6.3 %	5.4 %	6.1 %
	Carnet	21.9 %	24.0 %	20.6 %	22.9 %
	Travel pass	71.2 %	69.5 %	73.9 %	70.9 %
	Tourist ticket	0.1 %	0.2 %	0.1 %	0.1 %
Trip purpose	Work	58.3%	56.8 %	57.1 %	58.6 %
	Study	21.2 %	21.5 %	20.4 %	19.9 %
	Leisure	9.1 %	9.9 %	10.0 %	10.1 %
	Medical visit	4.2 %	4.0 %	3.6 %	3.8 %
	Shopping	1.3 %	1.6 %	1.4 %	1.5 %
	Other	5.8%	6.4 %	7.6 %	6.1 %
Weekly trips	<4	12.3%	11.6 %	10.8 %	11.4 %
	4-10	37.1 %	37.8 %	37.5 %	37.0 %
	11-20	34.6 %	34.7 %	34.2 %	34.5 %
	>20	16.0 %	15.9 %	17.5 %	17.0 %
Sample size		2,569	2,550	2,527	2,513

The survey was addressed to the passengers of the subway lines at stations using face-to-face interviews. The sample consists of 10,159 interviews, and each sub-sample counts about 2,500 interviews. The chi-squared test was performed on the data for understanding if there are statistical significant differences among the four sub-samples. In all analysed cases, the chi-squared test does not exceed the critical value with a 95% significance level (p-value < 0.05); consequently, the four sub-samples do not present statistical significant differences and can be considered as equal to each other. By observing table 1, we can see that the four sub-samples show more or less the same characteristics.

Table 2: Survey results per sample.

	Quality attributes	*Quality criteria*	*Av. evaluation score*
Sub-sample 1	Overall metro service		7.34
	Operation of automatic ticket vending machines	Accessibility	7.60
	Operation of elevators	Accessibility	6.94
	Temperature and ventilation system at stations	Comfort	6.87
	Noise level on the vehicles	Comfort	6.58
	Personal security from assault, theft, etc.	Security	7.30
	Attention to complaints and suggestions	Customer care	6.58
	Platform waiting time	Time	6.40
Sub-sample 2	Overall metro service		7.30
	Operation of turnstiles	Accessibility	7.62
	Information about incidents of the service	Information	7.17
	Kindness of the vigilantes	Customer care	7.09
	Cleanliness at stations	Comfort	7.40
	Cleanliness of vehicles	Comfort	7.44
	Speed of train travel	Time	7.44
Sub-sample 3	Overall metro service		7.30
	Escalator operation	Accessibility	7.20
	Abnormal stops of trains	Comfort	6.85
	Temperature and ventilation system on board	Comfort	6.98
	Maintenance of stations	Comfort	7.69
	Maintenance of vehicles	Comfort	7.59
	Signage in stations	Information	8.16
Sub-sample 4	Overall metro service		7.27
	Lighting of trains	Security	8.22
	Lighting at stations	Security	8.26
	Available space on board	Comfort	7.26
	Safety against accidents	Security	7.88
	Attention and kindness of employees	Customer care	7.36
	Accessibility for disabled	Accessibility	7.53

Regarding the gender, the interviewed passengers are mainly females (about 60.0 %). Concerning age, the largest groups are 16-24 years for sub-samples 2 and 3, and 25-34 for sub-samples 1 and 4. The major part of respondents claim to be national resident whereas the tourists interviewed are few. The ticket type that the majority of interviewees declared to use is the travel pass (about 70.0%). The most common purpose of the trip is work (about 60.0%), followed by study (about 20.0%) and leisure (about 10.0%). Most of the interviewed passengers declared to make about 4-10 trips for week (about 37.0%).

