The role of involvement with public transport in the relationship between service quality, satisfaction and behavioral intentions

By: Juan de Oña

This document is a **post-print version** (ie final draft post-refereeing) of the following paper:

Juan de Oña (2020) *The role of involvement with public transport in the relationship between service quality, satisfaction and behavioral intentions*. **Transportation Research Part A: Policy and Practice, 142, 296-318,** DOI: 10.1016/j.tra.2020.11.006

Direct access to the published version: http://dx.doi.org/10.1016/j.tra.2020.11.006

The role of involvement with public transport in the relationship between service quality, satisfaction and behavioral intentions

Juan de Oña

TRYSE research group, University of Granada, ETSI Caminos, Canales y Puertos, Campus de Fuentenueva, s/n, 18071 Granada (Spain)
email: jdona@ugr.es

Abstract:

Several studies have made manifest that involvement with public transport play a key role in the intentions of its use. However, conflicting models exist in the literature about involvement's role in the relationship between service quality, satisfaction and behavioral intentions or loyalty. Previous studies suggest all possible roles: antecedent, mediator, moderator and direct effects. A structural equation modeling approach is applied here to further understand the role of involvement with public transport, comparing eight alternative models and using data from a single survey carried out in five European cities (Madrid, Rome, Berlin, Lisbon and London). Later, the study uses a multiple indicators and multiple causes structural equation modeling approach (SEM-MIMIC) to analyze the effect of heterogeneity present in the data over the four constructs considered (service quality, satisfaction, involvement and behavioral intentions). This comprehensive methodological approach provides a number of noteworthy findings, including the empirical verification that satisfaction is a full mediator between service quality and involvement, and involvement is a full mediator between satisfaction and behavioral intentions. The results further suggest that involvement is the factor that contributes most to behavioral intentions or loyalty, followed by service quality perceptions and satisfaction. Lastly, this study demonstrates the relevance of controlling for heterogeneity in users' perceptions, so as to obtain more robust relations among factors and identify significant differences among market segments, which could prove useful for public transport operators or policy makers.

Keywords: attitudes towards transport, loyalty, cross-country, mediator, antecedent, transit, SEM-MIMIC, customer satisfaction

1.- Introduction

There is full agreement when it comes to considering public transport as a key element for achieving sustainable development in urban and metropolitan areas. In major cities, where distances call for the use of motorized vehicles and where there are huge volumes of daily displacements, public transport is fundamental for mobility that respects the environment. In the coming years, if current trends are maintained, it is foreseen that the percentage of the population living in urban settings will grow (UN 2019). For this reason, strategies fomenting the use of public transport to attain sustainable cities and communities constitute one of the 17 Sustainable Development Goals.

Increasing the use of public transport calls for current users to perceive it as a quality service and view it with satisfaction. If current users become loyal, they will help attract new users through their recommendations and attitudes towards this mode of transport. Therefore, there is growing concern as to better understanding the relationships existing between service quality perception, satisfaction and loyalty or behavioral intentions surrounding public transport.

Recent years have seen numerous studies approaching these aspects of public transport in the literature, as two recent review articles: one on service quality and satisfaction (de Ona and de Ona 2015) and another on satisfaction and loyalty (van Lierop, Badami and El-Geneidy 2018). The concepts of service quality, satisfaction and loyalty, as well as their relations, have been analyzed for decades in the field of marketing (Miller 1976, Oliver 1981, Parasuraman, Zeithaml and Berry 1985, Zeithaml, Berry and Parasuraman 1996), and since the beginning of this century, studies of service quality – customer satisfaction – loyalty or the behavioral intentions paradigm are frequent in the field of public transport (Park, Robertson and Wu 2004, Chen 2008, Lai and Chen 2011). Most of these studies assume that satisfaction exerts a mediator role between service quality and loyalty or behavioral intentions. However, the public transport field is still divided, roughly 50%/50%, as to whether this is a full mediator effect or a partial mediator effect. In the field of marketing, where specific comparative studies have been carried out (Cronin and Taylor 1992), both roles have been identified, and likewise no consensus exists.

Numerous studies use structural equation modeling (SEM) to explore the effect of other factors in this framework: perceived value, image, experience of critical incidents, and switching costs (Minser and Webb 2010, Jen, Tu and Lu 2011, Zhao, Webb and Shah 2014, de Ona et al. 2016, Allen et al. 2019a). Despite the fact that this concept has been a focus in marketing and behavioral research (Olsen 2007), the role of involvement in terms of loyalty or behavioral intentions regarding public transport has been less investigated. Furthermore, there is a greater degree of consensus in the field of marketing and behavioral research (Olsen 2007, Menidjel et al. 2019) in that involvement exerts a mediator role between satisfaction and loyalty or behavioral intentions; in the field of public transport, all possible roles have been identified: antecedent (Machado-Leon et al. 2018), mediator (Lai and Chen 2011, Irtema et al. 2018), moderator (Wei and Kao 2010, Machado-Leon, de Ona and de Ona 2016) and direct effects (Allen et al. 2019a).

Most of these studies moreover present two limitations that impede the extrapolation of their findings: (i) a single data sampling is used, so that results cannot be replicated; and (ii) no alternative or equivalent models that might account for the same pattern of observed covariances are considered (i.e., which explain the data just as well as the researcher's preferred model but make differing causal claims). These two limitations generate a high probability of confirmation bias, give an overly positive evaluation of the author's preferred model, and overlook other explanations for the data (Shah and Goldstein 2006, Kline 2015). As far as the author knows, within the public transport field, only Machado-Leon et al. (2016) have compared diverse alternative models to try and pin down the role of involvement in the service quality – satisfaction – behavioral intention framework. Their results were not conclusive, however, because they used only one sample, and their models' fit statistics were not very good.

This paper aims to shed further light on the role of involvement in public transport, comparing several alternative models and using data from one survey that was translated for use with the public transport users of the capitals of Germany, Spain, Italy, Portugal and the United Kingdom. Once the model that best fit the survey data was identified, a multiple indicators and multiple causes structural equation modeling approach (SEM-MIMIC) was applied to account for heterogeneity in all four latent factors used in the analysis (service quality, satisfaction, involvement and behavioral intentions). In this way, in addition to determining the relationship between these factors, the model controlled the effects of any possible heterogeneity that could affect each of the constructs differently. The potential sources of heterogeneity considered in this paper are tied to location of the user (country and household), the user´s sociodemographic characteristics (gender, age, educational level, dependent members in the family, or household income level) and travel patterns (frequency of public transport use).

The rest of the paper is organized as follows. Section 2 presents a literature review of three topics: (i) service quality, customer satisfaction and loyalty paradigm; (ii) role of involvement towards public transport; and (iii) different approaches for dealing with heterogeneity. Section 3 describes the survey used, including the main survey results. Section 4 focuses on the competing research models and methodology. Section 5 presents the main results of the analysis distributed in several subsections: (i) data preparation and screening; (ii) confirmatory factor analysis measurement model; (iii) comparison and selection of structural regression models; and (iv) SEM-MIMIC's results. The paper finishes with discussion of the results in Section 6 and a summary of the most important conclusions in the final section.

2.- Literature review

2.1.- Service quality, customer satisfaction and loyalty paradigm

A number of recent studies have looked into the effects of service quality and customer satisfaction upon behavioral intentions or loyalty within the public transport field. In some studies the concepts of service quality and satisfaction are blended, although there is wide agreement in the literature that the two constructs comprise different factors (de Oña et al. 2018). Generally, service quality is associated with specific service attributes (e.g., frequency, cleanliness, comfort, speed), while satisfaction is associated with more elaborated perceptions and affective judgements (liking, feeling, pleasure, etc.) (Oliver 2010).

Something similar is seen in the case of behavioral intentions and loyalty, though this lack of consensus is not exclusive to the transport field. In marketing research, where these latent factors are studied to a greater extent, some authors claim that behavioral intentions are a sub-construct of loyalty (Oliver 2010), whereas others hold that loyalty is a sub-construct of behavioral intentions (Zeithaml et al. 1996). At any rate, the two approaches share a point in common: both consider attitudinal and behavioral measures to define and assess loyalty or behavioral intentions. Because of this, the vast majority of studies in the field of transport —regardless of the construct used— employ indicators associated with the reusage intention and willingness to recommend the service. The present contribution, while acknowledging a distinction between the two constructs, will use the terms loyalty and behavioral intentions indistinctly, as do some previous studies (Allen et al. 2019a)

Just as there is no broad agreement about the definition of certain constructs, there is a lack of consensus regarding the relations existing among them. Service quality is generally regarded as an antecedent of satisfaction, but there is disagreement about the type of mediator effect of satisfaction between service quality and behavioral intentions. Specifically in the field of urban and metropolitan public transport, there is a prevalence of studies defending the partial mediator role of satisfaction (Minser and Webb 2010, Lai and Chen 2011, Zhao et al. 2014, de Ona et al. 2016, Fu and Juan 2017, Irtema et al. 2018, Machado-Leon et al. 2018, Fu, Zhang and Chan 2018, Li et al. 2018, Nguyen-Phuoc et al. 2020), yet a good number of studies suggest the full mediator role of satisfaction between service quality and behavioral intentions (Zhang et al. 2019, Yuan et al. 2019, Sun and Duan 2019, Allen et al. 2020). Some studies identify both effects, depending on the type of user (de Ona, Machado and de Ona 2015) or the data used (Allen et al. 2019a).

2.2.- Role of involvement with public transport

According to Olsen (2007), involvement is related to an individual's subjective sense of the concern, care, importance, personal relevance, and significance attached to an attitude. That is, involvement is an unobservable state of motivation existing in both product and service consumers. Zaichkowsky (1985) and Mittal (1995) see involvement as the perceived importance of a specific product or service based on customer requirements, values and interests. Flynn and Goldsmith (1993) suggested that highly involved customers are inclined to display more loyal buying behavior. Indeed, related research in the marketing field established that the level of involvement may influence the relationship between service quality, satisfaction and loyalty (Gordon, McKeage and Fox 1998, Kinard and Capella 2006), and that it works as an important determinant of consumer evaluations and behaviors (Chen and Tsai 2008). Olsen (2007) applied a comprehensive methodological approach to identify the role of involvement by comparing four different involvement-loyalty models and using a representative survey of Norwegian pupils in middle and high school. Using five constructs (satisfaction, social norms, perceived control, involvement and repurchase loyalty), he tested involvement as full mediator, partial mediator, direct effect and moderator, and concluded that it was a complete mediator between satisfaction and loyalty. In the public transport sector, Lai and Chen (2011) developed a scale to measure involvement, which was defined as the level of interest or importance of public transit to a passenger. For a further definition of involvement the reader is advised to consult Machado-Leon et al. (2016).

In the public transport field, a limited number of studies have analyzed the influence of involvement with public transport upon service quality, satisfaction and loyalty (Wei and Kao 2010, Lai and Chen 2011, Machado-Leon et al. 2016, Machado-Leon et al. 2018, Irtema et al. 2018, Allen et al. 2019a). Some studies consider other factors that are not exactly equivalents of involvement, but have elements in common. In transit research, image (or corporate image) is based on how an individual views the contribution of public transport to one´s own wellbeing, and to society at large (van Lierop and El-Geneidy 2018), and is one of the most used latent factors (Park et al. 2004, Minser and Webb 2010, Chou and Kim 2009, Chou and Yeh 2013, Kuo and Tang 2013, Chang and Yeh 2017, Fu et al. 2018). Further options as factors are attitudes (Borhan et al. 2014, Simsekoglu, Nordfjaern and Rundmo 2015) or feelings toward public transport (de Ona et al. 2016). Researchers and practitioners working in this area are divided as to whether involvement, image, attitude, or a different term best describes how a passenger is engaged with the public transport service (van Lierop and El-Geneidy 2018); still, most of the indicators used for involvement, image or attitudes are similar, having some conceptual overlap.

Most previous studies focusing on the role of involvement, image or attitudes through SEM indicate a partial mediator effect (Park et al. 2004, Chou and Kim 2009, Lai and Chen 2011, Kuo and Tang 2013, Chou and Yeh 2013, Borhan et al. 2014, Chang and Yeh 2017, Irtema et al. 2018, Fu et al. 2018). In second place are those studies suggesting that involvement, image, attitudes or feelings are antecedents of service quality, satisfaction or behavioral intentions (Minser and Webb 2010, Simsekoglu et al. 2015, de Ona et al. 2016, Machado-Leon et al. 2018). Finally, a reduced number of studies published point to a moderator role or a direct effect role for involvement. In their study of public transport, Allen et al. (2019a) found that loyalty presented a positive effect on involvement in the context of a metro system in Spain. However, they underlined the need for further research considering other constructs and more indicators to measure loyalty and involvement (they used only one indicator for each). Wei and Kao (2010) and Machado-Leon et al. (2016) analyzed involvement by comparing models for two different groups: users with low and with high involvement towards public transport. Wei and Kao (2010) identified significant differences between the behavioral intentions of the two groups. Using data from a light rail transit system in Spain, Machado-Leon et al. (2016) compared five different models for identifying the role of involvement with regard to service quality, customer satisfaction and behavioral intentions. They tested involvement as a mediator (full mediator, partial mediator or direct effect), as an antecedent, and as a moderator (considering low- and highly involved users). Although they support a moderating role for involvement, their results were not conclusive —none of the models could be validated, since all of them presented one or more non-significant relationships among the considered constructs.

In light of the above studies, it can be clearly stated that no consensus governs the role of involvement with respect to service quality, customer satisfaction and behavioral intentions or loyalty. It also emerges that, despite agreement about the mediator role played by satisfaction between service quality and behavioral intentions, it remains unclear whether the mediator role is full or partial. All these considerations should be taken into account when suggesting alternative competing models in Section 4.

2.3.- Dealing with heterogeneity

When analyzing service quality, satisfaction, involvement and behavioral intentions or loyalty, the researcher must be aware that he/she is dealing with highly heterogeneous subjective data. Their heterogeneity may be due to: data from different territorial contexts, the influence of the place of residence (urbanization degree), specific socio-demographic characteristics (e.g., gender, age, level of education, income level), or patterns of mobility (e.g., frequency of travel, travel reason).

