Understanding the mediator role of satisfaction in public transport: a cross-country analysis

By: Juan de Oña

This document is a **postprint version** (ie final draft after peer review) of the following paper:

Juan de Oña (2021) *Understanding the mediator role of satisfaction in public transport: a cross-country analysis*. **Transport Policy, 100, 129-149.**

Direct access to the published version: https://doi.org/10.1016/j.tranpol.2020.09.011

Understanding the mediator role of satisfaction in public transport: a cross-country analysis

Abstract:
Many studies have analyzed the relationship between service quality, satisfaction and behavioral intentions or loyalty in the field of public transport. Yet despite growing interest, there is a lack of consensus regarding a number of aspects, e.g.: the difference between service quality and satisfaction, between behavioral intentions and loyalty, or the mediating effect of satisfaction between service quality and behavioral intentions. The main objective of this article is to shed light on the type of mediator effect exerted by satisfaction between service quality and behavioral intentions or loyalty in the area of urban and metropolitan public transport. To this end, structural equation modeling (SEM) is used to compare two competitive models, one in which satisfaction plays a partial mediating role (i.e., service quality presents direct and indirect effects on behavioral intentions or loyalty), and another where satisfaction exerts a complete or full mediator effect (i.e., service quality presents only indirect effects). The comparison is based on data from a single (translated) survey of public transport users in five European cities: Madrid, Rome, Berlin, Lisbon and London. The results support the superiority of the full mediator model over the partial mediator one in the urban and metropolitan public transport sector. The use of five independent samples made it possible replicate results and generalize conclusions, as well as identify other methodological and practical aspects. From a methodological standpoint, this paper confirmed the need to consider service quality and satisfaction as different factors and to compare alternative models with different samples when applying SEM to the public transport field. From a practical standpoint, the results suggest that service quality, associated with specific attributes of service, exerts a total effect on behavioral intentions or loyalty, superior to the effect of satisfaction, a finding that has important implications for transport operators. Finally, this study confirms that in large cities, the intermodality of public transport is one of the attributes that contributes most to the appraisal of its service quality, together with frequency, punctuality and speed.

Keywords: quality of service, customer satisfaction, loyalty, behavioral intentions, mediator, public transport, structural equation model, urban, metropolitan

1.- Introduction

In urban settings, public transport is a key element, facilitating the mobility of citizens without threatening the sustainability of the city. Public transport has a lesser socio-economic and environmental impact than private transport, above all in areas with high densities of population. According to the United Nations (United Nations 2019), each year more people live in urban areas. While 55% of the world's population resided in urban areas in 2018, by 2050, 68% of the world's population is foreseen to be urban. In urban and especially in metropolitan areas, mobility in terms of public transport continues to be outweighed by mobility by private transport.

For this reason, there is broad consensus about the need to increase the market quota of public transport, retaining current users and attracting new ones away from the private vehicle. This makes practitioners and researchers increasingly interested in determining the main aspects of service that contribute to its perception as of high quality; how these aspects contribute to customer satisfaction; and how much influence such perceptions and satisfaction bear upon a person´s tendency to use the service again, recommend it to others, or become a frequent and loyal user.

Traditionally, operators´ and managers' interest in ensuring passenger satisfaction have been instrumentalized through customer satisfaction surveys, conducted periodically. The relevance of surveys has been made manifest in recent years, with customer satisfaction surveys successfully undertaken even in times of economic recession (De Oña et al. 2018).

Although supplying public transport systems with high levels of service quality is fundamental for retaining users and attracting new ones, a service of high objective quality (e.g. with high frequency and velocity) does not always imply high user satisfaction (Friman and Fellesson 2009). Likewise, high user satisfaction does not necessarily mean greater use, or more recommendations of use, or higher user fidelity (Oliver 1999).

The concepts of service quality, satisfaction and behavioral intentions, as well as their interrelations, have been widely studied in the field of marketing. Satisfaction was originally defined as disconfirmation (Miller 1976); later, disconfirmation was viewed as an antecedent to satisfaction (Oliver 1981). Based on the disconfirmation theory, (Parasuraman, Zeithaml and Berry 1985) proposed a conceptual model of service quality that has since been extensively used in a number of fields (i.e., SERVQUAL model). Later, (Zeithaml, Berry and Parasuraman 1996) explored the behavioral consequences of service quality.

In the past twenty years, many studies have analyzed service quality, customer satisfaction, and behavioral intentions (or loyalty) paradigm in the field of public transport (Chen 2008, Lai and Chen 2011, Park, Robertson and Wu 2004, Saha and Theingi 2009). This peaked interest has coincided with an apparent change in mentality on the part of public transport operators, above all in the urban and metropolitan realm, to the extent where the "public transport user" is now referred to as a "public transport customer".

Two recent peer-reviewed publications look into public transport and review the literature on the subject: one focuses on service quality and satisfaction (de Ona and de Ona 2015) and the other on satisfaction and loyalty (van Lierop, Badami and El-Geneidy 2018). Still, because some concepts in the field are not entirely clear, arriving at conceptual interrelations and generalizations can be problematic.

On the one hand, the concepts of service quality and satisfaction have been used interchangeably in much of the public transport literature (Cavana, Corbett and Lo 2007, Elkhani, Soltani and Jamshidi 2014, van Lierop and El-Geneidy 2016) even though the field of marketing makes a distinction between the two (Oliver 2010). The concepts of loyalty and behavioral intentions also merge in the area of public transport: some authors talk about loyalty (Wen et al. 2005, Chou, Lu and Chang 2014, Allen et al. 2019a), others about behavioral intentions (Machado-Leon, de Ona and de Ona 2016, de Ona, Machado and de Ona 2015, Lai and Chen 2011), using similar indicators for the two concepts. Furthermore, there are authors who hold that these two elements represent the same construct (Lai and Chen 2011, Allen et al. 2019a). Therefore, a primary aim of the present contribution is to try clarifying the differences between these two particular constructs.

The fact that there is no precise distinction between those constructs most likely adds confusion to the study of the relationship between them. Among studies specifically centered on public transport, some adopt satisfaction with service quality as an antecedent of loyalty (van Lierop and El-Geneidy 2016), although it is more common to postulate that satisfaction is a mediator between service quality and behavioral intentions or loyalty (Jen, Tu and Lu 2011). Again, there may be discrepancies as to whether this mediating effect is total or partial (Figure 1). As Table 1 shows, the number of studies in the public transport literature that postulate a partial mediator effect is similar to the number postulating a full mediator effect.

Nearly all the studies listed in Table 1 use structural equation modeling (SEM) to identify this mediating effect, since (Iacobucci, Saldanha and Deng 2007) demonstrated the superiority of SEM over other approaches. SEM allows for the confirmation of hypotheses regarding the relationships existing among the different constructs in light of covariance analysis. (Kline 2015) recommends considering alternative models to account for the pattern of observed covariances, not only the author's preferred model, as long as the patterns can be justified from a theoretical standpoint. Such alternative models are known as equivalent models, and ignoring them amounts to a form of confirmation bias, leading to an overly positive evaluation of the model employed, and a possible failure to consider other explanations behind the data (Shah and Goldstein 2006). According to (Kline 2015), the probability of confirmation bias is even greater if no replication of the analysis is produced using an independent sample, a cross-validation method, or a split-sample approach.

The need for large samples when using SEM conditions the replication of results, so that most SEM studies in Table 1 are "one-shot" studies. Among them, just one (Chou and Kim 2009) used two independent samples for the purposes of validation and generalization. Using alternative models and different samples is frequent in other disciplines, such as marketing, however. For example, (Cronin and Taylor 1992) evaluated two alternative models in their analysis of the relationship between overall service quality, consumer satisfaction and purchase intentions, with data from four different sectors (banks, pest control, dry cleaning and fast food); whereas (Cronin, Brady and Hult 2000) evaluated four alternative models for analyzing the relationship between five constructs (sacrifice, service quality, service value, satisfaction and behavioral intentions) with data from six different sectors (spectator sports, participative sports, entertainment, health care, long distance carriers and fast food).

A second intention of the present study is to validate and generalize the links between service quality, customer satisfaction and behavioral intentions or loyalty for urban and metropolitan public transport users. To this end, the two alternative models shown in Figure 1 (satisfaction as complete mediator versus satisfaction as partial mediator) are applied to data from a single survey on the perception of quality in the public transport system, but the survey was carried out in the capital cities of five European countries (Germany, Spain, Italy, Portugal and the United Kingdom).

And finally, taking advantage the availability of the five sample populations, the paper discusses similarities and differences in the perception of service quality, customer satisfaction and behavioral intentions in the five cities (Berlin, Madrid, Rome, Lisbon and London). Although previous studies have analyzed data from multiple cities with a common survey (Friman and Fellesson 2009, Allen, Munoz and Ortuzar 2019b), as far as the author knowledge, to date none has compared user perceptions of the entire public transport system of each city. (Friman and Fellesson 2009) looked into the relationship between objective performance measures of public transport services and the satisfaction perceived by travelers, with data collected from six different European cities (Barcelona, Copenhagen, Helsinki, Oslo, Stockholm and Vienna). They analyzed the correlation between three objective service performance measures and three subjective satisfaction attribute measures, finding that the relationship between satisfaction and service performance in public transport was far from perfect. (Allen et al. 2019b) encountered strong evidence confirming the existence of a Maslow's hierarchy of transit needs, with three types of attributes: functional, security and hedonic. They used different types of SEM and data collected by the public transport operators of four cities in Latin America through customer satisfaction surveys for specific bus rapid transit systems.

The rest of this paper is organized as follows. Section 2 presents a literature overview on three topics: (i) differences between service quality and satisfaction; (ii) differences between behavioral intentions and loyalty; and (iii) rival conceptualizations of customer satisfaction as a mediator between service quality and behavioral intentions. Section 3 presents competing research models, hypotheses, methodologies and modeling approaches. Section 4 offers a general presentation of the customer satisfaction benchmark survey used, followed by the samples, the main survey results, and data preparation and screening. Section 5 describes the process followed to confirm the satisfactory psychometric properties of the three factors considered in analysis. Section 6 presents the results of the structural regression models and discusses the competing models, while Section 7 discusses the main findings of the paper. Finally, Section 8 summarizes the most important conclusions and recommendations.

2.- Literature review, theoretical background and hypotheses

2.1. Similarities and differences between service quality and satisfaction.

The literature tends to discuss service quality and satisfaction without clearly distinguishing between the two concepts. Yet looking at them under different lenses is important for managers and researchers alike, as service providers need to know whether their objective is to have consumers who are satisfied with performance, or else to deliver the maximum level of perceived service quality (Cronin and Taylor 1992). Some of the reasons behind this confusion may be also found in other sectors, while some are specific to the public transport sector.

The concept of service quality is complex, fuzzy, and abstract, comprising three service properties: intangibility, heterogeneity, and inseparability (Parasuraman et al. 1985). In the literature, service quality usually accompanies satisfaction and the two are highly correlated. Both constructs derive from the disconfirmatory theory: satisfaction was originally defined as disconfirmation (Miller 1976), though some years later, disconfirmation was viewed as an antecedent to satisfaction (Oliver 1981). With a focus on public transport, studies of service quality and satisfaction are based on customer satisfaction surveys (de Ona and de Ona 2015) that register users´ assessment of a number of specific attributes of the service along with a global appraisal. Notwithstanding, the wording of the survey may use the terms satisfaction, perceptions, expectations, and importance to express largely the same notion. Sometimes the terminology is stirred up by public transport operators and managers who consider the concepts of service quality and satisfaction as equivalents (De Oña et al. 2018).

Meanwhile, authors in the area of marketing uphold service quality and satisfaction as closely related yet different constructs (Cronin and Taylor 1992, Mattsson 1992, Dabholkar, Shepherd and Thorpe 2000, Oliver 2010, Bansal and Taylor 2015). (Cronin and Taylor 1992) highlighted that service quality is an overall evaluation, whereas satisfaction is a more immediate reaction to a specific service experience. Based on other studies, (Bansal and Taylor 2015) suggested that service quality evaluations are expected to be more cognitive, and satisfaction is expected to be primarily affective in nature. This agrees with (Mattsson 1992) in that satisfaction occupies a "higher" attitude level with regard to service quality. (Oliver 2010) suggested that the dimensions underlying quality judgments are specific, yet satisfaction judgments can come from any dimension, quality-related or not; an expectation of quality is based on ideals or "excellence" perceptions, while non-quality referents are used in satisfaction judgments; experience is not required for quality judgment, while satisfaction is primarily experiential; and although service quality has fewer conceptual antecedents, satisfaction would be influenced by a number of cognitive and affective processes. Thus, he defined service quality as a cognitive judgment (thinking/judging) that summarizes the exceptionally good (or bad) elements of a service, especially when compared with other direct alternatives; in contrast, customers' satisfaction is a purely experiential affective judgment (liking/pleasure), defined as the "consumer's fulfillment response". Some of the above studies also analyzed data from different samples and identified service quality and satisfaction as distinct constructs (Cronin and Taylor 1992, Dabholkar et al. 2000, Bansal and Taylor 2015).

In most of the public transport literature, service quality is associated with specific service attributes (e.g., frequency, cleanliness, comfort, speed, etc.), while satisfaction is associated with more elaborated perceptions and affective judgements (e.g., liking, feeling, pleasure, etc.). The indicator most often used to assess this construct is "overall satisfaction", but there are other indicators in public transport studies: pleasure (Irtema et al. 2018, Li and Petrick 2010, Zhang et al. 2013, Kim and Lee 2011), enjoyable (Fu and Juan 2017, Jen et al. 2011, Rajaguru 2016), right or wise choice (Chou and Yeh 2013, Chou et al. 2014, Park et al. 2004, Zhang et al. 2013), attraction (de Ona et al. 2016, Machado-Leon et al. 2018), interesting and surprised (Jen et al. 2011), even contented and delighted (Li and Petrick 2010).

Altogether, the field of public transport can be said to conceive service quality and satisfaction are different constructs. Appraisals of the specific attributes of the service should be used as indicators of the construct "service quality", yet to define the construct "satisfaction", indicators more closely tied to affective judgements should be used (e.g., pleasure, enjoy, attraction, etc.).

2.2. Behavioral intentions and loyalty: different constructs or the same one?

There is no single criterion for using the concepts of loyalty or behavioral intentions in the transport literature. As seen in Table 1, the studies using loyalty or behavioral intentions are roughly equal in proportion, 50%. One reason may be that the concept of loyalty is not well defined in the transport literature, and given that it is a more recent topic of study, researchers have not yet agreed on how to measure it (van Lierop et al. 2018).

In this case, the controversy surrounding the two terms is not limited to the framework of transport, but extend into marketing. Authors (Zeithaml et al. 1996) contrasted, with four independent samples from different sectors (computer manufacturing, retail chains, automobile insurance and life insurance), that loyalty was a dimension of behavioral intentions. They defined loyalty using five indicators: saying positive things about the company, recommending the company to someone who seeks advice, encouraging friends and relatives to do business with the company, considering the company the first choice from which to buy services, and doing more business with the company in the next few years. Accordingly, to (Zeithaml et al. 1996) loyalty would be a behavioral intentions' sub-construct.

Going back to (Oliver 1981, Oliver 2010), loyalty can be determined from four aspects: cognitive loyalty, affective loyalty, conative loyalty, and behavioral/action loyalty. Following (Pedersen and Nysveen 2001), cognitive loyalty is based upon the service information available to the customer; affective loyalty is based on consumers' affect-based attitudes to a service, and attitudes to a service are based upon an established relationship between the consumer and the service; and, finally, conative loyalty, or consumers' behavioral intention to keep on using a service in the future, is argued to a stronger predictor of behavioral/action loyalty compared to cognitive and affective loyalty. Most researchers employ the behavioral intention measure to represent customer loyalty. Compared to action loyalty, conative loyalty is more measurable and thus widely adopted by researchers (Yang and Peterson 2004). In light of the above, behavioral intentions would be a loyalty's sub-construct.

Despite certain discrepancies, both approaches share a consideration of attitudinal and behavioral measures to define and assess loyalty or behavioral intentions. This has helped most of the studies listed in Table 1 —regardless of the construct used (loyalty or behavioral intentions)— define it using largely the same items associated with reusage intention (behavioral measure) and willingness of recommending the service to others (attitudinal measure).

In many studies using behavioral intentions as the construct, just the two items are given (reusage and recommendation), although some add others: price tolerance (Kim and Lee 2011), say positive things (Saha and Theingi 2009), first choice (Fu and Juan 2017), or complaint behavior (Kim and Lee 2011). Yet in studies using the construct loyalty, it is more frequent to encounter (aside from items associated with reusage and recommendation) such items as price tolerance (Chou and Kim 2009, Chou and Yeh 2013, Wen et al. 2005, Sun and Duan 2019), say positive things (Chou et al. 2014, Nguyen-Phuo et al. 2020), first choice (Chou et al. 2014, Fu, Zhang and Chan 2018, Li et al. 2018, Nguyen-Phuo et al. 2020, Sun and Duan 2019), and confidence (Allen et al. 2019b, Fu et al. 2018, Zhang et al. 2019). Thus, most of the additional items are repeated, regardless of the name of the construct (whether loyalty or behavioral intentions).

