Service quality, satisfaction and behavioral intentions towards public transport from the point of view of private vehicle users

By: Juan de Oña

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

Juan de Oña (2022) *Service quality, satisfaction and behavioral intentions towards public transport from the point of view of private vehicle users*. **Transportation,** DOI: 10.1007/s11116-021-10175-7

Direct access to the published version: http://dx.doi.org/10.1007/s11116-021-10175-7

Service quality, satisfaction and behavioral intentions towards public transport from the point of view of private vehicle users

Abstract:
In order to attract car users towards the public transport services in an urban and metropolitan context, contributing to a sustainable mobility in cities, it is fundamental to improve our knowledge of service quality perceptions, satisfaction and behavioral intentions toward transit from the point of view of private transport users. This paper is based on the data from a single survey —carried out in two European cities (Madrid and Lisbon) — of regular private vehicle users that use public transport at least occasionally. The questionnaire gathers information about 14 attributes of service quality, four indicators for satisfaction and four indicators for behavioral intentions; as well as several sociodemographic variables that are used in the models (household location, gender, age, education, dependent members in the family and income). The study uses confirmatory factor analysis (CFA) to identify the most important service quality attributes for the car users; structural equation modeling (SEM) for investigating the relationships among the three factors; and multi-group analysis (MGA) and a multiple-indicator and multiple-causes (MIMIC) approach to identify heterogeneity in the models because of geographical context or sociodemographic characteristics. Regular private vehicle users in both cities agree that punctuality, frequency, information and intermodality are among the five most important service quality attributes. Residents in Madrid also emphasize speed, while service hours would be a priority in Lisbon. The models for both cities agree on a complete mediator role of satisfaction between service quality and behavioral intentions. The MGA and MIMIC approaches show that the models do not present important differences tied to the sociodemographic characteristics, although differences are identified between Madrid and Lisbon. The MIMIC approach identified differences associated with city, household location and education for the pooled data; while household location, age and education were significant in Lisbon.

Keywords: public transport, potential user, non-user, full mediator, SEM-MIMIC, loyalty

Highlights:

- Cost, safety and security are the three least important public transport attributes for private vehicle users in Madrid and Lisbon
- Punctuality, frequency, information and intermodality are among the five most important attributes; speed should be added in Madrid, and service hours Lisbon
- Satisfaction exerts a complete mediator role between service quality and behavioral intentions for private transport users
- Madrid and Lisbon present significant differences regarding seven service quality attributes and two behavioral intentions indicators
- Sociodemographic characteristics are not significant in Madrid; but household location, age and education are significant in Lisbon

1.- Introduction

The sustainability of cities calls for rational and efficient use of means of transport, giving priority to the use of modes that generate fewer negative externalities on the urban environment. If distances traveled are short, non-motorized means should be encouraged or facilitated (e.g., walking or bicycle). When distances are great, as is frequently the case in large metropolitan areas, it is necessary to resort to motorized transport models. Many studies have made manifest that public transport is the most sustainable mode of transport in urban and metropolitan settings. Therefore, there is widespread consensus surrounding interventions to foment its use, to achieve a modal switch from the private vehicle.

Various studies have shown that improving the service quality perceptions and satisfaction towards public transport serves not only to retain existing users, but could also attract new users to the system (Bamberg, Rolle and Weber 2003, Beirao and Cabral 2007). In recent years, a number of articles have analyzed the influence of service quality perceptions and satisfaction over attitudes towards public transport and behavioral intentions (de Oña, Machado and de Oña 2015, de Oña et al. 2016, de Oña, Estevez and de Oña 2020, de Oña 2020, de Oña 2021, Machado-Leon, de Oña and de Oña 2016, Machado-Leon et al. 2018, Lai and Chen 2011). Yet most of these contributions look at these three factors and their relationships with a focus on the users of public transport, while the perception of private vehicle users and their satisfaction and behavioral intentions toward public transport are rarely analyzed (Abenoza, Cats and Susilo 2017, de Oña et al. 2020, Pedersen, Kristensson and Friman 2012, Redman et al. 2013). To generate a modal change, attracting private vehicle users to the public transport system, the perspective of these users in terms of these three factors (service quality, satisfaction and behavioral intentions) is key.

The present study contributes to the corpus of literature in the field of public transport in more than one sense. Firstly, it identifies the main attributes of the public transport service that influence the perception of quality on the part of private vehicle users. Secondly, it confirms that for private vehicle users, satisfaction also has a mediator role between service quality and behavioral intentions, just as it does for public transport users. Thirdly, it attempts to verify whether it is possible to generalize the relationships identified, or else there are differences tied to location of the respondent (country and household), or sociodemographic characteristics (gender, age, education, dependent members in the family, or income level).

To this end, the study relies on a survey of declared regular private vehicle users in Madrid and in Lisbon, in which the users were queried about their perceptions of the quality of the public transport system in their city, their satisfaction with it, and their behavioral intentions. In order to be able to ask the respondents about their service quality perceptions and satisfaction with public transport, it is fundamental that they have previously used the public transport system. The survey also gathered information about certain sociodemographic features so as to analyze the heterogeneity that such characteristics might introduce in the service quality perceptions, satisfaction and behavioral intentions. To identify the attributes with the greatest influence on the perception of quality, as well as the relationships existing among the three factors considered, a structural equation modeling (SEM) approach was used. By using two independent samples (Madrid and Lisbon), the results could be replicated in order to validate the model, and detect differences insofar as the location of the respondent or sociodemographic characteristics, using multi-group analyses (MGA). Finally, to control for the heterogeneity that the above characteristics simultaneously introduce, a multiple indicator and multiple causes (MIMIC) approach was applied.

The rest of the paper is organized as follows. Section 2 reviews the relevant literature. Section 3 focuses on the data used for the analysis, the survey and the sample characteristics. Section 4 describes the different models applied in analysis. Section 5 presents and discusses the results, while Section 6 summarizes the most important conclusions and recommendations.

2.- Literature review

2.1.- Satisfaction as mediator between service quality and behavioral intentions in public transport

In the general literature there is a reasonable agreement that service quality is an antecedent of satisfaction, but no agreement about which type of mediating effect satisfaction would exert between service quality and behavioral intentions. With data from four industries, Cronin and Taylor (1992) investigated whether satisfaction had a mediator effect between service quality and behavioral intentions, or if service quality had a mediator effect between satisfaction and behavioral intentions. Their results suggested that service quality is an antecedent of satisfaction and that satisfaction exerts a mediator effect between service quality and behavioral intentions. Their results were not conclusive about the type of mediator effect, as they found partial and full mediator effects, depending on the industry. Gotlieb et al. (1994) compared two competing models, where satisfaction and service quality were changed in order, and concluded that service quality is an antecedent of satisfaction, and behavioral intentions is affected by satisfaction. In their study only the complete mediator effect was considered, just like Dabholkar et al. (2000), who also found satisfaction to play a full mediating role on the effect of service quality on behavioral intentions. Cronin et al. (2000) compared four more complex models using data from six service environments. Exploring the direct relationship between service quality and behavioral intentions, they found that quality had a direct effect on consumers' intentions in four of the six industries. In all cases, service quality was an antecedent of satisfaction. They also found partial and full mediating roles on the effect of service quality on behavioral intentions.

The growth of literature in the field of quality of service, satisfaction and behavioral intentions towards urban and metropolitan public transport reflect concern about gaining a better understanding of the attributes or variables affecting these three factors. Recently, several interesting articles have been published: de Oña and de Oña (2015) focusing on service quality and satisfaction; van Lierop et al. (2018) dedicated to satisfaction and behavioral intentions or loyalty; and de Oña (2021) trying to shed light on the type of mediator effect exerted by satisfaction between service quality and behavioral intentions in the public transport sector. However, all of them only considered the perspective of public transport users.

There are many papers that explore the relationship among those three factors using data from urban or metropolitan public transport users and structural equation modeling. A few papers investigate the relationships among only three factors (de Oña 2021, Shao et al. 2020, de Oña et al. 2015), while most of them (de Oña 2020, Machado-Leon et al. 2018, Fu, Zhang and Chan 2018, Irtema et al. 2018, Li et al. 2018, Fu and Juan 2017, de Oña et al. 2016, Machado-Leon et al. 2016, Lai and Chen 2011, Minser and Webb 2010) include in their analysis other factors that could also influence service quality, satisfaction and behavioral intentions (e.g., involvement, attitudes, expectations, perceived value, perceived costs, complaints, habit, corporate image, etc.). Despite the fact the relationships existing among these three factors has received considerable attention in the public transport literature, there is also a lack of consensus regarding these relationships. 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. Although studies defending the partial mediator role of satisfaction prevail (Machado-Leon et al. 2018, Fu et al. 2018, Irtema et al. 2018, Li et al. 2018, Fu and Juan 2017, de Oña et al. 2016, Lai and Chen 2011, Minser and Webb 2010), a good number point to the complete mediator role of satisfaction between service quality and behavioral intentions (de Oña 2021, de Oña 2020, Allen et al. 2020, Sun and Duan 2019, Yuan et al. 2019, Zhang et al. 2019).

2.2.- Perception of public transport from the perspective of car users

Still, none of the previous papers considers the perspective of non-public transport users (i.e., car users, private vehicle or private transport users). To date, no publication has analyzed the relationships among quality of service, satisfaction and behavioral intentions towards transit from the standpoint of private transport users using structural equation modeling. However, there are some authors that have used other approaches for trying to shed light into this topic. Recent articles attempting to identify the main attributes of service that could help attract users to public transport have tended to focus on regular car users (Al-Ayyash and Abou-Zeid 2019, Kang et al. 2019, Li et al. 2019, Abou-Zeid and Fujii 2016, Mahmoud and Hine 2016, Abou-Zeid and Ben-Akiva 2014, Redman et al. 2013, Abou-Zeid et al. 2012, Abou-Zeid and Ben-Akiva 2012, Pedersen et al. 2012, Pedersen, Friman and Kristensson 2011, Hine and Scott 2000), while others broaden the field of analysis to private vehicle users (i.e., car, motorcycle, bicycle, etc.) (Bamberg et al. 2003, Woods and Masthoff 2017, de Oña et al. 2020), and others do not distinguish between potential users or non-users (Pedersen et al. 2011, dell'Olio, Ibeas and Cecin 2011, Bellizzi et al. 2020, Krizek and El-Geneidy 2007).

Hine and Scott (2000) reported the findings from a series of focus groups and in-depth interviews of both public transport and car users about the decision that they made in order to either facilitate a public transport journey or to push the traveler to car use. The car users identified as the most important aspects of public transport: trip flexibility, journey time, frequent services, bus stop located nearby, regularity, and discomfort during wet and cold seasons. Bamberg et al. (2003) studied the influence of two specific attributes of public transport (information and cost) on the possibility of producing a modal switch to public transit in a scenario of change of residence. Their study showed that both attributes were significant for the modal switch. Redman et al. (2013) published a literature review paper aiming to identify the quality attributes of public transport that could attract car users. For this purpose, 74 studies that collected public transport improvements were classified in terms of how they affected the main public transport physical attributes identified previously in the literature (i.e., reliability, frequency, speed, accessibility, price, comfort and convenience). One important conclusion was that public transport improvements have the potential to attract private car users, but exactly what improvements should be made depend on the context and particularities of each targeted sample and individual motivations for using private vehicles.