In order to obtain robust results, researchers should bear in mind these sources of heterogeneity and be able to control them in view of their needs. There are several means of dealing with this issue, but the ones most frequently used in the public transport field in conjunction with SEM are segmentation techniques. They allow to compare the results of the models among diverse population sectors, thereby detecting relevant differences. The most common variables used to segment the sample in the field at hand are: gender (Fu et al. 2018, de Ona, Estevez and de Ona 2020), age (de Ona et al. 2020, Allen et al. 2019a), frequency of travel (Chou, Lu and Chang 2014, de Ona et al. 2020, Allen et al. 2019a), income (Chou et al. 2014, de Ona et al. 2020), territorial context (Chou and Kim 2009, Allen, Munoz and Ortuzar 2019b), and type of service (Rajaguru 2016, An and Noh 2009, Koklic, Kukar-Kinney and Vegelj 2017). Cluster analysis is a more complex means of segmenting a sample, and has also been used in the public transport field (Machado-Leon et al. 2018, Sun and Duan 2019); it is based on heuristics that try to maximize the similarity between elements within a group (cluster) and derive the maximum difference between elements of different groups. Notwithstanding, most such studies, when comparing two segments (e.g., males vs. females) do not control for the rest of the possible causes of heterogeneity (e.g., mobility patterns).

More recently researchers have begun to use SEM-MIMIC models to analyze urban public transport (Zhao et al. 2014, Allen, Munoz and Ortuzar 2018, Ingvardson and Nielsen 2019, Allen et al. 2020). These models allow controlling for heterogeneity, considering several variables simultaneously. Zhao et al. (2014) analyzed the differences between captive and choice users of the Chicago public transport system for ten factors in their model, identifying significant differences in four constructs. Allen et al. (2018) investigated the urban bus system in Santiago de Chile, focusing on the effects of certain travel characteristics and sociodemographic attributes over ten latent factors linked to satisfaction with a specific bus line and with the global transit system. Using data from six European cities, Ingvardson and Nielsen (2019) studied the relationship between norms, satisfaction and public transport use. They controlled for heterogeneity using SEM-MIMIC and included dummy variables tied to socioeconomic characteristics (gender, age, occupation, etc.). Based on data from the regional railway transport services in a region in northern Italy, Allen et al. (2020) studied the effect of critical incidents on satisfaction and loyalty. Given that they used the full database, containing almost 100,000 records, they controlled for heterogeneity through many variables (date, day, time of day, access mode, service, line, use frequency, ticket type, type of user, gender, age, income and education).

3.- Case studies, survey and sample description

The case studies selected were the metropolitan area of five major western European cities (Madrid, Rome, Lisbon, Berlin and London). Appendix A provides detailed information about the case studies, survey and sample description. Table A1 (in appendix) provides some information for the metropolitan areas under study: area's definition, surface, population, density, and transport options in the area. Data supporting this research were collected through an online panel survey during May, June and July 2019. The survey, translated into the local language, was the same for the five sites. The questionnaire contained eight parts, and took an average of seven minutes' time to complete. This study only uses information from the following five parts: Part 1 focused on sociodemographic and mobility characteristics of the respondents (Table A2, in appendix); Part 2 referring to perceived quality of service, where users were asked to rate 14 service quality attributes (q1-q14) with a 5-point scale from "very low quality" to "very high quality"; Part 3 included four satisfaction statements (s1-s4); while Part 4 focused on eight indicators used for measuring involvement with the public transport service (a1-a8); and Part 5 referred to four behavioral intentions statements about the public transport system (b1-b4). Part 3 to 5 were rated also with a 5-point Likert scale from “completely disagree” to “completely agree”. Table 1 gives the average values for the pooled sample and for each city independently. A sampling stratified by gender and age was designed, with assignment proportional to the real size population of the strata for each city (EC 2019).

Table 1.- Average values for survey service quality, satisfaction, involvement and behavioral intention indicators

Table 1 shows the average values for each one of the attributes, indicators or statements utilized to evaluate service quality perception, customer satisfaction, involvement with public transport, and behavioral intentions. Similarities as well as differences are seen among the countries, varying according to the attribute, indicator or statement considered. Regarding service quality and satisfaction, London and Berlin present the highest appraisals, followed by Madrid. Based on the pooled data, *proximity*, *safety* and *accessibility* were the best rated indicators for service quality, while *individual space*, *punctuality* and *temperature* were the worst rated indicators. Total agreement is seen for all the cities in terms of *individual space*, one of the three lowest rated indicators. A fair degree of agreement is also found for *punctuality* (in London alone it was not one of the three lowest valued attributes). On the positive side, there is also a good level of agreement for *safety* and *proximity*. Some differences from city to city deserve mention, however: Madrid included among its best rated indicators *intermodality*, and among the worst, *security*; Rome gave a very high rating to *cost*, and a low one to *cleanliness*; among Berlin´s most appraised indicators is *service hours* but *cleanliness* gets a low rating; Lisbon had high ratings for *accessibility* and *intermodality*, but one of its lowest ratings went to *temperature*; and London included among its most valued indicators *service hours* and *frequency*, while among the least appraised were *temperature* and *cleanliness*. The respondents in Rome express the lowest involvement with public transport. For all the statements, except *judgement*, they present the lowest ratings. In the case of *low income*, the ratings are the highest, because the item is formulated inversely (i.e., a greater agreement with it indicates a lower involvement), as can be observed in Table A3 (in appendix). In contrast, the respondents of London and Lisbon present the highest values for a greater number of statements. Although Madrid only stands out in *save time and money*, along with Berlin, the average of its evaluations (excluding *low income* for reasons specified above) is the highest of all five cities (3.69), followed by London (3.68) and Berlin (3.58). Rome and Lisbon show the lowest average ratings (3.29 and 3.53, respectively). Madrid attains the highest values for the indicators of behavioral intentions. Also noteworthy is the fact that, while Rome and Lisbon are given low ratings in service quality, satisfaction and involvement with public transport, both cities get evaluations very similar to the other three cities for the indicators of behavioral intentions. Oliver (1999) likewise signaled that a greater degree of satisfaction does not necessarily entail greater intentions of using the service.

4.- Research models and structural equation modeling approach

4.1.- Competing models

To date, no consensus has been reached regarding the role of involvement among the constructs service quality, satisfaction and behavioral intentions or loyalty. This article aspires to help clarify this role using a comprehensive methodological approach like the ones applied in social psychological research (Olsen 2007).

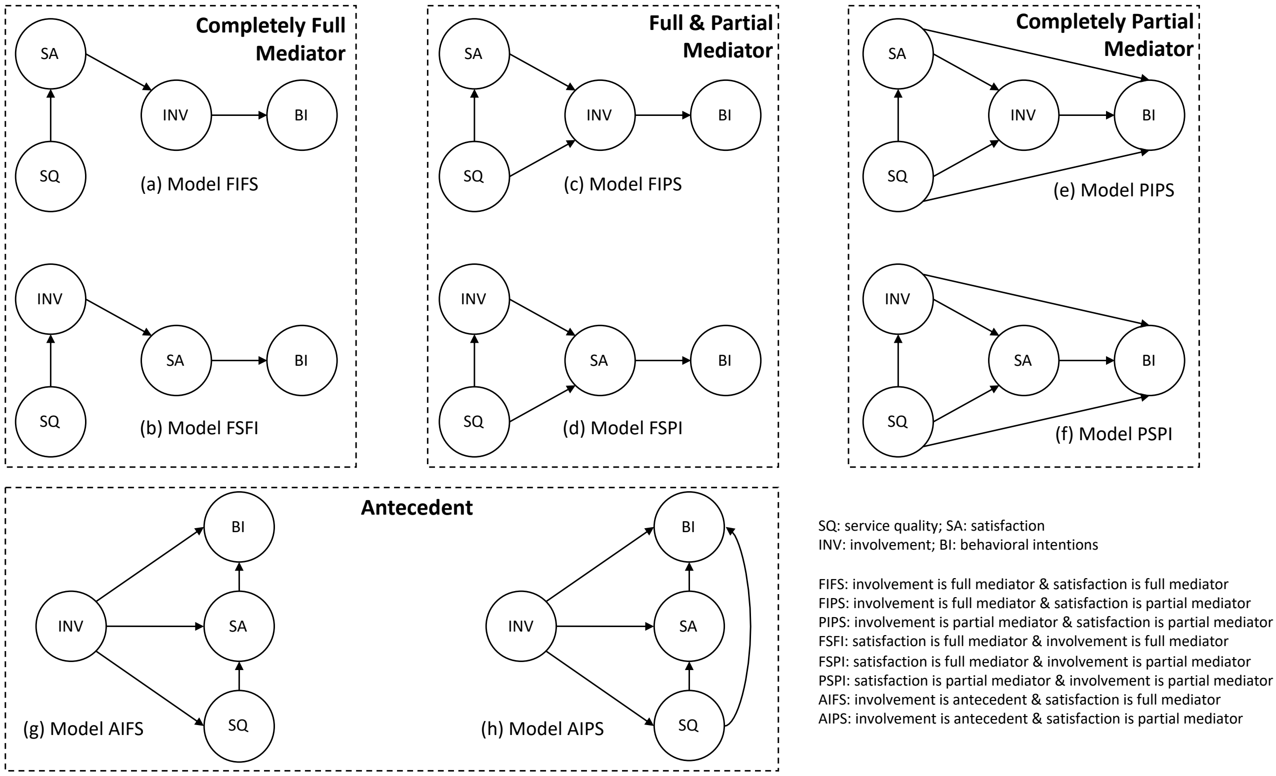


Figure 1.- Competing research models

As shown in Figure 1, eight competing models are considered in this paper. The model best supported in public transport literature (Park et al. 2004, Lai and Chen 2011, Borhan et al. 2014, Machado-Leon et al. 2016, Irtema et al. 2018) is the PIPS model (Figure 1.e), wherein satisfaction acts as a partial mediator between service quality and involvement and between service quality and behavioral intentions; and involvement also plays a partial mediator role between both constructs and behavioral intentions. Working with this family of models (completely partial mediator role), Fu et al. (2018) tested the PSPI model (Figure 1.f), where involvement acts as partial mediator between service quality and satisfaction and between service quality and behavioral intentions; and satisfaction also plays a partial mediator role between both constructs and behavioral intentions. There are also several studies that suggest a full and partial mediator role of satisfaction and involvement. Most of them support the FSPI model (Figure 1.d), where involvement plays a partial mediator role between service quality and satisfaction (Chou and Kim 2009, Kuo and Tang 2013, Chou and Yeh 2013, Chang and Yeh 2017). However, Machado-Leon et al. (2016) tested the FIPS model (Figure 1.c), wherein satisfaction plays a partial mediator role between service quality and involvement. This model was also proposed and tested by Olsen (2007). Finally, in the case of involvement as antecedent, most studies (Minser and Webb 2010, de Ona et al. 2016, Machado-Leon et al. 2016, Machado-Leon et al. 2018) uphold the AIPS model (Figure 1.h), where involvement is antecedent and satisfaction plays a partial mediator role. Still, Simsekoglu et al. (2015) proposed the AIFS model (Figure 1.g), where involvement is antecedent and satisfaction plays a full mediator role between service quality and behavioral intentions.
To complete this framework, Figure 1 depicts two other models analyzed here. They consider the possibility that the mediator effects of involvement and satisfaction be completely full in both cases. The difference between the two models resides in the order of the factors involvement and satisfaction between service quality and behavioral intentions. The FIFS model (Figure 1.a)

considers satisfaction to play a full mediator role between service quality and involvement, and involvement to play a full mediator role between satisfaction and behavioral intentions. In turn, model FSFI (Figure 1.b) holds that involvement exerts a full mediator role between service quality and satisfaction, and satisfaction plays a full mediator role between involvement and behavioral intentions. The FIFS model is a restricted version of the FIPS model, whereas model FSFI is a restricted version of model FSPI.

4.2.- Methodology

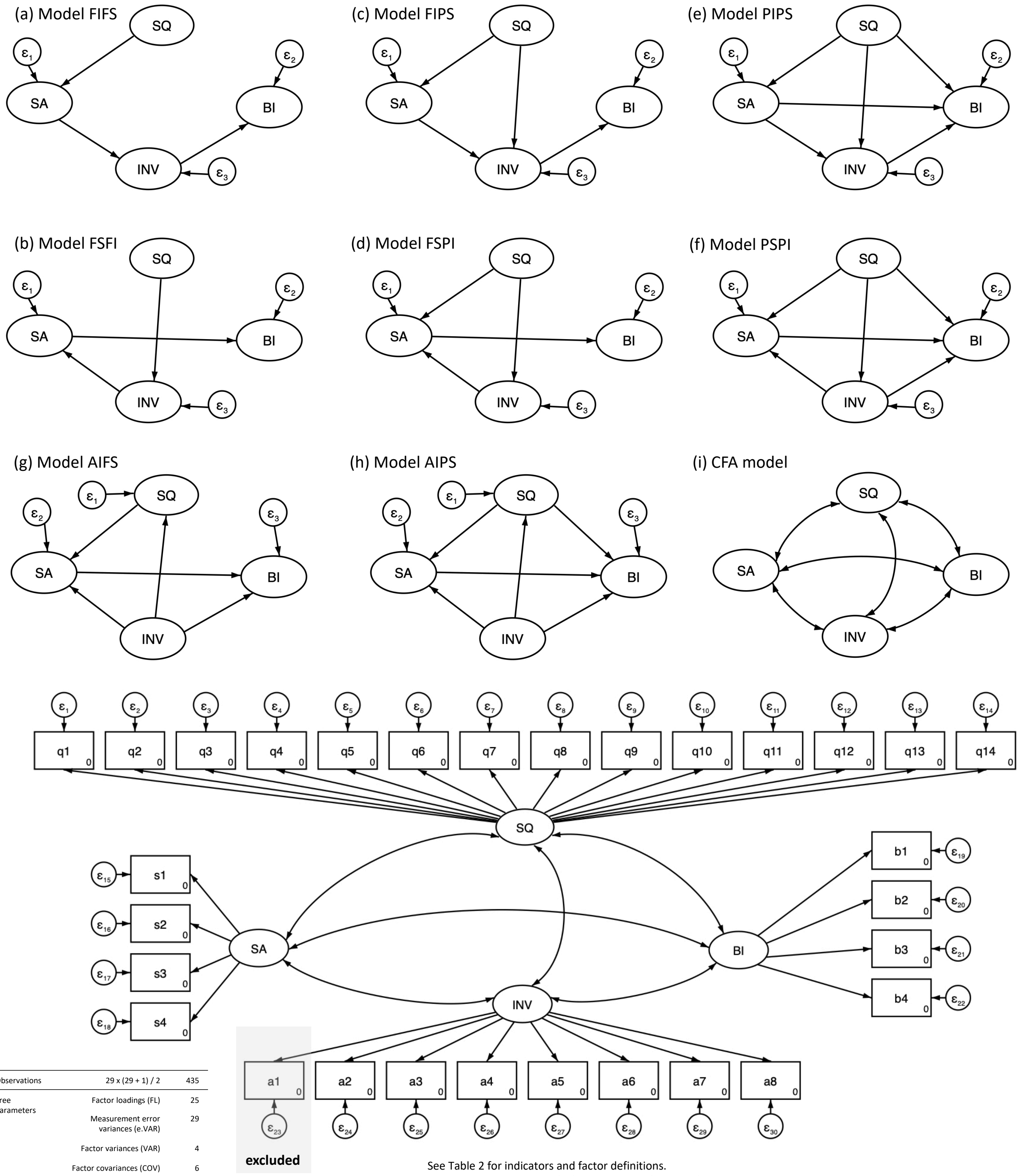
Like nearly all the papers analyzed in the literature review, this study uses structural equation modeling (SEM) to identify the effect of involvement and satisfaction among the constructs service quality and behavioral intentions or loyalty. SEM is a statistical tool that allows researchers to explain the relationship among different constructs by examining the covariance and mean structure. SEM includes a set of multivariate statistical approaches, allow for the examination of more than one relationship at a time, and it has two components: a structural regression (SR) model and a measurement model. The latter assesses unobserved latent factors as functions of observed indicators; the SR model shows the direction and strengths of the relationships between the latent factors.