The interviewees were asked to evaluate the overall metro service and the different factors of service quality on an 11-point scale, from 0 (the lowest level) to 10 (the highest level). The average evaluation scores referred to the overall metro service and calculated for each sub-sample are on fairly close values. In particular, the average evaluation score is equal to 7.34 for the sub-sample 1, 7.30 for both the sub-samples 2 and 3, and 7.27 for the sub-sample 4. Table 2 reports the users' average evaluation rates and the service quality attributes respectively evaluated by interviewees belonging to each sub-sample. The attributes can be grouped in according to the criteria reported in the EN13816 (CEN/TC 320, 2002). Specifically, the document refers to eight quality criteria: availability, accessibility, information, time, customer care, comfort, security, and environmental impact. After a deep studying of the European document and a careful observation of the service quality attributes investigating in the survey, we proposed the association reported in table 2. As shown in the table, not all the eight quality criteria reported in the EN13816 (CEN/TC 320, 2002) were considered for selecting the attributes of the survey. Specifically, the twenty-five service quality attributes belong to six quality criteria: accessibility, information, time, customer care, comfort, and security. In particular, the most investigated quality criterion is "comfort", represented by nine service quality attributes, followed by "accessibility", expressed by five attributes. On the contrary, "information" and "time" are represented by only two attributes. In addition, we can also observe that not all the quality criteria are present in each sub-sample. As an example, the first and the second groups are the most variegated in terms of quality criteria, including service attributes belonging to five criteria, while the third group includes for the major part attributes describing comfort (four out of six), and the fourth group is mostly oriented to "security" (three attributes out of six). Anyway, the respondents seem enough satisfied in average for all the investigated service quality factors; in fact, the rates vary from 6.40 for "Platform waiting time" to 8.26 for "Lighting at stations".

3. Data analysis and findings

We proposed OP models for analysing the service quality factors and discovering differences among the models. OP models are an extension of the logistic regression applied when the dependent variable Y is categorical and has a meaningful order with more than two categories (or levels). Unlike OLS regression, OP discerns unequal differences between ordinal categories in the dependent variable. For this reason, they have a good applicability for analysing public transport service quality. In the OP model (McKelvey and Zavoina, 1975), the observed ordinal variable Y is a function of another variable Y^* that is a continuous unmeasured latent variable, whose values determine various threshold points (Borooah, 2001). The value Y_i of the observed variable depends on whether or not it crossed a particular threshold. Particularly, in all the proposed models, the dependent variable is the evaluation score about the overall metro service. The independent variables, instead, are different among the four models and correspond to the evaluation scores of the different service quality factors asked to each sub-sample of users. The variables included in the models are measured on the 11-point scale adopted in the questionnaire, as specified above.

Table 3 reports the goodness-of-fit calculated for the four proposed models. We can see that the models are statistical significant, but the Mc Fadden Pseudo R-squared statistics are rather low if compared with reference values (Hair et al., 2010). We have really expected such a result, because each model cannot include all the service quality aspects (as independent variables) that influence the overall service quality (dependent variable), due to the fact that each sub-sample evaluated only a group of service quality factors. Consequently, the variables included in each model do not explain all the variation in the dependent variable, and this leads to a low value of R^2.

Table 3: Goodness-of-fit of OP models.

Goodness-of-fit	*Model 1*	*Model 2*	*Model 3*	*Model 4*
Log likelihood function	-4113.26642	-4146.11124	-4192.75132	-4196.09544
Restricted log likelihood	-4601.29999	-4631.14293	-4512.67168	-4522.59144
Chi squared [dof=6] (P= .000)	976.06714	970.06339	639.84073	652.99200
Significance level	0.00000	0.00000	0.00000	0.00000
McFadden Pseudo R-squared	0.1060643	0.1047326	0.0708938	0.0721922
AIC	8258.5	8322.2	8415.5	8422.2
No. of observations	2,569	2,550	2,527	2,513

By observing tables from 4 to 7, we can see that the sign of all the independent variables is positive. This means that the increase of the evaluation score of a service quality factor produces the increase of the evaluation score of the overall service quality.

Table 4: Results of OP model 1.

