PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
SCHOOL OF ENGINEERING

MODELLING SATISFACTION WITH PUBLIC TRANSPORT

JAIME ALLEN MONGE

Thesis submitted to the Office of Graduate Studies in partial fulfilment of the requirements for the Degree of Doctor of Science in Engineering (or Doctor in Engineering Sciences)

Supervisors:

JUAN DE DIOS ORTUZAR

JUAN CARLOS MUÑOZ

Santiago de Chile, November 2018

PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
ESCUELA DE INGENIERÍA

MODELACIÓN DE LA SATISFACCIÓN CON EL TRANSPORTE PÚBLICO

JAIME ALLEN MONGE

Tesis presentada a la Comisión integrada por los profesores:

JUAN DE DIOS ORTÚZAR

JUAN CARLOS MUÑOZ

LUIS IGNACIO RIZZI

JORGE MANZI

JUAN ANTONIO CARRASCO

GABRIELLA MAZZULLA

JORGE VÁSQUEZ

Para completar las exigencias del grado de
Doctor en Ciencias de la Ingeniería

Santiago de Chile, Noviembre 2018

A mi esposa, Priscila, que me apoyó y motivó durante todo el proceso. Gracias a ella, hemos disfrutado cada paso y momento del doctorado.

A mi familia, y a mis amigos.

AGRADECIMIENTOS

Primero que todo agradezco a mi esposa, Priscila; su encanto y amor incondicional siempre nos ayudó a salir adelante en esta etapa de nuestras vidas.

Agradezco a mi hermano Pablo y a mi madre, Elia Ana, mis dos fuentes de inspiración. Me enseñaron a perseverar, y me impulsaron a conseguir esta gran meta académica y personal.

Agradezco a mi padre Jaime, nuestros caminos se unieron en Bolivia una vez más. A él le agradezco por esas lecciones de vida que no enseña ningún manual.

Agradezco a mis tutores Juan de Dios y Juan Carlos, por ser valuartes y grandes fuentes de inspiración. Sin duda alguna, será difícil encontrar una dupla de tutores con el nivel de compromiso y el impulso por la investigación que ellos tienen. Gracias por sus enseñanzas.

Agradezco a Laura y a Gabriella, mis dos tutoras en la pasantía en Calabria, Italia, por haberme recibido tan calurosamente como lo hicieron.

Agradezco especialmente a César, Margareth, Carlos, Yerly, Luis, Thomas, y Ana. ¡Sin ellos los días de estudiante en la PUC no hubieran sido tan amenos!

Agradezco a César y a JuanCa (Herrera) por convencerme de aceptar el reto.

Agradezco a los profesores del Departamento, por su dedicación.

Agradezco a los Ticos en Chile, por acompañarnos todos estos años en los asados y paseos.

Agradezco a Adela, Margareth, Mariana, y Pame (García), por ser excelentes amigas.

Agradezco a Cristian Domarchi, por echarse el equipo al hombro con la Encuesta.

A Adri (Ramírez) le agradezco por la ayuda con el inglés. A Astrid y a Pri, por las figuras!

Agradezco a Jordi por su colaboración y a las entidades que brindaron acceso a los datos.

Agradezco a Alejandro, Luis Guillermo, José Pablo, y demás compañeros, por apoyarme desde el LanammeUCR para cumplir esta meta, que finalmente se logra.

Agradezco a Ana Batista, Carlos Bueno, Felipe González, Margareth Gutiérrez, Pablo Guarda, César Henao y David Palma, por múltiples ayudas y consejos durante el doctorado.
Gracias a Chile y a los chilenos por acogernos durante estos cuatro años.
Finalmente, agradezco a LanammeUCR, CEDEUS, BRT+ Centre of Excellence, VRI y DIIPEI de la PUC y, por supuesto, a CONICYT-Chile, por los patrocinios. ¡Sin los mismos, no se concreta este proyecto!

GENERAL INDEX

Page

TABLES

Page

FIGURES

Page

RESUMEN

Las autoridades responsables de los sistemas de transporte público requieren instrumentos para priorizar las inversiones destinadas a mejorar sus servicios, a fin de conservar y atraer más usuarios, principalmente del transporte privado que es menos sostenible. Además, la satisfacción influye en la aceptación política del sistema. Por esto, se requiere mejorar el nivel de satisfacción percibido por sus clientes. Las encuestas de satisfacción permiten detectar niveles de satisfacción con el sistema y con sus diferentes atributos, por ejemplo, frecuencia, confiabilidad, accesibilidad, seguridad y comodidad.