Table 1.- Public transport literature linking service quality, customer satisfaction and behavioral intentions or loyalty, based on the type of mediation

Source	Mediator Role	Factor (1)	Relevant attributes (2)	Context	Mode
(Allen et al. 2019a)	Both	Both	REC	Urban	metro
(Allen et al. 2020)	Full	LO	REC	Urban	train
(Allen, Munoz and Ortuzar 2019b)	Partial	BI	REC, CON, OTH	Urban	bus
(An and Noh 2009)	Partial	LO	REC, USE	Interuban	air
(Chang and Yeh 2017)	Full	LO	REC, USE	Interuban	bus
(Chen 2008)	Full	Both	REC, USE	Interuban	air
(Chou and Kim 2009)	Full	LO	REC, USE, PRI	Interuban	train
(Chou and Yeh 2013)	Full	LO	REC, USE, PRI, OTH	Interuban	train
(Chou, Lu and Chang 2014)	Partial	LO	REC, USE, CHO, SAY, OTH	Interuban	train
(de Ona et al. 2016)	Partial	BI	REC, USE	Urban	LRT
(de Ona, Machado and de Ona 2015)	Both	BI	REC, USE	Urban	LRT
(Elkhani, Soltani and Jamshidi 2014)	Full	LO	REC, USE	Interuban	air
(Fu and Juan 2017)	Partial	BI	USE, CHO	Urban	general
(Fu, Zhang and Chan 2018)	Partial	LO	REC, CHO, CON	Urban	bus
(Han 2013)	Full	BI	REC, USE	Interuban	air
(Irtema et al. 2018)	Partial	BI	USE	Urban	metro
(Jen, Tu and Lu 2011)	Full	BI	REC, USE	Interuban	bus
(Joewono and Kubota 2007)	Partial	LO	REC, USE	Urban	paratransit
(Kim and Lee 2011)	Full	BI	REC, USE, PRI, OTH	Interuban	air
(Koklic, Kukar-Kinney and Vegelj 2017)	Full	BI	REC, USE	Interuban	air
(Kuo and Tang 2013)	Full	BI	REC, USE	Interuban	train
(Lai and Chen 2011)	Partial	Both	REC, USE	Urban	metro
(Li et al. 2018)	Partial	LO	REC, USE, CHO, OTH	Urban	general
(Machado-Leon et al. 2018)	Partial	BI	REC, USE	Urban	LRT
(Machado-Leon, de Ona and de Ona 2016)	Both	BI	REC, USE	Urban	LRT
(Minser and Webb 2010)	Partial	LO	REC, USE	Urban	general
(Nguyen-Phuo et al. 2020)	Partial	LO	REC, USE, CHO, SAY, OTH	Urban	taxi
(Park, Robertson and Wu 2004)	Full	BI	REC, USE	Interuban	air
(Park, Robertson and Wu 2006)	Full	BI	REC, USE	Interuban	air
(Rahim 2016)	Partial	LO	n.a.	Interuban	air
(Rajaguru 2016)	Partial	BI	REC, USE	Interuban	air
(Saha and Theingi 2009)	Partial	BI	REC, USE, SAY	Interuban	air
(Sun and Duan 2019)	Full	LO	USE, PRI, CHO, OTH	Urban	bus
(Wen et al. 2005)	Full	LO	REC, USE, PRI	Interuban	bus
(Yuan et al. 2019)	Full	LO	REC, USE	Urban	bus
(Zhang et al. 2019)	Full	LO	REC, USE, CON	Urban	general
(Zhao, Webb and Shah 2014)	Partial	LO	REC, USE	Urban	general

Notes: (1) loyalty (LO), behavioral intentions (BI) and both factors or without defining (Both); (2) Attributes used for defining loyalty or behavioral intentions associated with recommendation (REC), reusage (USE), price tolerance (PRI), first or correct choice (CHO), confidence (CON), saying positive things or word-of-mouth (SAY), and others (OTH).

In view of the controversy, even outside the field of transport, it would be unwise to judge one term better than the other. We do believe, however, that whichever prevails (loyalty or behavioral intentions), indicators associated with reusage intention and willingness of recommending the service should be included, at the very least. It would likewise be advisable to include indicators associated with price tolerance, say positive things, and first choice. While acknowledging that the two constructs are not identical, from this point onward we shall employ loyalty and behavioral intentions indistinctly.

2.3. Customer satisfaction as mediator between service quality and behavioral intentions: different approaches.

Conflicting models exist in the general literature regarding the process through which service quality and satisfaction affect behavioral intentions or loyalty. (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, using data from four industries (banking, pest control, dry cleaning, and fast food). Their results suggest that service quality is an antecedent of satisfaction and that satisfaction in turn exerts a mediator effect between service quality and behavioral intentions. The results were not conclusive about the type of mediator effect, as they found partial and complete effects, depending on the industry. Along the same lines, (Gotlieb, Grewal and Brown 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. Authors (Cronin et al. 2000) compared four more complex models (service value and sacrifice were included as constructs) 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. Therefore, there is 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.

The lack of consensus is evident in transport literature. Table 1 shows there are practically the same number of studies defending a full mediator effect as those defending a partial mediator effect. For the sake of brevity, this literature review will focus on the studies to date surrounding urban and metropolitan public transport.

In the specific case of urban public transport, studies defending a partial mediator effect are seen to predominate. With data from the Chicago Transit Authority, (Minser and Webb 2010) explored the relationship between public image, problem experience, service quality, service value, satisfaction, and loyalty; service quality was found to have significant direct and indirect (through satisfaction) effects upon loyalty. (Lai and Chen 2011) analyzed the roles of service quality, perceived value, satisfaction and involvement in behavioral intentions based on passenger survey data from a metro line in Taiwan. They found that satisfaction, perceived value and involvement are partial mediators between service quality and behavioral intentions. (Irtema et al. 2018) proposed the same model for a metro line in Malaysia, and obtained the same results regarding the relationships between service quality, satisfaction and behavioral intentions. (Zhao, Webb and Shah 2014) studied customer loyalty differences between captive and choice transit riders in Chicago, considering as indicators of loyalty both future use and recommending to others. They found that the relationship between service quality and recommending to others was significant for both type of users, while the relationship between service quality and future use was only significant for choice riders. (de Ona et al. 2016) adopted a complex model featuring seven constructs (service quality, behavioral intentions, satisfaction, attractive alternatives, perceived benefits, perceived costs, and feeling towards transit) to analyze a light rail transit system in Spain. They determined that service quality had direct and indirect (through satisfaction) effects over behavioral intentions. Using the same constructs and database, (Machado-Leon et al. 2018) applied a model equivalent to that of (de Ona et al. 2016), splitting the sample into six user groups. The new model revalidated all the relationships previously identified, and added to them service quality as an antecedent of perceived benefits. This association also proved significant. With data from a transit system in China, (Fu and Juan 2017) proposed a model considering satisfaction, behavioral intentions, public transport use, habit, subjective norm, perceived behavioral control, cost, intangible service and information. In this study service quality was introduced in the model through three dimensions: cost, intangible service and information. The effect of each service quality's dimension on behavioral intentions was different: cost presented only a direct effect; information only an indirect effect, mediated by satisfaction; and intangible service presented both direct and indirect effects (mediated by satisfaction and habit). Based on data from Shanghai, (Li et al. 2018) explored the relationships between service quality, passenger satisfaction, switching costs, attraction of alternatives, and passenger loyalty. They confirmed the partial mediator effect of satisfaction and found significant correlations between service quality and car attraction and switching costs. With data from a bus service in China, (Fu et al. 2018) evaluated a comprehensive model that considers seven constructs (expectation, confirmation, service quality, perceived value, satisfaction, corporate image, and loyalty). They found that service quality had a direct effect on loyalty, but also presented indirect effects through satisfaction, confirmation, perceived value and corporate image. (Allen et al. 2019b) used different structural equation models to look into the relationships of four dimensions of service quality (reliability, comfort, customer service and safety) with regard to satisfaction and loyalty using data from four Latin American bus rapid transit systems. And finally, (Nguyen-Phuo et al. 2020) explored the passenger satisfaction and loyalty intention of a ride-hailing taxi system in Vietnam. They considered five constructs (perceived benefits of booking app, perceived sales promotion, and service quality, satisfaction and loyalty), finding that service quality presented a direct and indirect (through satisfaction) effect upon loyalty.

With pooled data from 58 public transport services of 13 cities in China, (Zhang et al. 2019) explored the relationship between service quality, perceived value, expectations, complaint, and loyalty. Their analysis suggests that satisfaction is a full mediator between service quality and loyalty, and that perceived value is a partial mediator between service quality and satisfaction. Using the same model, (Yuan et al. 2019) analyzed another bus system in China, and arrived at very similar results. (Sun and Duan 2019) evaluated the passengers' loyalty towards a bus system in China considering eight constructs (environment and facilities, operation and efficiency, convenience and safety, hedonic value, perceived service cost, perceived value, expectation, and satisfaction) and five customer segments based on their loyalty level. Instead of service quality they considered the constructs "environment and facilities" and "convenience and safety". They found that the effect of these constructs on loyalty was completely mediated by perceived value and satisfaction. And finally, (Allen et al. 2020) explored how critical incidents could affect loyalty. They used a database from customer satisfaction surveys in the railways services in the hinterland of Milan, and considered satisfaction as full mediator between nine service quality factors (safety, cleanliness on board, cleanliness at station, comfort, reliability and accessibility, additional services, information, personnel, and added-value services) affected by the critical incidents, and loyalty.

It should further be noted that some studies identify both types of effects, as complete and partial mediator. (de Ona et al. 2015) analyzed the relationship between perceived service quality, satisfaction and behavioral intentions for a light rail transit system in Spain for captive and for noncaptive users, finding the direct relationship between service quality and behavioral intentions was significant for noncaptive users but non-significant for captive ones. With the same database, (Machado-Leon et al. 2016) tested three possible roles of involvement regarding passenger perception (mediator, moderator and antecedent). In one of their valid models there was no direct relationship between service quality and behavioral intentions, because satisfaction and involvement acted as full mediators. Yet other valid models (partial mediator, antecedent and moderator) showed a significant direct relationship between service quality and behavioral intentions. And finally, (Allen et al. 2019a) analyzed the relationship between specific service quality attributes, satisfaction, loyalty and involvement, using data obtained from four different surveys of a metro system in Spain. Each survey collected users' perception about six (or seven) different service quality attributes alone. After building a model for each one of the surveys, they found that in one of the models there was no direct relationship between service quality and loyalty, while this relationship was indeed significant for the other three models.

3.- Research models, hypotheses, and structural equation modeling approach

3.1. Rival conceptualizations and hypothesis.

Thus far no consensus has been reached regarding the precise role of satisfaction between service quality and loyalty or behavioral intentions in the transport literature. This article is meant to help clarify the relationships among service quality, satisfaction and behavioral intentions or loyalty for the users of urban and metropolitan public transport, comparing the results obtained by means of two alternative models (Figure 1) with independent samples from five cities.

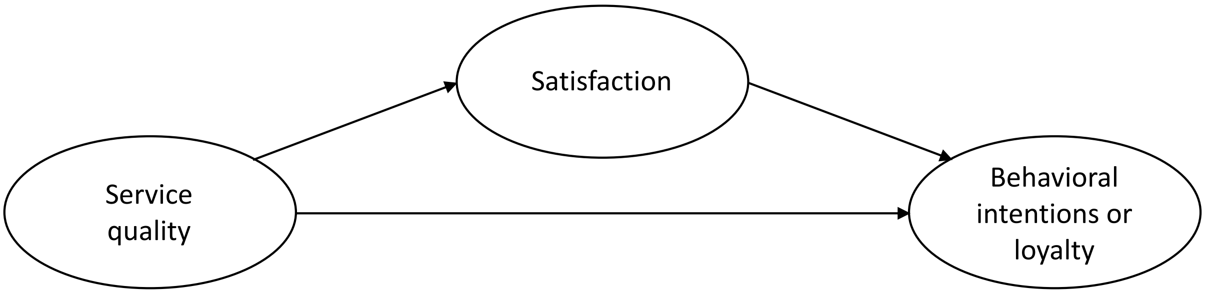
In model (A), satisfaction partially mediates the effect of service quality on behavioral intentions (or loyalty); so, service quality has direct and indirect effects on behavioral intentions through satisfaction. In model (B), satisfaction fully mediates the effect of service quality on behavioral intentions (or loyalty); hence, the effect of service quality on behavioral intentions is completely mediated by satisfaction, and service quality has only an indirect effect on behavioral intentions.

The following hypotheses are proposed:

- Model A (partial mediator model):
    - H1A: Service quality has a positive direct effect on satisfaction
    - H2A: Satisfaction has a positive direct effect on behavioral intentions or loyalty
    - H3A: Service quality has a positive direct effect on behavioral intentions or loyalty
    - H4A: Service quality has a positive indirect effect on behavioral intentions or loyalty
- Model B (full mediator model):
    - H1B: Service quality has a positive direct effect on satisfaction
    - H2B: Satisfaction has a positive direct effect on behavioral intentions or loyalty

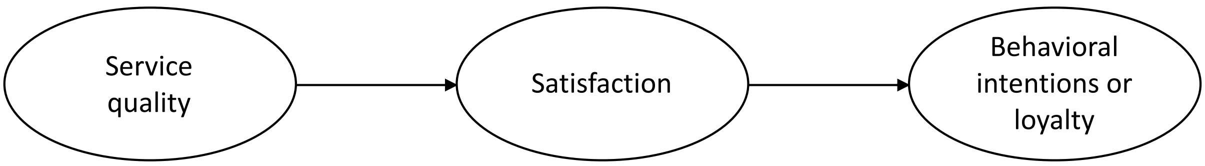
- H3B: Service quality has a positive indirect effect on behavioral intentions or loyalty

Model A. Satisfaction as partial mediator



H1A: Service quality has a positive direct effect on satisfaction
H2A: Satisfaction has a positive direct effect on behavioral intentions or loyalty
H3A: Service quality has a positive direct effect on behavioral intentions or loyalty
H4A: Service quality has a positive indirect effect on behavioral intentions or loyalty

Model B. Satisfaction as full mediator



H1B: Service quality has a positive direct effect on satisfaction
H2B: Satisfaction has a positive direct effect on behavioral intentions or loyalty
H3B: Service quality has a positive indirect effect on behavioral intentions or loyalty

Figure 1.- Competing research models: (A) satisfaction as partial mediator versus (B) satisfaction as full mediator

3.2. General procedure for validating the relationship between service quality, satisfaction and behavioral intentions or loyalty

Most previous studies apply structural equation modelling (SEM) to identify the mediator effect of satisfaction. Unlike other multivariate techniques, SEM examines more than one relationship at a time, considering both covariance structure and mean structure (Kline 2015). SEM is the synthesis of a structural regression (SR) model and a measurement model that assesses unobserved latent variables or factors as linear functions of observed variables. The SR model shows the direction and strengths of the relationships between the latent variables. Additionally, latent variables are classified as endogenous (dependent) or exogenous (independent). SEM offers certain advantages with respect to other statistical methods (Golob 2003): it treats endogenous and exogenous variables as random variables with measurement error; it explains latent variables with multiple indicators; it allows for overall testing of the model fit; and it is capable of handling non-normal data.

A valid measurement model is needed before proceeding to evaluate the SR part of the model. (Kline 2015) suggest a two-step modeling approach: in the first step, the SR model is re-specified as a Confirmatory Factor Analysis (CFA) measurement model; then, given an acceptable measurement model, the different 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.

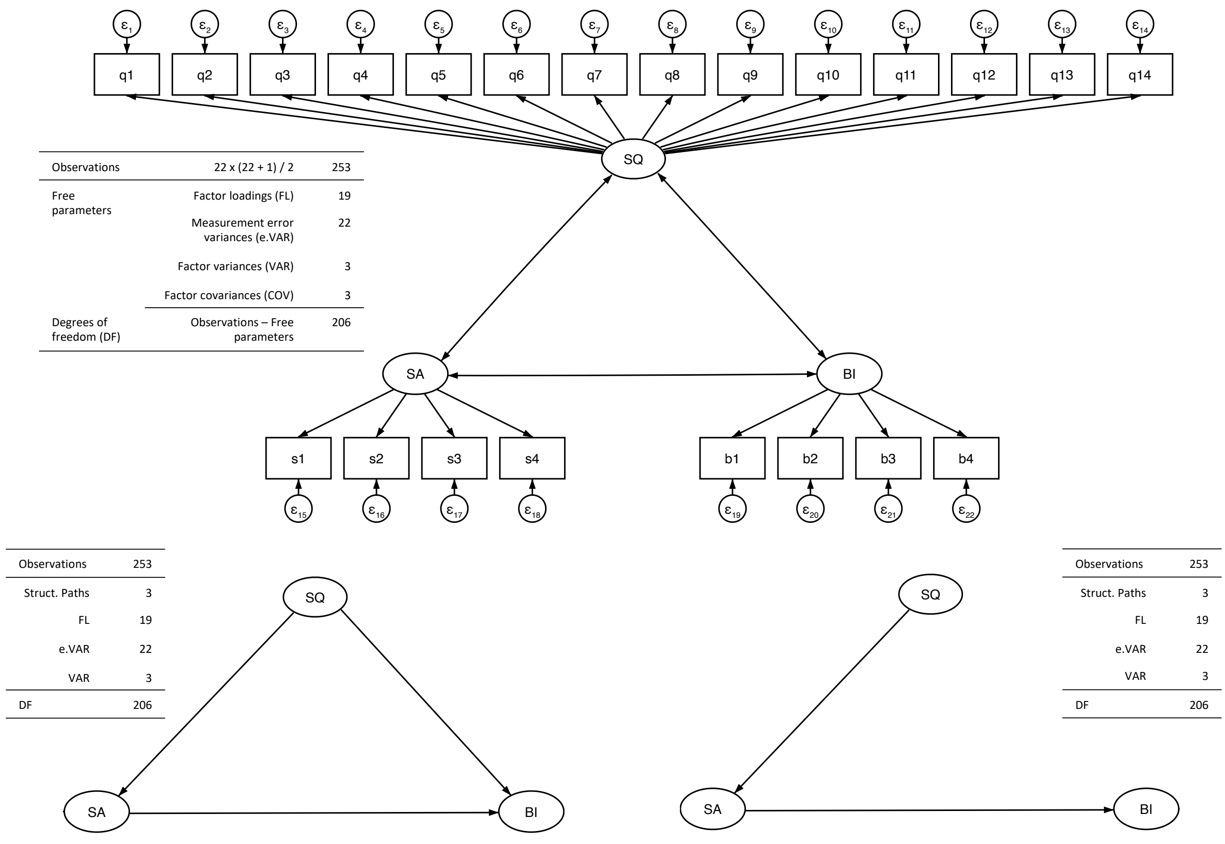
Table 2.- General tasks and specific steps for testing and validating the competing models

Task	Step Description	Step
Preliminary	Specify the model and evaluate model identification	1
	Operationalize the constructs for the measurement model	2
	Data collection, preparation and screening	3
Confirmatory Factor Analysis (CFA)	Measurement model estimation	4
	Evaluate model fit and interpret parameters estimates	5
	Respecification of measurement model	6
Structural Regression Model	Structural model estimation	7
	Evaluate model fit and interpret parameters estimates	8
	Compare competing models (Model A versus Model B) (Figure 1)	9
Validation	Replicate step 3 to 9 with the other independent samples	10

The steps followed in this research, as shown in Table 2, can be divided into four main endeavors. The first entails model specification and model identification evaluation (Step 1), that is, the theoretical proposal of the model seen in Figure 2. The three models evaluated at this stage are: the CFA measurement model, SR model A (partial mediator model) and SR model B (full mediator model). Before collecting data, it is necessary to verify if the models are identified —a model is identified if it is possible to derive a unique estimate of every model parameter. This is a property of the model itself, not of the sample. According to (Kline 2015), all models in Figure 2 are identified for the following reasons: (i) their degrees of freedom are positive (overidentified models); (ii) the CFA model has three factors with two or more indicators per factor; and (iii) both SR models are recursive (uncorrelated disturbances and all effects are unidirectional). In light of the literature review and a pilot survey, multiple-indicator measurements were considered —those for which more than one observed indicator is used to measure the same construct. In this way 14 indicators (q1-q14) were selected to operationalize the construct service quality (SQ), plus four items to define the constructs of satisfaction (s1-s4) and behavioral intentions (b1-b4) (Step 2). Although the minimum needed to ensure identification of the model is two indicators per each one of the latent factors (Kline 2015), is was deemed wise to increase them to four in case any of those items had to be discarded during analysis. These preliminary tasks led us to successfully finish data collection, preparation and screening (Step 3), though preparation and screening involved checking sample size, missing values, outliers, collinearity, relative variances, and univariate and multivariate normality. All the statistical analysis was performed using Stata/MP 16.1.