Variable	*Coefficient*		*S.E.*	*z*	*Prob. |z|>Z**
Operation of automatic ticket vending machines	0.09783	***	0.01165	8.40	0.0000
Operation of elevators	0.01040		0.01045	0.99	0.3198
Temperature and ventilation system at stations	0.06428	***	0.01287	4.99	0.0000
Noise level on the vehicles	0.03707	***	0.01249	2.97	0.0030
Personal security from assault, theft, etc.	0.04525	***	0.01223	3.70	0.0002
Attention to complaints and suggestions	0.00927		0.01230	0.75	0.4509
Platform waiting time	0.19115	***	0.01084	17.64	0.0000
μ1 (threshold)	0.06572	*	0.03557	1.85	0.0647
μ2 (threshold)	0.34226	***	0.05641	6.07	0.0000
μ3 (threshold)	0.69827	***	0.05347	13.06	0.0000
μ4 (threshold)	1.06449	***	0.04566	23.31	0.0000
μ5 (threshold)	1.58583	***	0.03640	43.56	0.0000
μ6 (threshold)	2.21633	***	0.02974	74.52	0.0000
μ7 (threshold)	3.11913	***	0.02748	113.53	0.0000
μ8 (threshold)	4.19610	***	0.03114	134.75	0.0000
μ9 (threshold)	4.82963	***	0.04044	119.42	0.0000

*** Significance at 1% level, ** Significance at 5% level, *Significance at 10% level.

The results of the model 1 (Table 4) shows that the variable having the greatest weight on the overall service quality is “Platform waiting time”, while the variables “Operation of elevators” and “Attention to complaints and suggestions” are not significant. In model 2 (Table 5), the variable that mainly influences overall satisfaction is "Speed of train travel"; whereas "Operation of turnstiles" and "Cleanliness at stations" do not appear as significant. On the other hand, all the variables included in model 3 (Table 6) are significant. The variables relating to comfort aspects, "Maintenance of stations" and "Maintenance of vehicles", most influence overall service quality.

Table 5: Results of OP model 2.

Variable	*Coefficient*		*S.E.*	*z*	*Prob.* $|z|>Z^*$
Operation of turnstiles	0.01588		0.01185	1.34	0.1804
Information about incidents of the service	0.06032	***	0.01306	4.62	0.0000
Kindness of the vigilantes	0.003894	***	0.01174	3.32	0.0009
Cleanliness at stations	0.02673		0.02345	1.14	0.2544
Cleanliness of vehicles	0.08311	***	0.02406	3.45	0.0006
Speed of train travel	0.22559	***	0.01324	17.04	0.0000
μ1 (threshold)	0.17615	***	0.06536	2.70	0.0070
μ2 (threshold)	0.35353	***	0.07163	4.94	0.0000
μ3 (threshold)	0.83172	***	0.06024	13.81	0.0000
μ4 (threshold)	1.33873	***	0.04578	29.24	0.0000
μ5 (threshold)	1.87338	***	0.03588	52.22	0.0000
μ6 (threshold)	2.50933	***	0.02928	85.69	0.0000
μ7 (threshold)	3.34612	***	0.02728	122.66	0.0000
μ8 (threshold)	4.38875	***	0.03101	141.53	0.0000
μ9 (threshold)	5.07295	***	0.04113	123.33	0.0000

*** Significance at 1% level, ** Significance at 5% level, *Significance at 10% level.

Table 6: Results of OP model 3.

Variable	*Coefficient*		*S.E.*	*z*	*Prob.* $|z|>Z^*$
Escalator operation	0.05023	***	0.01155	4.35	0.0000
Abnormal stops of trains	0.03397	***	0.01068	3.18	0.0015
Temperature and ventilation system on board	0.05071	***	0.01334	3.80	0.0001
Maintenance of stations	0.11534	***	0.01950	5.91	0.0000
Maintenance of vehicles	0.10147	***	0.01958	5.18	0.0000
Signage in stations	0.03717	**	0.01527	2.43	0.0149
μ1 (threshold)	0.12138	**	0.04751	2.55	0.0106
μ2 (threshold)	0.38162	***	0.05707	6.69	0.0000
μ3 (threshold)	0.58082	***	0.05412	10.73	0.0000
μ4 (threshold)	0.98320	***	0.04472	21.99	0.0000
μ5 (threshold)	1.53678	***	0.03445	44.60	0.0000
μ6 (threshold)	2.08298	***	0.02880	72.32	0.0000
μ7 (threshold)	2.93123	***	0.02690	108.99	0.0000
μ8 (threshold)	3.95565	***	0.03095	127.79	0.0000
μ9 (threshold)	4.58232	***	0.04080	112.31	0.0000