Aunque la literatura científica postula diversos modelos econométricos para determinar qué atributos son más relevantes en la tarea anterior, no hay consenso sobre cuál es la técnica de modelación más apropiada. No obstante, la literatura informa diferencias entre la importancia de diferentes atributos del sistema según modo de transporte y contexto de la ciudad en que opera. Además, pocos modelos capturan correctamente la heterogeneidad sociodemográfica de los usuarios y las variaciones en las características de sus viajes. También existe evidencia sobre relaciones no lineales entre algunos atributos de los viajes y la satisfacción, y que esto también podría depender del tipo de usuario. Estas dos últimas condiciones no se han investigado adecuadamente, ya que la población de usuarios se ha tratado como una masa homogénea, y solo se han considerado relaciones lineales en los parámetros entre atributos y satisfacción. Esto podría introducir sesgos en los modelos, invalidando su uso en aplicaciones prácticas. Finalmente, en la literatura del área de transporte, no se han propuesto modelos psicológico-conductuales que justifiquen los modelos de satisfacción resultantes. Consideramos que esta falta de fundamento teórico es una brecha significativa en esta literatura.

Para abordar estos problemas, esta tesis formula un modelo de satisfacción aplicable a cualquier ciudad y modo de transporte público. Se intenta replicar los resultados en diferentes ciudades y modos, y se realiza una evaluación comparativa para determinar

los atributos comunes más relevantes en todos los casos. Para este fin, se analizan las encuestas de autobuses y Metro en Santiago, utilizando dos instrumentos diferentes. Además, se analizan encuestas de satisfacción para autobuses en cuatro metrópolis de América Latina (Brasil, Chile y México), utilizando el mismo instrumento. Luego se explora tres bancos de datos en Europa, específicamente el sistema ferroviario de Milán, el Metro de Madrid y el sistema de autobuses de Barcelona. Todos estos análisis contribuyen individualmente a la literatura de transporte público desde el punto de vista metodológico y de formulación de políticas públicas. Además, como parte del trabajo de tesis, se diseñó una encuesta de satisfacción experimental basada en fundamentos teóricos, en nuestros resultados previos y en los resultados de dos grupos de enfoques diferentes, y se aplicó al sistema de transporte público (Metro y buses) Transantiago. En esta encuesta, también se exploró el vínculo entre satisfacción y evasión, que es un problema importante en la componente buses de Transantiago y no se ha abordado suficientemente en la literatura.

La tesis analiza si la teoría de la motivación humana de Maslow ayuda a determinar el orden de relevancia de los atributos de satisfacción en transporte público. Para esto se postula la existencia de tres tipos de atributos: funcional-utilitario, protección de seguridad y excitación hedónica. Esta, que es una de las principales contribuciones de la tesis, se basa en el banco de datos sobre sistemas de buses de América Latina. Además, se aplican las metodologías propuestas para responder cuatro preguntas específicas: (i) cómo afecta a la satisfacción la ocurrencia de incidentes críticos; (ii) si, y cómo, una reforma importante de la red de transporte público afecta la satisfacción de sus usuarios; (iii) si la inclusión de variables operativas mejora los modelos de satisfacción propuestos, y (iv) si existe un vínculo entre satisfacción y evasión. Finalmente, se discute la implicación de nuestros resultados respecto a políticas operativas.