This paper uses the two-step modeling approach suggested by Kline (2015): in the first step, the SR model is re-specified as a Confirmatory Factor Analysis (CFA) measurement model. Then, given an acceptable measurement model, in the second step the eight different SR models included in Figure 1 are compared to one another. As all the proposed competing models use the same latent factors, the CFA measurement model is the same for all of them.

Comparison is carried out in two successive phases. In the first, using the complete sample (pooled data) the models meant to interpret the parameter estimates are compared, their fits are evaluated, and the relative ability of each model to explain variation in behavioral intentions is assessed. If non-significant structural paths are identified, or paths whose signs are not consistent with the theory, this implies that the sample data do not support the model, and the model is concluded to be invalid.

In the event that more than one valid model is identified using the pooled data, the one presenting the best fit statistics and ability to explain variation in behavioral intentions is selected for the next stage. To confirm the model´s validity, in second place, the selected model(s) is estimated using, one by one, each of the five independent samples. The model affording consistent and statistically significant parameter estimates, reasonable fit statistics, and satisfactory explanation of behavioral intentions' variance for the independent samples as well is ultimately deemed valid.

Finally, in the third step, a SEM-MIMIC approach is applied to control for heterogeneity. SEM-MIMIC makes it possible to include in the analysis a set of attributes that could cause heterogeneity in the constructs considered. This enables to explicitly investigate hypotheses of invariance across subpopulations. In the present study, this approach is used to control the possible heterogeneity owing to the territorial setting (Madrid, Rome, Berlin, Lisbon or London), household location (urban vs. metropolitan area), specific sociodemographic characteristics of the public transport user (gender, age, level of education, dependent members in the family and income level) and frequency of use (frequent vs. occasional user). The approach applies for both the pooled data, which provides for a global perspective, and each one of the independent samples, thus permitting comparison of results across the cities.



Observations	29 x (29 + 1) / 2	435
Free parameters	Factor loadings (FL)	25
	Measurement error variances (e.VAR)	29
	Factor variances (VAR)	4
	Factor covariances (COV)	6
Degrees of freedom (DF)	Observations – Free parameters	371

See Table 2 for indicators and factor definitions.
The indicator low income (a1) was excluded from the factor involvement during preliminary steps because of negative bivariate correlations with other indicators included in this construct. See section 5.1 for more information.

Figure 2.- Research model specification: (a-h) competing structural regression models; and (i) CFA measurement model.

5.- Results

5.1.- Preliminary steps, data preparation and screening

The preliminary steps consisted of specification of the models, and checking that the models were identified. Model specification involved the theoretical establishment of the models, as shown in Figure 2. Initially 14 indicators were selected (q1-q14) for service quality (SQ), four indicators (s1-s4) for satisfaction (SA), eight indicators (a1-a8) for involvement (INV), and four (b1-b4) for behavioral intentions (BI). As the statement *low income* (a1) was formulated in an inverse form with respect to the rest of the statements, the original scores were reversed to introduce it in the model.

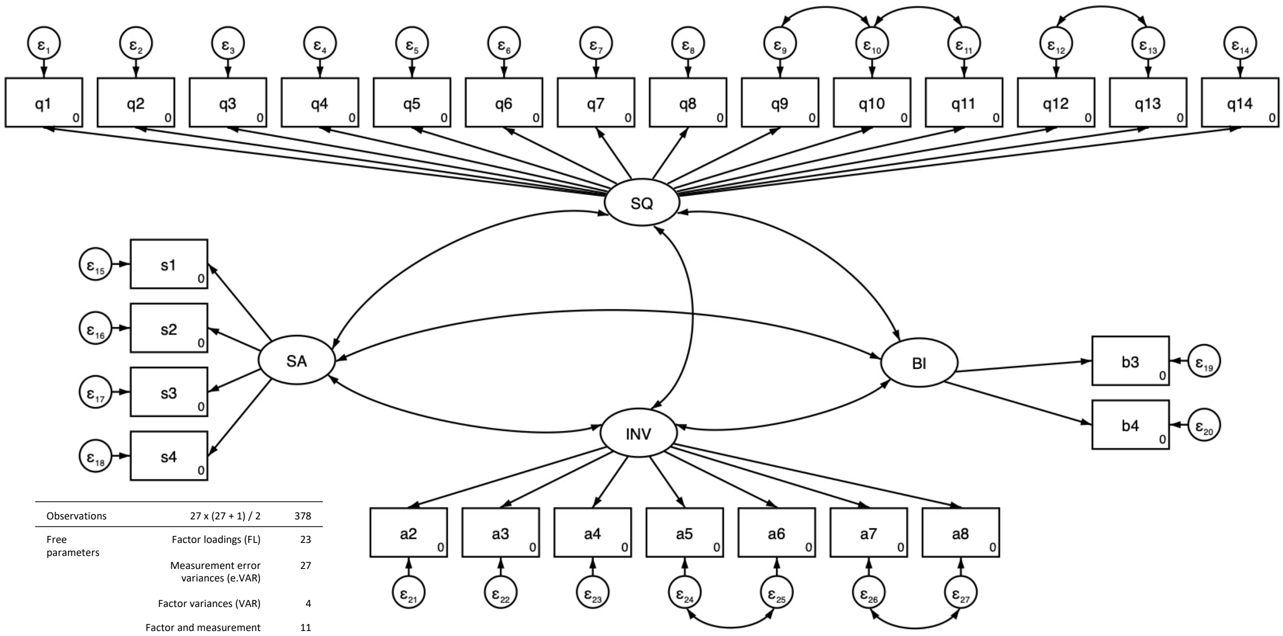
All the models that are depicted in Figure 1 and Figure 2 were identified, and it was possible to derive a unique estimate of every model parameter, for the following reasons (Kline 2015): (i) their degrees of freedom were positive; (ii) the CFA model had four factors with two or more indicators per factor; and (iii) all the SR models were recursive.

The data preparation and screening included several checks for each one of the independent samples (Madrid, Rome, Berlin, Lisbon and London). Appendix B shows a detailed description of all the results. All the statistical analysis showed appropriate results with the exception of univariate and multivariate normality. The results showed that most variables were not normal distributed. As univariate normality is a requirement for multivariate normality, the hypothesis that the data presented multivariate normality was also rejected. To address this issue, we used the Satorra-Bentler estimator, which controls for non-normality. The results below report the χ^2 corrected using this estimator, as well as all the corrected model fit indices that use χ^2.

5.2.- Measurement model

The measures' psychometric properties for all four scales were evaluated using a CFA measurement model where all factors were assumed to covary with each other. A CFA model was performed with the pooled data, considering 29 indicators and four factors (Figure 2). Each indicator was only allowed to load on one factor and could not cross-load on any other factors. Table C1 (in appendix) shows the parameter estimates and the values of selected fit indices for the initial CFA measurement model. Based on the results (Appendix C shows a detailed description), the CFA model was re-specified. The final CFA measurement model differs from the initial model in that the factor behavioral intentions (BI) was left comprising just two indicators (*Increase usage* and *I will recommend PT*); and five measurement error correlations were specified: (1) between *individual space* (q9) and *temperature* (q10); (2) between *temperature* (q10) and *cleanliness* (q11); (3) between *safety* (q12) and *security* (q13); (4) between *environment* (a5) and *reduce traffic* (a6); and (5) between *recommendation* (a7) and *judgement* (a8). All these correlations are plausible and theoretically justified. Figure 3 shows the final CFA measurement model and Table C2 (in appendix) shows the parameter estimates and the values of selected fit indices. All the approximate fit indices improved to excellent values (Hooper et al. 2008), with CFI equal to 0.958 (>0.95), TLI equal to 0.953 (>0.95), SRMR equal to 0.038 (<0.05) and RMSEA equal to 0.043 (<0.05).

The parameter estimates slightly improved in the final CFA model as compared to the initial one, with the exception of *cost* (0.497). Nonetheless, as the value was very close to 0.5, it was retained in the model. The construct validity of the model also improved. The four factors presented good values (above 0.7) for Construct Reliability (CR) and Cronbach's Alpha, ranging from 0.759 to 0.943 for CR, and 0.737 to 0.938 for Cronbach's Alpha. The Average Variance Extracted (AVE) was above the recommended threshold (0.50) in all cases, with the exception of involvement with public transport. However, Fornell and Larcker (1981) established that if AVE is less than 0.5, but CR is high, the convergent validity of the construct is still adequate. Finally, estimated factor correlations ranged from 0.513 to 0.881. Accordingly, the not excessively high factor correlations suggested discriminant validity (Kline 2015).



Observations	27 x (27 + 1) / 2	378
Free parameters	Factor loadings (FL)	23
	Measurement error variances (e.VAR)	27
	Factor variances (VAR)	4
	Factor and measurement errors covariances (COV)	11
Degrees of freedom (DF)	Observations – Free parameters	313

Figure 3.- Final CFA measurement model

5.3.- Comparison of structural regression models

Since enough evidence was found to assess the construct validity of the measurement model, the next step was to test the competing models (Figure 2) estimating eight SR models for the pooled data. Following Cronin, Brady and Hult (2000), comparison of the models was based on the values of the parameter estimates, the approximate fit indices and the relative ability of each model to explain variation in behavioral intentions (as measured by the R^2-value). Though some of these models were nested, the models were not compared using the χ^2 differences test. The estimation process relied on Satorra-Bentler statistics, which adjusted the value of χ^2 from a standard maximum likelihood estimation by an amount that reflected the degree of non-normality. The difference between the Satorra–Bentler statistics for two nested models fitted to the same data does not follow a chi-square distribution (Kline 2015).

All the models considered 27 indicators and four factors. Three models (PIPS, PSPI and AIPS) presented the same number of free parameters as the final CFA measurement model because all the correlations between latent factors were transformed into directional paths. These models were equivalent versions and generated the same predicted correlations, covariances and fit statistics (Kline 2015). The other five models were restricted versions of the previous three models because some of the paths were eliminated. Models FIPS and PSFI were restricted versions of models PIPS and PSPI, respectively, since both of them had restricted two paths. Their degrees of freedom thus increased by two, from 313 to 315. Model FIPS eliminated the direct paths between satisfaction and behavioral intentions, and between service quality and behavioral intentions; and model FSPI eliminated the direct paths between involvement and behavioral intentions, and between service quality and behavioral intentions. Model FIFS was a restricted version of models FIPS and PIPS, eliminating the direct path between service quality and involvement. Model FSFI was a restricted version of models FSPI and PSPI as it eliminated the direct path between service quality and satisfaction. Finally, model AIFS was a restricted version of model AIPS, because it eliminated the direct path between service quality and behavioral intentions. In this case the number of free parameters was reduced by one, while the degrees of freedom increased by one. The fit statistics of these five restricted versions were different from the CFA measurement model.

Table 2.- Structural paths and models' fit statistics for competing structural regression models

The eight models' estimation in Stata converged to admissible solutions. Table 2 shows the structural paths, models' fit statistics (using Satorra-Bentler estimation) for the eight competing SR models, and the ability of each model to explain variation in behavioral intentions. Like the CFA measurement model, all the models failed the exact-fit test, but all the approximate fit indices were acceptable or excellent, with CFI ranging from 0.931 to 0.958, TLI from 0.923 to 0.953, SRMR from 0.038 to 0.054, and RMSEA ranging from 0.043 to 0.055. The R^2-values ranged from 0.333 to 0.632.

Based on the parameter estimates, the data did not support four models. In models PIPS and PSPI the relationship between satisfaction and behavioral intentions was not statistically significant, whereas the relationship between service quality and behavioral intentions showed an inconsistent negative effect. Both models that considered involvement as an antecedent also presented inconsistent and non-significant parameter estimates. Satisfaction presented an inconsistent negative effect on behavioral intentions in model AIFS, and this relationship was not significant in model AIPS. Service quality furthermore presented an inconsistent negative effect on behavioral intentions in model AIPS. Thus, these four models were not retained for further analysis. As the other four models presented significant and consistent parameters estimates, their comparison was based on the approximate fit indices and on the ability of each model to explain variation in behavioral intentions.

All fit statistics suggested the superiority of models FIFS and FIPS over models FSFI and FSPI. The FSFI and FSPI models' approximate fit indices showed acceptable values, with CFI ranging from 0.931 to 0.945 (>0.90), TLI from 0.923 to 0.939 (>0.90) and SRMR from 0.052 to 0.054 (<0.08); the other two models (FIFS and FIPS) showed excellent values for all the approximate fit indices. In addition, the ability of FIFS and FIPS to explain variation in behavioral intentions (as measured by the R^2-value) was much higher than in the other two models.

Given that FIFS and FIPS showed very similar values for the approximate fit indices and R^2-values, the final decision was to retain both models for the following step, where they were estimated using each one of the five independent samples separately. The ten models' estimation in Stata converged to admissible solutions. Table 3 shows the structural paths, models' fit statistics (using Satorra-Bentler estimation) and R^2-values for behavioral intentions for the ten SR models using the data from Madrid, Rome, Berlin, Lisbon and London.