Gathering data through telephone interviews, Pedersen et al. (2011) analyzed the differences reported by public transport users and non-users about overall satisfaction with public transport and perceptions on 17 specific attributes that were grouped into three factors (comfort, reliability and safety). Users were more satisfied than non-users with regard to reliability and safety, as well as with regard to overall satisfaction, while the differences in comfort were not significant. Kang et al. (2019) studied a sample of car users in Malaysia to identify the predictors of drivers' intentions to switch from car to public transport and their behavioral readiness to use public transport. Convenience, flexible service and commute impedance were used as latent factors. Results showed the significant associations between convenience, flexible service, commute impedance, and intention to switch; but only convenience and flexible service presented medium effect sizes on intention to switch, whereas commute impedance showed small effect size on intention to switch to public transport.

Several types of discrete choice modeling approaches have been used previously to identify the main attributes that contribute to overall service quality or satisfaction. With data from a stated preference (SP) survey in Spain and using multinomial logit models, dell'Olio et al. (2011) investigated the perception of users and potential users about six public transport attributes (waiting time, journey time, occupancy, cleanliness, driver's kindness and comfort). Whereas for the public transport users all six variables were significant, for the potential users only waiting time, journey time and occupancy proved significant. Al-Ayyash and Abou-Zeid (2019) explored the satisfaction with public transport in Beirut from the point of view of car users. They used ordinal logit models and found that only three public transport attributes (cost, bus travel time and shared taxis travel time) and one socioeconomic variable (number of vehicles in the family) were significant for explaining satisfaction. Logistic regression models were also utilized to analyze public transport travel intentions of people who commute daily by car in China (Li et al. 2019). They studied car users' intentions of commuting by public transport through five aspects (comfort, timeliness, reliability, economics, and safety). Their results revealed that the difference in service quality between public transport and car travel is insignificant for safety and convenience, while comfort, reliability, and economics had significant influence on public transport travel intentions. Using random parameters mixed logit models and data from a SP survey in Spain, Bellizzi et al. (2020) investigated transit and potential user preferences. They considered five public transport attributes (waiting time, journey time, access time, occupancy and fare) and several socioeconomic and travel characteristics (gender, age, income level, driving license, owning a car, frequency of journey and trip purpose). They found that potential users are more demanding of public transport than current users, and detected several differences concerning socioeconomic characteristics (e.g., among public transport users, women are more exigent than men; while among potential users, men would be willing to pay more than women for improving the service).

The above studies are characterized by considering a reduced number of attributes (from three to six) of public transport as influential on the perceptions of private vehicle users. Nonetheless, other studies analyze a number of attributes similar to those used to evaluate the perceptions of public transport users. With data from a bus service in the United Kingdom and using binary logistic regression models, Mahmoud and Hine (2016) quantified the relationships between the perceptions of 29 indicators and the overall service quality. Eleven significant indicators were reported to have significant influence on the perception of users. They compared public transport users and regular car users who used public transport occasionally, and found that both groups had very similar service quality perceptions. Recently, de Oña et al. (2020) investigated the main attributes that influence the perception of regular private vehicle users about the public transport services in the capital of Spain. They used ordinal logit models to identify the contribution of 14 service quality attributes to overall satisfaction from the point of view of regular private vehicle users. To achieve a comprehensive analysis and to deal with heterogeneity in perceptions, they developed models for the entire sample and for 14 user segments. Their results indicated that the most important public transport attributes for private vehicle users were frequency, speed, and intermodality. This study followed a methodological approach very similar to the study of Abenoza et al. (2017), which aspired to explain travel satisfaction with public transport based on 12 service quality attributes, comparing five different groups of users (inactive travelers, long distance commuters, urban motorist commuters, rural motorist commuters and students). They found that the most important attributes for all the groups were customer interface, operation, network and length of trip time. In this study all five segments included public transport users.

A comparison of the above studies yields additional conclusions: while in the study by Abenoza et al. (2017) all the 12 service quality attributes exerted a significant effect upon satisfaction for all the groups, in the study by de Oña et al. (2020), which analyzed the perception of regular private vehicle users, only a limited number of attributes (from two to six) were identified as significant in each of the segments. Thus, previous results appear to indicate that the full board of attributes would be significant for public transport users while regular private vehicle users identify only a reduced number of important variables.

Previous results (less overall satisfaction and a narrow range of public transport attributes considered by private vehicle users or non-users) also agree with the studies of Pedersen et al. (2011, 2012) when analyzing car users. According to these authors, car users might underestimate their satisfaction with public transport due to a focusing illusion, which implies that car users pay attention to a very narrow range of aspects related to public transport.

All previous studies are based on surveys or interviews to private vehicle users. However, there are also some studies supported by real experiments to attempt to switch car users to public transport (Abou-Zeid and Fujii 2016, Abou-Zeid and Ben-Akiva 2014, Abou-Zeid et al. 2012, Abou-Zeid and Ben-Akiva 2012). Most of these studies use some kind of incentive (e.g. free public transport pass for a period) to encouraging the use of public transport. The majority of them find evidences for changes in psychological variables (e.g. correction of misperceptions about public transport attributes) (Abou-Zeid and Fujii 2016).

2.3.- Dealing with heterogeneity in perceptions: multi-group analysis (MGA) and multiple-indicator and multiple-causes (MIMIC) models

Previous studies have also shown that heterogeneity exists among the perceptions of different sectors or segments of the population (Oliver 2010). Most of these studies look into heterogeneity by comparing models built for different groups according to diverse criteria: territorial contexts, sociodemographic characteristics, mobility patterns, etc. (de Oña et al. 2020, Abenoza et al. 2017, Diana 2012, dell'Olio et al. 2011, Beirao and Cabral 2007, Krizek and El-Geneidy 2007). The main limitation of this type of analysis is that the construction of independent models impedes contrasting to differences among the parameters of the models to determine whether they are statistically significant.

To overcome this limitation, researchers that use structural equation modeling (SEM) have begun to use multi-group analysis (MGA) and multiple-indicator and multiple-causes (MIMIC) models.

MGA makes it possible to simultaneously fit a model to data from multiple groups. It is therefore a type of analysis based on segmentation of the sample. However, through the specification of cross-group equality constraints, group differences for any individual parameter or set of parameters can be directly tested. The fit of the constrained model is compared with that of the model without equality constraints; if the constrained model´s fit is much worse, then the parameters constrained may not be equal in both groups. The MIMIC approach has become more widely used in recent years. The main advantage of SEM-MIMIC models is that they do not need to partition the population into groups and allow for inclusion of the effect of several different grouping variables all at once.

Several studies have used MGA or MIMIC approaches in the field of service quality, satisfaction or behavioral intentions in the context of urban and metropolitan public transport. Antonucci et al. (2014) used MGA to investigate differences in passenger satisfaction based on gender, age, education, employment status, residence area, time slot of use, travel frequency and reason of use. Customer satisfaction and loyalty differences between captive and choice riders have been also analyzed using this approach with data from the Chicago Transit Authority (Zhao, Webb and Shah 2014). Fu et al. (2018) investigated the determinants of loyalty to public transit using a model that considered seven latent factors (expectation, confirmation, service quality, perceived value, satisfaction, corporate image and loyalty) and identified gender differences using MGA. Recently, Allen et al. (2019a) studied the role of critical incidents and involvement in transit satisfaction and loyalty in a metro system in Spain. They also analyzed possible sources of heterogeneity because of several categorical variables (time of day, gender, age, nationality, ticket type, trip purpose, and travel frequency). Their results showed that time, age and travel frequency introduced significant heterogeneity in the models. Allen et al. (2019b) likewise used MGA to identify differences in overall public transport satisfaction and loyalty because of several categorical variables related to case study (city), to travel characteristics (travel frequency, trip purpose, designated lane, number of buses, travel time, and time of day) or to sociodemographic characteristics (gender, age, education, occupation, driving license and income). They identified that city, designated lane, travel time, time of day, education and income required an MGA.

To date, SEM-MIMIC models have been used less in the context of urban and metropolitan public transport. Zhao et al. (2014) used this approach to identify the effect of the captive versus choice riders for each model factor. They arrived at statistically significant differences in safety, service value, problem experience and likelihood of future use. Using data from a bus system in Chile, Allen et al. (2018) investigated the heterogeneity generated by 13 variables (operator, date, socioeconomic status, age, gender, time of day, day, frequency of bus use, frequency of metro use, transfer, trip purpose, perceived waiting time, and perceived travel time) over the general satisfaction with the system and the satisfaction with a specific bus line. Allen et al. (2020) studied the effect of critical incidents on public transport satisfaction and loyalty using a SEM-MIMIC approach with data from railways services in the hinterland of Milan (Italy). They also considered 13 socioeconomic and travel characteristics (data, day, time of day, access mode, service, line, travel frequency, ticket type, type of user, gender, age, income and education) to explore heterogeneity in the models.

3.- Data

This study considers a regular private vehicle user as an individual who uses a motorized vehicle (i.e., car, motorcycle or scooter) for his/her regular journeys (understood to be trips made for work/occupation/daily activities). However, in order to evaluate the public transport system in the area, the private vehicle user should have a minimal knowledge about the local service (Diana 2012, Mahmoud and Hine 2016). For this reason, the survey focused on self-declared regular private vehicle users that were at least occasional public transport users.

The quality of service perception, satisfaction and behavioral intentions towards public transport was collected through an online panel survey in the metropolitan areas of Madrid and Lisbon from May to June 2019. The same survey, in Spanish and Portuguese, was used in both cities. Although the questionnaire, with an average duration of seven minutes, contained eight parts, this study only uses information from the following four parts: Part 1 focused on sociodemographic and mobility characteristics of the respondents (i.e., gender, age, length of time living in the area, household location, education, occupation status, dependent members in the family, household income and frequency of public transport use); Part 2 referred to perceived quality of service, where users were asked to rate 14 service quality attributes (q1-q14); Part 3 included four satisfaction statements (s1-s4); and Part 4 focused on four behavioral intentions statements (b1-b4). Parts 2 to 4 were rated with a 5-point Likert scale.