Secondly, the CFA measurement model is estimated (Step 4), evaluating not only the model fit but also interpreting the parameter estimates (Step 5). The maximum likelihood (ML) is the method used most often in the literature for estimation. However, it assumes multivariate normality for the population distributions of the endogenous variables. The distributions in customer satisfaction surveys, using ordinal scales, are generally non-normal and would need an alternative estimation method. Several have been developed to adjust ML estimators to account for non-normality (e.g., Satorra-Bentler). The robustness of ML estimation and the correction factors developed for non-normal data make SEM suitable for use with ML estimation in many situations with ordinal scales (Golob 2003). The model fit can be evaluated using absolute fit indices (χ^2, RMSEA and SRMR)—Stata does not provide GFI— and incremental or comparative fit indices (CFI and TLI), or parsimony fit indices (AIC and BIC), adopting the following suggested thresholds (Hooper, Coughlan and Mullen 2008): RMSEA < 0.08, CFI and TLI > 0.95 and SRMR < 0.05. The construct validity of the model is assessed by analyzing four components: convergent validity, average variance extracted, construct reliability, and discriminant validity (Hair et al. 2010). Convergent validity means that the indicators are related to a construct converge or share a high proportion of variance. The amount of convergent validity can be assumed as satisfactory if the factor loadings of the indicators related to a construct are statistically significant, and ideally higher than 0.7. Moreover, the average variance extracted (AVE) is calculated as the mean variance extracted for the indicator loading on a construct, its recommended value being 0.5 or higher. Good reliability, assessed by Cronbach's alpha and construct reliability, indicated by a value of 0.7 or more. Lastly, discriminant validity refers to the fact that a construct is unique and captures some phenomena that another measure does not explain. If the AVE for any two constructs is higher than the squared intercorrelation between these two constructs, we have discriminant validity. This task is completed with the possible re-specification of the CFA model (Step 6). The models presented in Figure 2 are particular in that each indicator loads on a single factor (restricted factor models) and the indicators' error terms are independent. Because this is a confirmatory analysis, for model re-specification it is not possible for each indicator to load on ≥ 2 factors. However, correlated error terms are considered for re-specification. An error correlation reflects the assumption that the two corresponding indicators have something in common that is not explicitly represented in the model, which occurs quite frequently (Kline 2015, Li and Petrick 2010). For re-specification purposes, the specific parameter estimates are evaluated based on the indicators' error variance, modification indices, and correlation residuals.

(1) CFA measurement model



(2) Satisfaction as partial mediator **(2) Satisfaction as full mediator**

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

Below, two SR models are estimated (Step 7), evaluating the model fit and interpreting the parameter estimates (Step 8). The following step is to compare the two competing models proposed (Step 9): the partial mediator model, where satisfaction partly mediates between service quality and behavioral intentions; and the complete mediator model, where satisfaction fully mediates the relationship (Figures 1 and 2). Finally, for validation purposes, steps three to nine are replicated with the data collected from the five cities and the pooled data.

4.- Case studies, data collection, preparation and screening

4.1. Case studies description in five European countries

The case studies selected were the metropolitan area in five European cities: Madrid, Rome, Lisbon, Berlin and London. Table 3 provides some key data and statistics for the metropolitan areas under study: area's definition, surface, population, density, transit authority, public transport options in the area, and modal split. The size of the metropolitan area ranges from 892 km^2 in Greater Berlin to 8,030 km^2 in Madrid Metropolitan Area; population ranges from 2.89 million inhabitants in Lisbon Metropolitan Area to 8.90 in Greater London; and population density ranges from 811 inhabitants per km^2 in Madrid Metropolitan Area to 5,672 in Greater London.

In all the metropolitan areas the public transport options include the commuter rail, metro, tram and bus services. Almost all of them also have public bicycle systems available, with the exception of Rome. Lisbon and London also include the ferry as another public transport option, although its participation in modal split is residual.

Table 3 shows that Greater Berlin and Greater London present the lowest private car use, with values ranging from 30% to 37%. Both areas also present the highest density, with values ranging from 3,948 inhabitants per km^2 in Greater Berlin to 5,672 in Greater London. Private car use in the other three areas ranges from 40% in Madrid Metropolitan Area to 60% in the Metropolitan City of Rome. Public transport use ranges from 20% in the Metropolitan City of Rome to 35% in Greater London. Greater Berlin, which is also the smallest area under study, presents the highest walking or bicycle usage (44%). In the other four areas this value ranges from 16% in Lisbon Metropolitan Area to 31% in Madrid Metropolitan Area.

4.2. Survey description

The data was collected through an online panel survey from May to July 2019. The questionnaire was designed based on an extensive literature review and was validated through a pilot survey. It was translated into the local language, and took an average duration of seven minutes to complete. The items were divided into eight modules, since it was carried out in the framework of a larger research project involving public transport users as well as users of private vehicles.

This study focuses on individuals who reported that they mainly use public transport for their regular daily trips (regular understood to be a trip that they carry out for work/occupation/daily activities). Thus, only the following four modules of the survey were analyzed:

Table 3.- Key data and statistics per city: Madrid, Rome, Berlin, Lisbon and London.

	Madrid*	Rome**	Berlin**	Lisbon**	London**
Definition of analysis area	Madrid Metropolitan Area	Metropolitan City of Rome	Greater Berlin	Lisbon Metropolitan Area	Greater London
Analysis area (km^2)	8,030	5,363	892	3,015	1,569
Population (inhab)	6.51M (2017)	4.34M (2019)	3.52 (2015)	2.89M (2014)	8.90 (2018)
Population density (inhab/km^2)	811	812	3,948	957	5,672
Major transit authority	Regional Transportation Consortium of Madrid	Rome Mobility Services	Berliner Verkehrsbetriebe	Metropolitano de Lisboa, Carris, Transtejo	Transport for London
Public transport options	Rail, metro, tram, bus, bike	Rail, metro, tram, bus	Rail, metro, tram, bus, bike	Rail, metro, tram, bus, ferry, bike	Rail, metro, tram, bus, ferry, bike
Modal split					
Private car	40%	60%	30%	54%	37%
Public transport	28%	20%	22%	29%	35%
Walking/Bicycle	31%	20%	44%	16%	28%
Other	1%	0%	4%	1%	0%

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

i. Customer and usage profile, which included sociodemographic characteristics (gender, age, length of time living in the area, household location, education level, occupation status, family size, dependent members in the family, and net household income) and frequency of public transport use.
ii. Service quality, which included 14 quality indicators about the public transport system (q1-q14): *service hours; proximity of stops to starting point or destination of the trip; frequency or number of daily services; punctuality; speed; cost; ease of entrance and exit from the vehicle and/or stations; ease of transfers/good connections with other modes of transport; individual space available inside the vehicle; temperature inside the vehicle; cleanliness of the vehicle and stations; safety on board (regarding accidents); safety regarding robbery and violence; and, information provided*. Indicators were rated on a 5-point scale from 1 (very low quality) to 5 (very high quality).
iii. Satisfaction, which included four satisfaction statements about the public transport system (s1-s4): *in general, I am satisfied with the public transport service provided in XYZ; the public transport service in XYZ meets my expectations; with the existing modes of transport in XYZ, I consider that the commuting needs of inhabitants are well covered; and, when I take public transport in XYZ, I feel very satisfied*. XYZ was the name of the city. Indicators were rated on a 5-point Likert scale from 1 (completely disagree) to 5 (completely agree).
iv. Behavioral intentions, which included four intentions statements about the public transport system (b1-b4): *in the next few weeks I will take public transport for one-off trips; in the next few weeks I will take public transport for my regular trips; I am sure I will increase the number of times I use public transport in the future; and, not only will I use public transport, but I will also recommend it to friends and family*. Indicators were rated on a 5-point Likert scale from 1 (completely disagree) to 5 (completely agree).

4.3. Samples and main survey statistics

The questionnaire was completed by 2,579 persons, with a minimum of 500 respondents in each city. Table 4 shows the sample data and summarizes the sample's main sociodemographic and travel characteristics. The number of surveys completed per city was very well balanced.

In the pooled sample there were more females (58%), the biggest age group was between 25 and 44 (43%), and the smallest age groups were under 24 (12%) and over 65 (9%). Most public transport users had been living in the same area for "a few years" (38%) or "all their life" (57%), and their household was located mainly in the metropolitan area (59%). However, this changed drastically in Berlin, where most respondents lived in the city center (78%). The educational level was very similar across cities, as most users had Secondary education (42%) or Higher education (52%). Berlin´s sample population scored somewhat higher than the average in Secondary education (56%) and London was the highest in Higher education (70%).

In all the cities, the public transport users tended to be a third-party employee (59%) or a professional (12%). In the case of Berlin, the number of retired people was also very high (20%). Most respondents had a family size of two (31%) or three (23%) people, except in Berlin, where the largest group was unipersonal households (40%). This would be due to the higher percentage of retired respondents in Berlin. Most public transport users (76%) did not have dependent members in the household (i.e., children or other dependent relatives), and they used the service every day or nearly every day (5-6 days per week, 65%) or frequently (3-4 days per week, 24%).

Table 4.- Sample, sociodemographic and travel characteristics

Category	Group	Madrid	Rome	Berlin	Lisbon	London	All
Sample (n)	n	525	509	508	530	507	2,579
	%	20.4%	19.7%	19.7%	20.6%	19.7%	100.0%
Gender	Male	40.2%	40.7%	44.9%	43.4%	41.2%	42.1%
	Female	59.8%	59.3%	55.1%	56.6%	58.8%	57.9%
Age (years)	18-24	9.7%	13.2%	10.6%	16.6%	7.1%	11.5%
	25-44	43.8%	43.0%	38.8%	43.0%	43.8%	42.5%
	45-64	35.2%	40.3%	37.2%	31.7%	41.8%	37.2%
	65 or more	11.2%	3.5%	13.4%	8.7%	7.3%	8.8%
Time living in the area	< 1 year	4.8%	4.3%	3.6%	4.4%	5.7%	4.6%
	A few years	30.0%	36.1%	45.0%	40.9%	38.3%	38.0%
	All my life	65.3%	59.6%	51.5%	54.8%	55.9%	57.4%
Household location	Metropolitan area	59.8%	69.4%	22.2%	69.6%	67.3%	57.8%
	City center	40.2%	30.6%	77.8%	30.4%	32.7%	42.2%
Education level	Primary schools or less	6.9%	4.3%	4.9%	7.7%	4.9%	5.8%
	Secondary schools	41.3%	47.3%	56.1%	40.0%	25.2%	42.0%
	Higher education	51.2%	47.5%	38.0%	51.7%	69.6%	51.6%
Occupation status	Professional	7.5%	20.9%	8.7%	12.3%	9.9%	11.8%
	Employed	66.9%	46.6%	53.7%	57.8%	68.0%	58.7%
	Student	8.2%	11.5%	7.0%	11.2%	3.2%	8.2%
	Retired/Pensioner	9.9%	4.2%	19.7%	9.1%	7.3%	10.0%
	Other	7.5%	16.8%	10.9%	9.7%	11.7%	11.3%
Family size	1	8.8%	10.8%	40.3%	11.0%	17.3%	17.6%
	2	31.2%	23.6%	36.3%	31.6%	30.0%	30.6%
	3	27.6%	28.2%	12.1%	27.7%	20.4%	23.3%
	4	24.9%	28.8%	9.1%	22.4%	21.6%	21.4%
	5 or more	7.5%	8.6%	2.2%	7.3%	10.6%	7.2%
Dependent members in the household	No	74.1%	64.8%	88.3%	70.8%	81.4%	75.7%
	Yes	25.9%	35.2%	11.7%	29.2%	18.6%	24.3%
Net household income	<2 minimum wages	31.6%	46.5%	71.1%	44.4%	45.5%	47.8%
	2-3 minimum wages	32.7%	24.5%	18.2%	29.4%	22.9%	25.6%
	>3 minimum wages	35.6%	29.0%	10.7%	26.2%	31.6%	26.6%
Frequency of use (days/week)	0-2	9.7%	12.0%	10.2%	11.3%	12.8%	11.2%
	3-4	18.5%	26.1%	22.2%	19.1%	34.1%	23.9%
	5-7	71.8%	61.9%	67.5%	69.6%	53.1%	64.9%

The most important differences across cities were in the net household income. The net incomes were calculated based on the minimum wages for each country in 2018: 900€ in Spain, 700€ in Portugal, 1,000€ in Italy, 1,300€ in UK and 1,560€ in Germany. Actually, because Italy has no established minimum wage, it was calculated with reference to Spain, bearing in mind the rent per capita of each country. The users in Madrid were distributed homogeneously among the three levels of net household income considered. In Rome, Lisbon and London, the users were also quite evenly distributed, though with a somewhat higher proportion in the lowest income level (around

45%). Yet in Berlin, the survey respondents were mostly in the groups of less than two minimum wages (71%) and between two and three minimum wages (18%).

It can therefore be stated that the user profiles of each one of the five cities analyzed had a number of aspects largely in common (e.g., gender, age, time living in the area, educational level, dependent members in the household and frequency of use). Only certain characteristics varied substantially from one city to the next, namely: household location, family size and net household income).

Table 5.- Survey service quality, satisfaction and behavioral intentions results (5-point scales)

Factor	Indicator	Madrid	Rome	Berlin	Lisbon	London	All
Service Quality	q1. Service hours	3.63	2.88	3.89	3.07	4.11	3.51
	q2. Proximity	3.91	3.56	4.09	3.65	3.93	3.83
	q3. Frequency	3.47	2.77	3.83	2.95	3.96	3.39
	q4. Punctuality	3.33	2.51	3.21	2.60	3.64	3.05
	q5. Speed	3.69	3.02	3.72	3.29	3.74	3.49
	q6. Cost	3.42	3.35	3.37	3.49	3.58	3.44
	q7. Accessibility	3.80	3.07	3.84	3.54	3.87	3.62
	q8. Intermodality	3.84	2.90	3.81	3.52	3.95	3.61
	q9. Individual space	3.18	2.71	3.10	2.79	3.29	3.01
	q10. Temperature	3.37	2.76	3.28	2.78	3.38	3.11
	q11. Cleanliness	3.49	2.50	3.18	3.03	3.47	3.14
	q12. Safety	3.84	3.14	3.96	3.38	3.96	3.66
	q13. Security	3.16	2.77	3.45	3.03	3.62	3.20
	q14. Information	3.55	2.95	3.47	3.07	3.84	3.37
Satisfaction	s1. General satisfaction	3.65	2.62	3.76	3.15	3.81	3.40
	s2. Expetations	3.60	2.57	3.60	2.98	3.82	3.31
	s3. Needs	3.65	2.47	3.48	3.01	3.78	3.28
	s4. Global experience	3.52	2.46	3.58	3.05	3.68	3.26
Behavioral Intentions	b1. I will use PT for one-off trips	4.06	3.73	3.92	3.73	3.99	3.89
	b2. I will use PT for regular trips	4.48	4.15	4.34	4.45	4.36	4.36
	b3. I will increase PT usage	3.72	3.59	3.34	3.54	3.72	3.58
	b4. I will recommend PT	3.82	3.33	3.40	3.59	3.79	3.59

Table 5 shows the survey results in terms of service quality, satisfaction and behavioral intentions for the pooled sample and for each city independently. London and Berlin presented the best appraisals of their public transport systems in nearly all the indicators, though Madrid was very close behind Berlin; Rome gave the poorest assessment.

London presented the highest average values for all the service quality indicators and satisfaction, followed by Berlin and Madrid. Among the pooled data, *proximity*, *safety* and *accessibility* were the best rated indicators for service quality, while *individual space*, *punctuality* and *temperature* were the worst rated indicators. There was total agreement overall in the cities´ perception of *individual space* as one of the three poorest rated indicators. There was also broad agreement regarding *punctuality* (only the public transport users of London did not give it one of the three lowest values). In the positive aspects there was also a good degree of agreement regarding *safety* and *proximity*. However, there were also some differences deserving mention here: Madrid gave a high rating to *intermodality*, but one of the lowest to *security*; one of Rome´s best appraised aspects was *cost*, and one of the worst, *cleanliness*; Berlin highly appraised *service hours* but disapproved of *cleanliness*; Lisbon gave high value to *accessibility* and *intermodality*, but *temperature* merited low appraisal; and London included within its best-appraised aspects *service hours* and *frequency*, while two of the worst-rated were *temperature* and *cleanliness*.