*** Significance at 1% level, ** Significance at 5% level, *Significance at 10% level.

The results of model 4 (Table 7) highlight a different frame because all the coefficients are rather low compared to the coefficients obtained from the other models.

Anyway, the variable "Available space on board" has the highest weight on the overall satisfaction.

Table 7: Results of OP model 4.

Variable	*Coefficient*		*S.E.*	*z*	*Prob.* $\|z\|>Z^*$
Lighting of trains	0.05119	**	0.02300	2.23	0.0260
Lighting at stations	0.04007	*	0.02288	1.75	0.0799
Available space on board	0.09916	***	0.01282	7.73	0.0000
Safety against accidents	0.05907	***	0.01545	3.82	0.0001
Attention and kindness of employees	0.06894	***	0.01239	5.56	0.0000
Accessibility for disabled	0.08202	***	0.01239	6.62	0.0000
μ1 (threshold)	0.17006	***	0.06257	2.72	0.0066
μ2 (threshold)	0.50084	***	0.06473	7.74	0.0000
μ3 (threshold)	0.80617	***	0.05540	14.55	0.0000
μ4 (threshold)	1.20903	***	0.04393	27.52	0.0000
μ5 (threshold)	1.73212	***	0.03417	50.70	0.0000
μ6 (threshold)	2.24333	***	0.02894	77.53	0.0000
μ7 (threshold)	3.10125	***	0.02692	115.20	0.0000
μ8 (threshold)	4.10450	***	0.03095	132.63	0.0000
μ9 (threshold)	4.79569	***	0.04244	113.01	0.0000

*** Significance at 1% level, ** Significance at 5% level, *Significance at 10% level.

By referring to the quality criteria, we can say that according to the results of models 1 and 2, "time" is the most important aspect for the users. On the other hand, if we consider models 3 and 4, "comfort" appears the aspects that impacts mainly on the overall service quality. We have to make a consideration regarding this last result, that is comfort could have been resulted the most important aspect because in these two models "time" is not represented by none attribute. However, the results of the four models show that comfort seems to be a relevant aspect for the users. "Accessibility" presents a fair weight in some cases; as an example, when it is represented by the "operation of automatic ticket vending machines" (model 1), it is the second attribute in terms of weight value, as well as in the fourth model when it is represented by "accessibility for disabled". The other quality criteria, such as "information", "customer care", and "security" resulted the least important in general.

4. Conclusions

The aim of this work was to verify if a particular structure of a CSS is suitable for collecting data about the perceptions of service quality from public transport passengers. Specifically, the analysed survey was characterized by proposing to each group of users a different questionnaire, characterized by including not all the investigated service attributes but only a part of them. In particular, four sub-samples of users of a metro service were selected, and to each of them a restricted number of a total list of twenty-five service quality attributes was proposed in order to obtain an evaluation in terms of rating (six or seven attributes per sub-sample). The desire to investigate on the suitability of such kind of survey arises from its possible convenience. In fact, by asking to users an evaluation of a restricted number of service attributes can guarantee to complete the questionnaire in brief time, avoiding the fatigue-effect to the user and allowing to save money for realizing the survey. For achieving our objectives and find the more opportune evidences, we firstly verify the representativeness of the sub-samples, and from the results we concluded that the sub-samples have similar characteristics, and for this reason the results of the survey could be retained as reliable. Afterwards, we have analysed in detail the whole list of the twenty-five attributes and their subdivision into four groups. By considering the EN13816 (CEN/TC 320, 2002) for associating each service attribute to the main quality criteria characterizing a public transport service, we verified that the subdivision of the service attributes was made without any criterion or logic, and this is the main problem registered in the analysis of the survey. We observed groups where few quality criteria were included, and consequently which included too service attributes of the same criterion. The results obtained from the OP models confirmed these disadvantages. As an example, we found that in the groups including the attributes linked to time, this aspect results as the most important, while in the other groups "comfort" resulted as the most important aspect but probably for the absence of "time". It was very difficult to compare the coefficient values, just for the absence of some aspects in certain groups. Definitively, the survey design and particularly the subdivision of the service attributes in different groups should be made following a specific criterion. One criterion could be assuring in each groups the presence of at least one service attribute belonging to a certain quality criterion. This solution could permit to establish the importance assigned by the users to the main quality criteria. Another criterion could be proposing to the same users all the service attributes included in a quality criterion. This solution could permit to establish the importance assigned by the users within a specific quality criterion, such as comfort, or time, and so on.