Table 1 shows the sample data and summarizes the sample's main sociodemographic and travel characteristics for both cities; as well as the average values for the service quality attributes, and statements on satisfaction and behavioral intentions. 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.- Sample data and survey results

Part 1. Sample, sociodemographic and travel characteristics			
		Madrid	**Lisbon**
Sample (n)	n	500	530
	%	48.5%	51.5%
Gender	Male	59.8%	51.3%
	Female	40.2%	48.7%
Age (years)	18-24	7.8%	6.6%
	25-44	45.8%	44.3%
	45-64	30.4%	40.0%
	65 or more	16.0%	9.1%
Time living in the area	< 1 year	1.4%	1.9%
	A few years	26.7%	36.6%
	All my life	71.9%	61.5%
Household location	Metropolitan area	77.6%	77.5%
	City center	22.4%	22.5%
Education level	Primary schools or less	4.6%	5.5%
	Secondary schools	35.6%	30.8%
	Higher education (university degree)	59.2%	63.8%
Occupation status	Professional	9.3%	16.6%
	Employed	63.8%	61.1%
	Student	4.8%	3.2%
	Retired/Pensioner	16.1%	11.3%
	Other	6.0%	7.7%

Dependent members in the family	No	67.0%	62.1%
	Yes	33.0%	37.9%
Household net income	<2 minimum wages	20.5%	32.3%
	2-3 minimum wages	31.7%	27.8%
	>3 minimum wages	47.9%	39.9%
Frequency of use (days/week)	0-2	86.0%	93.0%
	3-4	11.4%	4.0%
	5-7	2.6%	3.0%
Part 2. Service quality (5-point Likert scale from 1 "very low quality" to 5 "very high quality")*			
q1. Service hours	Service hours	3.39	2.91
q2. Proximity	Proximity of stops to starting point or destination of the trip	3.41	3.12
q3. Frequency	Frequency or number of daily services	3.30	2.87
q4. Punctuality	Punctuality	3.36	2.62
q5. Speed	Speed	3.35	2.99
q6. Cost	Cost	3.10	3.07
q7. Accessibility	Ease of entrance and exit from the vehicle and/or stations	3.67	3.37
q8. Intermodality	Ease of transfers/good connections with other modes of transport	3.53	3.22
q9. Individual space	Individual space available inside the vehicle	3.02	2.77
q10. Temperature	Temperature inside the vehicle	3.28	2.93
q11. Cleanliness	Cleanliness of the vehicle and stations	3.43	2.97
q12. Safety	Safety on board (regarding accidents)	3.74	3.21
q13. Security	Safety regarding robbery and violence	3.03	2.84
q14. Information	Information provided	3.48	3.00
Part 3. Satisfaction (5-point Likert scale from 1 "completely disagree" to 5 "completely agree")*			
s1. General satisfaction	In general, I am satisfied with the PT service provided in XYZ**	3.51	2.94
s2. Expectations	The PT service in XYZ meets my expectations	3.36	2.77
s3. Needs	With the existing modes of transport in XYZ, I consider that the commuting needs of inhabitants are well covered	3.35	2.76
s4. Global experience	When I take PT in XYZ, I feel very satisfied	3.39	2.86
Part 4. Behavioral intentions (5-point Likert scale from 1 "completely disagree" to 5 "completely agree")*			
b1. I will use PT for one-off trips	In the next few weeks I will take PT for one-off trips	3.48	2.94
b2. I will use PT for regular trips	In the next few weeks I will take PT for my regular trips	2.47	2.18
b3. I will increase PT usage	I am sure I will increase the number of times I use PT in the future	2.90	2.73
b4. I will recommend PT	Not only will I use PT, but I will also recommend it to friends and family	2.97	2.74

* Average values; ** XYZ was the name of the city (Madrid or Lisbon)

The questionnaire was completed by 500 regular private vehicle users in Madrid and 530 in Lisbon. Both samples showed very similar sociodemographic and travel characteristics. Males were predominant in both samples, ranging from 51% to 60%; with the largest age group between 25 and 44 (44% to 46%); and the smallest age group under 24 (7% to 8%). Almost all respondents had been living in the same area for "all their life" or "a few years", and their household was located mainly in the metropolitan area (around 78% in both places). Most private vehicle users (59% to 64%) had higher education, and tended to be a third-party employee (59% to 64%). Most respondents (62% to 67%) did not have dependent members in the household (i.e., children or other dependent relatives), and most of them (86% to 93%) used the public transport service occasionally (two or less days per week). The net incomes were calculated based on the minimum wages for each country in 2019 (900€ in Spain and 700€ in Portugal). The income distribution in both cities was also very similar. The private vehicle users were distributed more or less homogeneously among the three levels of net household income considered, though with a somewhat higher proportion in the highest income level (40% in Lisbon and 48% in Madrid).

Madrid presented the best appraisal of its public transport system. Private vehicle users in Madrid gave the highest ratings for all service quality attributes, satisfaction and behavioral intentions statements. All average values were greater for Madrid than for Lisbon. In Madrid, the best rated indicators for service quality were *safety*, *accessibility* and *proximity*; while *individual space*, *cost* and *temperature* were the worst rated indicators. In Lisbon, the best rated indicators for service quality were *safety*, *accessibility* and *intermodality*; while *punctuality*, *individual space*, and *security* were the worst rated indicators.

Regarding satisfaction, the *general satisfaction* statement harvested the best rating in both cities, followed by *global experience*. The worst-rated satisfaction's statement was "*with the existing modes of transport in XYZ, I consider that the commuting needs of inhabitants are well covered*". Almost all the average values of behavioral intentions indicators were below the scale's central value (3), with the exception of *I will use PT for one-off trips* in the case of Madrid. Finally, the statement *I will use PT for regular trips* received the lowest rating in both cities. This is not surprising as the respondents were regular private vehicle users.

4.- Hypothesis and modeling approach

4.1.- Structural equation modeling (SEM)

Almost all previous studies apply structural equation modeling (SEM) to identify the mediator effect of satisfaction between service quality and behavioral intentions. SEM helps explain the relationships among multiple variables by examining the structure of interrelationships expressed through a series of equations. Unlike other multivariate techniques, SEM examines more than one relationship at a time; hence structural equation modeling is a technique for testing a set of hypotheses that considers all possible information (Hair et al. 2010). SEM is the synthesis of a structural regression (SR) model and a measurement model that assesses unobserved latent variables or factors as functions of observed variables. A valid measurement model is needed before proceeding to evaluate the SR part of the model. Kline (2015) suggests a two-step modeling approach: first, the SR model is re-specified as a Confirmatory Factor Analysis (CFA) measurement model; then, given an acceptable measurement model, several plausible SR models are compared to one another. If the fit of the CFA model is poor, not only is the researcher's hypotheses about the measurement possibly wrong, but the fit of the original SR model may be even worse.

This paper uses the two-step modeling approach, that includes several tasks: (1) specify the model; (2) evaluate model identification; (3) operationalize the constructs and collect, prepare, and screen the data; (4) estimate the model, including evaluating the model fit, interpret parameter estimates and consider equivalent or near equivalent models; (5) re-specify the model (if necessary, returning to task 4); and (6) replicate the results with an independent sample. The study follows this process, initially for the CFA measurement model, and later for the SR model.

4.2.- Measurement invariance and structural invariance

SEM offers the possibility of analyzing whether values of model parameters vary appreciably across different subgroups or different samples. Multi-Group Analysis (MGA) allows to simultaneously fit a model to data from multiple samples or groups, in order to compare them. MGA involves a two-step procedure: measurement invariance and structural invariance. First, measurement invariance needs to be examined. It concerns whether scores from the definition of a construct have the same meaning under different conditions. Kline (2015) identifies four basic kinds of measurement invariance, with increasingly restrictive hypotheses: (1) configural invariance, (2) weak invariance, (3) strong invariance, and (4) strict invariance. As the modeler increases the constraints, the fit indices will worsen. In the SEM literature, the most used fit index to assess for invariance when comparing nested models is the difference of the Comparative Fit Index (ΔCFI). When the ΔCFI for two models is greater than 0.01, then invariance is probably untenable (Cheung and Rensvold 2002).

Configural invariance is the least restrictive level, and it is tested by specifying the same measurement model in each group. In this model, both the number of factors and the correlations between factors and indicators are the same, but all parameters are freely estimated in each group. Weak invariance (or metric invariance) assumes configural invariance and it also requires equality of the unstandardized coefficients. This hypothesis is tested by imposing an equality constraint over groups on the unstandardized coefficient of each indicator, and comparing this model with the previous one. If the fit of the weak invariance model is not appreciably worse than that of the configural invariance model, the more restrictive hypothesis is retained. This allows for a comparison of the latent variance or covariances over groups. Strong invariance assumes weak invariance and equal unstandardized intercepts over the groups. Finally, strict invariance is the highest level measured, and it assumes strong invariance and equality in error variances and covariances across the groups.

The previous description corresponds to a model trimming strategy in which an initial unconstraint model (configural invariance) is gradually restricted by adding cross-group equality constraints. Failure at a particular step means that more restricted models should not be considered. However, as most of those constraints are very restrictive, Byrne et al. (1989) described partial measurement invariance as an intermediate state of invariance. For example, weak invariance assumes cross-group equality of each unstandardized pattern coefficient. If some, but not all, pattern coefficients are invariant, then only partial weak invariance holds. In this case, testing for intercept equality (strong invariance) could still be performed because noninvariant pattern coefficients are freely estimated in each group, which controls for these differences. There are no clear guidelines for determining the degree of partial invariance that would be acceptable in all situations. Hair et al. (2010) suggested a minimum of two loadings per construct that should remain invariant for assuring partial weak invariance in order to test the equivalence of theoretical relationships across two groups.

Once measurement invariance has been achieved, the final step is testing for structural invariance, or to determine whether the unstandardized coefficients for direct effects or disturbance variances and covariances are equal across groups. Such tests are conducted by imposing cross-group equality constraints on the corresponding parameter estimates for the SR model and then comparing the relative fits of the constrained model with that of the model without equality constraints. If the fit of the constrained model is not much worse than that of the unrestricted one, there is evidence of structural invariance. Several authors (Hair et al. 2010, Koklic, Kukar-Kinney and Vegelj 2017) suggest that testing for structural invariance requires only partial weak invariance. Therefore, if at least partial weak invariance is achieved in this study, the structural invariance will be tested next.

4.3.- Multiple-indicator and multiple-causes (MIMIC) approach

Another way to estimate group differences on latent variables is through the specification of a MIMIC model (Hauser and Goldberger 1971, Joreskog and Goldberger 1975). This approach allows for controlling heterogeneity in the measurement of latent constructs between different population segments, considering the restriction of a group-invariant covariance matrix for several observed indicators, conditional on grouping variables represented by regressors, which could be binary, categorical or numeric values.

Including a relevant set of explanatory variables provides MIMIC modeling extra information about the measurement model and enables the explicit investigation of hypotheses of invariance across subpopulations (Muthen 1989). The main advantages of this approach are the following (Kline 2015, Allen et al. 2018): (i) unlike with MGA, there is no need to partition the population into subsamples at the modeling stage; (ii) it does not present special identification requirements beyond the usual ones for single-sample analysis; and (iii) it is possible to include the effect of several different grouping variables all at once, instead of performing a MGA with just one variable at a time.

However, the MIMIC approach assumes measurement invariance across the groups, and it does not consider heterogeneity in the cross-sectional preferences of different segments of the population, which could be analyzed by means of MGA. Following previous studies (Zhao et al. 2014), this paper uses both methods as complementary.

4.4.- Hypothesis and model specification

Based on the literature review for public transport users (see Section 2.1), this study considered two competing SR models: in Model 1 satisfaction plays a complete mediator role between service quality and behavioral intentions; while in Model 2 satisfaction plays a partial mediator role between the two constructs, and service quality has direct and indirect effects on behavioral intentions through satisfaction.

Figure 1 shows the model specification. This figure also shows the CFA measurement model that has to be tested in the first place. The three models are identified for the following reasons (Kline 2015): (i) they are over-identified models (i.e., positive degrees of freedom); (ii) the CFA model has three factors with two or more indicators per factor; and (iii) both SR models are recursive. The factor service quality (SQ) was operationalized using 14 indicators (q1-q14), while satisfaction (SA) used four indicators (s1-s4), the same number as behavioral intentions (BI, b1-b4).