Although London and Berlin were the cities harvesting the best ratings for the indicators service quality and satisfaction, Madrid gave the highest ratings of all to the indicators of behavioral intentions. Also noteworthy is the fact that Rome and Lisbon, presenting the lowest ratings in service quality and satisfaction, secured appraisals very similar to the other three cities for the indicators of behavioral intentions. Indeed, for some indicators (e.g., *I will increase PT usage* or *I will recommend PT*) their average values were greater than those of Berlin. These findings agree with the results of (Oliver 1999), who underlined that a greater degree of satisfaction does not necessarily imply more frequent use or recommendation of the service.

4.4. Data preparation and screening

The data from each country were considered separately for all the checks required for data preparation and screening described in this section.

Maximum likelihood (ML) estimation requires a sufficient sample size, particularly when non-normal data are involved. The literature provides several sample size recommendations: (i) sample size for ML estimation should amount to at least five times the number of free parameters in the model (Bentler and Chou 1987); (ii) with strongly kurtotic data, the minimum sample size should be 10 times the number of free parameters (Hoogland and Boomsma 1998); (iii) sample size for ML estimation should be at least 15 times the number of observed variables (Stevens 2009); and (iv) a minimum sample of 200 is needed to reduce biases to an acceptable level for any type of SEM estimation (Kline 2015). Table 4 shows that the minimum sample size was 507 (London), and that it was over the limit for the most restrictive criteria (Hoogland and Boomsma 1998) involved in the CFA measurement model (47 free parameters, Figure 2).

The screening for outliers, missing values and relative variances followed the procedure recommended by (Kline 2015). No extreme outliers were identified when checking for univariate and multivariate outliers. For all countries, the indicator with the highest number of missing values (4.5% on average) was *I will increase PT usage*. As this value was below 5% and did not show a systematic pattern, this data loss was of little concern and could be considered as missing completely at random. Finally, as all the attributes used the 5-point scale, the covariance matrices were not ill-scaled (i.e., when the ratio of the largest to the smallest variance is greater than 10).

Extreme collinearity may occur when what appear to be separate variables measure the same thing. (Kline 2015) recommend eliminating variables or combining redundant ones into a composite if extreme collinearity appears. In our study, three methods were used to detect multi-collinearity. First, we calculated bivariate correlation among all the variables, finding that the greatest values corresponded to the satisfaction indicators in Rome (around 0.85). Second, we calculated squared multiple correlations (R^2) between each variable and all the rest, running several multiple regressions, each with a different variable as the criterion and the rest as predictors, and we observed that $R^2 < 0.83$ in all cases. Bivariate and multivariate correlations below 0.90 suggest no multi-collinearity. Furthermore, we calculated the variance inflator factor (VIF) after each regression and found all the values to be below six (a variable with a VIF>10 may be redundant). Therefore, the sample data did not present multi-collinearity problems.

Finally, we checked for univariate and multivariate normality. Univariate normality was tested using the Shapiro-Wilk test. Results showed that most of the variables, above all for Madrid, Berlin and London, were not normally distributed —they presented negative skewness owing to the sound appraisals by users. 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, we used the Satorra-Bentler estimator to control for non-normality. The following results report the corrected χ^2 using this estimator, as well as all the corrected model fit indices that use χ^2. We also explored the possibility of using the SEM-ordinal Probit framework, recently suggested by (Allen et al. 2020) as ideal for ordinal style ratings. However, this approach involves much more parameters and its dataset requirements were not compatible with the sample size of each one of the independent samples considered in this study (around 500 observations).

5.- Confirmatory Factor Analysis measurement models

Before performing the CFA measurement model, the dimensionality of the service quality factor was examined. Although Figure 1 proposes the unidimensionality of the 14 items based on previous studies in the service quality field (Cronin and Taylor 1992, Chou and Kim 2009), it is best practice to run an exploratory principal component analysis (PCA) to identify the dimensionality (multidimensional or unidimensional) of service quality. We performed a PCA using the VARIMAX rotation with Stata/MP 16.1 for the pooled sample. All the items loaded on a single factor with positive loadings, ranging from 0.538 for *cost* to 0.790 for *punctuality*, and a Cronbach's alpha value of 0.930 (good reliability). These results suggested that the service quality scale can be treated as unidimensional.

At the following step, before testing for significant relationships in the SR model, one must demonstrate that the measures have satisfactory psychometric properties. These properties were evaluated using a CFA measurement model via the use of Stata/MP 16.1. All three scales were tested simultaneously in one confirmatory factor model, where all factors involved were assumed to covary with each other (Kline 2015).

Six CFA measurement models were performed (data from the five cities and the pooled data), considering all the 22 indicators and three factors (Figure 2). Each scale indicator was only allowed to load on one factor and could not cross-load on any other factors. The unstandardized loading of *service hours*, *general satisfaction* and *I will use PT for one-off trips* were each fixed to 1.0 to scale each of the factors. With 22 indicators, there are 253 observations available to estimate a total of 47 free parameters, including 25 variances of exogenous variables (three factors and 22 measurement errors), three factor covariances and 19 factor loadings, so $df_M = 206$. The six models' estimations in Stata converged to admissible solutions.

Table A.1 (see Appendix) displays the values of selected fit indices for each one of the models. The initial CFA measurement model (Figure 2) failed the exact-fit test, with $\chi^2(206)$ ranging from 409.22 to 1,437.59 ($p<0.001$). As this is generally the case in the literature, we considered other approximate fit indices to diagnose the possible sources of misfit. The approximate fit indices likewise showed some misfit, with CFI ranging from 0.913 to 0.959 (>0.95), TLI from 0.902 to 0.953

(>0.95) and SRMR from 0.036 to 0.060 (<0.05). RMSEA was the only fit index below the suggested threshold (<0.08) in all cases.

Even though the three factors considered were made up of indicators similar to those often used in the literature, it was deemed necessary to prove their validity. The construct validity was assessed by analyzing convergent validity, average variance extracted, construct reliability and discriminant validity (Table A.1). This table also shows that all factor loadings were statistically significant, presenting the correct sign (positive), and that their values were higher than the 0.5 suggested by (Hair et al. 2010) —most were above the ideal threshold of 0.7, the exceptions being *cost*, *I will use PT for one-off trips* and *I will use PT for regular trips*. The behavioral intention indicators gave very low factor loadings, ranging from 0.234 to 0.424 for *I will use PT for regular trips*, and from 0.283 to 0.517 for *I will use PT for one-off trips*. Table A.1 also shows that standardized error variances for both these indicators were very high, ranging from 0.733 to 0.945. Moreover, the factor behavioral intentions presented very low values for Construct Reliability (CR), Average Variance Extracted (AVE) and Cronbach's Alpha. CR ranged from 0.573 to 0.737, AVE from 0.292 to 0.429, and Cronbach's Alpha from 0.556 to 0.725. Most of the parameters estimated for service quality and satisfaction were above the suggested thresholds, presenting very high CR: values ranging from 0.911 to 0.950 for service quality, and 0.897 to 0.957 for satisfaction. Cronbach's alpha ranges from 0.903 to 0.948 for service quality and from 0.892 to 0.955 for satisfaction. Service quality presents only one factor loading below 0.5 (*cost* ranges from 0.439 and 0.603) for two samples (Lisbon and the pooled data). However, as no other service quality indicators were below the threshold for factor loadings, and both values were above 0.4 (recently suggested by (Allen et al. 2019a)), *cost* was retained in the model. Satisfaction presented AVE values ranging from 0.684 and 0.847, while service quality values ranged from 0.424 to 0.580. While some of these values were below the recommended threshold (0.5), (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 0.500 to 0.892. Thus, not excessively high factor correlations (e.g., below 0.90) suggest discriminant validity (Kline 2015).

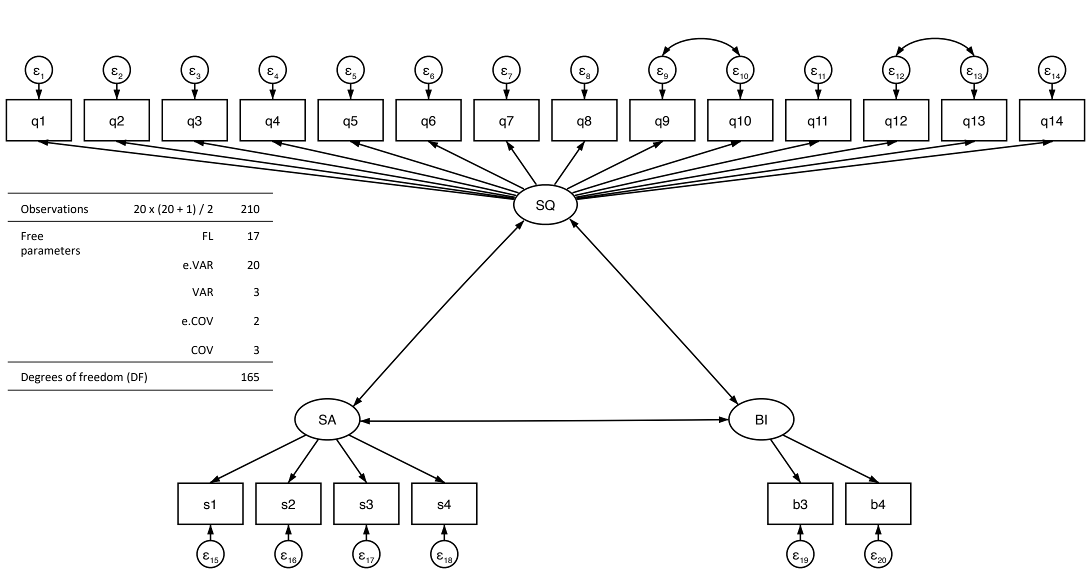


Figure 3.- Final CFA measurement model

In view of the limitations and the adjustment of the models, along with the problems of validity and reliability identified in the factor behavioral intentions, we proceeded to re-specify the model. As the first step, the possible exclusion of the indicators *I will use PT for one-off* trips (b1) and *I will use PT for regular trips* (b2) was considered, after which inspection of the correlation residuals and modification indexes was resumed. Although the model's re-specification was undertaken step-by-step, as the misfit sources can vary when changes are made to the model, because of space limitations only the results of the final CFA measurement model are offered. Figure 3 shows the final CFA measurement model.

The final CFA measurement model differs from the initial model in the following ways:

- The factor behavioral intentions comprised just two indicators (*Increase usage* and *I will recommend PT*). Both factor loadings were higher than 0.5, and CR, AVE and Cronbach's alpha improved upon the initial model, with almost all CR and Cronbach's alpha above the 0.7 threshold, and all AVE above 0.5.
- The specification of two error correlations for all the samples: (i) between *individual space* (q9) and *temperature* (q10); and (ii) between *safety* (q12) and *security* (q13). It is logical that *individual space* and *temperature*, closely tied to comfort in public transport, shared something that was not explicitly represented in the model. The same occurs with *safety* and *security*. Both correlations were thus plausible and theoretically justified.

Figure 3 shows the final CFA measurement model, considering only 20 indicators and three factors. In this case, there were 210 observations available to estimate a total of 45 free parameters, including 23 variances of exogenous variables (three factors and 20 measurement errors), three factor covariances, two measurement error covariances and 17 factor loadings, so $df_M = 165$. The six models' estimations in Stata converged to admissible solutions.

All the parameters for checking the validity and reliability of the scales were kept or improved in the final CFA measurement model when compared to the initial one. Table A.2 (see Appendix) shows that construct validity and reliability scores were maintained in the case of service quality and satisfaction, and improved for behavioral intentions. All factor loadings were statistically significant, presented the correct sign, and were higher than 0.5 (most of them even above 0.7), with the exception of *cost* for Lisbon and the pooled data. Table A.2 also indicates measurement error variances, factor variances and covariances for the final CFA measurement model. In this model, the estimated factor correlations decreased, ranging from 0.415 to 0.890. The correlations between the measurement errors were significant and had values ranging from 0.139 to 0.351 between *individual space* and *temperature*, and from 0.133 to 0.347 between *safety* and *security*.

Although the final CFA measurement model also failed the exact-fit test, with $\chi^2(165)$ ranging from 295.93 to 923.81 ($p<0.001$), all the approximate fit indices improved: RMSEA ranged from 0.043 to 0.054, CFI from 0.941 to 0.974, TLI from 0.932 to 0.970, and SRMR from 0.028 to 0.047.

6.- Structural Regression models

Since enough evidence was found to assess the construct validity of the measurement model, the next step was to test the competing models (Figure 2), estimating two SR models for each one of the six datasets (pooled data and five independent samples). For each sample, the comparison of the models was based on BIC statistics and the values of the parameter estimates. Although both models were nested, the SR models were not compared using the χ^2 differences test. The estimation process entailed the use of Satorra-Bentler statistics, which adjusted the value of χ^2 from the standard maximum likelihood estimation by an amount reflecting the degree of non-normality. The difference between the Satorra–Bentler statistics for two nested models fitted to the same data does not follow a χ^2 distribution (Kline 2015).

The partial mediator model considered 20 indicators and three factors. It presented the same number of free parameters as the final CFA measurement model because all the correlations between latent factors were changed to directional paths. These models were equivalent versions and generated the same predicted correlations, covariances and fit statistics (Kline 2015). Table 6 shows the following results for the structural regression model with satisfaction as partial mediator: structural paths, factor loadings, measurement error variances, factor variances and covariances, and the model's fit statistics (using Satorra-Bentler estimation).

Table 7 shows the same results for the structural regression model with satisfaction as full mediator. The full mediator model also considered 20 indicators and three factors. In this case the number of free parameters was reduced by one, as there were only two directional paths instead of three factor correlations. Thus, the degrees of freedom increased by one, and the fit statistics were different from the CFA measurement model.

The twelve-model estimation in Stata converged to admissible solutions. Although all the full mediator models failed the exact-fit test, with $\chi^2(166)$ ranging from 301.80 to 924.01 ($p<0.001$), the approximate fit indices showed very similar values to those of the partial mediator models (RMSEA ranged from 0.044 to 0.056, CFI from 0.941 to 0.974, TLI from 0.932 to 0.970 and SRMR from 0.028 to 0.048). Based on a comparison of models' fit statistics, the full mediator model was preferred over the partial mediator model for the pooled data and four samples (Madrid, Rome, Berlin and Lisbon), as the differences in BIC provided strong support of evidence (Raftery 1996) favoring the full mediator approach rather than the partial mediator one for the pooled sample, and positive support of evidence for Madrid, Rome, Berlin and Lisbon. In the case of London, the difference in BIC provided only weak evidence in favor of the partial mediator model.

The two models presented very similar factor variances (Tables 6 and 7). As for the path estimates, the results across the five cities were consistent with those for the pooled data in the full mediator model and all of them were also significant. Service quality had a positive effect on satisfaction, ranging from 0.823 to 0.891, and satisfaction had a positive effect on behavioral intentions, ranging from 0.523 to 0.676. In addition to the direct effect, we further examined service quality to determine whether it was indirectly related to behavioral intentions. So, the indirect relationship between service quality and behavioral intentions via satisfaction was tested and it was found to be significant in all the samples. Table 8 shows that all the hypothesis tested for the full mediator model were supported in all data samples.

Table 6 shows that the partial mediator model's results across the five cities were consistent with those for the pooled data in terms of the relationships between service quality and satisfaction, and between satisfaction and behavioral intentions. They were moreover significant ($p < 0.001$). The service quality had a positive effect on satisfaction, ranging from 0.821 to 0.890, while satisfaction had a positive effect on behavioral intentions, ranging from 0.401 to 0.637.