Dado que la satisfacción se mide a través de la evaluación de los atributos por parte de los usuarios, y por lo tanto tiene un componente psicológico, se formulan modelos de ecuaciones estructurales (SEM) con el fin de incorporar constructos latentes usando todos los bancos de datos disponibles. El objetivo principal es determinar cuáles son los principales atributos que influyen en la satisfacción y, al mismo tiempo, determinar si las características sociodemográficas y de viaje influyen en las percepciones de satisfacción. Como los modelos SEM-MIMIC permiten examinar cómo la heterogeneidad afecta la medición de variables latentes, se incluye la heterogeneidad transversal de los usuarios debido a diferencias en sus características socioeconómicas y de viaje. Además, se propone modelos SEM-Multigrupo, donde los subgrupos especificados pueden tener diferentes modelos de satisfacción. A continuación, se propone modelos donde las relaciones entre los atributos y la satisfacción pueden ser no lineales. Para esto, se usa la técnica SEM con efectos cuadráticos e interacciones. Esta aproximación nos permite distinguir entre diferentes tipos de atributos de acuerdo con su relación con la satisfacción. También se formulan modelos de clases latentes, con el objetivo de distinguir entre subpoblaciones y capturar preferencias heterogéneas entre ellas. Para esto, se usa el enfoque SEM de mixtura finita. Finalmente, los modelos no lineales y de clases latentes se combinan para capturar preferencias heterogéneas y enlaces no lineales en las subpoblaciones, simultáneamente. De todo este trabajo, se seleccionan los modelos que proporcionan mayor interpretabilidad a los resultados y el mejor conocimiento relacionado con las políticas.

La última contribución de la tesis es una discusión de las implicaciones de los resultados, respecto a políticas operacionales en sistemas de transporte público. Para esto, se analizan las ventajas de nuestros modelos y los resultados propuestos en relación con políticas estratégicas y tácticas. Nuestra hipótesis principal establece que debe haber un orden de preferencia definido según el tipo de atributo: funcional, de seguridad o hedónico, de acuerdo a la jerarquía de necesidades de Maslow. Esto se prueba empíricamente en los diferentes contextos de modo-ciudad. Nuestros resultados, logrados en diferentes contextos urbanos, pueden generalizarse mediante la base teórica propuesta.

Palabras Claves: satisfacción del transporte público; SEM-MIMIC; análisis Multigrupo SEM; jerarquía de necesidades de Maslow; incidentes críticos; evasión

ABSTRACT

To conserve and attract more users, mainly from private transport which is a less sustainable option, public transport systems' authorities require instruments to prioritize investments aimed at improving their services. This requires improving the level of satisfaction perceived by their clients. Satisfaction surveys can detect levels of global satisfaction with the system and its different attributes; for example, frequency, reliability, accessibility, security and comfort.

Although the scientific literature postulates various econometric models to determine which attributes are more relevant in the previous task, there is no consensus about the most appropriate modelling technique. However, the literature reports differences between the importance of different system's attributes depending on the specific mode of transport and the city context where it operates. In addition, few models correctly capture the socio-demographic heterogeneity of users and the variations in the characteristics of their trips. There is also evidence of non-linear relationships between some attributes and satisfaction and that this may also depend on the type of user. These last two conditions have not been properly investigated, since the user population has been treated as a homogeneous mass, and only linear relationships have been considered between attributes and satisfaction. This could introduce biases in the models, invalidating their use in practical applications. Finally, in the transport literature, psychological-behavioural models have not been proposed to justify the resulting satisfaction models; we consider this a significant gap in this literature.

To address these problems, this thesis formulates a satisfaction model applicable to any city and public transport mode. We strive to replicate results in different cities and modes, and carry out a comparative evaluation to determine the most relevant common attributes in all cases. For this purpose, bus and Metro surveys in Santiago are analysed, using two different instruments. In addition, satisfaction surveys are analysed for buses in four metropolises in Latin America (Brazil, Chile and Mexico), using the same

instrument. Furthermore, we explore three data banks in Europe, specifically the Milan rail system, the Madrid Metro, and the Barcelona bus system. All these analyses contribute individually to the public transport literature, both from a methodological point of view and in the context of formulating public policies. In addition, as part of the thesis, we designed an experimental satisfaction survey based on theoretical foundations, our previous results and the results of two different focus groups, and applied it to the Transantiago public transport system (Metro and buses). In this survey, the link between satisfaction and evasion was also explored, which is a major problem of the bus component of Transantiago and has not been sufficiently addressed in the literature.

The thesis also analyses whether Maslow's theory of human motivation helps determining the order of relevance of satisfaction attributes in public transport. For this, the existence of three types of attributes is postulated: functional-utilitarian, security protection and hedonic excitation. This is one of the main contributions of the thesis, and was based on the data bank for bus systems in Latin America. Additionally, the proposed methodologies were applied to answer four specific questions: (i) how satisfaction is affected by the occurrence of critical incidents; (ii) if, and how, a major network reform affects satisfaction; (iii) whether the inclusion of operational variables improves the proposed satisfaction models, and (iv) whether there is a link between satisfaction and evasion. The implication of our results regarding operational policies is discussed.