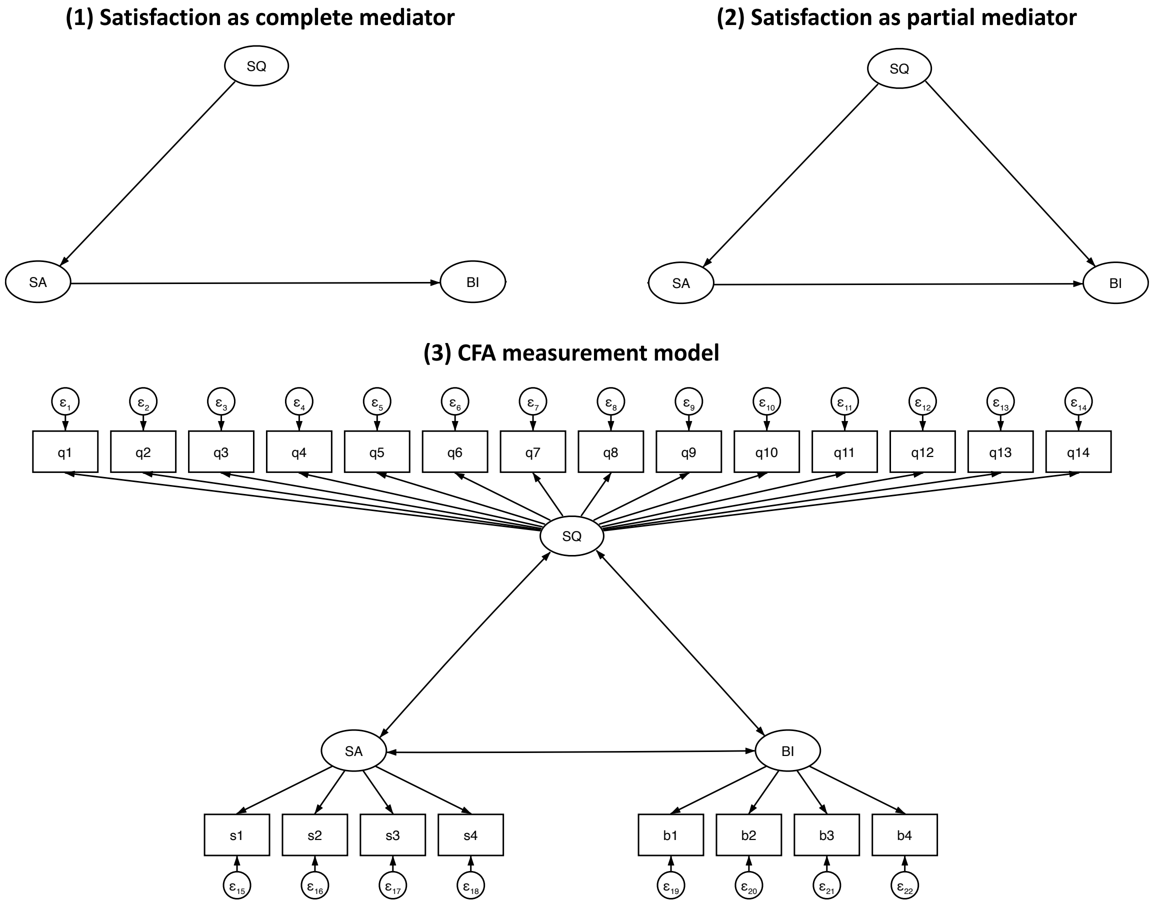

Figure 1.- Research models' specification: (1) Satisfaction as complete mediator; (2) Satisfaction as partial mediator; and (3) initial CFA measurement model

5.- Results and discussion

5.1.- Initial steps

The following step was data collection, preparation and screening (data collection is described in Section 3). Data preparation and screening involved checking sample size, missing values, outliers, collinearity, relative variances, and univariate and multivariate normality. The data from each country were considered separately for all the checks. All the statistical analysis was performed using Stata/MP 16.1.

Table 1 shows that the minimum sample size was 500 (Madrid), which was over the limit for one of the most restrictive criteria when using non-normal data. Hoogland and Boomsma (1998) suggested that the minimum sample size should be 10 times the number of free parameters. In this study the CFA measurement model has 47 free parameters (Figure 1). However, more recently Kline (2015) proposed a minimum sample of 200 to reduce biases to an acceptable level for any type of SEM estimation.

No extreme outliers were identified when checking for univariate and multivariate outliers. In Madrid, the indicator with the highest number of missing values (5.2%) was *I will increase PT usage*; while in Lisbon, it was *I will use PT for one-off trips* (3.2%). As the values were very low and did not show a systematic pattern, this data loss was of little concern. Finally, as all the attributes used the same Likert scale, the covariance matrices were not ill-scaled. This study used three methods to detect multi-collinearity: (i) calculating bivariate correlation among all the variables, identifying the greatest values for the satisfaction indicators (below 0.8); (ii) running squared multiple correlations between each variable and all the others using several multiple regressions, finding $R^2 < 0.78$ in all cases; and (iii) calculating the variance inflator factor (VIF) after each regression, obtaining values below five. Therefore, the sample data did not present multi-collinearity problems.

Finally, univariate normality was tested using the Shapiro-Wilk test. Results showed that most of the variables were not normally distributed. As univariate normality is a requirement for multivariate normality, the hypothesis that the data would present multivariate normality was also rejected. In addressing this issue, this study used the Satorra-Bentler estimator to control for non-normality (Satorra and Bentler 1994). The following results report the corrected χ^2 using this estimator, as well as all the corrected model fit indices that use χ^2.

5.2.- CFA measurement model

Two initial CFA measurement models (Madrid and Lisbon) were estimated in Stata, both of them converged to admissible solutions. Figure 1 shows that the initial CFA models considered three factors and 22 indicators that were allowed to load on only one factor. Table 2 displays the values of several fit indices provided by Stata. As the initial CFA models failed the exact-fit test, with $\chi^2(206)$ ranging from 499.51 to 509.05 ($p<0.001$), it was necessary to use approximate fit indices to evaluate the models. The approximate fit indices showed acceptable values (Hooper, Coughlan and Mullen 2008) with CFI ranging from 0.930 to 0.932 (>0.90), TLI ranging from 0.921 to 0.924 (>0.90), SRMR ranging from 0.050 to 0.051 (<0.08) and RMSEA equal to 0.057 (<0.08).

Table 2.- Results for initial CFA model

Parameters	**Madrid** **Unst. (SE) St.**	**Lisbon** **Unst. (SE) St.**
	Factor loadings	
Service Quality (SQ)		
SQ->Service hours (q1)	1 (*) 0.633	1 (*) 0.687
SQ->Proximity (q2)	0.918 (0.086) 0.587	0.785 (0.069) 0.54
SQ->Frequency (q3)	1.144 (0.072) 0.733	0.969 (0.062) 0.687
SQ->Punctuality (q4)	1.131 (0.077) 0.75	1.08 (0.069) 0.747
SQ->Speed (q5)	1.138 (0.082) 0.738	0.811 (0.066) 0.593
SQ->Cost (q6)	0.86 (0.082) 0.522	0.698 (0.077) 0.447
SQ->Accessibility (q7)	0.833 (0.068) 0.614	0.834 (0.067) 0.637
SQ->Intermodality (q8)	1.022 (0.076) 0.684	0.938 (0.063) 0.677
SQ->Individual space (q9)	0.989 (0.09) 0.647	0.942 (0.076) 0.677
SQ->Temperature (q10)	0.86 (0.086) 0.568	0.874 (0.071) 0.622
SQ->Cleanliness (q11)	0.752 (0.08) 0.558	0.857 (0.065) 0.645
SQ->Safety (q12)	0.743 (0.082) 0.531	0.761 (0.068) 0.552
SQ->Security (q13)	0.772 (0.086) 0.528	0.851 (0.064) 0.594
SQ->Information (q14)	0.989 (0.075) 0.702	1.027 (0.068) 0.705
Satisfaction (SA)		
SA->General satisfaction (s1)	1 (*) 0.911	1 (*) 0.914
SA->Meet expectations (s2)	1.008 (0.029) 0.898	0.984 (0.032) 0.889
SA->Covered needs (s3)	0.915 (0.04) 0.783	0.934 (0.033) 0.82
SA->I feel satisfied (s4)	0.918 (0.03) 0.854	0.818 (0.032) 0.822

Behavioral Intentions (BI)		
BI->I will use PT for one-off trips (b1)	1 (*) 0.604	1 (*) 0.602
BI-> I will use PT for regular trips (b2)	1.039 (0.109) 0.577	1.098 (0.079) 0.692
BI->Increase usage (b3)	1.346 (0.116) 0.775	1.364 (0.094) 0.863
BI->I will recommend PT (b4)	1.483 (0.117) 0.852	1.293 (0.09) 0.845
Factor variances and covariances		
var(e.SA)	0.512 (0.067) 1	0.531 (0.061) 1
var(e.BI)	0.924 (0.073) 1	0.915 (0.064) 1
var(SQ)	0.505 (0.076) 1	0.559 (0.074) 1
cov(SQ,SA)	0.584 (0.054) 0.849	0.551 (0.049) 0.79
cov(SQ,BI)	0.286 (0.037) 0.563	0.215 (0.036) 0.394
cov(SA,BI)	0.386 (0.048) 0.565	0.345 (0.045) 0.482
Construct validity and reliability		
Service Quality		
Construct Reliability (CR)	0.902	0.903
Average Variance Extracted (AVE)	0.401	0.402
Cronbach's Alpha	0.898	0.899
Satisfaction		
Construct Reliability (CR)	0.921	0.920
Average Variance Extracted (AVE)	0.745	0.743
Cronbach's Alpha	0.916	0.918
Behavioral Intentions		
Construct Reliability (CR)	0.799	0.841
Average Variance Extracted (AVE)	0.506	0.575
Cronbach's Alpha	0.788	0.835
Model's fit statistics		
df	206	206
chi-square	499.51	509.05
p-value	0.000	0.000
RMSEA	0.057	0.057
CFI	0.930	0.932
TLI	0.921	0.924
SRMR	0.050	0.051
AIC	24,660.9	24,919.7
BIC	24,942.6	25,204.1

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

Although the indicators used for defining the factors are similar to those used in the public transport literature, it is convenient to prove their validity, taking into account that this study focuses on private vehicle users. In general, the results showed convergent validity, construct reliability and discriminant validity. All factor loadings were statistically significant, presenting the correct sign (positive), and their values were higher than the 0.5 value suggested by Hair et al. (2010), with the exception of *cost* in Lisbon. As no other service quality indicator was below the threshold, *cost* was retained in the model. Both models presented good values for Construct Reliability (CR) and Cronbach's Alpha for the three factors (over 0.7). Satisfaction and behavioral intentions also presented Average Variance Explained (AVE) values over the recommended threshold (0.5), while this value was around 0.4 for service quality. 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 0.482 to 0.849. Not excessively high factor correlations (e.g., below 0.90) suggest discriminant validity (Kline 2015).

As the approximate fit indices were only acceptable, the decision was to study other plausible model specifications. The re-specification was undertaken step-by-step (as misfit sources can vary when changes are made to the model) inspecting the correlation residuals and modification indexes. For the sake of brevity however, only the results of the final CFA measurement model are presented. The final CFA models differs from the initial ones in the specification of six error correlations for both samples: (i) between *individual space* (q9) and *temperature* (q10); (ii) between *individual space* (q9) and *security* (q13); (iii) between *temperature* (q10) and *cleanliness* (q11); (iv) between *cleanliness* (q11) and *safety* (q12); (v) between *cleanliness* (q11) and *security* (q13); and (vi) between *safety* (q12) and *security* (q13). All these correlations are plausible and theoretically justified. Figure 2 shows the final CFA measurement model, including six measurement error covariances. The degrees of freedom of this model decrease by six, as compared to the initial CFA model.