Table 3.- Structural paths and fit statistics for the five independent samples

The FIFS model's results across the independent samples from the five cities were consistent with those of the pooled data for all the relationships. The five models presented consistent and significant parameter estimates for the three relationships included in the models, and the approximate fit statistics were around the thresholds. Hence, the pooled data and the independent samples from the five European cities supported the FIFS model. The data from Madrid and Lisbon, however, did not support a direct relationship between service quality and involvement in the FIPS model. Both cities gave non-significant unstandardized estimates for this parameter. For all cities with the exception of London, the ability of model FIFS to explain variation in behavioral intentions was slightly superior than the ability of model FIPS.

In light of this comprehensive comparison to try and explain the relationship between service quality, satisfaction, involvement and behavioral intentions, it can be said that FIFS (Figure 1.a) is the best model. This model indicates that satisfaction is a completely full mediator between service quality and involvement, and involvement is a completely full mediator between satisfaction and behavioral intentions.

5.4.- SEM-MIMIC results

The SEM-MIMIC approach —used both for the pooled data and for each one of the independent samples— made it possible to identify, control and correct for possible bias in parameter estimation due to user heterogeneity behind their service quality perceptions, satisfaction, involvement with public transport and behavioral intentions.

To capture heterogeneity, the SEM-MIMIC model considered the following dummy variables:
- City: understood as the geographical area of residence, differentiating between residents in the city center and in the metropolitan area (reference).
- Male: the influence of gender, using female as the reference value.
- Old: because of differences in age, users being divided into two age groups —from 18 to 44 years old (reference) and 45 or older.
- Frequent: the influence of public transport use frequency, either occasional users (two or fewer trips per week, reference) and frequent users (more than two trips per week).
- University: people with or without (reference) a university degree.
- Dependent: the possible effect of having a dependent member in the family (i.e., children or other dependent relatives) as opposed to not having one (reference).
- High income: people with income levels above the sum of two minimum wages, as opposed to people below that threshold (reference).
- Rome, Berlin, Lisbon and London were also dummy variables, taking Madrid as reference.

All the dummy variables acted as regressors onto the four latent factors (service quality, satisfaction, involvement and behavioral intentions). Yet the dummy variables Rome, Berlin, Lisbon and London were only used in the model with the pooled data to capture heterogeneity because of the location.

Table 4 shows the structural paths, statistically significant regressor influence, selected fit statistics and the ability of each model to explain variation in behavioral intentions for the SEM-MIMIC models. All the dummy variables were significant on one or more factors, demonstrating the importance of correcting for heterogeneity. Although nearly all the model fit statistics showed a slight deterioration when compared to the original FIFS model (Table 2 for the pooled data and Table 3 for the independent samples), most of the values were still excellent (CFI and TFL above 0.95; RMSEA and SRMR below 0.05) or acceptable (CFI and TFL above 0.90; RMSEA and SRMR below 0.08). Remarkably, in all cases the SEM-MIMIC models showed higher R^2-values (ranging from 0.478 to 0.804), indicating that they outperformed the original FIFS in explaining variation in behavioral intentions.

Table 4.- SEM-MIMIC results: structural paths, statistically significant regressor influence and selected fit statistics

Table 4 shows that the models for the pooled data identified several regressors as being significant for each one of the latent factors. The main findings for service quality are: living in the city center increased the service quality perception, as did being male or having income levels above two minimum wages; respondents in London presented significantly higher service quality perceptions than residents in Madrid; but residents in Rome and Lisbon presented significantly lower service quality perceptions. In the case of satisfaction, heterogeneity for the pooled data was lower. Males expressed higher satisfaction than females, while people with university degrees and the residents of Rome expressed significantly lower satisfaction. Involvement with public transport is the construct with the highest number of significant regressors: people 45 years of age or older, those having a university degree, and frequent users showed better involvement with public transport, male involvement being lower than that of females. Residents of Rome and Lisbon presented a significantly higher involvement than residents of Madrid, while London residents reflected the lowest involvement. Behavioral intentions also showed low levels of heterogeneity: people 45 years of age or older and people in Rome expressed significantly better behavioral intentions towards public transport, in light of their reference categories. Contrariwise, Berlin residents showed worse behavioral intentions than residents of Madrid.

Analysis of the results of the models for the independent samples per country also point to interesting findings. On the one hand, the latent factors with a greater number of significant regressors are service quality and involvement; the least number of significant regressors corresponds to satisfaction. This result is largely in line with those obtained for the pooled data. On the other hand, Table 4 illustrates how Rome and Lisbon have greater heterogeneity in that they are identified with more significant regressors. Notwithstanding, in London and Madrid, homogeneity is greater, being identified with a lesser number of significant regressors. In London, significant differences were only detected for one dummy variable and involvement: involvement with public transport improved in the case of respondents with dependent members in their families. In Madrid, significant differences were found for three variables: living in the city center and being male increased service quality perceptions; and people 45 or older presented significantly better behavioral intentions towards public transport. In Berlin and Lisbon heterogeneity was identified in all the latent factors: Berlin´s respondents 45 or older showed higher service quality perceptions; being male increased satisfaction; people with a university degree had better involvement with public transport; and having income levels above two minimum wages contributed to better behavioral intentions towards public transport. Still, females presented higher involvement than males. In the case of Lisbon, having a university degree contributed negatively to satisfaction and behavioral intentions; people 45 years of age or older presented significantly better behavioral intentions towards the public transport; being male contributed negatively to involvement, and being a frequent user contributed positively. Lastly, and unlike the other cities, living in the city center of Lisbon decreased service quality perceptions. In Rome, meanwhile, even though significant variables were not found for satisfaction and behavioral intentions, it is the city exhibiting the greatest heterogeneity regarding service quality: living in the city center and having income levels above two minimum wages contributed to increased service quality perceptions, yet frequent users and people 45 years old or older presented significantly worse perceptions. Finally, people 45 or older in Rome and having income levels above two minimum wages presented better involvement with public transport, when compared with their reference categories.

Table 4 also helps one to clearly identify the regressors where greater heterogeneity exists, where latent factors are involved in such heterogeneity, and if a consistent pattern in the behavior of the regressors exists (i.e., the direction of the influence is stable). The place of residence is found to only affect service quality: generally, living in the city center contributes positively to the service quality perception, the exception in this case being Lisbon, where the opposite trend is observed. Significant differences are seen between men and women for three of the four factors analyzed: men perceive the quality of service as better, and are more satisfied than women; but they present poorer involvement with public transport. Age also influences in three of the four factors, albeit with diverse effects. In general, people 45 years old or older present more positive involvement and behavioral intentions toward public transport than the younger population, and no significant differences are observed in terms of satisfaction. As for service quality perception, in most cities the elevated perception remains for the older age bracket, and is significant in Berlin, but the lowest perception for this age group in Rome is noteworthy. Overall, the frequency of public transport use contributes to a greater involvement with it, but also to a worse service quality perception in the majority of the five cities surveyed —above all in Rome. In most of the cities, having a university degree contributes negatively to satisfaction and behavioral intentions. However, it contributes positively to having better involvement with public transport. The income level also bears some influence on three of the four factors analyzed, and a higher income tends to imply better service quality perceptions, involvement and behavioral intentions —these differences proving significant in the specific cases of service quality and involvement in Rome, and behavioral intentions in Berlin.

This table also highlights (if compared to Table 2 and Table 3) changes in the parameter estimates of the structural paths, though in most cases such variations are of scant importance (less than 5%). London's model shows the lowest changes (ranging from -1.1% for the path between service quality and satisfaction to +0.2% for the path between involvement and behavioral intentions). On the other hand, the model with the pooled data shows the largest changes (from -5.1% for the path between service quality and satisfaction to +6.6% for the path between satisfaction and involvement). The reliability of these parameter estimates (Table 4) is greater than for the original FIFS model because the SEM-MIMIC approach allows one to control for heterogeneity, and all the models enhance their ability to explain variation in behavioral intentions (as measured by the R^2-value) when compared to the original FIFS model.

6.- Discussion of results

Reviewing the state-of-the-art makes manifest that, despite the lack of consensus regarding the role of involvement in the relationship between service quality, satisfaction and behavioral intentions, two predominant trends stand out: studies suggesting a mediator role are more numerous, yet studies pointing to involvement as an antecedent of service quality, satisfaction and behavioral intentions are also fairly abundant (Minser and Webb 2010, Simsekoglu et al. 2015, de Ona et al. 2016, Machado-Leon et al. 2016, Machado-Leon et al. 2018). What is more, studies supporting the mediator role of involvement do not agree on whether that role is full or partial, or if the mediator role lies between service quality and satisfaction, or else between satisfaction and behavioral intentions. A slight majority proposes that involvement exerts a partial mediator role between satisfaction and service quality with respect to behavioral intentions (Figure 1.e) (Park et al. 2004, Lai and Chen 2011, Borhan et al. 2014, Machado-Leon et al. 2016, Irtema et al. 2018), and somewhat fewer hold that there is a partial mediator role between service quality and satisfaction (Figure 1.d) (Chou and Kim 2009, Chou and Yeh 2013, Kuo and Tang 2013, Chang and Yeh 2017). A minority of studies propose other types of mediator role.

This study uses a comprehensive methodological approach to identify the role of involvement comparing eight alternative models and using data from independent surveys carried out in five European cities (Madrid, Rome, Berlin, Lisbon and London). Excepting the study by Machado-Leon et al. (2016), which compares several alternative models with data from a single sampling, this is the only study to date in the field of public transport that compares a significant number of alternative models using data from independent samples.

The results expounded in the preceding section indicate that the FIFS model fits well and outperforms the competing models. In this model, satisfaction is a full mediator between service quality and involvement, and involvement is a full mediator between satisfaction and behavioral intentions. As far as the author knows, this model has never before been suggested in the field of public transport, even though it is one of the simplest of the evaluated models, and has a number of traits that come to support its candidacy:

- It is a restricted version of the model PIPS (Figure 1.e), the one best supported by the literature; and of the model FIPS (Figure 1.c), well defended in two studies involving alternative models: Olsen (2007) outside the field of public transport, and Machado-Leon et al. (2016) within the field. In comparison with PIPS and FIPS, the results put forth here show it to be the only of the three models that would be valid for all the cities and for the pooled data.
- There are many studies in the field of public transport that defend the full mediator role of satisfaction between service quality and behavioral intentions or loyalty (Zhang et al. 2019, Yuan et al. 2019, Sun and Duan 2019, de Ona et al. 2015, Allen et al. 2020). This comes to lend theoretical support to model FIFS over models FIPS and PIPS.
- In comparison with the models in which satisfaction is a full mediator between involvement and behavioral intentions (FSFI and FSPI), the findings expounded here show that the capacity of the latter for explaining the variability of behavioral intentions is much lower, and the fit parameters are poorer than those of the FIFS model.
- The other three models analyzed here —PSPI, AIFS and AIPS— are not considered valid because they present non-significant or non-consistent parameter estimates. Despite the fact that these models present good fit parameters, they need to be re-specified, eliminating the non-significant structural paths (Kline 2015).

To summarize, all the competing models analyzed share some positive features. Their fit parameters are quite sound (RMSEA, CFI, TLI and SRMR having excellent values) and they provide a high value in terms of explaining the variability of behavioral intentions. Only models FSFI and FSPI show lower values. The total effects (sum of direct and indirect effects) of service quality on satisfaction and behavioral intentions, of satisfaction on behavioral intentions, and of involvement on behavioral intentions are significant, of a high magnitude, and with a proper sign (positive) in all cases. In the event that only one model and one sample were analyzed, this could give rise to a validation of models that explain the data of a specific study case quite well, but would not prove to be the best model, or a generally valid one. This disadvantage highlights the importance of considering alternative models that might account for the same pattern of observed covariances. It is essential to avoid confirmation bias —e.g., giving an overly positive evaluation of the author's preferred model, and failing to consider other explanations of the data (Shah and Goldstein 2006, Kline 2015).

The use of SEM-MIMIC models has become increasingly frequent in the field of public transport for identifying, controlling and correcting for possible biases in parameter estimation due to users' heterogeneity (Zhao et al. 2014, Allen et al. 2018, Ingvardson and Nielsen 2019, Allen et al. 2020). This study adopted the MIMIC approach to analyze the effects of territorial context, frequency of use of public transport and other sociodemographic characteristics on the service quality perceptions, satisfaction, involvement with public transport and behavioral intentions.

If one compares the results of Table 2 (FIFS model for the pooled sample), and Table 3 (FIFS models per city) with those of Table 4 (SEM-MIMIC results for pooled data and per city), the models with the MIMIC approach are seen to increase their explicative capacity regarding the variability of behavioral intentions (as measured by the R^2-value), although a slight decrease in some of the parameters of global fit of the model (CFI and TLI) are seen, going from excellent to acceptable; whereas RMSEA and SRMR remain stable. Notwithstanding, there is evidence that increasing the number of variables in the model, and also the degrees of freedom, tends to worsen the CFI and the TLI (Kenny and McCoach 2003).

The standardized values of the structural paths in Table 4 are slightly different from those of Table 2 and Table 3, as they include the control of heterogeneity of the MIMIC approach. The fact that the changes are not accentuated means that there is not excessive heterogeneity, above all in the case of considering the samples independently.

Table 5.- SEM-MIMIC results: total effects between factors

Table 5 indicates that, for the pooled sample and for each one of the cities analyzed, involvement with public transport is the factor that contributes most to behavioral intentions, with total effects ranging from 0.613 to 0.976; it is followed, in second place, by service quality, with total effects ranging from 0.440 to 0.808; and in third place comes satisfaction, with total effects ranging from 0.418 to 0.571. This same order of importance was reported by Lai and Chen (2011) in their analysis of a mass rapid transit system in Taiwan. It was furthermore observed that in all the models, service quality has a greater effect on involvement —with total effects ranging from 0.553 to 0.828— than the effect of satisfaction upon this factor, with total effects ranging from 0.490 to 0.749. The reason why the indirect effect of service quality on involvement is greater than the effect of satisfaction on involvement lies in the elevated direct effect of service quality upon satisfaction, which is superior to one in all cases, ranging from 1.047 to 1.414. This means that a 1-point increase in the service quality variable predicts a more than 1-point increase in satisfaction. From a practical standpoint this result is highly relevant for the management of public transport, as an increase in perceived quality by users would lead to a more than proportional increase in satisfaction.