Table 6.- Structural regression model with satisfaction as partial mediator

Parameters	All	Madrid	Rome	Berlin	Lisbon	London
	Unst. (SE) St.	**Unst. (SE) St.**	**Unst. (SE) St.**	**Unst. (SE) St.**	**Unst. (SE) St.**	**Unst. (SE) St.**
Structural paths						
SQ->Satisfaction (SA)	1.091 (0.026) 0.882	1.097 (0.06) 0.877	1.034 (0.045) 0.89	1.095 (0.097) 0.821	1.127 (0.075) 0.839	1.413 (0.127) 0.831
SA->Behavioral intentions (BI)	0.346 (0.043) 0.513	0.41 (0.096) 0.637	0.342 (0.078) 0.495	0.473 (0.115) 0.472	0.481 (0.103) 0.61	0.325 (0.095) 0.401
SQ->Behavioral intentions (BI)	**0.027 (0.048) 0.033**	**-0.049 (0.093) -0.061**	**0.103 (0.092) 0.128**	**0.097 (0.141) 0.073**	**-0.103 (0.113) -0.098**	0.428 (0.172) 0.311
Factor loadings						
Service Quality (SQ)						
SQ->Service hours (q1)	1 (*) 0.744	1 (*) 0.73	1 (*) 0.777	1 (*) 0.654	1 (*) 0.669	1 (*) 0.586
SQ->Proximity (q2)	0.725 (0.027) 0.615	0.796 (0.061) 0.636	0.645 (0.048) 0.553	0.963 (0.093) 0.648	0.792 (0.074) 0.586	1.19 (0.114) 0.656
SQ->Frequency (q3)	1.082 (0.023) 0.784	1.11 (0.058) 0.754	1.049 (0.041) 0.832	1.153 (0.093) 0.686	1.127 (0.058) 0.749	1.261 (0.127) 0.686
SQ->Punctuality (q4)	1.113 (0.026) 0.788	1.098 (0.06) 0.764	1.085 (0.045) 0.837	1.149 (0.096) 0.672	1.213 (0.076) 0.756	1.425 (0.139) 0.707
SQ->Speed (q5)	0.945 (0.025) 0.769	0.921 (0.054) 0.739	1.015 (0.043) 0.836	1.102 (0.092) 0.713	1.003 (0.075) 0.723	1.258 (0.132) 0.676
SQ->Cost (q6)	0.652 (0.028) 0.491	0.893 (0.063) 0.607	0.647 (0.045) 0.572	0.969 (0.102) 0.541	0.694 (0.082) 0.437	1.184 (0.148) 0.507
SQ->Accessibility (q7)	0.868 (0.025) 0.705	0.863 (0.055) 0.683	0.888 (0.045) 0.743	1.078 (0.085) 0.689	0.878 (0.069) 0.64	1.134 (0.112) 0.606
SQ->Intermodality (q8)	0.96 (0.024) 0.762	0.825 (0.053) 0.674	1.039 (0.041) 0.828	1.118 (0.08) 0.731	0.959 (0.073) 0.684	1.21 (0.122) 0.689
SQ->Individual space (q9)	0.929 (0.027) 0.691	0.892 (0.063) 0.622	0.994 (0.045) 0.798	1.135 (0.103) 0.64	0.984 (0.074) 0.664	1.559 (0.165) 0.678
SQ->Temperature (q10)	0.899 (0.026) 0.681	0.889 (0.062) 0.645	0.908 (0.045) 0.768	1.037 (0.102) 0.591	0.962 (0.069) 0.64	1.361 (0.146) 0.634
SQ->Cleanliness (q11)	0.953 (0.027) 0.707	0.828 (0.061) 0.644	1.049 (0.047) 0.809	1.1 (0.11) 0.634	0.877 (0.074) 0.598	1.392 (0.14) 0.652
SQ->Safety (q12)	0.828 (0.026) 0.675	0.657 (0.061) 0.55	0.824 (0.047) 0.694	0.922 (0.089) 0.614	0.846 (0.074) 0.607	1.257 (0.115) 0.682
SQ->Security (q13)	0.898 (0.028) 0.679	0.811 (0.065) 0.58	0.974 (0.05) 0.762	1.009 (0.099) 0.624	0.892 (0.081) 0.601	1.329 (0.144) 0.654
SQ->Information (q14)	0.942 (0.026) 0.722	0.951 (0.06) 0.702	0.892 (0.046) 0.737	1.137 (0.095) 0.675	1.052 (0.075) 0.699	1.209 (0.14) 0.633
Satisfaction (SA)						
SA->General satisfaction (s1)	1 (*) 0.91	1 (*) 0.907	1 (*) 0.926	1 (*) 0.817	1 (*) 0.915	1 (*) 0.871
SA->Meet expectations (s2)	1.007 (0.012) 0.907	0.958 (0.033) 0.877	1.019 (0.027) 0.928	1.087 (0.047) 0.859	0.958 (0.03) 0.893	0.998 (0.037) 0.86
SA->Covered needs (s3)	0.993 (0.014) 0.862	0.904 (0.038) 0.822	1.002 (0.027) 0.903	1.069 (0.06) 0.771	0.966 (0.032) 0.833	1.005 (0.039) 0.838
SA->I feel satisfied (s4)	0.997 (0.012) 0.9	0.973 (0.031) 0.911	1.01 (0.025) 0.923	1.097 (0.051) 0.853	0.943 (0.029) 0.87	0.982 (0.041) 0.834
Behavioral Intentions (BI)						
BI->Increase usage (b3)	1 (*) 0.632	1 (*) 0.599	1 (*) 0.7	1 (*) 0.696	1 (*) 0.647	1 (*) 0.634
BI->I will recommend PT (b4)	1.579 (0.09) 0.931	1.664 (0.227) 0.948	1.399 (0.13) 0.85	1.355 (0.146) 0.872	1.535 (0.165) 0.967	1.265 (0.149) 0.817
Measurement error variances						
var(e.q1)	0.624 (0.023) 0.447	0.586 (0.045) 0.467	0.68 (0.052) 0.396	0.508 (0.035) 0.573	0.695 (0.05) 0.552	0.473 (0.035) 0.657
var(e.q2)	0.668 (0.022) 0.622	0.624 (0.04) 0.596	0.98 (0.059) 0.694	0.485 (0.034) 0.58	0.677 (0.042) 0.657	0.463 (0.034) 0.57

var(e.q3)	0.566 (0.021) 0.385	0.626 (0.045) 0.432	0.507 (0.041) 0.308	0.565 (0.053) 0.529	0.559 (0.045) 0.438	0.442 (0.034) 0.529
var(e.q4)	0.584 (0.021) 0.379	0.576 (0.038) 0.416	0.522 (0.04) 0.299	0.607 (0.047) 0.548	0.621 (0.044) 0.428	0.504 (0.043) 0.501
var(e.q5)	0.478 (0.018) 0.409	0.471 (0.031) 0.454	0.462 (0.035) 0.302	0.446 (0.034) 0.492	0.518 (0.037) 0.477	0.465 (0.039) 0.543
var(e.q6)	1.032 (0.03) 0.759	0.913 (0.059) 0.632	0.893 (0.057) 0.673	0.859 (0.058) 0.707	1.149 (0.066) 0.809	1.005 (0.071) 0.743
var(e.q7)	0.59 (0.021) 0.503	0.57 (0.044) 0.534	0.666 (0.049) 0.448	0.488 (0.037) 0.526	0.627 (0.042) 0.591	0.547 (0.046) 0.632
var(e.q8)	0.513 (0.017) 0.419	0.546 (0.035) 0.545	0.515 (0.043) 0.315	0.412 (0.032) 0.465	0.589 (0.035) 0.532	0.401 (0.028) 0.526
var(e.q9)	0.728 (0.023) 0.522	0.845 (0.057) 0.613	0.585 (0.044) 0.363	0.703 (0.043) 0.59	0.691 (0.046) 0.559	0.705 (0.049) 0.54
var(e.q10)	0.72 (0.023) 0.536	0.742 (0.051) 0.584	0.595 (0.04) 0.41	0.76 (0.047) 0.651	0.751 (0.049) 0.59	0.682 (0.042) 0.598
var(e.q11)	0.7 (0.022) 0.5	0.647 (0.043) 0.586	0.605 (0.043) 0.346	0.683 (0.041) 0.598	0.778 (0.042) 0.642	0.648 (0.048) 0.575
var(e.q12)	0.633 (0.022) 0.545	0.666 (0.047) 0.698	0.757 (0.051) 0.518	0.53 (0.038) 0.623	0.693 (0.057) 0.632	0.449 (0.035) 0.535
var(e.q13)	0.728 (0.026) 0.539	0.867 (0.063) 0.663	0.71 (0.046) 0.419	0.605 (0.042) 0.61	0.793 (0.064) 0.639	0.584 (0.044) 0.572
var(e.q14)	0.631 (0.022) 0.479	0.622 (0.044) 0.507	0.695 (0.053) 0.457	0.586 (0.04) 0.545	0.653 (0.042) 0.511	0.541 (0.042) 0.599
var(e.s1)	0.245 (0.013) 0.172	0.226 (0.028) 0.178	0.232 (0.024) 0.142	0.335 (0.031) 0.332	0.198 (0.019) 0.163	0.227 (0.029) 0.241
var(e.s2)	0.258 (0.015) 0.177	0.288 (0.033) 0.231	0.234 (0.032) 0.139	0.284 (0.031) 0.263	0.236 (0.031) 0.202	0.251 (0.033) 0.26
var(e.s3)	0.403 (0.018) 0.257	0.41 (0.044) 0.325	0.318 (0.036) 0.185	0.524 (0.044) 0.405	0.418 (0.032) 0.306	0.306 (0.027) 0.298
var(e.s4)	0.275 (0.014) 0.189	0.204 (0.019) 0.171	0.249 (0.027) 0.148	0.304 (0.034) 0.273	0.291 (0.029) 0.244	0.303 (0.032) 0.305
var(e.b3)	0.807 (0.04) 0.601	0.771 (0.076) 0.641	0.693 (0.074) 0.509	0.72 (0.1) 0.516	0.878 (0.095) 0.582	0.696 (0.073) 0.598
var(e.b4)	0.206 (0.065) 0.133	0.134 (0.135) 0.101	0.501 (0.121) 0.277	0.392 (0.126) 0.24	0.104 (0.129) 0.065	0.373 (0.08) 0.332
Factor variances and covariances						
var(e.SA)	0.263 (0.016) 0.223	0.241 (0.03) 0.231	0.291 (0.035) 0.207	0.219 (0.028) 0.325	0.3 (0.039) 0.295	0.222 (0.031) 0.31
var(e.BI)	0.379 (0.029) 0.706	0.284 (0.05) 0.658	0.418 (0.063) 0.626	0.484 (0.078) 0.716	0.453 (0.069) 0.718	0.25 (0.046) 0.534
var(SQ)	0.773 (0.037) 1	0.669 (0.073) 1	1.038 (0.085) 1	0.379 (0.061) 1	0.563 (0.07) 1	0.247 (0.048) 1
cov(e.q9,e.q10)	0.178 (0.017) 0.246	0.164 (0.035) 0.207	0.082 (0.03) 0.139	0.251 (0.039) 0.343	0.101 (0.032) 0.14	0.243 (0.037) 0.351
cov(e.q12,e.q13)	0.165 (0.019) 0.243	0.193 (0.042) 0.255	0.171 (0.038) 0.233	0.075 (0.028) 0.133	0.258 (0.051) 0.347	0.099 (0.032) 0.194
Model's fit statistics (Satorra-Bentler estimation)						
df (sb)	165	165	165	165	165	165
chi-square (sb)	923.81	371.93	329.79	369.68	410.68	295.93
p-value	0.000	0.000	0.000	0.000	0.000	0.000
RMSEA (sb)	0.045	0.052	0.047	0.054	0.056	0.043
CFI (sb)	0.968	0.953	0.974	0.941	0.945	0.959
TLI (sb)	0.963	0.946	0.970	0.932	0.937	0.952
SRMR	0.028	0.039	0.030	0.047	0.043	0.041
AIC	112,745.5	23,107.9	22,860.2	21,139.6	24,389.5	19,740.7
BIC	113,117.3	23,377.1	23,127.8	21,403.9	24,660.8	20,003.7

* Not tested for statistical significance. All other unstandardized estimates are statistically significant at p < 0.001, except the values in **bold**.

Table 7.- Structural regression model with satisfaction as full mediator

Parameters	All	Madrid	Rome	Berlin	Lisbon	London
	Unst. (SE) St.	Unst. (SE) St.	Unst. (SE) St.	Unst. (SE) St.	Unst. (SE) St.	Unst. (SE) St.
Structural paths						
SQ->Satisfaction (SA)	1.091 (0.026) 0.882	1.096 (0.06) 0.876	1.035 (0.044) 0.891	1.096 (0.097) 0.823	1.126 (0.075) 0.838	1.412 (0.127) 0.838
SA->Behavioral intentions (BI)	0.366 (0.024) 0.543	0.373 (0.061) 0.581	0.424 (0.042) 0.613	0.541 (0.077) 0.537	0.411 (0.057) 0.523	0.547 (0.064) 0.676
Factor loadings						
Service Quality (SQ)						
SQ->Service hours (q1)	1 (*) 0.744	1 (0 * 0.73	1 (*) 0.777	1 (*) 0.654	1 (*) 0.669	1 (*) 0.589
SQ->Proximity (q2)	0.725 (0.027) 0.615	0.796 (0.061) 0.636	0.644 (0.048) 0.552	0.962 (0.093) 0.648	0.793 (0.074) 0.586	1.184 (0.109) 0.657
SQ->Frequency (q3)	1.082 (0.023) 0.784	1.11 (0.058) 0.753	1.049 (0.041) 0.832	1.153 (0.093) 0.686	1.128 (0.059) 0.749	1.257 (0.123) 0.688
SQ->Punctuality (q4)	1.113 (0.026) 0.788	1.098 (0.06) 0.764	1.085 (0.045) 0.837	1.149 (0.096) 0.672	1.213 (0.075) 0.756	1.421 (0.136) 0.708
SQ->Speed (q5)	0.945 (0.025) 0.769	0.921 (0.054) 0.739	1.015 (0.043) 0.836	1.101 (0.092) 0.712	1.004 (0.075) 0.723	1.25 (0.128) 0.676
SQ->Cost (q6)	0.652 (0.028) 0.491	0.893 (0.063) 0.607	0.647 (0.045) 0.572	0.968 (0.102) 0.541	0.696 (0.082) 0.438	1.178 (0.145) 0.507
SQ->Accessibility (q7)	0.868 (0.025) 0.705	0.864 (0.055) 0.683	0.888 (0.045) 0.743	1.078 (0.085) 0.689	0.878 (0.069) 0.64	1.128 (0.109) 0.607
SQ->Intermodality (q8)	0.96 (0.024) 0.762	0.826 (0.053) 0.674	1.039 (0.041) 0.828	1.117 (0.08) 0.731	0.96 (0.072) 0.684	1.202 (0.117) 0.688
SQ->Individual space (q9)	0.928 (0.027) 0.691	0.893 (0.063) 0.622	0.994 (0.045) 0.798	1.134 (0.103) 0.64	0.984 (0.074) 0.664	1.544 (0.16) 0.676
SQ->Temperature (q10)	0.898 (0.026) 0.681	0.889 (0.062) 0.645	0.908 (0.045) 0.768	1.036 (0.101) 0.59	0.963 (0.069) 0.64	1.348 (0.142) 0.631
SQ->Cleanliness (q11)	0.953 (0.027) 0.707	0.828 (0.061) 0.644	1.05 (0.047) 0.809	1.1 (0.109) 0.634	0.877 (0.074) 0.598	1.383 (0.138) 0.652
SQ->Safety (q12)	0.827 (0.026) 0.675	0.658 (0.061) 0.55	0.824 (0.047) 0.694	0.921 (0.089) 0.615	0.847 (0.074) 0.607	1.248 (0.111) 0.681
SQ->Security (q13)	0.898 (0.028) 0.679	0.811 (0.065) 0.58	0.974 (0.05) 0.763	1.009 (0.099) 0.624	0.892 (0.081) 0.601	1.319 (0.14) 0.653
SQ->Information (q14)	0.942 (0.026) 0.722	0.951 (0.059) 0.702	0.892 (0.046) 0.737	1.138 (0.095) 0.675	1.052 (0.075) 0.699	1.2 (0.135) 0.632
Satisfaction (SA)						
SA->General satisfaction (s1)	1 (*) 0.91	1 (*) 0.907	1 (*) 0.926	1 (*) 0.817	1 (*) 0.916	1 (*) 0.868
SA->Meet expectations (s2)	1.007 (0.012) 0.907	0.958 (0.033) 0.877	1.019 (0.027) 0.928	1.087 (0.047) 0.858	0.958 (0.03) 0.893	0.998 (0.037) 0.857
SA->Covered needs (s3)	0.993 (0.014) 0.862	0.904 (0.038) 0.822	1.002 (0.027) 0.903	1.069 (0.06) 0.771	0.965 (0.032) 0.833	1.009 (0.039) 0.838
SA->I feel satisfied (s4)	0.997 (0.012) 0.9	0.973 (0.031) 0.911	1.01 (0.025) 0.923	1.097 (0.051) 0.852	0.942 (0.029) 0.87	0.987 (0.041) 0.835
Behavioral Intentions (BI)						
BI->Increase usage (b3)	1 (*) 0.632	1 (*) 0.598	1 (*) 0.702	1 (*) 0.699	1 (*) 0.644	1 (*) 0.632
BI->I will recommend PT (b4)	1.577 (0.089) 0.931	1.67 (0.23) 0.95	1.394 (0.127) 0.849	1.342 (0.143) 0.868	1.546 (0.169) 0.97	1.274 (0.155) 0.82