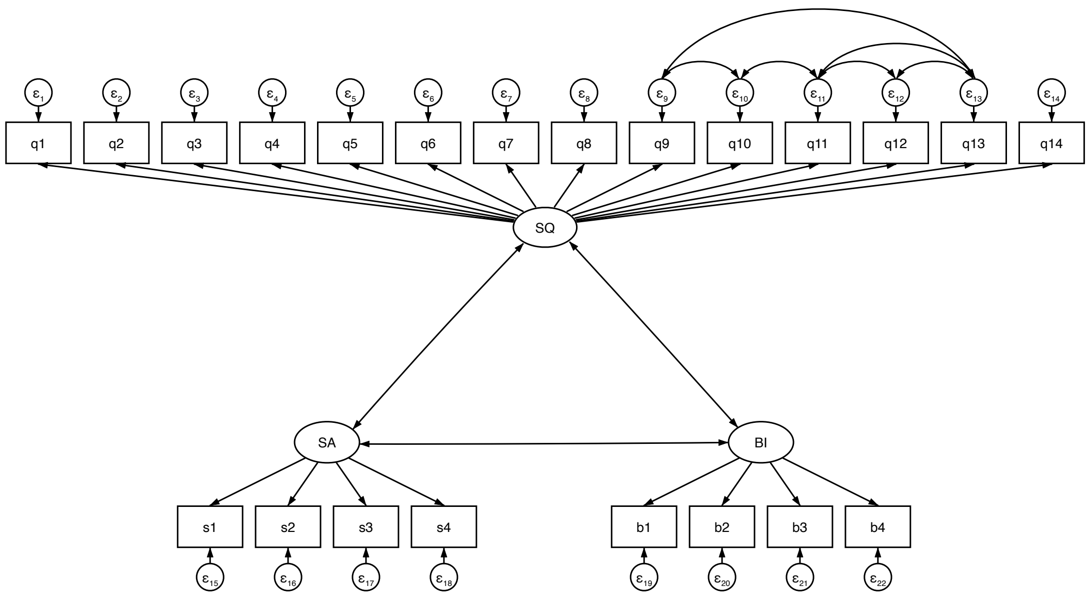

Figure 2.- Final CFA measurement model

Table 3 shows the final CFA model's results. The approximate fit indices in the final CFA model improved to excellent values in almost all cases, with CFI ranging from 0.953 to 0.962 (>0.95), TLI ranging from 0.945 to 0.957 (>0.95), SRMR ranging from 0.042 to 0.045 (<0.05) and RMSEA ranging from 0.043 to 0.048 (<0.05). The BIC values also provided strong support of preference for the final CFA model if compared to the initial one for both cities. Although both models were nested, they were not compared using the χ^2 differences test, because the estimation process entailed the use of Satorra-Bentler statistics. The difference between the Satorra–Bentler statistics for two nested models fitted to the same data does not follow a χ^2 distribution (Kline 2015). The other results remained in similar values. The main change was that the factor loading for *security* in Madrid decreased below 0.5. As no other service quality indicator in Madrid was below the threshold, *security* was retained in the model.

Factor loadings estimate the direct effects of factors on indicators and are interpreted as regression coefficients (Kline 2015). As in this study all indicators load on a single factor, standardized factor loadings are estimated correlations between the indicator and its factor. This makes it possible to identify, for each city, which attributes are more and less correlated with each of the factors. The three attributes presenting less correlation with service quality in both cities are *cost*, *safety* and *security*. These attributes can be considered the least important within the perception of quality on the part of regular private vehicle users. In turn, *punctuality* and *frequency* are the two most important attributes for both cities, while positions four and five in the ranking are for *information* and *intermodality*. In third place for Madrid is *speed*, but for Lisbon *service hours*.

Table 3.- Results for final CFA model

Parameters	**Madrid** **Unst. (SE) St.**	**Lisbon** **Unst. (SE) St.**
	Factor loadings	
Service Quality (SQ)		
SQ->Service hours (q1)	1 (*) 0.645	1 (*) 0.702
SQ->Proximity (q2)	0.916 (0.085) 0.597	0.784 (0.069) 0.552
SQ->Frequency (q3)	1.145 (0.071) 0.747	0.985 (0.063) 0.715
SQ->Punctuality (q4)	1.112 (0.076) 0.752	1.076 (0.068) 0.761
SQ->Speed (q5)	1.132 (0.08) 0.748	0.808 (0.065) 0.605
SQ->Cost (q6)	0.843 (0.08) 0.522	0.678 (0.076) 0.444
SQ->Accessibility (q7)	0.808 (0.066) 0.607	0.804 (0.067) 0.628
SQ->Intermodality (q8)	1.021 (0.074) 0.696	0.929 (0.064) 0.686
SQ->Individual space (q9)	0.929 (0.088) 0.621	0.884 (0.075) 0.651
SQ->Temperature (q10)	0.796 (0.084) 0.539	0.808 (0.07) 0.59
SQ->Cleanliness (q11)	0.695 (0.077) 0.528	0.772 (0.063) 0.598
SQ->Safety (q12)	0.694 (0.08) 0.506	0.674 (0.066) 0.5
SQ->Security (q13)	0.703 (0.084) 0.493	0.756 (0.062) 0.542
SQ->Information (q14)	0.967 (0.073) 0.699	0.987 (0.067) 0.694
Satisfaction (SA)		
SA->General satisfaction (s1)	1 (*) 0.911	1 (*) 0.914
SA->Meet expectations (s2)	1.01 (0.029) 0.899	0.985 (0.032) 0.889
SA->Covered needs (s3)	0.917 (0.04) 0.785	0.937 (0.033) 0.821
SA->I feel satisfied (s4)	0.916 (0.031) 0.852	0.817 (0.032) 0.82
Behavioral Intentions (BI)		
BI->I will use PT for one-off trips (b1)	1 (*) 0.603	1 (*) 0.602
BI-> I will use PT for regular trips (b2)	1.041 (0.108) 0.577	1.099 (0.079) 0.692
BI->Increase usage (b3)	1.347 (0.116) 0.775	1.364 (0.095) 0.864
BI->I will recommend PT (b4)	1.483 (0.117) 0.852	1.293 (0.091) 0.845
	Factor variances and covariances	
var(e.SA)	0.532 (0.069) 1	0.556 (0.063) 1
var(e.BI)	0.923 (0.073) 1	0.915 (0.064) 1
var(SQ)	0.505 (0.076) 1	0.559 (0.075) 1
cov(e.q9,e.q10)	0.187 (0.039) 0.241	0.118 (0.033) 0.186
cov(e.q9,e.q13)	0.165 (0.036) 0.214	0.065 (0.028) 0.097
cov(e.q10,e.q11)	0.157 (0.037) 0.212	0.131 (0.03) 0.207
cov(e.q11,e.q12)	0.112 (0.036) 0.159	0.203 (0.033) 0.302
cov(e.q11,e.q13)	0.123 (0.037) 0.167	0.19 (0.032) 0.282
cov(e.q12,e.q13)	0.177 (0.039) 0.227	0.29 (0.033) 0.382
cov(SQ,SA)	0.597 (0.055) 0.853	0.577 (0.05) 0.809
cov(SQ,BI)	0.288 (0.038) 0.555	0.226 (0.037) 0.405
cov(SA,BI)	0.385 (0.048) 0.564	0.344 (0.045) 0.482

Construct validity and reliability		
Service Quality		
Construct Reliability (CR)	0.899	0.898
Average Variance Extracted (AVE)	0.395	0.391
Cronbach's Alpha	0.898	0.899
Satisfaction		
Construct Reliability (CR)	0.921	0.920
Average Variance Extracted (AVE)	0.745	0.743
Cronbach's Alpha	0.916	0.918
Behavioral Intentions		
Construct Reliability (CR)	0.799	0.841
Average Variance Extracted (AVE)	0.506	0.575
Cronbach's Alpha	0.788	0.835
Model's fit statistics **		
df	200	200
chi-square	398.56	368.59
p-value	0.000	0.000
RMSEA	0.048	0.043
CFI	0.953	0.962
TLI	0.945	0.957
SRMR	0.045	0.042
AIC	24,551.1	24,767.1
BIC	24,857.3	25,076.3

* Not tested for statistical significance. All other unstandardized estimates are statistically significant at $p < 0.001$.
** SR Model (partial mediator) present the same model's fit statistics

Previous studies focusing on public transport users usually find that punctuality, frequency, comfort, and speed are the most important factors affecting service quality (Quddus et al. 2019). The results of this study make manifest that *punctuality* and *frequency* are furthermore the most important attributes for the regular private vehicle users surveyed, just as for public transport users. Even so, there are differences regarding the rest of the attributes: the importance of *intermodality* can be justified in light of the territorial context, as the surveys were carried out in large cities where it is often necessary to use more than one mode of public transport for a single journey; regular private vehicle users, not being frequent users of the public transport of their own cities, logically are not as familiar with its characteristics (e.g., lines, timetables, etc.) and therefore lend greater importance to *information*; and the fact that they use the public transport service little would relegate to lesser importance aspects related with comfort, highly appraised by regular users. *Speed* and *service hours*, respectively identified as important in Madrid and Lisbon, are attributes often found to be very highly valued by users of public transport.

Previous studies dealing with private vehicle users have arrived at similar results. *Frequency* is also identified by several authors (Hine and Scott 2000, Redman et al. 2013, de Oña et al. 2020) as one of the most important attributes for attracting car users. Krizek and El-Geneidy (2007) also found that travel time (linked to *frequency* and *speed*) presented notable similarities in the preferences between choice riders and potential users. Travel time is probably the attribute most frequently identified as relevant for non-users in numerous studies (de Oña et al. 2020, Bellizzi et al. 2020, Al-Ayyash and Abou-Zeid 2019, dell'Olio et al. 2011, Hine and Scott 2000). Several authors (Pedersen et al. 2011, Redman et al. 2013, Li et al. 2019) identified reliability (linked to *punctuality*) as one of the most important attributes. Rahman et al. (2013) found that current car users and infrequent transit users showed a high interest in real-time *information*. Bamberg et al. (2003) also identified *information* as a significant attribute to promote the modal switch from the car to public transport. *Intermodality* has been also reported in previous studies (de Oña et al. 2020, Hine and Scott 2000) as one of the most important attributes. Although the attribute *service hours* has not been previously identified in the literature, several studies (Redman et al. 2013, Kang et al. 2019) have reported convenience as one of the most important factors for non-users.

Comparison of service quality's factor loadings (Table 3) with the service quality evaluations given in Table 1 provides some interesting findings. In general, in Madrid, the indicators receiving the highest and lowest assessments are not the ones presenting the highest or lowest correlations — with the exception of *safety*, which received the best assessment although it presents a very low correlation with the factor service quality. The situation is different in the case of Lisbon. *Intermodality* received one of the three highest assessments, and it is also among the five indicators with the highest correlation, making it positive for attracting private vehicle users. Nonetheless, in Lisbon, the most correlated indicator (*punctuality*) received the lowest assessment, making it negative for attracting private vehicle users. This could explain why Lisbon shows very low values for all the service quality indicators, with an average value of 2.99 (considering the 14 indicators). Finally, although *safety* and *security* are among the attributes with the lowest correlations, *safety* was among the best rated while *security* received one of the three lowest assessments.