Table 5 also shows that London is the city where the highest values for all the effects among factors can be found, except for the effect of satisfaction on involvement, which pertains to Rome and Berlin. The lowest values for the total effects among factors is distributed among Madrid, Rome and Lisbon. The lowest total effects for the relations between service quality and behavioral intentions and satisfaction and behavioral intentions is found in Madrid. The lowest effects between service quality and satisfaction, and between involvement and behavioral intentions, would correspond to Rome. And the lowest values for total effects on involvement of service quality and satisfaction correspond to Lisbon.

The model with the pooled data makes it possible to capture the influence of the geographical context controlling for all the other sociodemographic factors and travel patterns, with the following findings (Table 4):

- The differences between Madrid and Rome are significant for all the factors, being positive for involvement and behavioral intentions, and negative for service quality and satisfaction.
- The differences between Madrid and Lisbon are only significant for service quality (negative) and involvement (positive).
- The opposite is true for London, where the differences are significant for service quality (positive) and involvement (negative).
- Between Madrid and Berlin, significant differences are only identified for behavioral intentions (negative).

These results agree with the average values for the indicators of service quality, satisfaction, involvement and behavioral intentions presented in Table 1. The table shows that the perception of the quality and the satisfaction in Rome is much lower than in Madrid; yet the differences in terms of involvement and behavioral intentions are minimal. Between Lisbon and Madrid, there are important differences in the appraisal of the attributes of service quality (lower in Lisbon); nonetheless, for the other three factors —even though Lisbon´s values are below those of Madrid— the values are not so different, some attributes of involvement even presenting higher values. Contrariwise, in London the values of service quality are above those of Madrid, whereas the indicators of the other three factors present similar appraisals, the figures for Madrid being higher for most of the indicators associated with involvement. The evaluations of Madrid and Berlin are very similar for all the factors, though somewhat lower for Berlin when it comes to the indicators associated with behavioral intentions.

Finally, the MIMIC approach provides some insight as to which sociodemographic characteristics and travel habits of users affected the perception of different latent factors. All the regressors produced significant results for at least one factor, summed up below.

- Household location: In general, living in the city center has a positive effect on service quality perception. This effect is significant for the pooled data, and for Madrid and Rome; but an opposite effect was identified in Lisbon. This effect could be linked to the fact that the city center usually presents better transport services (e.g., higher frequency and punctuality).
- Gender: Although no significant differences on behavioral intentions are identified, males have a positive effect upon service quality and satisfaction, yet a negative effect on involvement. Such a finding about service quality and satisfaction has been reported in previous studies

(Allen et al. 2018, Allen et al. 2020, Ingvardson and Nielsen 2019). (Allen et al. 2018) justify it based on the fact that: (i) women tend to travel with children and are more prone to make shopping trips, which can be more uncomfortable; (ii) women perceive more insecurity than men, especially in crowded spaces; and (iii) women tend to be captive users. The higher involvement of women with public transport may be due to the fact that they feel a stronger environmental commitment than men (Golob and Hensher 1998). Authors Ingvardson and Nielsen (2019) and Allen et al. (2020) found no significant differences in terms of gender over behavioral intentions or loyalty.

- Age: In general, people 45 years old or older present higher involvement and behavioral intentions toward public transport than younger users. No significant differences were found for their satisfaction, and in the case of service quality the effect varies from city to city: while the perception of older users is more positive in most cities, in Rome and Lisbon the perception of young users is more positive. This lack of consensus for the satisfaction factor is also identified in the literature. Allen et al. (2018) found that older people tend to be less satisfied, whereas Allen et al. (2020) arrived at the opposite conclusion at another location: users above 65 years old expressed greater satisfaction with public transport and higher loyalty. This positive effect of age on involvement may be traced to a greater awareness of the importance of this service to society, a stronger environmental commitment on the part of the elderly (Golob and Hensher 1998), or a greater dependence upon public transport as one grows old and drives less, meaning older respondents could have higher involvement and behavioral intentions.
- Public transport use frequency: The frequency of use of public transport is not significant in most cases, and does not present a constant pattern for any of the latent factors. This finding also coincides with that of a previous study (Allen et al. 2020) analyzing the effect of this attribute on satisfaction and loyalty, which was not found to be significant in either case. However, more insight is warranted in Rome and Lisbon to identify the causes of significant differences.
- Education level: The educational level is not significant for service quality, though it is indeed significant for the other three factors. The effect is diverse: people with a higher educational level present higher involvement with public transport, but their satisfaction and behavioral intentions are worse. The effect upon satisfaction agrees with the study of Allen et al. (2020) in that they found that having a university degree contributed negatively to overall satisfaction. Golob and Hensher (1998) likewise identified highly educated people as having stronger environmental commitment, which could imply higher involvement.
- Dependent members in the family: Generally, having dependent members in the family is not significant for any of the factors. Nonetheless, in London a significantly positive effect was found upon the involvement with public transport. As no previous studies have analyzed this effect, more insight is warranted to determine the causes of differences in involvement in London.
- Income level: In general, people with an income level amounting to over two minimum wages present higher appraisals of all the latent factors, being significant for service quality and involvement in Rome, and for behavioral intentions in Berlin. This finding agrees with Allen et al. (2020), who found that satisfaction and loyalty increased for passengers with middle-high incomes. The reason may be that low-income people must make a substantial economic effort to use the public transport; hence they perceive a higher price than middle-high income users. Regarding involvement, Golob and Hensher (1998) found that having a high household income contributes to stronger environmental commitment.

7.- Conclusions and recommendations

The main objective of this study was to clarify the relationship between service quality, satisfaction, involvement with public transport, and behavioral intentions or loyalty. The comprehensive methodological approach served to compare eight alternative structural equation models based on five independent data samples, from the European cities of Madrid, Rome, Berlin, Lisbon and London. The results support that satisfaction is a complete mediator between service quality and involvement, and involvement is a complete mediator between satisfaction and behavioral intentions. Therefore, this contribution can be said to uphold the mediator role of involvement in the relationship among service quality, satisfaction, and behavioral intentions or loyalty. Unlike most previous studies, which suggest a partial mediator role, the evidence presented here supports a complete mediator role.

It is important to underline that most previous studies do not compare alternative models or use different samples, which might give rise to a confirmation bias. The probability that this could happen is high, since all the competing models analyzed in this article present factors of global adjustment that are acceptable. Even though it can be difficult to obtain access to more than one independent sampling in order to contrast models, if the sample size is sufficient, one should apply a cross-validation method or a split-sample approach to validate results. At any rate, this type of analysis should not be limited to a single model, but rather is recommended for the comparison of all the alternative models that may be plausible from a theoretical standpoint.

From a practical standpoint, some relevant conclusions can be drawn and extrapolated to all the cases analyzed. Involvement with public transport would be the factor contributing the most to behavioral intentions or loyalty, followed by service quality perceptions and satisfaction. Public transport operators, transport authorities and policy makers would be wise to remain concerned about offering quality service, because it is high correlated with involvement, but they should increase efforts in campaigns to encourage an overall change of involvement with public transport. Awareness campaigns should be oriented to engage the entire population, from an early age, since today´s youth will be the public transport users of tomorrow.

In closing, the importance of bearing in mind the heterogeneity in user perceptions should also be underlined. The MIMIC approach makes it possible to identify significant differences in all the market segments considered (city, household location, gender, age, frequency of travel, education level, dependent members in the family, and income level). By controlling for this heterogeneity, models with more robust parameter estimates can be derived, better able to explain variation in behavioral intentions. This study has shown that service quality perceptions and involvement with public transport are the latent factors presenting the most heterogeneity. Therefore, if the sample size is large enough, the use of methodologies that allow for control of the heterogeneity in the sample would be advisable. In this case dummy variables were used to control for heterogeneity, with just two categories for each of the regressors; yet if the sample size had been greater, more categories could have been considered (e.g., several age ranges).

However, as all the five sample cities are located in western countries with similar social and economic background, further studies should be performed before being able to generalize these results to other cities with different contexts.

Acknowledgements
Support from Spanish Ministry of Economy and Competitiveness (Research Project TRA2015-66235-R) is gratefully acknowledged.

Appendix A.- Detailed information about the case studies, survey and sample description

The case studies selected were the metropolitan area of five major western European cities (Madrid, Rome, Lisbon, Berlin and London). Table A1 shows that the size of the metropolitan area ranges from 892 km2 in Greater Berlin to 8,030 km2 in Madrid Metropolitan Area; population ranges from 2.89 million inhabitants in Lisbon Metropolitan Area to 8.90 in Greater London; and population density ranges from 811 inhabitants per km2 in Madrid Metropolitan Area to 5,672 in Greater London. In all the metropolitan areas the public transport options include the commuter rail, metro, tram and bus services. Almost all of them also have public bicycle systems available, with the exception of Rome. Lisbon and London also include the ferry as another public transport option.

Table A1.- Basic data per city (Madrid, Rome, Berlin, Lisbon and London)

Table A2 shows that the questionnaire was completed by 2,579 public transport users (525 from Madrid, 509 from Rome, 508 from Berlin, 530 from Lisbon and 507 from London), giving us a well-balanced set of data. In general, the questionnaire was completed by more women than men, above all in the age brackets of 25-44, and 45-64 years of age. This was true of the full (pooled) sample and for each one of the cities. Most of the participants had been living for quite a long time in the same area, making them very familiar with the local system of public transport. Excepting the case of Berlin, most respondents lived in the metropolitan area. The educational level was high in all cases, with a prevalence of respondents having higher education, except in Berlin, where secondary education prevailed. The majority of users queried tended to be employees or professionals. In all cities except Rome, the predominant living situation was a nuclear family with two members, followed by three members; whereas the families with five or more members made up the smallest percentage of the samples. In Rome, the highest percentage corresponded to families having four members. Most respondents had no dependent persons in the nuclear family (i.e., children or other dependent relatives) and used public transport regularly (from five to seven times per week). Finally, except in Madrid, the predominant net family income among the respondents was below the sum of two minimal wages. To homogenize income, and thus make them comparable among the five respective countries, the minimum wages for each country in 2018 was taken as reference. In the case of Italy, where there is no established minimum wage, Spain´s figure was used, normalized according to the rent per capita of the two countries. Notwithstanding, it must be stressed that this is the sociodemographic characteristics presenting the highest number of lost values, ranging from 48 in London to 70 in Madrid.

Table A2.- Sociodemographic characteristics and mobility patterns

Table A3 offers the description of each statement or indicator, and Table 1 (in the main text) gives the average values for the pooled sample and for each city independently.

Table A3.- Description of indicators and constructs

Appendix B.- Detailed description of data preparation and screening

The data preparation and screening included several checks for each one of the independent samples: relative variances, missing values, outliers, collinearity, sample size, univariate and multivariate normality. All the statistical analysis was performed using Stata/MP 16.1.

As all the attributes used the same scale (5-point Likert scale), the covariance matrices were not ill scaled (i.e., when the ratio of the largest to the smallest variance is greatest than 10). For all countries, the indicator with the highest number of missing values (9.2% in average) was *I think that by using public transport I can improve the way that relatives and friends judge me (judgement)*. As this value was below 10% and did not show a systematic pattern, this data loss was of little concern. No extreme outliers were identified when checking for univariate and multivariate outliers.

Collinearity was analyzed using three methods: (i) calculating bivariate correlation between all the variables; (ii) calculating squared multiple correlations between each variable and all the rest; and (iii) calculating the variance inflator factor (VIF). As all the results respected the thresholds, the sample data did not present multi-collinearity problems. However, in the case of the construct involvement, negative bivariate correlation was identified between *low income* (already reversed) and four of the other indicators utilized by this construct. It was therefore decided to not include that indicator in the models; hence 14 indicators were finally used for service quality, four indicators for satisfaction, seven for involvement and four for behavioral intentions (Figure 2, in the main text).

The minimum sample size was 507 (London), over the limit according to several thresholds recommended in the literature: (i) at least five times the number of free parameters in the model (Bentler and Chou 1987); (ii) at least 15 times the number of observed variables (Stevens 2009); and (iii) a minimum sample of 200 (Kline 2015). The CFA measurement model in Figure 2 has 64 free parameters and 29 observed variables.

Univariate normality was tested using the Shapiro-Wilk test for normality. The results showed that most variables were not normal distributed. As univariate normality is a requirement for multivariate normality, the hypothesis that the data presented multivariate normality was also rejected. To address this issue, we used the Satorra-Bentler estimator, which controls for non-normality. The results report the χ^2 corrected using this estimator, as well as all the corrected model fit indices that use χ^2.

Appendix C.- Detailed description of CFA measurement model

The measures' psychometric properties for all four scales were evaluated using a CFA measurement model where all factors were assumed to covary with each other. A CFA model was performed with the pooled data, considering 29 indicators and four factors (Figure 2, in the main text). Each indicator was only allowed to load on one factor and could not cross-load on any other factors. The unstandardized loading of *service hours*, *general satisfaction*, *freedom* and *I will use PT for one-off trips* were fixed to 1.0 to scale each one of the factors. With 29 indicators, there were 435 observations available to estimate a total of 64 free parameters, including 33 variances of exogenous variables (four factors and 29 measurement errors), six factor covariances and 25 factor loadings, so $df_M = 371$. The model's estimation converged to admissible solutions.

Table C1 shows the parameter estimates and the values of selected fit indices for the initial CFA measurement model. This model failed the exact-fit test with $\chi^2(371)$ equal to 2,872.43 ($p<0.001$). As this is generally the case in the literature, we considered other approximate fit indices to diagnose the possible sources of misfit. The approximate fit indices showed acceptable values (Hooper, Coughlan and Mullen 2008), with CFI equal to 0.914 (>0.90), TLI equal to 0.906 (>0.90), SRMR equal to 0.050 (<0.08) and RMSEA equal to 0.058 (<0.08). Table C1 also shows that all factor loadings were statistically significant, presented the correct sign (positive), and their values were higher than 0.50 as suggested by Hair et al. (2010), most being above the ideal threshold of 0.7, with the exception of *I will use PT for one-off trips* (0.453) and *I will use PT for regular trips* (0.379). Moreover, the factor behavioral intentions presented very low values for Construct Reliability (CR), Average Variance Extracted (AVE) and Cronbach's Alpha. Therefore, in view of the parameters of the global adjustment and the problems of validity and reliability identified in the factor behavioral intentions, the model was re-specified. First, the possible exclusion of the indicators *I will use PT for one-off* trips (b1) and *I will use PT for regular trips* (b2) was considered, then the correlation residuals and modification indexes were inspected. Although the model's re-specification was done step-by-step, as the misfit sources can vary as changes are introduced in the model, due to space considerations only the results of the final CFA measurement model are depicted (Figure 3, in the main text).