Since satisfaction is measured by the evaluation of attributes by users, and therefore has a psychological component, structural equation models (SEM) were formulated to allow the incorporation of latent constructs for all the available data banks. The main objective was to determine the main attributes that influence satisfaction and, at the same time, whether sociodemographic and travel characteristics may influence the perception of satisfaction. As SEM-MIMIC models allow to examine how heterogeneity affects the measurement of latent variables, users' heterogeneity was included by considering differences in users socioeconomic and travel characteristics. In addition, SEM-Multigroup models were proposed, where the specified subgroups could have different satisfaction models. Next, we proposed models with non-linear relationships between attributes and satisfaction. For this, the SEM technique with quadratic effects and interactions was used. Therefore, we could distinguish between different types of attributes according to their relationship with satisfaction. In addition, latent class models were formulated with the aim of distinguishing between subpopulations and to capture heterogeneous preferences among them. For this, SEM finite mixture models were used. Finally, non-linear and latent class models were combined to capture heterogeneous preferences and non-linear links in the subpopulations, simultaneously. From all this work, the models that provided greater interpretability for the results and better knowledge related to policy formulation were selected.

The last contribution of the thesis is a discussion of the implications of the results, regarding operational policies of public transport systems. For this, the advantages of our models and proposed results were analysed in relation to strategic and tactical policies. Our main hypothesis states that there must be a preference order depending on the type of attribute: functional, security or hedonic, according to Maslow's hierarchy of needs. This was proven empirically in the different city-mode contexts. Our results, for different urban contexts, can be generalized with the theoretical underpinning proposed.

Keywords: public transport satisfaction; SEM-MIMIC; SEM multigroup analysis; Maslow's hierarchy of needs; critical incidents; evasion

1. INTRODUCTION

1.1 Motivation

Central governments and transport authorities are increasingly under pressure to design systems that meet the population's accessibility and mobility requirements, and also reduce the externalities of congestion, pollution and accidents (Schiller et al., 2010). There are multiple references to the demand of users for sustainable and active modes of transport, such as walking and cycling, and for efficient public transport (PT) regarding energy. Litman (2007), for example, presents a detailed summary. In addition to this, most governments are interested in encouraging private modes' users to transfer to PT (Beirão and Sarsfield Cabral, 2007) to mitigate congestion (Anable, 2005, Steg, 2005). In effect, by reducing vehicle flow, there will be greater mobility, concerning higher speeds and shorter travel times, and, at the same time, the amount of polluting emissions will decrease.

In marketing science, satisfaction is defined as a measure of how the products and services provided by a company meet or exceed the expectations of customers or users (Grönroos, 1984, Oliver, 1980). Sales, or the market shares of a company are a good indicator of its current performance. However, satisfaction is the best indicator of how likely customers will rebuy from any given company. Also, when customers are satisfied with a product or service, they can recommend it to their acquaintances, a useful marketing tool. There is an extensive literature on the relationship between satisfaction and client retention (Oliver, 2010, Olsen, 2007). Companies invest considerable resources in the measurement, control and communication of customer satisfaction with their respective products and services. It is one of the most prolific research areas in marketing.

In recent decades and with the aim of providing better public services, transport authorities in multiple cities have resorted to user satisfaction surveys (de Oña and de Oña, 2014, Fellesson and Friman, 2012). These allow users to report a subjective assessment of their levels of satisfaction both with the complete transport system, with specific services, and with the individual attributes of the service (Donoso et al., 2013). The measurement of satisfaction with PT is a useful tool that allows authorities to obtain metrics of the subjective assessment of users on the fulfilment of their expectations with the service (de Oña and de Oña, 2014). The analysis of satisfaction surveys also allows auscultation of satisfaction with some attributes of a service, such as frequency, waiting time, reliability, access time, security, and comfort. Thus, using econometric models, it is possible to statistically determine which attributes correlate most with the overall satisfaction of the PT mode under scrutiny (Bordagaray et al., 2014, Hensher et al., 2003). By establishing which attributes affect satisfaction more significantly and considering the specific levels of satisfaction of each one, the authority can design operational policy plans associated with each attribute (Mahmoud and Hine, 2016, Nathanail, 2008). It is even possible to design specific policies, for example, maintaining all the relevant attributes at a high level of satisfaction, maintaining those that are not as relevant at a medium level, and not devoting significant efforts to attributes with little relevance to users.