General satisfaction is the indicator that shows the highest correlation with satisfaction in both cities, while *covered needs* and *I feel satisfied* present the lowest correlations. In the case of behavioral intentions, the indicators *I will recommend PT* and *increase usage* present the highest correlation in both cities. Such results suggest that indicators about recommending the service and increasing usage should always be included when defining the factor behavioral intention, just as a statement about general satisfaction should be part of the indicators meant to define the factor satisfaction.

5.3.- Structural equation modeling

The next step was to test and compare the two SR models (complete and partial mediator role of satisfaction) (Figure 1) for each one of the two databases. Following Cronin et al. (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 satisfaction and behavioral intentions (as measured by the R^2-value). Although both models were nested, they were not compared using the χ^2 differences test, because the estimation process entailed the use of Satorra-Bentler statistics.

The partial mediator model presented the same number of free parameters as the final CFA model because all the correlations between latent factors were changed to directional paths. The two models were therefore equivalent versions and generated the same fit statistics. The complete mediator model was a restricted version, because the path between service quality and behavioral intentions was eliminated. The two models' estimation in Stata converged to admissible solutions. Table 3 shows fit statistics for the partial mediator model and Table 4 shows the fit statistics for the complete mediator model. As previously stated, the partial mediator model's fit statistics are the same as those of the final CFA model. For both cities, the fit statistics show very similar values regardless of the model used. The approximate fit indices were excellent in almost all cases, with CFI ranging from 0.951 to 0.963 (>0.95), TLI ranging from 0.944 to 0.957 (>0.95), SRMR ranging from 0.042 to 0.048 (<0.05) and RMSEA ranging from 0.043 to 0.048 (<0.05). However, while the complete mediator model was validated in both cities (all parameter estimates were consistent with the literature and significant), the data did not support the partial mediator model in Lisbon — the relationship between service quality and behavioral intentions was not significant. Following Kline (2015), the partial mediator model should not be considered valid because it should be re-specified, eliminating the non-significant structural paths. For the sake of brevity, Table 4 just shows the structural paths for the complete mediator model, the only one retained for the following steps.

Table 4.- Results for structural regression (SR) model (complete mediator)

Parameters	**Madrid** **Unst. (SE) St.**	**Lisbon** **Unst. (SE) St.**
	Structural paths	
SQ->Satisfaction (SA)	1.121 (0.08) 0.857	1.038 (0.065) 0.809
SA->Behavioral intentions (BI)	0.426 (0.048) 0.576	0.378 (0.045) 0.484
	Factor loadings	
Service Quality (SQ)		
SQ->Service hours (q1)	1 (*) 0.648	1 (*) 0.703
SQ->Proximity (q2)	0.908 (0.084) 0.596	0.783 (0.068) 0.551
SQ->Frequency (q3)	1.141 (0.07) 0.749	0.985 (0.062) 0.715
SQ->Punctuality (q4)	1.107 (0.075) 0.753	1.075 (0.068) 0.761
SQ->Speed (q5)	1.124 (0.079) 0.747	0.808 (0.065) 0.605
SQ->Cost (q6)	0.838 (0.079) 0.522	0.678 (0.076) 0.444
SQ->Accessibility (q7)	0.805 (0.066) 0.608	0.804 (0.066) 0.629
SQ->Intermodality (q8)	1.017 (0.073) 0.697	0.928 (0.064) 0.686
SQ->Individual space (q9)	0.92 (0.087) 0.618	0.884 (0.075) 0.651
SQ->Temperature (q10)	0.791 (0.084) 0.538	0.808 (0.07) 0.59
SQ->Cleanliness (q11)	0.688 (0.077) 0.526	0.772 (0.063) 0.598
SQ->Safety (q12)	0.688 (0.08) 0.504	0.674 (0.066) 0.5
SQ->Security (q13)	0.696 (0.084) 0.491	0.756 (0.062) 0.543
SQ->Information (q14)	0.961 (0.072) 0.699	0.987 (0.067) 0.694
Satisfaction (SA)		
SA->General satisfaction (s1)	1 (*) 0.91	1 (*) 0.913
SA->Meet expectations (s2)	1.009 (0.029) 0.897	0.985 (0.032) 0.889
SA->Covered needs (s3)	0.917 (0.04) 0.784	0.937 (0.033) 0.821
SA->I feel satisfied (s4)	0.918 (0.03) 0.853	0.817 (0.032) 0.82
Behavioral Intentions (BI)		
BI->I will use PT for one-off trips (b1)	1 (*) 0.603	1 (*) 0.601
BI-> I will use PT for regular trips (b2)	1.038 (0.107) 0.575	1.099 (0.079) 0.692
BI->Increase usage (b3)	1.347 (0.116) 0.775	1.364 (0.094) 0.864
BI->I will recommend PT (b4)	1.486 (0.118) 0.853	1.294 (0.091) 0.845
	Factor variances and covariances	
var(e.SA)	0.245 (0.029) 0.266	0.315 (0.035) 0.345
var(e.BI)	0.337 (0.053) 0.668	0.428 (0.063) 0.766
var(SQ)	0.537 (0.069) 1	0.556 (0.063) 1
cov(e.q9,e.q10)	0.188 (0.039) 0.242	0.118 (0.033) 0.186
cov(e.q9,e.q13)	0.167 (0.036) 0.215	0.065 (0.028) 0.097
cov(e.q10,e.q11)	0.157 (0.037) 0.212	0.131 (0.03) 0.207

cov(e.q11,e.q12)	0.113 (0.036) 0.161	0.203 (0.033) 0.302
cov(e.q11,e.q13)	0.124 (0.037) 0.168	0.19 (0.032) 0.282
cov(e.q12,e.q13)	0.179 (0.039) 0.229	0.29 (0.033) 0.382
Model's fit statistics		
df	201	201
chi-square	404.01	368.97
p-value	0.000	0.000
RMSEA	0.048	0.043
CFI	0.951	0.963
TLI	0.944	0.957
SRMR	0.048	0.043
AIC	24,555.3	24,765.3
BIC	24,857.4	25,070.4
R^2 (Satisfaction)	0.734	0.655
R^2 (Behavioral intentions)	0.332	0.234

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

The complete mediator model presented a high ability to explain variation in satisfaction (as measured by the R^2-value), ranging from 0.655 in Lisbon to 0.734 in Madrid; but a low ability to explain variation in behavioral intentions, ranging from 0.234 in Lisbon to 0.332 in Madrid. The values obtained for satisfaction are similar to those of previous studies of public transport users (Allen et al. 2019b), but the values for behavioral intentions are much lower. This finding is logical in the sense that behavioral intentions towards public transport on the part of regular private vehicle users do not depend solely on their perceptions of the quality of service and on their satisfaction with it; they are conditioned to some extent by other aspects not accounted for in the model (e.g., attitudes toward public transport, satisfaction with the use of one's private vehicle, etc.). If any of those omitted variables is highly correlated with the variables included in the model, this could generate bias in parameter estimates. This is possible in the case of attitudes toward public transport that could be correlated with satisfaction and behavioral intentions. So, parameters estimates should be analyzed and interpreted with caution given the possibility of omitted variable bias.

5.4.- Multi-group analysis (MGA)

Once confirmed that, for regular private vehicle users, satisfaction exerts a complete mediator role between service quality and behavioral intentions, analysis was focused on determining if there were differences tied to the location of the respondent or to his/her sociodemographic characteristics. The objective was to determine whether different groups in the population had different models or if the same model could be generalized. This study used the following seven categorical variables for group definition: city (Madrid vs. Lisbon); household location (city center vs. metropolitan area); gender (female vs. male); age (from 18 to 44 vs. 45 or older); education level (with vs. without university degree); dependent (with vs. without dependent members in the family); and income (low vs. high; with the cut-off point equivalent to two minimum wages).

MGA started with a measurement invariance analysis on the previous seven categorical variables; it continued with a structural invariance analysis according to the same variables. According to Kline (2015), the second step should only be analyzed if a strict invariance is verified. Yet other authors (Hair et al. 2010, Koklic et al. 2017) suggest that merely partial weak invariance is required for analyzing structural invariance (see Section 4.2 for details). This study uses the partial weak invariance criterion. When the different levels of invariance were analyzed (i.e., configural, weak, strong, strict, structural), if any model presented a statistical worsening of the CFI index (ΔCFI>0.01) (Cheung and Rensvold 2002), then the MGA was assumed to be warranted.

Table 5.- Measurement and structural invariance

Invariance level	**CITY**		**LOCATION**		**GENDER**		**AGE**		**EDUCATION**		**DEPENDENT**		**INCOME**	
	CFI	**ΔCFI**	**CFI**	**ΔCFI**	**CFI**	**ΔCFI**	**CFI**	**ΔCFI**	**CFI**	**ΔCFI**	**CFI**	**ΔCFI**	**ΔCFI**	**ΔCFI**
Configural (parameters freely estimated)	0.957		0.956		0.955		0.958		0.956		0.955		0.954	
Weak (all loadings constrained)	0.949	0.008	0.951	0.005	0.950	0.005	0.951	0.007	0.954	0.002	0.950	0.005	0.947	0.007
Strong (intercepts constrained)	**0.938**	**0.011**	0.949	0.002	0.949	0.001	0.950	0.001	0.954	0.000	0.949	0.001	0.946	0.001
Partial strong (most intercepts constrained)	0.949	0.000												
Strict (error variances and covariances constrained)	0.945	0.004	0.948	0.001	0.948	0.001	0.949	0.001	0.953	0.001	0.949	0.000	0.946	0.000
Structural (structural paths constrained)	0.944	0.001	0.948	0.000	0.948	0.000	0.949	0.000	0.951	0.002	0.949	0.000	0.946	0.000
Thresholds		*<0.01*		*<0.01*		*<0.01*		*<0.01*		*<0.01*		*<0.01*		*<0.01*

Table 5 shows the measurement invariance analysis. Results pointed out that almost all categorical variables guaranteed measurement and structural invariance. The exception was the variable city, which achieved weak invariance, implying equality of the unstandardized factor loadings; but it did not guarantee strong invariance, which also requires equal unstandardized intercepts over the groups. Score tests were performed to identify whether parameters constrained to be equal across groups should be relaxed (Sorbom 1989). These tests identified that the unstandardized intercepts of seven service quality attributes (*punctuality*, *cost*, *intermodality*, *individual space*, *cleanliness*, *safety* and *security*) and two behavioral intentions indicators (*I will use PT for one-off trips* and *increase usage*) were different in Madrid and Lisbon. Partial strong invariance was achieved when the previous nine unstandardized intercepts were freely estimated across cities. Table 5 shows that, once the previous constraints were relaxed, strict invariance (equal error variances and covariances) and structural invariance (equal structural paths across latent factors) were also guaranteed for city.

In summary, the above results indicate that the relationships in the measurement model and in the structural regression model are invariant, regardless of the city considered, the household location, gender, age, education level, dependent members in the family and income level. Yet this does not mean that there are no differences between the cities; for the variable city no strong invariance was achieved, meaning that the unstandardized intercepts are not the same in the two cities for all the indicators. The score tests made it possible to identify the nine indicators marking the differences between the two cities.

Focusing on public transport users, Allen et al. (2019a) used MGA to try to identify differences in service quality and satisfaction with Madrid's Metro system in view of gender and age, among other variables. They found that gender presented measurement and structural invariance, while age only supported weak measurement invariance. These authors identified differences on the unstandardized intercepts (strong invariance) and on the structural paths using five age groups. In another study, using data from the public transport in four Latin American cities, Allen et al.