Measurement error variances						
var(e.q1)	0.624 (0.023) 0.447	0.587 (0.045) 0.468	0.68 (0.052) 0.396	0.507 (0.035) 0.572	0.695 (0.05) 0.553	0.471 (0.035) 0.653
var(e.q2)	0.668 (0.022) 0.622	0.623 (0.04) 0.595	0.981 (0.06) 0.695	0.485 (0.034) 0.58	0.676 (0.042) 0.657	0.463 (0.035) 0.569
var(e.q3)	0.566 (0.021) 0.385	0.627 (0.045) 0.432	0.507 (0.041) 0.308	0.565 (0.053) 0.529	0.559 (0.045) 0.439	0.44 (0.034) 0.526
var(e.q4)	0.584 (0.021) 0.379	0.576 (0.037) 0.417	0.521 (0.04) 0.299	0.607 (0.047) 0.548	0.622 (0.044) 0.429	0.502 (0.043) 0.498
var(e.q5)	0.478 (0.018) 0.409	0.471 (0.031) 0.454	0.462 (0.035) 0.302	0.446 (0.034) 0.493	0.518 (0.037) 0.477	0.465 (0.039) 0.543
var(e.q6)	1.033 (0.03) 0.759	0.913 (0.059) 0.631	0.894 (0.057) 0.673	0.859 (0.058) 0.707	1.148 (0.066) 0.808	1.005 (0.071) 0.743
var(e.q7)	0.59 (0.021) 0.503	0.57 (0.044) 0.533	0.666 (0.049) 0.448	0.487 (0.037) 0.525	0.628 (0.042) 0.591	0.546 (0.046) 0.632
var(e.q8)	0.513 (0.017) 0.419	0.546 (0.035) 0.545	0.515 (0.043) 0.315	0.413 (0.032) 0.466	0.59 (0.035) 0.532	0.402 (0.028) 0.526
var(e.q9)	0.729 (0.023) 0.523	0.845 (0.057) 0.613	0.586 (0.045) 0.363	0.703 (0.043) 0.59	0.691 (0.046) 0.559	0.71 (0.049) 0.544
var(e.q10)	0.72 (0.023) 0.536	0.742 (0.051) 0.584	0.595 (0.04) 0.41	0.761 (0.047) 0.652	0.751 (0.049) 0.59	0.686 (0.042) 0.601
var(e.q11)	0.7 (0.022) 0.5	0.647 (0.043) 0.586	0.604 (0.043) 0.346	0.683 (0.041) 0.598	0.778 (0.042) 0.642	0.648 (0.048) 0.575
var(e.q12)	0.633 (0.022) 0.545	0.666 (0.047) 0.697	0.758 (0.051) 0.518	0.53 (0.038) 0.622	0.693 (0.057) 0.632	0.45 (0.035) 0.536
var(e.q13)	0.728 (0.026) 0.539	0.867 (0.063) 0.663	0.709 (0.046) 0.419	0.605 (0.042) 0.611	0.793 (0.064) 0.639	0.586 (0.044) 0.574
var(e.q14)	0.631 (0.022) 0.479	0.622 (0.044) 0.507	0.695 (0.053) 0.457	0.585 (0.04) 0.544	0.653 (0.042) 0.511	0.543 (0.042) 0.601
var(e.s1)	0.246 (0.013) 0.172	0.225 (0.028) 0.177	0.232 (0.024) 0.142	0.336 (0.031) 0.333	0.196 (0.019) 0.162	0.232 (0.029) 0.246
var(e.s2)	0.258 (0.015) 0.177	0.287 (0.033) 0.23	0.235 (0.032) 0.139	0.285 (0.031) 0.264	0.236 (0.031) 0.202	0.256 (0.033) 0.265
var(e.s3)	0.403 (0.018) 0.257	0.41 (0.044) 0.324	0.319 (0.036) 0.185	0.525 (0.043) 0.406	0.418 (0.032) 0.306	0.305 (0.027) 0.297
var(e.s4)	0.275 (0.014) 0.189	0.204 (0.019) 0.171	0.249 (0.027) 0.148	0.305 (0.034) 0.273	0.291 (0.029) 0.244	0.301 (0.032) 0.303
var(e.b3)	0.807 (0.04) 0.6	0.772 (0.076) 0.642	0.691 (0.073) 0.508	0.714 (0.1) 0.511	0.883 (0.096) 0.585	0.7 (0.075) 0.601
var(e.b4)	0.207 (0.065) 0.134	0.13 (0.137) 0.098	0.505 (0.119) 0.28	0.404 (0.124) 0.247	0.094 (0.131) 0.059	0.367 (0.082) 0.327
Factor variances and covariances						
var(e.SA)	0.262 (0.016) 0.222	0.243 (0.03) 0.232	0.288 (0.035) 0.206	0.217 (0.028) 0.323	0.303 (0.039) 0.298	0.211 (0.029) 0.297
var(e.BI)	0.379 (0.029) 0.705	0.285 (0.051) 0.662	0.418 (0.063) 0.624	0.485 (0.079) 0.711	0.456 (0.069) 0.727	0.253 (0.046) 0.543
var(SQ)	0.773 (0.037) 1	0.668 (0.073) 1	1.038 (0.085) 1	0.379 (0.061) 1	0.563 (0.069) 1	0.25 (0.047) 1
cov(e.q9,e.q10)	0.179 (0.017) 0.246	0.164 (0.035) 0.207	0.082 (0.03) 0.139	0.252 (0.039) 0.344	0.101 (0.032) 0.14	0.248 (0.037) 0.355
cov(e.q12,e.q13)	0.165 (0.019) 0.243	0.193 (0.042) 0.254	0.171 (0.038) 0.233	0.075 (0.028) 0.132	0.258 (0.052) 0.348	0.101 (0.032) 0.196
Model's fit statistics (Satorra-Bentler estimation)						
df (sb)	166	166	166	166	166	166
chi-square (sb)	924.01	372.31	330.85	370.26	411.27	301.80
p-value	0.000	0.000	0.000	0.000	0.000	0.000
RMSEA (sb)	0.045	0.052	0.047	0.053	0.056	0.044
CFI (sb)	0.968	0.953	0.974	0.941	0.945	0.957
TLI (sb)	0.963	0.946	0.970	0.932	0.937	0.951

SRMR	0.028	0.039	0.031	0.048	0.043	0.044
AIC	112,743.9	23,106.2	22,859.3	21,138.1	24,388.6	19,747.2
BIC	113,110.0	23,371.3	23,122.9	21,398.3	24,655.7	20,006.1

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

Table 8.- Direct and indirect effects and hypothesis testing

	All	**Madrid**	**Rome**	**Berlin**	**Lisbon**	**London**
Model A: Satisfaction as partial mediator						
Paths						
SQ -> SA	1.091	1.097	1.034	1.095	1.127	1.413
SA -> BI	0.346	0.410	0.342	0.473	0.481	0.325
SQ -> BI	**0.027**	**-0.049**	**0.103**	**0.097**	**-0.103**	0.428
SQ -> SA -> BI	0.377	0.449	0.353	0.518	0.542	0.459
Hypothesis						
H1A	Supported	Supported	Supported	Supported	Supported	Supported
H2A	Supported	Supported	Supported	Supported	Supported	Supported
H3A	***Not supported***	***Not supported***	***Not supported***	***Not supported***	***Not supported***	Supported
H4A	Supported	Supported	Supported	Supported	Supported	Supported
Model B: Satisfaction as full mediator						
Paths						
SQ -> SA	1.091	1.096	1.035	1.096	1.126	1.412
SA -> BI	0.366	0.373	0.424	0.541	0.411	0.547
SQ -> SA -> BI	0.399	0.408	0.439	0.593	0.462	0.772
Hypothesis						
H1B	Supported	Supported	Supported	Supported	Supported	Supported
H2B	Supported	Supported	Supported	Supported	Supported	Supported
H3B	Supported	Supported	Supported	Supported	Supported	Supported

All values are significant ($p<0.001$), except the values in **bold**.

However, the results for the direct relationship between service quality and behavioral intentions were neither consistent nor significant across the different samples. Service quality bore an inconsistent negative effect on behavioral intentions in Madrid and Lisbon, and this relationship was not significant for all the samples (the exception being London). We also tested the indirect relationship between service quality and behavioral intentions via satisfaction. In all cases this relationship was significant. Table 8 shows that the hypothesis “service quality has a positive direct effect on behavioral intentions” (H3A) was not supported by all the samples except London. All the other hypotheses tested were supported by all the samples.

7.- Discussion

Based on the literature review and on the empirical results presented, this paper tried to answer three important questions in the public transport field:

- Are service quality and satisfaction the same construct or different ones?
- What is the mediating role of satisfaction between service quality and behavioral intentions or loyalty?
- Which are the main similarities and differences in service quality, customer satisfaction, and behavioral intentions or loyalty across the five European capitals of Madrid, Rome, Berlin, Lisbon and London?

In response to the first question, both the literature review and empirical results suggest that service quality and satisfaction should be conceptualized and measured as different factors. As in previous studies in the public transport field (Table 1), this study relied on evaluations of the specific attributes of the service as indicators of the factor service quality. To define the factor satisfaction, the indicator *general satisfaction* (s1) was used together with other statements having more affective connotations (i.e., expectations, needs, feelings). In all the independent samples analyzed in this study, the CFA measurement model gives service quality and satisfaction very high construct reliability and Cronbach alpha values; satisfactory and statistically significant factor loadings; and not excessively high estimated correlation between the factors. These results agree with previous studies close to the transport sector (Dabholkar et al. 2000), in which different alternative models for the leisure cruise industry were compared, and the importance of measuring satisfaction separately form service quality when trying to determine customer evaluations of service was underlined. The above authors moreover point to satisfaction as a much better predictor of behavioral intentions, whereas service quality is more closely related to specific factor evaluations about the service.

The second question investigated is the mediating role of satisfaction between service quality and behavioral intentions or loyalty. In the literature specifically dedicated to public transport, there is a lack of consensus about the type of mediating role of satisfaction, as both models are still used with similar frequency (Table 1). The results presented in the preceding section indicate that the full mediator model fits very well for all the samples and outperforms the partial mediator model in nearly all cases, with the exception of London. Following (Kline 2015), the partial mediator model should not be considered valid for the pooled data and four cities (Madrid, Rome, Berlin and Lisbon) because it should be re-specified, eliminating the non-significant structural paths. At the same time, as the hypothesis that “service quality has a positive direct effect on behavioral intentions” (H3A) is not supported in four of the five independent samples, and in the pooled sample, the partial mediator model is not supported in those five samples. Notwithstanding, the results from the city of London support both models, with a weak evidence favoring partial mediator against full mediator.

In summary, as the complete mediator model has a very good fit in all the samples, and the partial mediator model only has a good fit for London, these findings are interpreted as additional support for the superiority of the full mediator approach over the partial mediator one, in urban and metropolitan public transport systems. These results agree with previous studies in other sectors (Cronin and Taylor 1992, Cronin et al. 2000), that likewise used various samples to compare alternative models. (Cronin and Taylor 1992) investigated a similar model using data from four industries and found that the full mediator model was clearly supported in one of them (dry cleaning), though the other industries showed inconclusive results. Using data from six industries, (Cronin et al. 2000) evaluated a more complex model and found that the relationship between service quality and behavioral intentions was fully mediated by satisfaction and service value in the case of long distance carriers and health care.

Analyzing in greater detail the results of the full mediator model, Table 8 shows that the indirect effect of service quality over behavioral intentions is always superior than the direct effect of satisfaction over behavioral intentions in all the cities. These substantial effects of service quality over behavioral intentions, that it is consistent with previous studies (Minser and Webb 2010, Zhao et al. 2014) that analyzed the public transport system as a whole, may have contributed to the lack of agreement in the field of public transport as to which model to consider, complete versus partial. In addition, it should be noted that most public transport studies ignore the evaluation of alternative models, failing to consider other possible explanations behind their data, which is a form of confirmation bias (Kline 2015).

The reason why the indirect effect of service quality is greater than the direct effect of satisfaction upon behavioral intentions can be traced to the high direct effect of service quality on satisfaction, which is superior to one in all cases, ranging from 1.035 to 1.412 (Table 8). This means that a 1-point increase in the service quality variable predicts a greater than 1-point increase in satisfaction. From a practical standpoint this result is highly relevant for administrators and decision-makers in the public transport sector, since an increase in the quality perceived by users would produce a more than proportional increase in satisfaction. However, this point also lends itself to an inverse interpretation, that is, a 1-point decrease in service quality predicts a more than 1-point decrease in satisfaction.

In finishing this section, the last question about the similarities and differences across the five European cities is discussed below. The discussion excludes the pooled data and is limited to the complete mediator model. Table 7 shows that *frequency*, *punctuality*, *speed* and *intermodality* are almost always (with the exception of *punctuality* in Berlin and *intermodality* in Madrid) among the five indicators having the highest correlation with service quality. Other indicators highly correlated to service quality are *information* in Madrid, Berlin and Lisbon; *safety* and *individual space* in London; *accessibility* in Berlin; *cleanliness* in Rome; and *service hours* in Madrid. Contrariwise, *cost* is the indicator having the lowest correlation with service quality, except in Madrid and Rome, where *safety* and *proximity* are respectively the lowest ones. Previous studies that have analyzed the public transport system in a city as a whole showed similar results: in Chicago, (Minser and Webb 2010) identified that the factors most correlated to service quality were comfort (including individual space and temperature as indicators) and cleanliness, while the less correlated were associated to safety, accessibility, and service reliability (including frequency, speed and punctuality); in the same city, (Zhao et al. 2014) found the same results for comfort and safety, although they disagreed about service reliability that was placed among the most important factors; in Sweden, (Abenoza, Cats and Susilo 2017) identified that speed (length of the trip time) and safety (freedom from crime) were among the most important attributes for different segment of travelers, while individual space (ride comfort), cleanliness (on board conditions) and information were among the less important attributes; finally, in Shanghai, (Li et al. 2018) have recently found that comfort (including punctuality), safety, convenience (including intermodality, information and service hours) and timeliness (including speed and accessibility) are the most important factors associated to service quality, while economics (that includes ticket price as one indicator) is among the less important.

Table 7 also shows that *general satisfaction* and *meet expectations* are almost always (with the exception of Madrid in the case of *meet expectations* and Berlin in the case of *general satisfaction*) among the two indicators showing the highest correlation with satisfaction. In Madrid and Berlin, the indicator *I feel satisfied* also shows a high correlation with satisfaction. In Shanghai, (Li et al. 2018) also found that overall satisfaction and expectations were the most correlated items to customer satisfaction. In the case of behavioral intentions, the indicator *I will recommend PT* presents the highest correlation in all the cities. In Chicago, (Minser and Webb 2010) also found that the likelihood of recommending to others was the indicator most correlated to customer loyalty. Such results suggest that an indicator about recommending the service should always be included when defining the factor behavioral intention or loyalty, just as statements about general satisfaction, expectations and feelings should be part of the indicators meant to define the factor satisfaction.

Comparison of service quality's factor loadings (Table 7) with the service quality evaluations given in Table 5 provides some interesting findings. In general, the indicators receiving the highest and the lowest assessments in each city of study are not the ones that present the highest correlations with service quality. The exception to this general rule is seen for the indicator *punctuality* in Madrid, Rome and Lisbon, where *punctuality* is one of the aspects least appraised, but better correlated with service quality. The opposite happens in London, where the indicator *frequency* is one of the most valued and best correlated with service quality. *Safety* and *proximity*, two of the best appraised indicators in most of the cities studied, are not found among those highly related with service quality; indeed, *safety* in Madrid and *proximity* in Rome present a weak correlation. Finally, *individual space*, one of the least valued indicators across the board, presents intermediate correlations for all five cities.

To enhance understanding of the specific situation of each of the five cities, a cross analysis was performed between the five indicators best correlated to service quality (Table 7) and the three that received the best and the worst evaluations in the survey (Table 5). The best situation is that of London, which places two of the top indicators (*frequency* and *safety*) among the four most correlated with service quality, though one of its worst (*individual space*) is also highly correlated. It is followed by Berlin, where no poorly valued indicator is among the five presenting best correlation with service quality. In turn, Lisbon has an intermediate situation: one of its poorest rated aspects (*punctuality*) and one of the best rated (*intermodality*) are among the best correlated. In Madrid, just one poorly rated indicator (*punctuality*) is among those best correlated. Last but not least, Rome places, in this highly correlated group, two out of the three worst rated indicators (*punctuality* and *cleanliness*).

Table 8 shows that London yields the highest direct and indirect effects for all relationships. Rome presents the lowest relationship between service quality and satisfaction, and Madrid presents the lowest direct relationship between satisfaction and behavioral intentions, as well as the lowest indirect relationship between service quality and behavioral intentions.

This analysis also provides some interesting insights, with policy and managerial implications, about the public transport systems in five European capitals. Studies conducted in developed countries usually find that punctuality, frequency, comfort, and speed are the most important factors affecting service quality (Quddus et al. 2019). Because of the size of the cities considered in this study, intermodality has showed up as a factor more important than others related to comfort attributes. Meanwhile, in all five cities the effect of service quality over behavioral intentions is always superior than the effect of satisfaction over behavioral intentions. This is of great relevance for public transport managers, since the factor service quality is associated with specific attributes of the service, which can be controlled by public transport operators, whereas the factor satisfaction is subjective in nature and may be influenced by factors beyond service providers' control (Baker and Crompton 2000). If public transport operators focus on improving the perception of attributes bearing a closer correlation with the quality of service —*frequency*, *punctuality*, *speed* and *intermodality* in large cities— they could introduce substantial changes in behavioral intentions or loyalty towards public transportation. In London for example, a 1-point increase in service quality could produce a 1.41-point increase in satisfaction and a 0.77-point increase in favorable behavioral intentions regarding the public transport system (Table 8).

8.- Conclusions and recommendations

The main objective of this article was to analyze, within the context of the urban and metropolitan public transport sector, the mediating role of customer satisfaction between service quality and behavioral intentions or loyalty. Given that there is no consensus in the literature, in order to validate and generalize satisfaction's role, two alternative models were compared, one with satisfaction as a partial mediator, the other with satisfaction as a complete mediator. Independent samples were obtained through surveys conducted in five major European cities: Madrid, Rome, Berlin, Lisbon and London. The same customer satisfaction survey was translated for use in all five cities, and the participants were asked to evaluate the public transport system in its overall functioning. The results support the superiority of the full mediator approach over the partial mediator one in the urban and metropolitan public transport sector. The full mediator model was supported by the data from all cities, while the partial mediator model was not supported in four cities and in the pooled sample.

Nonetheless, this research has also made manifest other important methodological aspects to be considered when attempting to validate models that involve the concepts of service quality, satisfaction, behavioral intentions and loyalty. Service quality and satisfaction are different constructs and should be measured separately, as different factors. Service quality should include indicators associated with the assessments of attributes specific to the service under evaluation, whereas satisfaction ought to include indicators more linked to affective judgements. Behavioral intentions and loyalty are also different constructs, yet the two are broadly used as equivalents in the public transport field. Further research should be dedicated to identifying the differences existing between these two constructs, and their implications. In the meantime, when either is included in a model, it is essential to make sure that indicators reflecting the intention to reutilize the service (behavioral measure) and to recommended it (attitudinal measure) are also included. Another point to be highlighted is the relevance of comparing alternative models and using independent samples, or else using the cross-validation or split sample approach. This sound practice, more common in other disciplines, should be extended to the field of transport so as to reduce the probability of confirmation bias.

From a practical point of view, analysis of public transport systems —in this case those of five European capital cities— may also lead to conclusions of interest for policy makers and public transport operators around the world. In all five cities, even though satisfaction was found to act as a mediator between service quality and behavioral intentions, it was found that service quality exerts an overall positive effect on behavioral intentions superior to the effect of satisfaction. This can be seen as beneficial for transport operators, because service quality is associated with specific attributes of service that can be more easily or directly controlled. Notwithstanding, a vigilant attitude is needed to avoid deterioration in the perception of these attributes, which might lead to a greater reduction in behavioral intentions. The attributes of service that play the most important role in service quality in the five cities studied here are frequency, punctuality, speed and intermodality. Intermodality is positioned among the most relevant attributes in major cities, where it is often necessary to use more than one mode of public transport for a single journey.

Acknowledgments
Support from Spain´s Ministry of Economy and Competitiveness (Research Project TRA2015-66235-R) is gratefully acknowledged.

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.

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.

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

Baker, D. A. & J. L. Crompton (2000) Quality, satisfaction and behavioral intentions. *Annals of Tourism Research,* 27**,** 785-804.