Table C1.- Initial CFA measurement model

The final CFA measurement model differs from the initial model in that the factor behavioral intentions (BI) was left comprising just two indicators (*Increase usage* and *I will recommend PT*); and five measurement error correlations were specified: (1) between *individual space* (q9) and *temperature* (q10); (2) between *temperature* (q10) and *cleanliness* (q11); (3) between *safety* (q12) and *security* (q13); (4) between *environment* (a5) and *reduce traffic* (a6); and (5) between *recommendation* (a7) and *judgement* (a8). All these correlations are plausible and theoretically justified. Figure 3 shows the final CFA measurement model, considering only 27 indicators and four factors. In this case, there were 378 observations available to estimate a total of 65 free parameters, including 31 variances of exogenous variables (four factors and 27 measurement errors), four factor covariances, five measurement error covariances and 23 factor loadings, so $df_M = 313$.

Table C2 shows the parameter estimates and the values of selected fit indices for the final CFA measurement model. Although the final CFA model also failed the exact-fit test with $\chi^2(313)$ equal to 1,487.87 ($p<0.001$), all the approximate fit indices improved to excellent values (Hooper et al. 2008), with CFI equal to 0.958 (>0.95), TLI equal to 0.953 (>0.95), SRMR equal to 0.038 (<0.05) and RMSEA equal to 0.043 (<0.05).

Table C2.- Final CFA measurement model

The parameter estimates slightly improved in the final CFA model as compared to the initial one, with the exception of *cost* (0.497). Nonetheless, as the value was very close to 0.5, it was retained in the model. The construct validity of the model also improved. The four factors presented good values (above 0.7) for Construct Reliability (CR) and Cronbach's Alpha, ranging from 0.759 to 0.943 for CR, and 0.737 to 0.938 for Cronbach's Alpha. The Average Variance Extracted (AVE) was above the recommended threshold (0.50) in all cases, with the exception of involvement with public transport. However, Fornell and Larcker (1981) established that if AVE is less than 0.5, but CR is high, the convergent validity of the construct is still adequate. Finally, estimated factor correlations ranged from 0.513 to 0.881. Accordingly, the not excessively high factor correlations suggested discriminant validity (Kline 2015).

Acknowledgements

Support from Spanish Ministry of Economy and Competitiveness (Research Project TRA2015-66235-R) is gratefully acknowledged.

References

Allen, J., L. Eboli, C. Forciniti, G. Mazzulla & J. D. Ortuzar (2019a) The role of critical incidents and involvement in transit satisfaction and loyalty. *Transport Policy,* 75**,** 57-69.

Allen, J., L. Eboli, G. Mazzulla & J. D. Ortuzar (2020) Effect of critical incidents on public transport satisfaction and loyalty: an Ordinal Probit SEM-MIMIC approach. *Transportation,* 47**,** 827-863.

Allen, J., J. C. Munoz & J. D. Ortuzar (2019b) Understanding public transport satisfaction: Using Maslow's hierarchy of (transit) needs. *Transport Policy,* 81**,** 75-94.

Allen, J., J. C. Munoz & J. d. D. Ortuzar (2018) Modelling service-specific and global transit satisfaction under travel and user heterogeneity. *Transportation Research Part a-Policy and Practice,* 113**,** 509-528.

An, M. & Y. Noh (2009) Airline customer satisfaction and loyalty: impact of in-flight service quality. *Service Business,* 3**,** 293-307.

Bentler, P. M. & C. P. Chou (1987) Practical Issues in Structural Modeling. *Sociological Methods and Research,* 16**,** 78-117.

Borhan, M. N., D. Syamsunur, N. M. Akhir, M. R. M. Yazid, A. Ismail & R. A. Rahmat (2014) Predicting the use of public transportation: A case study from Putrajaya, Malaysia. *Scientific World Journal,* 2014.

Chang, Y. H. & C. H. Yeh (2017) Corporate social responsibility and customer loyalty in intercity bus services. *Transport Policy,* 59**,** 38-45.

Chen, C.-F. & M.-H. Tsai (2008) Perceived value, satisfaction, and loyalty of TV travel product shopping: Involvement as a moderator. *Tourism Management,* 29**,** 1166-1171.

Chen, C. F. (2008) Investigating structural relationships between service quality, perceived value, satisfaction, and behavioral intentions for air passengers: Evidence from Taiwan. *Transportation Research Part a-Policy and Practice,* 42**,** 709-717.

Chou, J. S. & C. Kim (2009) A structural equation analysis of the QSL relationship with passenger riding experience on high speed rail: An empirical study of Taiwan and Korea. *Expert Systems with Applications,* 36**,** 6945-6955.

Chou, J. S. & C. P. Yeh (2013) Influential constructs, mediating effects, and moderating effects on operations performance of high speed rail from passenger perspective. *Transport Policy,* 30**,** 207-219.

Chou, P. F., C. S. Lu & Y. H. Chang (2014) Effects of service quality and customer satisfaction on customer loyalty in high-speed rail services in Taiwan. *Transportmetrica a-Transport Science,* 10**,** 917-945.

Cronin, J. J., M. K. Brady & G. T. M. Hult (2000) Assessing the effects of quality, value, and customer satisfaction on consumer behavioral intentions in service environments. *Journal of Retailing,* 76**,** 193-218.
Cronin, J. J. & S. A. Taylor (1992) Measuring service quality - A reexamination and extension. *Journal of Marketing,* 56**,** 55-68.
de Ona, J. & R. de Ona (2015) Quality of service in public transport based on customer satisfaction surveys: A review and assessment of methodological approaches. *Transportation Science,* 49**,** 605-622.
de Ona, J., R. de Ona, L. Eboli, C. Forciniti & G. Mazzulla (2016) Transit passengers' behavioural intentions: the influence of service quality and customer satisfaction. *Transportmetrica a-Transport Science,* 12**,** 385-412.
de Ona, J., E. Estevez & R. de Ona (2020) Perception of public transport quality of service among regular private vehicle users in Madrid, Spain. *Transportation Research Record,* 2674**,** 213-224.
de Ona, R., J. L. Machado & J. de Ona (2015) Perceived service quality, customer satisfaction, and behavioral intentions Structural Equation Model for the metro of Seville, Spain. *Transportation Research Record,* 2538**,** 76-85.
de Oña, R., E. Estevez, L. Garach, F. J. Calvo & J. de Oña (2018) Transit quality evaluation: Processes conducted by managers and operators. *Transportation Research Procedia,* 33**,** 28-34.
EC. 2019. Dataset population. Urban Audit Cities. https://ec.europa.eu/eurostat/web/cities/data/database. European Commission. Eurostat.
Flynn, L. R. & R. E. Goldsmith (1993) Identifying innovators in consumer service markets. *Service Industries Journal,* 13**,** 97-109.
Fornell, C. & D. F. Larcker (1981) Evaluating Structural Equation Models with unobservable variables and measurement error. *Journal of Marketing Research,* 18**,** 39-50.
Fu, X. M. & Z. C. Juan (2017) Understanding public transit use behavior: integration of the theory of planned behavior and the customer satisfaction theory. *Transportation,* 44**,** 1021-1042.
Fu, X. M., J. H. Zhang & F. T. S. Chan (2018) Determinants of loyalty to public transit: A model integrating Satisfaction-Loyalty Theory and Expectation-Confirmation Theory. *Transportation Research Part a-Policy and Practice,* 113**,** 476-490.
Golob, T. F. & D. A. Hensher (1998) Greenhouse gas emissions and Australian commuters' attitudes and behavior concerning abatement policies and personal involvement. *Transportation Research Part D-Transport and Environment,* 3**,** 1-18.
Gordon, M. E., K. McKeage & M. A. Fox (1998) Relationship marketing effectiveness: The role of involvement. *Psychology & Marketing,* 15**,** 443-459.
Hair, J. F., W. C. Black, B. J. Babin & R. E. Anderson. 2010. *Multivariate Data Analysis*. Pearson.
Hooper, D., J. Coughlan & M. Mullen (2008) Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods,* 6**,** 53-60.
Ingvardson, J. B. & O. A. Nielsen (2019) The relationship between norms, satisfaction and public transport use: A comparison across six European cities using structural equation modelling. *Transportation Research Part a-Policy and Practice,* 126**,** 37-57.
Irtema, H. I. M., A. Ismail, M. N. Borhan, A. M. Das & A. B. Z. Alshetwi (2018) Case study of the behavioural intentions of public transportation passengers in Kuala Lumpur. *Case Studies on Transport Policy,* 6**,** 462-474.
Jen, W. L., R. T. Tu & T. Lu (2011) Managing passenger behavioral intention: an integrated framework for service quality, satisfaction, perceived value, and switching barriers. *Transportation,* 38**,** 321-342.

Kenny, D. A. & D. B. McCoach (2003) Effect of the number of variables on measures of fit in structural equation modeling. *Structural Equation Modeling,* 10**,** 333-351.
Kinard, B. R. & M. L. Capella (2006) Relationship marketing: the influence of consumer involvement on perceived service benefits. *Journal of Services Marketing,* 20**,** 359-368.
Kline, R. B. 2015. *Principles and Practice of Structural Equation Modeling*. Guilford Press.
Koklic, M. K., M. Kukar-Kinney & S. Vegelj (2017) An investigation of customer satisfaction with low-cost and full-service airline companies. *Journal of Business Research,* 80**,** 188-196.
Kuo, C. W. & M. L. Tang (2013) Relationships among service quality, corporate image, customer satisfaction, and behavioral intention for the elderly in high speed rail services. *Journal of Advanced Transportation,* 47**,** 512-525.
Lai, W. T. & C. F. Chen (2011) Behavioral intentions of public transit passengers. The roles of service quality, perceived value, satisfaction and involvement. *Transport Policy,* 18**,** 318-325.
Li, L. B., Y. F. Bai, Z. Q. Song, A. Chen & B. Wu (2018) Public transportation competitiveness analysis based on current passenger loyalty. *Transportation Research Part a-Policy and Practice,* 113**,** 213-226.
Machado-Leon, J. L., R. de Ona & J. de Ona (2016) The role of involvement in regards to public transit riders' perceptions of the service. *Transport Policy,* 48**,** 34-44.
Machado-Leon, J. L., R. de Ona, F. Diez-Mesa & J. de Ona (2018) Finding service quality improvement opportunities across different typologies of public transit customers. *Transportmetrica a-Transport Science,* 14**,** 761-783.
Menidjel, C., A. Benhabib, A. Bilgihan & M. Madanoglu (2019) Assessing the role of product category involvement and relationship proneness in the satisfaction–loyalty link in retailing. *International Journal of Retail and Distribution Management,* 48**,** 207-226.
Miller, J. A. 1976. Exploring some alternative measures of cunsumer satisfaction. In *Marketing: 1776-1976 and Beyond,* ed. K. L. Bernhardt, 661-664. Chicago, IL: American Marketing Association.
Minser, J. & V. Webb (2010) Quantifying the benefits application of customer loyalty modeling in public transportation context. *Transportation Research Record,* 2144**,** 111-120.
Mittal, B. (1995) A comparative analysis of 4 scales of consumer involvement. *Psychology & Marketing,* 12**,** 663-682.
Nguyen-Phuoc, D. Q., D. N. Su, P. T. K. Tran, D. T. T. Le & L. W. Johnson (2020) Factors influencing customer's loyalty towards ride-hailing taxi services - A case study of Vietnam. *Transportation Research Part a-Policy and Practice,* 134**,** 96-112.
Oliver, R. L. (1981) Measurement and evaluation of satisfaction processes in retail settings. *Journal of Retailing,* 57**,** 25-48.
--- (1999) Whence consumer loyalty? *Journal of Marketing,* 63**,** 33-44.
---. 2010. *Satisfaction. A behavioral perspective on the consumer*. Armonk, New York: JW. E. Sharpe.
Olsen, S. O. (2007) Repurchase loyalty: The role of involvement and satisfaction. *Psychology & Marketing,* 24**,** 315-341.
Parasuraman, A., V. A. Zeithaml & L. L. Berry (1985) A conceptual model of service quality and its implications for future research. *Journal of Marketing,* 49**,** 41-50.
Park, J. W., R. Robertson & C. L. Wu (2004) The effect of airline service quality on passengers' behavioural intentions: a Korean case study. *Journal of Air Transport Management,* 10**,** 435-439.
Rajaguru, R. (2016) Role of value for money and service quality on behavioural intention: A study of full service and low cost airlines. *Journal of Air Transport Management,* 53**,** 114-122.

Shah, R. & S. M. Goldstein (2006) Use of structural equation modeling in operations management research: Looking back and forward. *Journal of Operations Management,* 24**,** 148-169.
Simsekoglu, O., T. Nordfjaern & T. Rundmo (2015) The role of attitudes, transport priorities, and car use habit for travel mode use and intentions to use public transportation in an urban Norwegian public. *Transport Policy,* 42**,** 113-120.
Stevens, J. P. 2009. *Applied Multivariate Statistics for the Social Sciences*. New York: Routledge.
Sun, S. C. & Z. Y. Duan (2019) Modeling passengers' loyalty to public transit in a two-dimensional framework: A case study in Xiamen, China. *Transportation Research Part a-Policy and Practice,* 124**,** 295-309.
UN. 2019. *World Urbanization Prospect. The 2018 Revision (ST/ESA/SER.A/420)*. United Nations, Department of Economic and Social Affairs, Population Division. New York: United Nations.
van Lierop, D., M. G. Badami & A. M. El-Geneidy (2018) What influences satisfaction and loyalty in public transport? A review of the literature. *Transport Reviews,* 38**,** 52-72.
van Lierop, D. & A. El-Geneidy (2018) Is having a positive image of public transit associated with travel satisfaction and continued transit usage? An exploratory study of bus transit. *Public Transport,* 10**,** 241-256.
Wei, C. H. & C. Y. Kao (2010) Measuring traveler involvement in urban public transport services: The case of Kaohsiung. *Transport Policy,* 17**,** 444-453.
Yuan, Y. L., M. Yang, J. X. Wu, S. Rasouli & D. Lei (2019) Assessing bus transit service from the perspective of elderly passengers in Harbin, China. *International Journal of Sustainable Transportation,* 13**,** 761-776.
Zaichkowsky, J. L. (1985) Measuring the involvement construct. *Journal of Consumer Research,* 12**,** 341-352.
Zeithaml, V. A., L. L. Berry & A. Parasuraman (1996) The behavioral consequences of service quality. *Journal of Marketing,* 60**,** 31-46.
Zhang, C. Q., Y. Liu, W. T. Lu & G. N. Xiao (2019) Evaluating passenger satisfaction index based on PLS-SEM model: Evidence from Chinese public transport service. *Transportation Research Part a-Policy and Practice,* 120**,** 149-164.
Zhao, J. H., V. Webb & P. Shah (2014) Customer loyalty differences between captive and choice transit riders. *Transportation Research Record,* 2415**,** 80-88.