(2019b) aimed to discover which variables brought heterogeneity to satisfaction and loyalty. Through a comparison of four cities (Santiago de Chile, Mexico City and two Brazilian cities), four education levels and four income levels, they found that the variables city, level of education and income required an MGA because they did not show measurement and structural invariance, while MGAs analysis was not necessary for age and gender. Also in the urban public transport field, with data from a bus system in China, Fu et al. (2018) proposed a complex model with seven constructs (expectation, service quality, confirmation, perceived value, satisfaction, corporate image, and loyalty) for investigating the determinants of loyalty to public transport. They compared males and females using an MGA and found invariance in the measurement models, but they did not achieve structural invariance, identifying two structural relationships that significantly differ between male and female groups.

The results of the present study do not coincide with previous ones for most of the variables, leading one to certain reflections. The variable city exhibits differences, just as in the study by Allen et al. (2019b). Differences identified in the literature in terms of age, education and income, not identified in our case, may be due to the number of categories considered. This study uses just two categories, while previous studies use four or five (Allen et al. 2019a, Allen et al. 2019b). The literature reveals contradictory results regarding gender: whereas Fu et al. (2018) arrived at significant differences between male and female groups, other studies (Allen et al. 2019a, Allen et al. 2019b) did not identify any differences. It was not possible to compare the other two variables (household location and dependent members in the family) with previous studies because they had not been analyzed previously using MGA. Finally, but more importantly, the present study is unlike all previous ones in that it analyzes regular private vehicle users instead of public transport users.

Although MGA enables one to identify which variables bring heterogeneity to the service quality, satisfaction and behavioral intentions models for regular private vehicle users, this analysis entails the limitation of not permitting several variables to be considered at the same time. For instance, this implies that when the effect of potential heterogeneity between male and female is analyzed, the effect of age is not simultaneously controlled. A multiple-indicator and multiple-causes (MIMIC) approach allows for inclusion of the effect of several different grouping variables all at once. Hence the analysis of regular private vehicle users is completed with a MIMIC analysis.

5.5.- MIMIC results

The MIMIC approach —used both for the pooled data and for each of the two cities— made it possible to identify, control and correct for possible bias in parameter estimation due to respondent heterogeneity behind service quality perceptions, satisfaction and behavioral intentions. All the variables considered in the MGA acted as regressors on the three latent factors in Madrid and Lisbon, as well as in the pooled sample. Yet the location variable was used in the model with the pooled data to capture heterogeneity because of this variable. The reference categories for the seven dummy variables were: Madrid, metropolitan area, female, from 18 to 44 years old, without university degree, without dependent members in the family, and low-income level (below the sum of two minimum wages).

This type of analysis has one prerequisite and one limitation. First, the MIMIC approach assumes measurement invariance across the groups, as verified in Section 5.4. Then again, the MIMIC approach only allows for estimation of group differences on latent variables (i.e., group differences cannot be identified in paths). In Section 5.4 it was also verified that no differences existed in the paths across the groups considered in this study.

Table 6 shows the structural paths, statistically significant regressor influence, selected fit statistics and the ability of each model to explain variation in satisfaction and behavioral intentions for the SEM-MIMIC models. Although the seven dummy variables were included in the analysis, Table 6 shows only those variables that proved significant in at least one model. In the case of Madrid, none of the dummy variables were significant; in Lisbon, three of them (household location, age and education level) were significant. For the pooled data, the variables city, household location and education level were found to be significant. Gender, dependent members in the family and income level were not significant in any model.

Table 6.- SEM-MIMIC results

	All			Madrid			Lisbon		
Path / Regressor / Fit index	**Unst.**	**SE**	**Std.**	**Unst.**	**SE**	**Std.**	**Unst.**	**SE**	**Std.**
Structural Paths									
Service quality (SQ) -> Satisfaction (SA)	1.075	0.049	0.832	1.114	0.080	0.852	1.053	0.067	0.820
Satisfaction (SA) -> Behavioral intentions (BI)	0.420	0.033	0.550	0.430	0.049	0.574	0.385	0.045	0.492
Service quality (SQ)									
Center	0.136	0.065	0.074	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Old	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	-0.168	0.072	-0.113
Lisbon	-0.416	0.054	-0.269						
Satisfaction (SA)									
Center	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	-0.142	0.070	-0.063
University	-0.112	0.044	-0.055	n.s.	n.s.	n.s.	-0.179	0.060	-0.090
Lisbon	-0.132	0.041	-0.066						
Behavioral intentions (BI)									
University	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	0.144	0.067	0.093
Model's fit statistics									
df		296			277			277	
chi-square		805.54			533.84			485.65	
p-value		0.000			0.000			0.000	
RMSEA		0.044			0.046			0.041	
CFI		0.947			0.941			0.956	
TLI		0.939			0.933			0.949	
SRMR		0.038			0.047			0.041	
AIC		55,346.0			26,751.3			27,218.9	
BIC		55,868.4			27,158.9			27,631.1	
R^2 (Satisfaction)		0.726			0.736			0.669	
R^2 (Behavioral intentions)		0.299			0.338			0.249	

n.s. non-significant

Although all the model fit statistics showed a slight deterioration when compared to the original model (Table 6), most of these values were still excellent (RMSEA and SRMR below 0.05) or acceptable (CFI and TFL above 0.90). 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). In addition, the MIMIC models for Madrid and Lisbon showed slightly higher R^2-values for satisfaction and behavioral intentions, indicating that they outperformed the original models in explaining variation in the latent factors. In the case of Lisbon, where the respondents presented more heterogeneity, improvement was greater than for Madrid.

For the pooled data, the regressors were only significant for service quality and satisfaction. Respondents in Lisbon presented a significant lower service quality perception and satisfaction than those of Madrid. These results agree with the average values for the indicators of service quality, satisfaction and behavioral intentions presented in Table 1. Table 1 shows that the perception of quality and satisfaction in Lisbon is much lower than in Madrid; yet the differences in terms of behavioral intentions are not so important. In general, living in the city center increased the service quality perception, while people with university degrees expressed significantly lower satisfaction. The household location finding could be linked to the fact that the city center usually has better transport services (e.g., higher frequency and punctuality). The effect upon satisfaction agrees with the study of Allen et al. (2020) in that having a university degree contributed negatively to overall satisfaction among public transport users.

In the case of Lisbon, significant regressors were identified in all the latent variables. Young people (from 18 to 44 years of age) expressed better service quality perceptions. An in-depth study of Lisbon data shows statistically significant differences between the two age groups when rating the attributes *speed*, *intermodality*, *temperature* and *information*. As three of the attributes (*speed*, *intermodality* and *information*) are of the highest importance for private vehicles users, the fact that the younger age group scores them with higher rates would explain this result. In the public transport user's literature, there is no consensus about age's effect on service quality and satisfaction. 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. Residents in the city center expressed significantly lower satisfaction in our study. This unexpected result points to the need for more insight to identify the causes of significant differences in Lisbon. At the same time, however, this result is in line with a recent qualitative study about perceptions of the public transport service in the metropolitan area of Lisbon (Ramos et al. 2019). The authors analyzed transcriptions of interviews and focus groups using content analysis techniques, and found the main terms associated with the public transport inside Lisbon municipality and its metropolitan area. “Traffic”, “late” and “complicated” were the three words that stood out for road transport inside Lisbon municipality, whereas the expressions most used by transit users to describe above ground trains were “canned sardines”, to indicate overcrowding, and “strike”. Transit users outside Lisbon municipality mainly used the words “pollutant”, “traffic” and “expensive” to describe road transport, whereas for the trains outside Lisbon municipality, the most prominent words were “hygiene”, “reliable”, “comfortable”, “expensive” and “punctual”. Therefore, while all the words associated with public transport inside Lisbon municipality were negative, most of the words associated with public transport outside Lisbon municipality were positive. Finally, having a university degree in Lisbon contributed negatively to satisfaction, as for the pooled sample, but positively to behavioral intentions. This positive effect of having a university degree on behavioral intentions may be traced to a greater awareness of the importance of this service to society, given a stronger environmental commitment on the part of more educated people (Golob and Hensher 1998).

Most of the standardized values of the regressors in Table 6 are very low (below 0.1). Only age for Lisbon's model (-0.11) and location for the model with the pooled data and the service quality latent factor (-0.27) presented values that could be considered as low (between 0.1 and 0.3). These results come to confirm the results of MGA discussed in Section 5.4: except for location, the sociodemographic variables considered in the analysis did not introduce much heterogeneity in the perceptions of regular private vehicle users towards public transport service quality, satisfaction and behavioral intentions. To some extent this contributes to not producing significant changes in the parameter estimates of the structural paths. Lisbon's model showed the largest changes (ranging from +1.4% for the path between service quality and satisfaction to +1.7% for the path between satisfaction and behavioral intentions). Madrid's model showed the lowest changes (ranging from -0.6% for the path between service quality and satisfaction to -0.3% for the path between satisfaction and behavioral intentions).

In sum, the results show that SEM-MIMIC analysis is relevant for investigating the pooled data and Lisbon´s data, but it is not worthwhile in the case of Madrid, as it increases the complexity of the model without providing substantial improvement. As the MGA shows that there are differences between Madrid and Lisbon, the city-specific models would be preferable over the pooled model.

6.- Conclusions

This article is the first to analyze service quality perceptions, satisfaction and behavioral intentions towards urban and metropolitan public transport from the point of view of regular private vehicle users using surveys and a structural equation modeling approach. These three factors, along with their relationships, have been widely studied in the literature from the standpoint of public transport users. In order to attract private vehicle users to urban and metropolitan public transport systems, it is essential to have in-depth knowledge of users' and potential users' perceptions, level of satisfaction, and behavioral intentions regarding them.

The results expounded here are based on 1,030 questionnaires answered by regular private vehicle users in Madrid (500) and Lisbon (530), which were separately analyzed to replicate the models, validate the results, and identify similarities and differences between the two cities so as to try and generalize the results obtained. Out of the 14 attributed considered to define the factor service quality, in both cities, *cost*, *safety* and *security* were identified as the least important for the regular private vehicle user. Of the five most important attributes, both cities identified *punctuality*, *frequency*, *information* and *intermodality*. *Speed* was one of the most important attributes in Madrid, and *service hours* was very highly appraised in Lisbon. These attributes more or less coincide with the ones that public transport users tend to identify as most important: *punctuality*, *frequency*, *comfort* and *speed*. The fact that private vehicle users hardly use public transport on a regular basis contributes to the identification of *information* as important, and *comfort* as relatively unimportant. The emphasis on intermodality can be explained by the territorial context, since the data are from two large European cities, where it is often necessary to use more than one mode of public transport for a single journey.

The results also show that there is not much correspondence between the importance of the attributes and the score they are given. In Madrid, the best rated attributes were *safety*, *accessibility* and *proximity*; while *individual space*, *cost* and *temperature* were the worst rated. In Lisbon, the best rated were *safety*, *accessibility* and *intermodality*; while *punctuality*, *individual space*, and *security* were the worst rated indicators. Similar findings have been reported in other studies focusing on public transport users.