Bansal, H. S. & S. Taylor. 2015. Investigating the Relationship Between Service Quality, Satisfaction and Switching Intentions. In *Proceedings of the 1997 Academy of Marketing Science (AMS) Annual Conference. Developments in Marketing Science: Proceedings of the Academy of Marketing Science,* ed. H. Wilson E., Jr. J. Springer, Cham.

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

Cavana, R. Y., L. M. Corbett & Y. L. Lo (2007) Developing zones of tolerance for managing passenger rail service quality. *International Journal of Quality & Reliability Management,* 24**,** 7-+.

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

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

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

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

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

Cronin, J. J., M. K. Brady & G. T. M. Hult (2000) Assessing the effects of quality, value, and customer satisfaction on consumer behavioral intentions in service environments. *Journal of Retailing,* 76**,** 193-218.

Cronin, J. J. & S. A. Taylor (1992) Measuring Service Quality - A Reexamination and Extension. *Journal of Marketing,* 56**,** 55-68.

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 Ona, J. & R. de Ona (2015) Quality of Service in Public Transport Based on Customer Satisfaction Surveys: A Review and Assessment of Methodological Approaches. *Transportation Science,* 49**,** 605-622.

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

de Ona, R., J. L. Machado & J. de Ona (2015) Perceived Service Quality, Customer Satisfaction, and Behavioral Intentions Structural Equation Model for the Metro of Seville, Spain. *Transportation Research Record,* 2538**,** 76-85.

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

Elkhani, N., S. Soltani & M. H. M. Jamshidi (2014) Examining a hybrid model for e-satisfaction and e-loyalty to e-ticketing on airline websites. *Journal of Air Transport Management,* 37**,** 36-44.

Fornell, C. & D. F. Larcker (1981) Evaluating Structural Equation Models with unobservable variables and measurement error. *Journal of Marketing Research,* 18**,** 39-50.

Friman, M. & M. Fellesson (2009) Service Supply and Customer Satisfaction in Public Transportation: The Quality Paradox. *Journal of Public Transportation,* 12**,** 13.

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. (2003) Structural equation modeling for travel behavior research. *Transportation Research Part B: Methodological,* 37**,** 1-25.
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.
Han, H. (2013) Effects of in-flight ambience and space/function on air travelers' decision to select a low-cost airline. *Tourism Management,* 37**,** 125-135.
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.
Iacobucci, D., N. Saldanha & X. Deng (2007) A meditation on mediation: Evidence that structural equations models perform better than regressions. *Journal of Consumer Psychology,* 17**,** 139-153.
Irtema, H. I. M., A. Ismail, M. N. Borhan, A. M. Das & A. B. Z. Alshetwi (2018) Case study of the behavioural intentions of public transportation passengers in Kuala Lumpur. *Case Studies on Transport Policy,* 6**,** 462-474.
Jen, W. L., R. T. Tu & T. Lu (2011) Managing passenger behavioral intention: an integrated framework for service quality, satisfaction, perceived value, and switching barriers. *Transportation,* 38**,** 321-342.
Joewono, T. B. & H. Kubota (2007) Exploring negative experiences and user loyalty in paratransit. *Transportation Research Record,* 2034**,** 134-142.
Kim, Y. K. & H. R. Lee (2011) Customer satisfaction using low cost carriers. *Tourism Management,* 32**,** 235-243.
Kline, R. B. 2015. *Principles and Practice of Structural Equation Modeling*. Guilford Press.
Koklic, M. K., M. Kukar-Kinney & S. Vegelj (2017) An investigation of customer satisfaction with low-cost and full-service airline companies. *Journal of Business Research,* 80**,** 188-196.
Kuo, C. W. & M. L. Tang (2013) Relationships among service quality, corporate image, customer satisfaction, and behavioral intention for the elderly in high speed rail services. *Journal of Advanced Transportation,* 47**,** 512-525.
Lai, W. T. & C. F. Chen (2011) Behavioral intentions of public transit passengers-The roles of service quality, perceived value, satisfaction and involvement. *Transport Policy,* 18**,** 318-325.
Li, L. B., Y. F. Bai, Z. Q. Song, A. Chen & B. Wu (2018) Public transportation competitiveness analysis based on current passenger loyalty. *Transportation Research Part a-Policy and Practice,* 113**,** 213-226.
Li, X. & J. F. Petrick (2010) Towards an Integrative Model of Loyalty Formation: The Role of Quality and Value. *Leisure Sciences,* 32**,** 201-221.
Machado-Leon, J. L., R. de Ona & J. de Ona (2016) The role of involvement in regards to public transit riders' perceptions of the service. *Transport Policy,* 48**,** 34-44.
Machado-Leon, J. L., R. de Ona, F. Diez-Mesa & J. de Ona (2018) Finding service quality improvement opportunities across different typologies of public transit customers. *Transportmetrica a-Transport Science,* 14**,** 761-783.
Mattsson, J. (1992) A Service Quality Model Based on an Ideal Value Standard. *International Journal of Service Industry Management,* 3**,** 18-33.

Miller, J. A. 1976. Exploring some alternative measures of cunsumer satisfaction. In *Marketing: 1776-1976 and Beyond,* ed. K. L. Bernhardt, 661-664. Chicago, IL: American Marketing Association.
Minser, J. & V. Webb (2010) Quantifying the Benefits Application of Customer Loyalty Modeling in Public Transportation Context. *Transportation Research Record,* 2144**,** 111-120.
Nguyen-Phuo, D. Q., D. N. Su, P. T. K. Tran, D. T. T. Le & L. W. Johnson (2020) Factors influencing customer's loyalty towards ride-hailing taxi services - A case study of Vietnam. *Transportation Research Part a-Policy and Practice,* 134**,** 96-112.
Oliver, R. L. (1981) Measurement and evaluation of satisfaction processes in retail settings. *Journal of Retailing,* 57**,** 25-48.
--- (1999) Whence consumer loyalty? *Journal of Marketing,* 63**,** 33-44.
---. 2010. *Satisfaction. A behavioral perspective on the consumer*. Armonk, New York: JW. E. Sharpe.
Parasuraman, A., V. A. Zeithaml & L. L. Berry (1985) A conceptual model of service quality and its implications for future research. *Journal of Marketing,* 49**,** 41-50.
Park, J. W., R. Robertson & C. L. Wu (2004) The effect of airline service quality on passengers' behavioural intentions: a Korean case study. *Journal of Air Transport Management,* 10**,** 435-439.
--- (2006) Modelling the impact of airline service quality and marketing variables on passengers' future behavioural intentions. *Transportation Planning and Technology,* 29**,** 359-381.
Pedersen, E. & H. Nysveen (2001) Shopbot banking: an exploratory study of customer loyalty effects. *International Journal of Bank Marketing,* 19**,** 146-155.
Quddus, M., F. Rahman, F. Monsuur, J. de Ona & M. Enoch (2019) Analyzing Bus Passenger's Satisfaction in Dhaka using Discrete Choice Models. *Transportation Research Record,* 2673**,** 758-768.
Raftery, A. E. (1996) Approximate Bayes factors and accounting for model uncertainty in generalised linear models. *Biometrika,* 83**,** 251-266.
Rahim, A. G. (2016) Perceived Service Quality and Customer Loyalty: The Mediating Effect of Passenger Satisfaction in the Nigerian Airline Industry. *International Journal of Management and Economics,* 52**,** 94-117.
Rajaguru, R. (2016) Role of value for money and service quality on behavioural intention: A study of full service and low cost airlines. *Journal of Air Transport Management,* 53**,** 114-122.
Saha, G. C. & Theingi (2009) Service quality, satisfaction, and behavioural intentions A study of low-cost airline carriers in Thailand. *Managing Service Quality,* 19**,** 350-370.
Shah, R. & S. M. Goldstein (2006) Use of structural equation modeling in operations management research: Looking back and forward. *Journal of Operations Management,* 24**,** 148-169.
Stevens, J. P. 2009. *Applied Multivariate Statistics for the Social Sciences*. New York: Routledge.
Sun, S. C. & Z. Y. Duan (2019) Modeling passengers' loyalty to public transit in a two-dimensional framework: A case study in Xiamen, China. *Transportation Research Part a-Policy and Practice,* 124**,** 295-309.
United Nations, D. o. E. a. S. A., Population Division. 2019. *World Urbanization Prospect. The 2018 Revision (ST/ESA/SER.A/420)*. New York: United Nations.
van Lierop, D., M. G. Badami & A. M. El-Geneidy (2018) What influences satisfaction and loyalty in public transport? A review of the literature. *Transport Reviews,* 38**,** 52-72.
van Lierop, D. & A. El-Geneidy (2016) Enjoying loyalty: The relationship between service quality, customer satisfaction, and behavioral intentions in public transit. *Research in Transportation Economics,* 59**,** 50-59.

Wen, C. H., L. W. Lan, H. L. Cheng & Trb (2005) Structural equation modeling to determine passenger loyalty toward intercity bus services. *Transit: Planning, Management and Maintenance, Technology, Marketing and Fare Policy, and Capacity and Qualtiy of Sevice,* 1927**,** 249-255.

Yang, Z. & R. T. Peterson (2004) Customer perceived value, satisfaction, and loyalty: The role of switching costs. *Psychology and Marketing,* 21**,** 799-822.

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.

Zeithaml, V. A., L. L. Berry & A. Parasuraman (1996) The behavioral consequences of service quality. *Journal of Marketing,* 60**,** 31-46.

Zhang, C. Q., Y. Liu, W. T. Lu & G. N. Xiao (2019) Evaluating passenger satisfaction index based on PLS-SEM model: Evidence from Chinese public transport service. *Transportation Research Part a-Policy and Practice,* 120**,** 149-164.

Zhang, M., Y. Y. Xie, L. L. Huang & Z. He (2013) Service quality evaluation of car rental industry in China. *International Journal of Quality & Reliability Management,* 31**,** 82-+.

Zhao, J. H., V. Webb & P. Shah (2014) Customer Loyalty Differences Between Captive and Choice Transit Riders. *Transportation Research Record,* 2415**,** 80-88.

Appendix

Table A.1.- Initial CFA measurement model

Parameters	All Unst. (SE) St.	Madrid Unst. (SE) St.	Rome Unst. (SE) St.	Berlin Unst. (SE) St.	Lisbon Unst. (SE) St.	London Unst. (SE) St.
Factor loadings						
Service Quality (SQ)						
SQ->Service hours (q1)	1 (*) 0.744	1 (*) 0.726	1 (*) 0.78	1 (*) 0.641	1 (*) 0.672	1 (*) 0.581
SQ->Proximity (q2)	0.727 (0.026) 0.617	0.813 (0.061) 0.636	0.659 (0.046) 0.568	0.956 (0.101) 0.63	0.782 (0.072) 0.58	1.162 (0.109) 0.645
SQ->Frequency (q3)	1.08 (0.023) 0.782	1.13 (0.058) 0.757	1.039 (0.04) 0.829	1.177 (0.1) 0.684	1.117 (0.057) 0.743	1.227 (0.122) 0.672
SQ->Punctuality (q4)	1.111 (0.026) 0.785	1.12 (0.062) 0.768	1.075 (0.043) 0.833	1.183 (0.096) 0.668	1.203 (0.074) 0.752	1.398 (0.135) 0.694
SQ->Speed (q5)	0.945 (0.025) 0.767	0.93 (0.055) 0.731	1.006 (0.042) 0.834	1.129 (0.098) 0.709	0.998 (0.074) 0.719	1.259 (0.129) 0.68
SQ->Cost (q6)	0.657 (0.028) 0.495	0.898 (0.061) 0.603	0.652 (0.041) 0.579	0.986 (0.111) 0.533	0.696 (0.081) 0.439	1.22 (0.154) 0.519
SQ->Accessibility (q7)	0.868 (0.025) 0.704	0.879 (0.054) 0.686	0.881 (0.044) 0.742	1.086 (0.092) 0.676	0.873 (0.069) 0.637	1.146 (0.115) 0.608
SQ->Intermodality (q8)	0.957 (0.023) 0.759	0.835 (0.052) 0.67	1.03 (0.039) 0.827	1.129 (0.085) 0.72	0.95 (0.071) 0.677	1.216 (0.125) 0.687
SQ->Individual space (q9)	0.942 (0.027) 0.702	0.913 (0.065) 0.627	1.001 (0.043) 0.807	1.213 (0.107) 0.662	0.993 (0.071) 0.674	1.596 (0.169) 0.695
SQ->Temperature (q10)	0.917 (0.026) 0.694	0.94 (0.061) 0.666	0.908 (0.043) 0.772	1.104 (0.112) 0.611	0.982 (0.068) 0.655	1.406 (0.148) 0.654
SQ->Cleanliness (q11)	0.962 (0.027) 0.713	0.849 (0.061) 0.65	1.051 (0.047) 0.813	1.142 (0.118) 0.638	0.896 (0.073) 0.612	1.393 (0.14) 0.653
SQ->Safety (q12)	0.839 (0.025) 0.684	0.701 (0.059) 0.573	0.827 (0.045) 0.7	0.926 (0.097) 0.603	0.871 (0.07) 0.626	1.255 (0.112) 0.685
SQ->Security (q13)	0.914 (0.028) 0.691	0.845 (0.066) 0.593	0.976 (0.047) 0.771	1.063 (0.106) 0.635	0.913 (0.078) 0.62	1.355 (0.144) 0.668
SQ->Information (q14)	0.948 (0.026) 0.724	0.977 (0.06) 0.705	0.886 (0.043) 0.736	1.188 (0.101) 0.683	1.047 (0.074) 0.696	1.248 (0.145) 0.651
Satisfaction (SA)						
SA->General satisfaction (s1)	1 (*) 0.912	1 (*) 0.906	1 (*) 0.929	1 (*) 0.821	1 (*) 0.915	1 (*) 0.871
SA->Meet expectations (s2)	1.006 (0.012) 0.908	0.964 (0.034) 0.879	1.012 (0.025) 0.927	1.078 (0.046) 0.857	0.955 (0.031) 0.891	1 (0.037) 0.86
SA->Covered needs (s3)	0.991 (0.014) 0.863	0.905 (0.039) 0.821	0.996 (0.026) 0.903	1.068 (0.057) 0.775	0.967 (0.032) 0.834	0.997 (0.039) 0.835
SA->I feel satisfied (s4)	0.994 (0.012) 0.901	0.981 (0.031) 0.911	1.003 (0.024) 0.922	1.092 (0.051) 0.853	0.945 (0.029) 0.875	0.975 (0.041) 0.83
Behavioral Intentions (BI)						
BI->I will use PT for one-off trips (b1)	1 (*) 0.446	1 (*) 0.507	1 (*) 0.517	1 (*) 0.353	1 (*) 0.51	1 (*) 0.283
BI->I will use PT for regular trips (b2)	0.616 (0.051) 0.344	0.548 (0.105) 0.356	0.614 (0.094) 0.375	0.615 (0.149) 0.259	0.631 (0.081) 0.424	0.636 (0.177) 0.234
BI->Increase usage (b3)	1.554 (0.088) 0.707	1.352 (0.156) 0.679	1.425 (0.139) 0.785	2.228 (0.374) 0.72	1.487 (0.137) 0.765	2.096 (0.448) 0.659
BI->I will recommend PT (b4)	1.964 (0.115) 0.832	1.799 (0.207) 0.853	1.662 (0.168) 0.79	2.73 (0.467) 0.808	1.649 (0.158) 0.832	2.409 (0.521) 0.773
Measurement error variances						
var(e.q1)	0.621 (0.023) 0.447	0.578 (0.044) 0.473	0.674 (0.051) 0.391	0.507 (0.034) 0.589	0.688 (0.051) 0.549	0.481 (0.035) 0.662
var(e.q2)	0.663 (0.021) 0.62	0.626 (0.039) 0.595	0.954 (0.056) 0.677	0.492 (0.035) 0.604	0.683 (0.041) 0.664	0.465 (0.034) 0.584