Table 1.- Average values for survey service quality, satisfaction, involvement and behavioral intentions indicators

Construct	Item	Madrid	Rome	Berlin	Lisbon	London	All
Service quality (SQ)	q1. Service hours	3.63	2.88	3.89	3.07	4.11	3.51
	q2. Proximity	3.91	3.56	4.09	3.65	3.93	3.83
	q3. Frequency	3.47	2.77	3.83	2.95	3.96	3.39
	q4. Punctuality	3.33	2.51	3.21	2.60	3.64	3.05
	q5. Speed	3.69	3.02	3.72	3.29	3.74	3.49
	q6. Cost	3.42	3.35	3.37	3.49	3.58	3.44
	q7. Accessibility	3.80	3.07	3.84	3.54	3.87	3.62
	q8. Intermodality	3.84	2.90	3.81	3.52	3.95	3.61
	q9. Individual space	3.18	2.71	3.10	2.79	3.29	3.01
	q10. Temperature	3.37	2.76	3.28	2.78	3.38	3.11
	q11. Cleanliness	3.49	2.50	3.18	3.03	3.47	3.14
	q12. Safety	3.84	3.14	3.96	3.38	3.96	3.66
	q13. Security	3.16	2.77	3.45	3.03	3.62	3.20
	q14. Information	3.55	2.95	3.47	3.07	3.84	3.37
Satisfaction (SA)	s1. General satisfaction	3.65	2.62	3.76	3.15	3.81	3.40
	s2. Expectations	3.60	2.57	3.60	2.98	3.82	3.31
	s3. Needs	3.65	2.47	3.48	3.01	3.78	3.28
	s4. Global experience	3.52	2.46	3.58	3.05	3.68	3.26
Involvement (INV)	a1. Low income	1.93	2.58	2.05	1.69	2.32	2.11
	a2. Freedom	3.96	3.11	3.88	3.82	4.07	3.77
	a3. Save time and money	3.57	3.24	3.57	3.44	3.46	3.46
	a4. Lifestyle	3.84	3.44	3.84	3.62	3.88	3.72
	a5. Environment	4.27	3.84	4.07	4.33	4.00	4.11
	a6. Reduce traffic	4.14	3.81	4.00	4.18	3.97	4.02
	a7. Recommendation	3.28	2.79	3.08	2.96	3.36	3.09
	a8. Judgement	2.74	2.80	2.59	2.36	3.00	2.69
Behavioral intentions (BI)	b1. I will use PT for one-off trips	4.06	3.73	3.92	3.73	3.99	3.89
	b2. I will use PT for regular trips	4.48	4.15	4.34	4.45	4.36	4.36
	b3. I will increase PT usage	3.72	3.59	3.34	3.54	3.72	3.58
	b4. I will recommend PT	3.82	3.33	3.40	3.59	3.79	3.59

Table 2.- Structural paths and models' fit statistics for competing structural regression models

Path / Fit index	FIFS (a) Unst.	SE	St.	FIPS (c) Unst.	SE	St.	PIPS (e) Unst.	SE	St.
Structural paths									
Service quality (SQ) -> Satisfaction (SA)	1.093	0.026	0.885	1.090	0.026	0.881	1.089	0.026	0.881
Satisfaction (SA) -> Involvement (INV)	0.612	0.019	0.774	0.440	0.040	0.557	0.418	0.041	0.532
Service quality (SQ) -> Involvement (INV)	-	-	-	0.232	0.049	0.238	0.268	0.050	0.276
Satisfaction (SA) -> Behavioral intentions (BI)	-	-	-	-	-	-	**0.032**	0.041	0.046
Involvement (INV) -> Behavioral intentions (BI)	0.678	0.034	0.770	0.675	0.034	0.768	0.800	0.051	0.899
Service quality (SQ) -> Behavioral intentions (BI)	-	-	-	-	-	-	*-0.171*	0.051	-0.198
Model's fit statistics (Satorra-Bentler estimation)									
df (sb)	316			315			313		
chi-square (sb)	1,527.14			1,506.11			1,487.87		
p-value	0.000			0.000			0.000		
RMSEA (sb)	0.043			0.043			0.043		
CFI (sb)	0.957			0.958			0.958		
TLI (sb)	0.952			0.953			0.953		
SRMR	0.041			0.039			0.038		
AIC	138,841.81			138,816.93			138,798.72		
BIC	139,341.74			139,322.48			139,315.51		
R^2 (Behavioral intentions)	0.593			0.589			0.632		

Path / Fit index	FSFI (b) Unst.	SE	St.	FSPI (d) Unst.	SE	St.	PSPI (f) Unst.	SE	St.
Structural paths									
Service quality (SQ) -> Satisfaction (SA)	-	-	-	0.815	0.035	0.660	0.843	0.037	0.682
Involvement (INV) -> Satisfaction (SA)	1.223	0.031	0.926	0.366	0.032	0.298	0.341	0.037	0.268
Service quality (SQ) -> Involvement (INV)	0.840	0.026	0.897	0.749	0.026	0.744	0.723	0.026	0.745
Satisfaction (SA) -> Behavioral intentions (BI)	0.400	0.024	0.589	0.391	0.025	0.577	**0.032**	0.041	0.046
Involvement (INV) -> Behavioral intentions (BI)	-	-	-	-	-	-	0.800	0.051	0.899
Service quality (SQ) -> Behavioral intentions (BI)	-	-	-	-	-	-	*-0.171*	0.051	-0.198
Model's fit statistics (Satorra-Bentler estimation)									
df (sb)	316			315			313		
chi-square (sb)	2,273.44			1,854.68			1,487.87		
p-value	0.000			0.000			0.000		
RMSEA (sb)	0.055			0.049			0.043		
CFI (sb)	0.931			0.945			0.958		
TLI (sb)	0.923			0.939			0.953		
SRMR	0.054			0.052			0.038		
AIC	139,780.25			139,255.66			138,798.72		
BIC	140,280.18			139,761.22			139,315.51		
R^2 (Behavioral intentions)	0.346			0.333			0.632		

Path / Fit index	AIFS (g) Unst.	SE	St.	AIPS (h) Unst.	SE	St.
Structural paths						
Involvement (INV) -> Service quality (SQ)	0.754	0.028	0.735	0.768	0.028	0.745
Service quality (SQ) -> Satisfaction (SA)	0.837	0.036	0.676	0.843	0.037	0.682
Involvement (INV) -> Satisfaction (SA)	0.355	0.035	0.279	0.341	0.037	0.268
Service quality (SQ) -> Behavioral intentions (BI)	-	-	-	*-0.171*	0.051	-0.198
Satisfaction (SA) -> Behavioral intentions (BI)	*-0.075*	0.028	-0.108	**0.032**	0.041	0.046
Involvement (INV) -> Behavioral intentions (BI)	0.767	0.049	0.866	0.800	0.051	0.899
Model's fit statistics (Satorra-Bentler estimation)						
df (sb)	314			313		
chi-square (sb)	1,499.67			1,487.87		
p-value	0.000			0.000		
RMSEA (sb)	0.043			0.043		
CFI (sb)	0.958			0.958		
TLI (sb)	0.953			0.953		

SRMR	0.039	0.038
AIC	138,811.26	138,798.72
BIC	139,322.43	139,315.51
R^2 (Behavioral intentions)	0.616	0.632

All unstandardized estimates are statistically significant at $p < 0.01$, except the values in **bold**. A dashed line (-) indicates that the path is not specified in that model

Table 3.- Structural paths and fit statistics for the five independent samples

	Madrid			Rome			Berlin			Lisbon			London		
Path / Fit index	**Unst.**	**SE**	**St.**	**Unst.**	**SE**	**St.**	**Unst.**	**SE**	**St.**	**Unst.**	**SE**	**St.**	**Unst.**	**SE**	**St.**
FIFS (a)															
Service quality (SQ) -> Satisfaction (SA)	1.089	0.061	0.875	1.048	0.045	0.898	1.070	0.098	0.825	1.146	0.077	0.844	1.388	0.132	0.853
Satisfaction (SA) -> Involvement (INV)	0.621	0.042	0.786	0.767	0.032	0.844	0.673	0.061	0.730	0.454	0.042	0.672	0.594	0.056	0.806
Involvement (INV) -> Behavioral intentions (BI)	0.630	0.089	0.760	0.602	0.050	0.805	0.721	0.080	0.691	1.064	0.099	0.835	0.977	0.107	0.889
df (sb)		316			316			316			316			316	
chi-square (sb)		552.35			626.62			635.60			622.04			544.48	
p-value		0.000			0.000			0.000			0.000			0.000	
RMSEA (sb)		0.042			0.048			0.052			0.047			0.044	
CFI (sb)		0.958			0.960			0.925			0.940			0.943	
TLI (sb)		0.954			0.956			0.917			0.933			0.937	
SRMR		0.043			0.045			0.059			0.055			0.052	
AIC		27,742.42			29,233.27			25,238.95			30,108.13			24,772.76	
BIC		28,101.37			29,593.91			25,588.68			30,470.43			25,123.90	
R^2 (Behavioral intentions)		0.577			0.648			0.477			0.698			0.791	
FIPS (c)															
Service quality (SQ) -> Satisfaction (SA)	1.087	0.061	0.873	1.044	0.045	0.894	1.052	0.097	0.811	1.146	0.077	0.842	1.382	0.130	0.835
Satisfaction (SA) -> Involvement (INV)	0.541	0.082	0.685	0.566	0.067	0.623	0.393	0.079	0.427	0.388	0.074	0.573	0.321	0.084	0.439
Service quality (SQ) -> Involvement (INV)	**0.109**	0.099	0.111	0.253	0.079	0.239	0.414	0.112	0.346	**0.103**	0.099	0.112	0.496	0.159	0.410
Involvement (INV) -> Behavioral intentions (BI)	0.630	0.089	0.759	0.601	0.050	0.803	0.710	0.082	0.683	1.057	0.099	0.832	0.979	0.108	0.891
df (sb)		315			315			315			315			315	
chi-square (sb)		551.29			619.70			624.22			621.35			530.54	
p-value		0.000			0.000			0.000			0.000			0.000	
RMSEA (sb)		0.042			0.048			0.051			0.047			0.042	
CFI (sb)		0.958			0.961			0.928			0.940			0.947	
TLI (sb)		0.953			0.957			0.920			0.933			0.941	
SRMR		0.043			0.044			0.055			0.055			0.047	
AIC		27,743.11			29,227.66			25,226.90			30,108.87			24,754.79	
BIC		28,106.09			29,592.35			25,580.56			30,475.23			25,109.88	
R^2 (Behavioral intentions)		0.576			0.645			0.466			0.693			0.794	

All unstandardized estimates are statistically significant at $p < 0.01$, except the values in **bold**.

Table 4.- SEM-MIMIC results: structural paths, statistically significant regressors' influence and selected fit statistics

	All			Madrid			Rome			Berlin			Lisbon			London		
Path / Regressor / Fit index	**Unst.**	**SE**	**Std.**	**Unst.**	**SE**	**Std.**	**Unst.**	**SE**	**Std.**	**Unst.**	**SE**	**Std.**	**Unst.**	**SE**	**Std.**	**Unst.**	**SE**	**Std.**
Structural Paths																		
Service quality (SQ) -> Satisfaction (SA)	**1.047**	0.030	0.840	**1.052**	0.072	0.839	**1.047**	0.052	0.885	**1.061**	0.106	0.827	**1.128**	0.083	0.824	**1.414**	0.139	0.843
Satisfaction (SA) -> Involvement (INV)	**0.658**	0.021	0.825	**0.611**	0.049	0.780	**0.749**	0.036	0.837	**0.745**	0.062	0.754	**0.490**	0.044	0.680	**0.585**	0.058	0.804
Involvement (INV) -> Behavioral intentions (BI)	**0.727**	0.035	0.803	**0.685**	0.083	0.761	**0.613**	0.054	0.813	**0.739**	0.086	0.684	**0.953**	0.092	0.828	**0.976**	0.109	0.891
Service quality (SQ)																		
City	0.098	0.041	0.056	***0.246***	0.079	0.157	0.249	0.113	0.111	0.101	0.083	0.071	-0.169	0.084	-0.104	0.050	0.062	0.046
Male	***0.105***	0.039	0.060	0.183	0.084	0.118	0.150	0.105	0.071	0.042	0.071	0.035	0.104	0.079	0.069	0.056	0.056	0.054
Old	-0.063	0.039	-0.037	0.065	0.082	0.042	**-0.398**	0.103	-0.190	0.136	0.069	0.111	-0.045	0.078	-0.030	0.028	0.056	0.028
Frequent	-0.019	0.093	-0.004	-0.045	0.289	-0.010	***-0.583***	0.216	-0.093	0.319	0.177	0.102	0.097	0.191	0.030	-0.137	0.111	-0.057
High income	**0.238**	0.047	0.123	0.122	0.091	0.076	**0.677**	0.112	0.303	0.101	0.147	0.051	0.070	0.090	0.042	0.071	0.064	0.065
Rome	**-0.712**	0.068	-0.329															
Lisbon	**-0.540**	0.059	-0.255															
London	0.112	0.056	0.051															
Satisfaction (SA)																		
Male	***0.086***	0.029	0.039	0.095	0.062	0.049	0.024	0.062	0.009	0.133	0.055	0.085	0.105	0.062	0.051	0.093	0.057	0.054
University	-0.068	0.031	-0.031	-0.057	0.059	-0.029	-0.017	0.060	-0.007	-0.065	0.062	-0.041	-0.151	0.070	-0.074	0.001	0.070	0.001
Rome	**-0.321**	0.048	-0.119															
Involvement (INV)																		
Male	***-0.083***	0.031	-0.048	-0.101	0.059	-0.066	0.080	0.069	0.036	***-0.211***	0.072	-0.136	-0.116	0.058	-0.078	0.003	0.049	0.002
Old	0.080	0.032	0.046	0.049	0.061	0.033	0.172	0.069	0.078	0.008	0.071	0.005	0.104	0.067	0.070	0.010	0.051	0.008
Frequent	0.173	0.079	0.039	0.035	0.211	0.007	0.093	0.194	0.014	0.047	0.189	0.012	**0.532**	0.135	0.165	-0.099	0.129	-0.034
University	***0.091***	0.033	0.053	0.107	0.062	0.071	-0.017	0.067	-0.008	0.133	0.067	0.085	0.071	0.069	0.048	0.084	0.060	0.058
Dependent	0.000	0.038	0.000	-0.123	0.069	-0.071	0.079	0.070	0.035	-0.038	0.107	-0.016	-0.029	0.068	-0.018	***0.194***	0.064	0.116
High income	0.046	0.035	0.024	0.020	0.067	0.013	0.164	0.078	0.069	-0.099	0.122	-0.040	0.077	0.069	0.046	0.034	0.053	0.026
Rome	0.110	0.050	0.051															
Lisbon	**0.216**	0.049	0.102															
London	-0.112	0.046	-0.052															
Behavioral intentions (BI)																		
Old	0.078	0.031	0.050	***0.195***	0.063	0.143	0.117	0.069	0.070	-0.051	0.078	-0.030	***0.198***	0.069	0.115	-0.026	0.058	-0.019
University	-0.056	0.033	-0.036	-0.005	0.063	-0.004	-0.064	0.068	-0.038	0.029	0.073	0.017	-0.165	0.074	-0.097	-0.067	0.072	-0.042
High income	0.058	0.033	0.034	-0.002	0.064	-0.002	0.109	0.071	0.061	0.221	0.100	0.082	0.018	0.069	0.010	0.061	0.060	0.042
Rome	**0.179**	0.047	0.092															
Berlin	**-0.249**	0.051	-0.124															
Model's fit statistics (Satorra-Bentler estimation)																		
df (sb)		569			477			477			477			477			477	
chi-square (sb)		2421.98			728.20			805.44			841.51			852.97			821.68	
p-value		0.000			0.000			0.000			0.000			0.000			0.000	
RMSEA (sb)		0.043			0.038			0.044			0.048			0.046			0.046	