The structural equation modeling approach made it possible to demonstrate that for the regular private vehicle users, satisfaction exerts a complete mediator role between service quality and behavioral intentions. This implies that there is no direct relationship between service quality and behavioral intentions, although the indirect effect of service quality over behavioral intentions is slightly superior than the direct effect of satisfaction over behavioral intentions. Such results, obtained for both cities of study, have been reported in studies of public transport users elsewhere. The complete mediator model presented a high ability to explain variation in satisfaction; but a low ability to explain variation in behavioral intentions. This is logical, as the behavioral intentions of regular private vehicle users also depend on aspects that were not considered in the present study (e.g., attitudes toward public transport, or satisfaction with the use of one's private vehicle). Many previous studies that analyze the relationship between these three constructs (service quality, satisfaction and behavioral intentions), in the field of public transport or in other areas, do not include attitudes towards service in their models. However, if the attitudes are correlated with satisfaction and behavioral intentions, this could generate bias in parameter estimates (i.e., omitted variable bias). This is the main limitation of the models tested in this study. So, it is recommended that future research include the attitudes toward public transport in the models when analyzing service quality, satisfaction and behavioral intentions to eliminate possible bias and to better explain the behavioral intentions of private vehicle users, as well as public transport users.

The MGA and the MIMIC approaches make it possible to identify similarities and differences between the models. It can be said that overall the models do not present important differences tied to the respondents´ sociodemographic characteristics, yet differences associated with the city of residence were identified. MGA's results showed that all sociodemographic characteristics guaranteed measurement and structural invariance. This means that the models are equal regardless of the household location, gender, age, education level, dependent members in the family, and income. In Madrid, the MIMIC approach further showed that none of these variables were significant, suggesting that for this city the sociodemographic variables considered do not introduce heterogeneity in the models. Gender, dependent members in the family and income level were not found to be significant in any model when using the MIMIC approach.

MGA's results also showed that the models are not exactly equal for Madrid and for Lisbon. Yet the analysis made it possible to determine that the differences between the two models lies in the unstandardized intercepts of seven service quality attributes (*punctuality*, *cost*, *intermodality*, *individual space*, *cleanliness*, *safety* and *security*) and two behavioral intentions indicators (*I will use PT for one-off trips* and *increase usage*), while the rest of the parameter estimates can be considered as equal. These differences are likewise reflected in the results of the SEM-MIMIC. For the pooled data, the differences reside in service quality and satisfaction, where the model shows that: residents in Lisbon present significantly lower service quality perception and satisfaction than those of Madrid; living in the city center increases the service quality perception; and people with university degrees express significantly lower satisfaction. In the case of Lisbon, significant differences were identified in the three factors due to household location, age and education level: residents in the city center expresses significantly lower satisfaction; people from 18 to 44 years of age have better service quality perceptions; and having a university degree contributes negatively to satisfaction, but positively to behavioral intentions.

Acknowledgments
Support from Spain´s Ministry of Economy and Competitiveness (Research Project TRA2015-66235-R) is gratefully acknowledged. I would like to thanks the two anonymous reviewers whose comments and suggestions helped to improve and clarify this manuscript.

Conflict of interest

The author states that there is no conflict of interest.

Authors contributions

Conceptualization: Juan de Oña; Methodology: Juan de Oña; Formal analysis and investigation: Juan de Oña; Writing - original draft preparation: Juan de Oña; Writing - review and editing: Juan de Oña; Funding acquisition: Juan de Oña; Resources: Juan de Oña; Supervision: Juan de Oña

References

Abenoza, R. F., O. Cats & Y. O. Susilo (2017) Travel satisfaction with public transport: Determinants, user classes, regional disparities and their evolution. *Transportation Research Part a-Policy and Practice,* 95**,** 64-84.

Abou-Zeid, M. & M. Ben-Akiva (2012) Travel mode switching: Comparison of findings from two public transportation experiments. *Transport Policy,* 24**,** 48-59.

---. 2014. Satisfaction and Travel Choices. In *Handbook of Sustainable Travel*, 53-65. Springer Netherlands.

Abou-Zeid, M. & S. Fujii (2016) Travel satisfaction effects of changes in public transport usage. *Transportation,* 43**,** 301-314.

Abou-Zeid, M., R. Witter, M. Bierlaire, V. Kaufmann & M. Ben-Akiva (2012) Happiness and travel mode switching: Findings from a Swiss public transportation experiment. *Transport Policy,* 19**,** 93-104.

Al-Ayyash, Z. & M. Abou-Zeid (2019) Investigating commute satisfaction differences of private car users and public transport users in a developing country context. *Transportation,* 46**,** 515-536.

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.

Antonucci, L., C. Crocetta, F. D. d'Ovidio & E. Toma (2014) Passenger satisfaction: a multi-group analysis. *Quality & Quantity,* 48**,** 337-345.

Bamberg, S., D. Rolle & C. Weber (2003) Does habitual car use not lead to more resistance to change of travel mode? *Transportation,* 30**,** 97-108.

Beirao, G. & J. A. S. Cabral (2007) Understanding attitudes towards public transport and private car: A qualitative study. *Transport Policy,* 14**,** 478-489.

Bellizzi, M. G., L. dell'Olio, L. Eboli & G. Mazzulla (2020) Heterogeneity in desired bus service quality from users' and potential users' perspective. *Transportation Research Part a-Policy and Practice,* 132**,** 365-377.

Byrne, B. M., R. J. Shavelson & B. Muthen (1989) Testing for the equivalence of factor covariance and mean structures - The issue of partial measurement invariance. *Psychological Bulletin,* 105**,** 456-466.

Cheung, G. W. & R. B. Rensvold (2002) Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling-a Multidisciplinary Journal,* 9**,** 233-255.
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.
Dabholkar, P. A., C. D. Shepherd & D. I. Thorpe (2000) A comprehensive framework for service quality: An investigation of critical conceptual and measurement issues through a longitudinal study. *Journal of Retailing,* 76**,** 139-173.
de Oña, J. (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.
--- (2021) Understanding the mediator role of satisfaction in public transport: A cross-country analysis. *Transport Policy,* 100**,** 129-149.
de Oña, J. & R. de Oña (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 Oña, J., R. de Oña, 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 Oña, J., E. Estevez & R. de Oña (2020) Perception of public transport quality of service among regular private vehicle users in Madrid, Spain. *Transportation Research Record,* 2674**,** 213-224.
de Oña, R., J. L. Machado & J. de Oña (2015) Perceived service quality, customer satisfaction, and behavioral intentions Structural Equation Model for the metro of Seville, Spain. *Transportation Research Record,* 2538**,** 76-85.
dell'Olio, L., A. Ibeas & P. Cecin (2011) The quality of service desired by public transport users. *Transport Policy,* 18**,** 217-227.
Diana, M. (2012) Measuring the satisfaction of multimodal travelers for local transit services in different urban contexts. *Transportation Research Part a-Policy and Practice,* 46**,** 1-11.
EC. 2019. Dataset population. Urban Audit Cities. https://ec.europa.eu/eurostat/web/cities/data/database. European Commission. Eurostat.
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.
Gotlieb, J. B., D. Grewal & S. W. Brown (1994) Consumer satisfaction and perceived quality: complementary or divergent constructs. *Journal of Applied Psychology,* 79**,** 875-885.
Hair, J. F., W. C. Black, B. J. Babin & R. E. Anderson. 2010. *Multivariate Data Analysis*. Pearson.

Hauser, R. M. & A. S. Goldberger (1971) The treatment of unobservable variables in path analysis. *Sociological Methodology,* 3**,** 81-117.

Hine, J. & J. Scott (2000) Seamless, accessible travel: users' views of the public transport journey and interchange. *Transport Policy,* 7**,** 217-226.

Hoogland, J. J. & A. Boomsma (1998) Robustness Studies in Covariance Structure Modeling: An Overview and a Meta-Analysis. *Sociological Methods and Research,* 26**,** 329-367.

Hooper, D., J. Coughlan & M. Mullen (2008) Structural Equation Modelling: Guidelines for Determining Model Fit. *Electronic Journal of Business Research Methods,* 6**,** 53-60.

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.

Joreskog, K. G. & A. S. Goldberger (1975) Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association,* 70**,** 631-639.

Kang, A. S., K. Jayaraman, K.-L. Soh & W. P. Wong (2019) Convenience, flexible service, and commute impedance as the predictors of drivers' intention to switch and behavioral readiness to use public transport. *Transportation Research Part F-Traffic Psychology and Behaviour,* 62**,** 505-519.

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.

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.

Krizek, K. J. & A. El-Geneidy (2007) Segmenting Preferences and Habits of Transit Users and Non-Users. *Journal of Public Transportation,* 10**,** 71-94.

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., M. Cao, Y. Bai & Z. Song (2019) Analysis of Public Transportation Competitiveness Based on Potential Passenger Travel Intentions: Case Study in Shanghai, China. *Transportation Research Record,* 2673**,** 823-832.

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 Oña & J. de Oña (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 Oña, F. Diez-Mesa & J. de Oña (2018) Finding service quality improvement opportunities across different typologies of public transit customers. *Transportmetrica a-Transport Science,* 14**,** 761-783.

Mahmoud, M. & J. Hine (2016) Measuring the influence of bus service quality on the perception of users. *Transportation Planning and Technology,* 39**,** 284-299.

Minser, J. & V. Webb (2010) Quantifying the benefits application of customer loyalty modeling in public transportation context. *Transportation Research Record,* 2144**,** 111-120.

Muthen, B. O. (1989) Latent variable modeling in heterogeneous populations. *Psychometrika,* 54**,** 557-585.

Oliver, R. L. 2010. *Satisfaction. A behavioral perspective on the consumer*. Armonk, New York: JW. E. Sharpe.

Pedersen, T., M. Friman & P. Kristensson (2011) Affective Forecasting: Predicting and Experiencing Satisfaction With Public Transportation. *Journal of Applied Social Psychology,* 41**,** 1926-1946.

Pedersen, T., P. Kristensson & M. Friman (2012) Counteracting the focusing illusion: Effects of defocusing on car users' predicted satisfaction with public transport. *Journal of Environmental Psychology,* 32**,** 30-36.

Quddus, M., F. Rahman, F. Monsuur, J. de Oña & M. Enoch (2019) Analyzing bus passenger's satisfaction in Dhaka using Discrete Choice Models. *Transportation Research Record,* 2673**,** 758-768.

Rahman, M. M., S. C. Wirasinghe & L. Kattan (2013) Users' views on current and future real-time bus information systems. *Journal of Advanced Transportation,* 47**,** 336-354.

Ramos, S., P. Vicente, A. M. Passos, P. Costa & E. Reis (2019) Perceptions of the public transport service as a barrier to the adoption of public transport: A qualitative study. *Social Sciences,* 8**,** 150.

Redman, L., M. Friman, T. Garling & T. Hartig (2013) Quality attributes of public transport that attract car users: A research review. *Transport Policy,* 25**,** 119-127.

Satorra, A. & P. M. Bentler. 1994. Corrections to test statistics and standard errors in covariance structure analysis. In *Latent variables analysis: Applications for developmental research.*, 399-419. Thousand Oaks, CA, US: Sage Publications, Inc.

Shao, Z., X. T. Li, Y. Guo & L. Zhang (2020) Influence of service quality in sharing economy: Understanding customers' continuance intention of bicycle sharing. *Electronic Commerce Research and Applications,* 40.

Sorbom, D. (1989) Model modification. *Psychometrika,* 54**,** 371-384.

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.

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.

Woods, R. & J. Masthoff (2017) A comparison of car driving, public transport and cycling experiences in three European cities. *Transportation Research Part a-Policy and Practice,* 103**,** 211-222.

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.

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.