var(e.q3)	0.569 (0.021) 0.388	0.613 (0.044) 0.427	0.514 (0.039) 0.313	0.556 (0.049) 0.532	0.573 (0.045) 0.448	0.448 (0.032) 0.548
var(e.q4)	0.59 (0.02) 0.383	0.561 (0.037) 0.41	0.532 (0.039) 0.305	0.614 (0.047) 0.554	0.63 (0.044) 0.435	0.515 (0.043) 0.518
var(e.q5)	0.48 (0.017) 0.411	0.486 (0.032) 0.466	0.464 (0.035) 0.304	0.446 (0.034) 0.497	0.526 (0.037) 0.483	0.451 (0.037) 0.537
var(e.q6)	1.024 (0.03) 0.755	0.908 (0.06) 0.636	0.882 (0.055) 0.665	0.864 (0.058) 0.715	1.153 (0.065) 0.808	0.988 (0.07) 0.73
var(e.q7)	0.589 (0.021) 0.504	0.56 (0.044) 0.529	0.663 (0.049) 0.449	0.495 (0.038) 0.543	0.632 (0.043) 0.594	0.551 (0.047) 0.631
var(e.q8)	0.519 (0.017) 0.424	0.552 (0.036) 0.552	0.514 (0.044) 0.316	0.419 (0.031) 0.481	0.604 (0.035) 0.542	0.406 (0.028) 0.528
var(e.q9)	0.704 (0.022) 0.508	0.828 (0.056) 0.607	0.562 (0.042) 0.349	0.667 (0.038) 0.562	0.672 (0.044) 0.546	0.667 (0.046) 0.516
var(e.q10)	0.695 (0.022) 0.518	0.714 (0.049) 0.556	0.586 (0.039) 0.404	0.723 (0.044) 0.626	0.725 (0.049) 0.57	0.65 (0.038) 0.573
var(e.q11)	0.688 (0.022) 0.491	0.635 (0.043) 0.578	0.595 (0.043) 0.339	0.671 (0.041) 0.593	0.76 (0.043) 0.626	0.64 (0.046) 0.573
var(e.q12)	0.617 (0.019) 0.532	0.648 (0.041) 0.672	0.744 (0.048) 0.51	0.532 (0.033) 0.637	0.667 (0.042) 0.608	0.438 (0.032) 0.531
var(e.q13)	0.702 (0.023) 0.522	0.849 (0.058) 0.649	0.681 (0.041) 0.406	0.592 (0.041) 0.597	0.756 (0.052) 0.616	0.559 (0.04) 0.554
var(e.q14)	0.627 (0.022) 0.476	0.622 (0.043) 0.503	0.697 (0.052) 0.459	0.572 (0.041) 0.534	0.66 (0.042) 0.515	0.52 (0.04) 0.576
var(e.s1)	0.242 (0.013) 0.169	0.225 (0.027) 0.179	0.224 (0.024) 0.137	0.33 (0.032) 0.326	0.199 (0.019) 0.164	0.227 (0.029) 0.242
var(e.s2)	0.257 (0.015) 0.176	0.282 (0.033) 0.227	0.235 (0.03) 0.14	0.286 (0.032) 0.265	0.242 (0.031) 0.207	0.251 (0.033) 0.261
var(e.s3)	0.401 (0.018) 0.255	0.407 (0.045) 0.326	0.317 (0.034) 0.185	0.516 (0.042) 0.399	0.418 (0.032) 0.305	0.307 (0.027) 0.303
var(e.s4)	0.273 (0.014) 0.189	0.203 (0.019) 0.17	0.251 (0.027) 0.15	0.304 (0.034) 0.272	0.278 (0.026) 0.235	0.305 (0.032) 0.311
var(e.b1)	1.114 (0.04) 0.801	0.876 (0.072) 0.743	1.123 (0.088) 0.733	0.999 (0.073) 0.875	1.144 (0.08) 0.74	1.327 (0.104) 0.92
var(e.b2)	0.781 (0.028) 0.882	0.624 (0.056) 0.873	0.944 (0.059) 0.86	0.75 (0.068) 0.933	0.732 (0.051) 0.821	0.803 (0.071) 0.945
var(e.b3)	0.668 (0.036) 0.5	0.646 (0.064) 0.539	0.518 (0.06) 0.384	0.655 (0.093) 0.481	0.629 (0.08) 0.414	0.659 (0.072) 0.565
var(e.b4)	0.475 (0.041) 0.308	0.366 (0.083) 0.272	0.678 (0.087) 0.375	0.563 (0.102) 0.347	0.486 (0.074) 0.308	0.451 (0.068) 0.403
Factor variances and covariances						
var(e.SA)	0.77 (0.037) 1	0.644 (0.07) 1	1.048 (0.083) 1	0.354 (0.056) 1	0.566 (0.069) 1	0.245 (0.048) 1
var(e.BI)	1.19 (0.037) 1	1.031 (0.08) 1	1.408 (0.084) 1	0.682 (0.076) 1	1.017 (0.066) 1	0.711 (0.069) 1
var(SQ)	0.276 (0.03) 1	0.302 (0.065) 1	0.409 (0.076) 1	0.142 (0.045) 1	0.402 (0.073) 1	0.115 (0.047) 1
cov(SQ,SA)	0.842 (0.031) 0.88	0.713 (0.059) 0.875	1.083 (0.069) 0.892	0.405 (0.048) 0.826	0.631 (0.053) 0.832	0.343 (0.049) 0.822
cov(SQ,BI)	0.233 (0.018) 0.505	0.222 (0.036) 0.503	0.344 (0.049) 0.526	0.112 (0.025) 0.5	0.189 (0.032) 0.396	0.114 (0.027) 0.676
cov(SA,BI)	0.323 (0.023) 0.564	0.338 (0.052) 0.605	0.425 (0.055) 0.56	0.177 (0.038) 0.569	0.332 (0.045) 0.519	0.195 (0.044) 0.68
Contruct validity and reliability						
Service Quality						
Contruct Reliability (CR)	0.933	0.920	0.950	0.911	0.912	0.911
Average Variance Extracted (AVE)	0.502	0.453	0.580	0.424	0.428	0.424
Cronbach's Alpha	0.930	0.916	0.948	0.910	0.910	0.903
Satisfaction						
Contruct Reliability (CR)	0.942	0.932	0.957	0.897	0.931	0.912

Average Variance Extracted (AVE)	0.803	0.774	0.847	0.684	0.773	0.721
Cronbach's Alpha	0.938	0.931	0.955	0.892	0.927	0.898
Behavioral Intentions						
Contruct Reliability (CR)	0.685	0.703	0.721	0.635	0.737	0.573
Average Variance Extracted (AVE)	0.377	0.393	0.412	0.341	0.429	0.292
Cronbach's Alpha	0.680	0.694	0.725	0.639	0.714	0.556
Fit statistics (Satorra-Bentler estimation)						
df (sb)	206	206	206	206	206	206
chi-square (sb)	1,437.59	499.53	478.10	509.47	600.82	409.22
p-value	0.000	0.000	0.000	0.000	0.000	0.000
RMSEA (sb)	0.052	0.056	0.054	0.059	0.064	0.049
CFI (sb)	0.949	0.936	0.958	0.913	0.917	0.938
TLI (sb)	0.943	0.928	0.953	0.902	0.906	0.930
SRMR	0.036	0.050	0.056	0.060	0.054	0.047
AIC	123,571.3	25,017.0	25,268.5	22,900.6	26,826.9	22,000.4
BIC	123,964.9	25,301.4	25,552.0	23,179.4	27,114.0	22,278.8

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

Table A.2.- Final CFA measurement model

Parameters	All Unst. (SE) St.	Madrid Unst. (SE) St.	Rome Unst. (SE) St.	Berlin Unst. (SE) St.	Lisbon Unst. (SE) St.	London Unst. (SE) St.
Factor loadings						
Service Quality (SQ)						
SQ->Service hours (q1)	1 (*) 0.744	1 (*) 0.73	1 (*) 0.777	1 (*) 0.654	1 (*) 0.669	1 (*) 0.586
SQ->Proximity (q2)	0.725 (0.027) 0.615	0.796 (0.061) 0.636	0.645 (0.048) 0.553	0.963 (0.093) 0.648	0.792 (0.074) 0.586	1.19 (0.114) 0.656
SQ->Frequency (q3)	1.082 (0.023) 0.784	1.11 (0.058) 0.754	1.049 (0.041) 0.832	1.153 (0.093) 0.686	1.127 (0.058) 0.749	1.261 (0.127) 0.686
SQ->Punctuality (q4)	1.113 (0.026) 0.788	1.098 (0.06) 0.764	1.085 (0.045) 0.837	1.149 (0.096) 0.672	1.213 (0.076) 0.756	1.425 (0.139) 0.707
SQ->Speed (q5)	0.945 (0.025) 0.769	0.921 (0.054) 0.739	1.015 (0.043) 0.836	1.102 (0.092) 0.713	1.003 (0.075) 0.723	1.258 (0.132) 0.676
SQ->Cost (q6)	0.652 (0.028) 0.491	0.893 (0.063) 0.607	0.647 (0.045) 0.572	0.969 (0.102) 0.541	0.694 (0.082) 0.437	1.184 (0.148) 0.507
SQ->Accessibility (q7)	0.868 (0.025) 0.705	0.863 (0.055) 0.683	0.888 (0.045) 0.743	1.078 (0.085) 0.689	0.878 (0.069) 0.64	1.134 (0.112) 0.606
SQ->Intermodality (q8)	0.96 (0.024) 0.762	0.825 (0.053) 0.674	1.039 (0.041) 0.828	1.118 (0.08) 0.731	0.959 (0.073) 0.684	1.21 (0.122) 0.689
SQ->Individual space (q9)	0.929 (0.027) 0.691	0.892 (0.063) 0.622	0.994 (0.045) 0.798	1.135 (0.103) 0.64	0.984 (0.074) 0.664	1.559 (0.165) 0.678
SQ->Temperature (q10)	0.899 (0.026) 0.681	0.889 (0.062) 0.645	0.908 (0.045) 0.768	1.037 (0.102) 0.591	0.962 (0.069) 0.64	1.361 (0.146) 0.634
SQ->Cleanliness (q11)	0.953 (0.027) 0.707	0.828 (0.061) 0.644	1.049 (0.047) 0.809	1.1 (0.11) 0.634	0.877 (0.074) 0.598	1.392 (0.14) 0.652
SQ->Safety (q12)	0.828 (0.026) 0.675	0.657 (0.061) 0.55	0.824 (0.047) 0.694	0.922 (0.089) 0.614	0.846 (0.074) 0.607	1.257 (0.115) 0.682
SQ->Security (q13)	0.898 (0.028) 0.679	0.811 (0.065) 0.58	0.974 (0.05) 0.762	1.009 (0.099) 0.624	0.892 (0.081) 0.601	1.329 (0.144) 0.654
SQ->Information (q14)	0.942 (0.026) 0.722	0.951 (0.06) 0.702	0.892 (0.046) 0.737	1.137 (0.095) 0.675	1.052 (0.075) 0.699	1.209 (0.14) 0.633
Satisfaction (SA)						
SA->General satisfaction (s1)	1 (*) 0.91	1 (*) 0.907	1 (*) 0.926	1 (*) 0.817	1 (*) 0.915	1 (*) 0.871
SA->Meet expectations (s2)	1.007 (0.012) 0.907	0.958 (0.033) 0.877	1.019 (0.027) 0.928	1.087 (0.047) 0.859	0.958 (0.03) 0.893	0.998 (0.037) 0.86
SA->Covered needs (s3)	0.993 (0.014) 0.862	0.904 (0.038) 0.822	1.002 (0.027) 0.903	1.069 (0.06) 0.771	0.966 (0.032) 0.833	1.005 (0.039) 0.838
SA->I feel satisfied (s4)	0.997 (0.012) 0.9	0.973 (0.031) 0.911	1.01 (0.025) 0.923	1.097 (0.051) 0.853	0.943 (0.029) 0.87	0.982 (0.041) 0.834
Behavioral Intentions (BI)						
BI->Increase usage (b3)	1 (*) 0.632	1 (*) 0.599	1 (*) 0.7	1 (*) 0.696	1 (*) 0.647	1 (*) 0.634
BI->I will recommend PT (b4)	1.579 (0.09) 0.931	1.664 (0.227) 0.948	1.399 (0.13) 0.85	1.355 (0.146) 0.872	1.535 (0.165) 0.967	1.265 (0.149) 0.817
Measurement error variances						
var(e.q1)	0.624 (0.023) 0.447	0.586 (0.045) 0.467	0.68 (0.052) 0.396	0.508 (0.035) 0.573	0.695 (0.05) 0.552	0.473 (0.035) 0.657
var(e.q2)	0.668 (0.022) 0.622	0.624 (0.04) 0.596	0.98 (0.059) 0.694	0.485 (0.034) 0.58	0.677 (0.042) 0.657	0.463 (0.034) 0.57
var(e.q3)	0.566 (0.021) 0.385	0.626 (0.045) 0.432	0.507 (0.041) 0.308	0.565 (0.053) 0.529	0.559 (0.045) 0.438	0.442 (0.034) 0.529
var(e.q4)	0.584 (0.021) 0.379	0.576 (0.038) 0.416	0.522 (0.04) 0.299	0.607 (0.047) 0.548	0.621 (0.044) 0.428	0.504 (0.043) 0.501

var(e.q5)	0.478 (0.018) 0.409	0.471 (0.031) 0.454	0.462 (0.035) 0.302	0.446 (0.034) 0.492	0.518 (0.037) 0.477	0.465 (0.039) 0.543
var(e.q6)	1.032 (0.03) 0.759	0.913 (0.059) 0.632	0.893 (0.057) 0.673	0.859 (0.058) 0.707	1.149 (0.066) 0.809	1.005 (0.071) 0.743
var(e.q7)	0.59 (0.021) 0.503	0.57 (0.044) 0.534	0.666 (0.049) 0.448	0.488 (0.037) 0.526	0.627 (0.042) 0.591	0.547 (0.046) 0.632
var(e.q8)	0.513 (0.017) 0.419	0.546 (0.035) 0.545	0.515 (0.043) 0.315	0.412 (0.032) 0.465	0.589 (0.035) 0.532	0.401 (0.028) 0.526
var(e.q9)	0.728 (0.023) 0.522	0.845 (0.057) 0.613	0.585 (0.044) 0.363	0.703 (0.043) 0.59	0.691 (0.046) 0.559	0.705 (0.049) 0.54
var(e.q10)	0.72 (0.023) 0.536	0.742 (0.051) 0.584	0.595 (0.04) 0.41	0.76 (0.047) 0.651	0.751 (0.049) 0.59	0.682 (0.042) 0.598
var(e.q11)	0.7 (0.022) 0.5	0.647 (0.043) 0.586	0.605 (0.043) 0.346	0.683 (0.041) 0.598	0.778 (0.042) 0.642	0.648 (0.048) 0.575
var(e.q12)	0.633 (0.022) 0.545	0.666 (0.047) 0.698	0.757 (0.051) 0.518	0.53 (0.038) 0.623	0.693 (0.057) 0.632	0.449 (0.035) 0.535
var(e.q13)	0.728 (0.026) 0.539	0.867 (0.063) 0.663	0.71 (0.046) 0.419	0.605 (0.042) 0.61	0.793 (0.064) 0.639	0.584 (0.044) 0.572
var(e.q14)	0.631 (0.022) 0.479	0.622 (0.044) 0.507	0.695 (0.053) 0.457	0.586 (0.04) 0.545	0.653 (0.042) 0.511	0.541 (0.042) 0.599
var(e.s1)	0.245 (0.013) 0.172	0.226 (0.028) 0.178	0.232 (0.024) 0.142	0.335 (0.031) 0.332	0.198 (0.019) 0.163	0.227 (0.029) 0.241
var(e.s2)	0.258 (0.015) 0.177	0.288 (0.033) 0.231	0.234 (0.032) 0.139	0.284 (0.031) 0.263	0.236 (0.031) 0.202	0.251 (0.033) 0.26
var(e.s3)	0.403 (0.018) 0.257	0.41 (0.044) 0.325	0.318 (0.036) 0.185	0.524 (0.044) 0.405	0.418 (0.032) 0.306	0.306 (0.027) 0.298
var(e.s4)	0.275 (0.014) 0.189	0.204 (0.019) 0.171	0.249 (0.027) 0.148	0.304 (0.034) 0.273	0.291 (0.029) 0.244	0.303 (0.032) 0.305
var(e.b3)	0.807 (0.04) 0.601	0.771 (0.076) 0.641	0.693 (0.074) 0.509	0.72 (0.1) 0.516	0.878 (0.095) 0.582	0.696 (0.073) 0.598
var(e.b4)	0.206 (0.065) 0.133	0.134 (0.135) 0.101	0.501 (0.121) 0.277	0.392 (0.126) 0.24	0.104 (0.129) 0.065	0.373 (0.08) 0.332
Factor variances and covariances						
var(e.SA)	0.773 (0.037) 1	0.669 (0.073) 1	1.038 (0.085) 1	0.379 (0.061) 1	0.563 (0.07) 1	0.247 (0.048) 1
var(e.BI)	1.182 (0.037) 1	1.045 (0.082) 1	1.4 (0.083) 1	0.673 (0.074) 1	1.015 (0.066) 1	0.715 (0.069) 1
var(SQ)	0.536 (0.043) 1	0.432 (0.087) 1	0.668 (0.094) 1	0.676 (0.105) 1	0.631 (0.1) 1	0.468 (0.079) 1
cov(e.q9,e.q10)	0.178 (0.017) 0.246	0.164 (0.035) 0.207	0.082 (0.03) 0.139	0.251 (0.039) 0.343	0.101 (0.032) 0.14	0.243 (0.037) 0.351
cov(e.q12,e.q13)	0.165 (0.019) 0.243	0.193 (0.042) 0.255	0.171 (0.038) 0.233	0.075 (0.028) 0.133	0.258 (0.051) 0.347	0.099 (0.032) 0.194
cov(SQ,SA)	0.843 (0.031) 0.882	0.733 (0.061) 0.877	1.073 (0.07) 0.89	0.415 (0.05) 0.821	0.635 (0.054) 0.839	0.35 (0.048) 0.831
cov(SQ,BI)	0.312 (0.024) 0.485	0.268 (0.047) 0.498	0.473 (0.06) 0.568	0.233 (0.039) 0.46	0.247 (0.042) 0.415	0.22 (0.035) 0.645
cov(SA,BI)	0.431 (0.031) 0.542	0.392 (0.066) 0.584	0.589 (0.068) 0.609	0.359 (0.055) 0.532	0.423 (0.064) 0.528	0.382 (0.055) 0.66
Contruct validity and reliability						
Service Quality						
Contruct Reliability (CR)	0.932	0.918	0.949	0.912	0.910	0.910
Average Variance Extracted (AVE)	0.497	0.448	0.575	0.426	0.424	0.420
Cronbach's Alpha	0.930	0.916	0.948	0.910	0.910	0.903
Satisfaction						
Contruct Reliability (CR)	0.942	0.932	0.957	0.895	0.931	0.913

Average Variance Extracted (AVE)	0.801	0.774	0.847	0.682	0.771	0.724
Cronbach's Alpha	0.938	0.931	0.955	0.892	0.927	0.898
Behavioral Intentions						
Contruct Reliability (CR)	0.769	0.763	0.753	0.765	0.801	0.694
Average Variance Extracted (AVE)	0.633	0.629	0.607	0.622	0.676	0.535
Cronbach's Alpha	0.737	0.710	0.746	0.736	0.771	0.690
Model's fit statistics (Satorra-Bentler estimation)						
df (sb)	165	165	165	165	165	165
chi-square (sb)	923.81	371.93	329.79	369.68	410.68	295.93
p-value	0.000	0.000	0.000	0.000	0.000	0.000
RMSEA (sb)	0.045	0.052	0.047	0.054	0.056	0.043
CFI (sb)	0.968	0.953	0.974	0.941	0.945	0.959
TLI (sb)	0.963	0.946	0.970	0.932	0.937	0.952
SRMR	0.028	0.039	0.030	0.047	0.043	0.041
AIC	112,745.5	23,107.9	22,860.2	21,139.6	24,389.5	19,740.7
BIC	113,117.3	23,377.1	23,127.8	21,403.9	24,660.8	20,003.7

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