References

Allen J., Eboli L., Forciniti C., Mazzulla G., Ortúzar JdD. (2019) "The Role of Critical Incidents and Involvement in Transit Satisfaction and Loyalty", *Transport Policy* 75, pp. 57-69.

Allen J., Eboli L., Mazzulla G., Ortúzar JdD. (2018) "Effect of Critical Incidents on Public Transport Satisfaction and Loyalty: An Ordinal Probit SEM-MIMIC approach", *Transportation* (https://doi.org/10.1007/s11116-018-9921-4)

Borooah, V.K. (2001) "Logit and Probit: Ordered and Multinomial Models", Sage University Papers Series on Quantitative Applications in Social Sciences, *Serie no. 07-138*, Thousand Oaks, CA, Sage.

Calvo F., Eboli L., Forciniti C., Mazzulla G. (2019), "Factors impacting trip generation on Metro system in Madrid (Spain)", *Transportation Research Part D, Transport and Environment* 67, pp.156-172.

CEN/TC 320 (2002) *Transportation – Logistics and services. European Standard EN 13816: Public passenger transport – Service quality definition, targeting and measurement*, European Committee for Standardization, Brussels, Belgium.

de Oña J., de Oña R., Eboli L., Forciniti C., Mazzulla G. (2016) "Transit passengers' behavioural intentions: the influence of service quality and customer satisfaction", *Transportmetrica A: Transport Science* 12(5), pp. 385-412.

de Oña, R., Eboli, L., Forciniti, C., Mazzulla, G. (2019) "Analisi delle tecniche utilizzate dalle aziende di trasporto pubblico italiane per valutare la qualità percepita dei servizi offerti /Analysis of the techniques used by public transport operators for evaluating perceived service quality", *Ingegneria Ferroviaria* 74(5), pp. 401-415.

de Oña, R., Estevez, E., Garach, L., Calvo F.J., de Oña, J. (2018) "Transit quality evaluation: processes conducted by managers and operators", *Transportation Research Procedia* 33, pp. 28-34.

Eboli, L., Forciniti, C., Mazzulla, G. (2018) "Spatial variation of the perceived transit service quality at rail stations", *Transportation Research Part A: Policy and Practice* 114, pp. 67-83.

Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E. (2010) *Multivariate Data Analysis: A Global Perspective,* Prentice Hall, New Jersey.

McKelvey, R.D., Zavoina W. (1975) "A statistical model for the analysis of ordinal level dependent variables", *Journal of Mathematical Sociology* 4, pp.103-120.

Metro de Madrid (2018) "Evolucion de la demanda. Cierre año 2018. Servicio Planificacion y Estudios de Operacion - Area Control y Planificacion de la Operacion", *www. metrodemadrid.es*.

Rossi, P.H., Wright, J.D., Anderson, A.B. (1983) *Handbook of Survey Research*, Academic Press.

Transit Cooperative Research Program, Synthesis 63 (2005) *On Board and Intercept Transit Survey Techniques – A Synthesis of Transit Practice*, Transportation Research Board, Washington, DC.