CFI (sb)	0.934	0.949	0.954	0.906	0.924	0.919
TLI (sb)	0.924	0.942	0.948	0.893	0.914	0.909
SRMR	0.039	0.043	0.042	0.055	0.053	0.052
AIC	139,587.9	26,717.2	27,057.3	24,272.5	29,278.6	25,267.4
BIC	140,739.7	27,309.2	27,647.1	24,850.4	29,877.1	140,739.7
R^2 (Behavioral intentions)	0.642	0.625	0.684	0.478	0.716	0.804

Unstandardized values in **bold** (p<0.001); in ***bold and italics*** (p<0.01); and underlined (p<0.05). All other values are not statistically significant.

Table 5.- SEM-MIMIC results: total effects between factors

	All T. Effect	Madrid T. Effect	Rome T. Effect	Berlin T. Effect	Lisbon T. Effect	London T. Effect
SQ -> SA	1.047	1.052	1.047	1.061	1.128	1.414
SQ -> SA -> INV	0.689	0.643	0.784	0.790	0.553	0.828
SA -> INV	0.658	0.611	0.749	0.745	0.490	0.585
SQ -> SA -> INV -> BI	0.501	0.440	0.481	0.584	0.527	0.808
SA -> INV -> BI	0.479	0.418	0.459	0.550	0.467	0.571
INV -> BI	0.727	0.685	0.613	0.739	0.953	0.976

Service quality (SQ); Satisfaction (SA); Involvement (INV); Behavioral intentions (BI)

Table A1.- Basic data per city (Madrid, Rome, Berlin, Lisbon and London)

	Madrid*	Rome**	Berlin**	Lisbon**	London**
Definition of analysis area	Madrid Metropolitan Area	Metropolitan City of Rome	Greater Berlin	Lisbon Metropolitan Area	Greater London
Analysis area (km^2)	8,030	5,363	892	3,015	1,569
Population (inhab)	6.51M (2017)	4.34M (2019)	3.52 (2015)	2.89M (2014)	8.90 (2018)
Population density (inhab/km^2)	811	812	3,948	957	5,672
Public transport options	Rail, metro, tram, bus, bike	Rail, metro, tram, bus	Rail, metro, tram, bus, bike	Rail, metro, tram, bus, ferry, bike	Rail, metro, tram, bus, ferry, bike

Sources: * Observatorio de la movilidad metropolitana (www.observatoriomovilidad.es); ** Deloitte City Mobility Index 2018 (Berlin) and 2020 (Lisbon, London and Rome) (www2.deloitte.com/xe/en/insights/focus/future-of-mobility/deloitte-urban-mobility-index-for-cities.html)

Table A2.- Sociodemographic characteristics and mobility patterns

Category	Group	Madrid	Rome	Berlin	Lisbon	London	All
Sample (n)	n	525	509	508	530	507	2,579
Gender	Male	211	207	228	230	209	1,085
	Female	314	302	280	300	298	1,494
Age (years)	18-24	51	67	54	88	36	296
	25-44	230	219	197	228	222	1,096
	45-64	185	205	189	168	212	959
	65 or more	59	18	68	46	37	228
Time living	< 1 year	25	22	18	23	29	117
in the area	A few years	157	183	228	215	194	977
	All my life	342	302	261	288	283	1,476
Household location	Metropolitan area	314	353	113	369	341	1,490
	City center	211	156	395	161	166	1,089
Education level	Primary schools or less	36	22	25	41	25	149
	Secondary schools	217	241	285	212	128	1,083
	Higher education	269	242	193	274	353	1,331
Occupation status	Professional	39	106	44	65	50	304
	Employed	350	236	270	305	344	1,505
	Student	43	58	35	59	16	211
	Retired/Pensioner	52	21	99	48	37	257
	Other	39	85	55	51	59	289
Family size	1	46	54	203	57	85	445
	2	163	118	183	164	147	775
	3	144	141	61	144	100	590
	4	130	144	46	116	106	542
	5 or more	39	43	11	38	52	183
Dependent members	No	386	320	424	364	398	1,892
in the family	Yes	135	174	56	150	91	606
Household net	<2 minimum wages	144	207	324	205	209	1,089
income	2-3 minimum wages	149	109	83	136	105	582
	>3 minimum wages	162	129	49	121	145	606
Frequency of use	0-2	51	61	52	60	65	289
(days/week)	3-4	97	133	113	101	173	617
	5-7	377	315	343	369	269	1,673

Table A3.- Description of indicators and constructs

Construct	Indicators	Description
Service quality (SQ)	q1. Service hours	Service hours
	q2. Proximity	Proximity of stops to starting point or destination of the trip
	q3. Frequency	Frequency or number of daily services
	q4. Punctuality	Punctuality
	q5. Speed	Speed
	q6. Cost	Cost
	q7. Accessibility	Ease of entrance and exit from the vehicle and/or stations
	q8. Intermodality	Ease of transfers/good connections with other modes of transport
	q9. Individual space	Individual space available inside the vehicle
	q10. Temperature	Temperature inside the vehicle
	q11. Cleanliness	Cleanliness of the vehicle and stations
	q12. Safety	Safety on board (regarding accidents)
	q13. Security	Safety regarding robbery and violence
	q14. Information	Information provided
Satisfaction (SA)	s1. General satisfaction	In general, I am satisfied with the public transport service provided in XYZ
	s2. Expectations	The public transport service in XYZ meets my expectations
	s3. Needs	With the existing modes of transport in XYZ, I consider that the commuting needs of inhabitants are well covered
	s4. Global experience	When I take public transport in XYZ, I feel very satisfied
Involvement (INV)	a1. Low income	Public transport is only for citizens with low income
	a2. Freedom	Public transport gives me the freedom to move around Madrid easily
	a3. Save time and money	Although it is an effort for me to use public transport, I am rewarded because I save time and money
	a4. Lifestyle	I feel that using public transport is in line with my lifestyle
	a5. Environment	When using public transport, I am helping towards improving the environment
	a6. Reduce traffic	I feel that by travelling on public transport I am helping to reduce problems derived from traffic (in other words, traffic jams, noise, pollution, etc.)
	a7. Recommendation	The people that are most important to me recommend that I use public transport
	a8. Judgement	I think that by using public transport I can improve the way that relatives and friends judge me
Behavioral intentions (BI)	b1. I will use PT for one-off trips	In the next few weeks I will take public transport for one-off trips.
	b2. I will use PT for regular trips	In the next few weeks I will take public transport for my regular trips.
	b3. I will increase PT usage	I am sure I will increase the number of times I use public transport in the future.
	b4. I will recommend PT	Not only will I use public transport, but I will also recommend it to friends and family.

XYZ was changed to Madrid, Rome, Berlin, Lisbon or London

Table C1.- Initial CFA measurement model

	Unst.	SE	St.
SQ->Service hours (q1)	1.000	*	0.752
SQ->Proximity (q2)	0.741	0.027	0.632
SQ->Frequency (q3)	1.072	0.024	0.786
SQ->Punctuality (q4)	1.113	0.026	0.788
SQ->Speed (q5)	0.945	0.025	0.772
SQ->Cost (q6)	0.657	0.029	0.501
SQ->Accessibility (q7)	0.848	0.025	0.699
SQ->Intermodality (q8)	0.956	0.024	0.765
SQ->Individual space (q9)	0.955	0.028	0.716
SQ->Temperature (q10)	0.917	0.027	0.702
SQ->Cleanliness (q11)	0.955	0.028	0.714
SQ->Safety (q12)	0.840	0.025	0.692
SQ->Security (q13)	0.916	0.028	0.697
SQ->Information (q14)	0.932	0.026	0.721
SA->General satisfaction (s1)	1.000	*	0.914
SA->Meet expectations (s2)	1.002	0.013	0.909
SA->Covered needs (s3)	0.995	0.014	0.870
SA->I feel satisfied (s4)	0.990	0.013	0.901
INV->Freedom (a2)	1.000	*	0.757
INV->Save time & money (a3)	0.844	0.031	0.623
INV->Lifestyle (a4)	0.890	0.029	0.704
INV->Environment (a5)	0.766	0.030	0.625
INV->Reduce traffic (a6)	0.785	0.031	0.611
INV->Recommendation (a7)	0.941	0.031	0.641
INV->Judgement (a8)	0.837	0.034	0.531
BI->I will use PT for one-off trips (b1)	1.000	*	0.453
BI->I will use PT for regular trips (b2)	0.674	0.053	0.379
BI->Increase usage (b3)	1.488	0.082	0.690
BI->I will recommend PT (b4)	1.926	0.108	0.829
cov(SQ,SA)	0.857	0.032	0.879
cov(SQ,INV)	0.539	0.028	0.716
cov(SQ,BI)	0.250	0.019	0.525
cov(SA,INV)	0.694	0.031	0.744
cov(SA,BI)	0.342	0.025	0.580
cov(INV,BI)	0.376	0.025	0.827
df (sb)	371		
chi-square (sb)	2872.43		
p-value	0.000		
RMSEA (sb)	0.058		
CFI (sb)	0.914		
TLI (sb)	0.906		
SRMR	0.050		
AIC	149,731.9		
BIC	150,253.3		

* Not tested for statistical significance. All other unstandardized estimates are statistically significant at $p < 0.001$.

Table C2.- Final CFA measurement model

	Unst.	SE	St.
SQ->Service hours (q1)	1.000	*	0.753
SQ->Proximity (q2)	0.744	0.027	0.635
SQ->Frequency (q3)	1.075	0.024	0.790
SQ->Punctuality (q4)	1.112	0.027	0.790
SQ->Speed (q5)	0.943	0.026	0.772
SQ->Cost (q6)	0.649	0.029	0.497
SQ->Accessibility (q7)	0.846	0.025	0.700
SQ->Intermodality (q8)	0.959	0.024	0.770
SQ->Individual space (q9)	0.939	0.028	0.704
SQ->Temperature (q10)	0.887	0.027	0.682
SQ->Cleanliness (q11)	0.936	0.028	0.701
SQ->Safety (q12)	0.828	0.026	0.683
SQ->Security (q13)	0.896	0.028	0.682
SQ->Information (q14)	0.923	0.027	0.717
SA->General satisfaction (s1)	1.000	*	0.914
SA->Meet expectations (s2)	1.002	0.013	0.908
SA->Covered needs (s3)	0.997	0.014	0.869
SA->I feel satisfied (s4)	0.991	0.013	0.900
INV->Freedom (a2)	1.000	*	0.771
INV->Save time & money (a3)	0.824	0.031	0.618
INV->Lifestyle (a4)	0.869	0.029	0.699
INV->Environment (a5)	0.692	0.030	0.574
INV->Reduce traffic (a6)	0.704	0.031	0.558
INV->Recommendation (a7)	0.905	0.031	0.627
INV->Judgement (a8)	0.792	0.035	0.510
BI->Increase usage (b3)	1.000	*	0.663
BI->I will recommend PT (b4)	1.448	0.058	0.891
cov(e.q9,e.q10)	0.165	0.017	0.233
cov(e.q10,e.q11)	0.147	0.019	0.205
cov(e.q12,e.q13)	0.152	0.019	0.228
cov(e.a5,e.a6)	0.418	0.024	0.543
cov(e.a7,e.a8)	0.405	0.030	0.361
cov(SQ,SA)	0.860	0.033	0.881
cov(SQ,INV)	0.571	0.029	0.745
cov(SQ,BI)	0.349	0.023	0.513
cov(SA,INV)	0.735	0.032	0.776
cov(SA,BI)	0.479	0.029	0.569
cov(INV,BI)	0.521	0.027	0.787
df (sb)	313		
chi-square (sb)	1487.87		
p-value	0.000		
RMSEA (sb)	0.043		
CFI (sb)	0.958		
TLI (sb)	0.953		
SRMR	0.038		
AIC	138,798.7		
BIC	139,315.5		

* Not tested for statistical significance. All other unstandardized estimates are statistically significant at $p < 0.001$.