1.2 The Problem

Obtaining the relative importance of attributes concerning user satisfaction is a subject that has been widely developed in the last three decades. De Oña and de Oña (2014) present a comprehensive review that addresses the econometric techniques used so far. Additional reviews are reported by Eboli and Mazzulla (2010) and Hensher et al. (2003). It is important to note that, in the transport literature, the terms *perception of service quality* and *service satisfaction* are used interchangeably. Some of the most frequently used approaches in the PT satisfaction literature are exploratory and confirmatory factor analysis (Fellesson and Friman, 2012), linear regression models (del Castillo and Benitez, 2013), ordinal-Probit models with systematic taste variations (Bordagaray et al., 2014), structural equations models (Eboli and Mazzulla, 2007), factorial analysis and ordinal-logit models (Tyrinopoulos and Antoniou, 2008), hybrid discrete choice models with latent variables (Habib et al., 2011), and binary-logit models (Mahmoud and Hine, 2016).

Most studies reviewed agree that users value the *reliability* of the service over any other attribute; this is consistent with other studies where it is emphasised that the reliabilities of travel and waiting time are the most relevant attributes for users (Currie and Wallis, 2008, Wachs, 1976). In the last two decades, the technique of structural equations models (Chen, 2016, Jen and Hu, 2003, Yilmaz and Ari, 2016) has been preferred over other methods, but still, other methods have also been used. We review the essential gaps found in a literature review on this specific problem.

First, there is no consensus on how to model satisfaction in PT systems. In marketing, SEM have been used more frequently because they allow to define latent constructs (Brady and Cronin, 2001, Szymanski and Henard, 2001), and this agrees with the subjective aspect of satisfaction (Oliver, 2010). Although the transport satisfaction literature tends towards SEM also, the studies reviewed include a variety of econometric techniques.

Second, the heterogeneity of travel and sociodemographic traits has not been incorporated into the models. None of the studies reviewed directly integrates the heterogeneity of users or the characteristics of their trips. Implicitly, it has been assumed that all users have homogeneous characteristics. Some studies note that the difference in the perception of satisfaction may depend on sociodemographic characteristics (Bordagaray et al., 2014, Mouwen, 2015, Yaya et al., 2014). However, the models do not incorporate ways of correcting for this heterogeneity. Additionally, building from the marketing literature, it is possible to postulate the existence of two or more heterogeneous populations that have entirely different tastes and priorities (Anderson et al., 2008, Eisenbeiss et al., 2014). Thus, the assumption of a single homogenous population will hardly be fulfilled. In some contexts, it would be valid to apply latent class models to differentiate between subpopulations. Furthermore, there may be differences in the perception of satisfaction constructs, across groups.

Third, a linear relationship between the attributes of PT and global satisfaction has been assumed and none of the studies examined tested non-linear relationships in this sense. Intuitively, it can be assumed that people have tolerance thresholds for specific attributes that differ from others (Bettman, 1974, Simon, 1956); that is, a service will be satisfactory if it meets certain minimum thresholds regarding the values of these attributes. In the marketing literature, four types of attributes are mentioned in relation to overall satisfaction: (i) dissatisfaction, (ii) critical, (iii) satisfactory, and (iv) neutral (Herzberg, 1959, Johnston, 1995, Kano et al., 1984, Matzler, et al., 2004). The only type that maintains a linear relationship with satisfaction are the critical attributes. Thus, we postulate that the linearity assumption would not be fulfilled for several attributes, so that the relative importance of each one should depend on their current levels.

Fourth, the results obtained have not been theoretically justified. None of the studies reviewed justifies their results, in relation to a psychological theory of behavioural motivation. Given that the dependent variable, satisfaction, is a construct that includes both functional and affective evaluations, we consider plausible to obtain a theoretical justification of causality in a psychological context. In marketing, for example, there is literature (Falk et al., 2009, Kano et al., 1984) that refers to Maslow's theory of human motivation (1943). In particular, Falk et al. (2009) show that for a specific service, shopping online, the relationships between attributes and satisfaction depend on the level of experience of the person; that is, depending on the type of user (new or experienced), the levels they seek in the different attributes and their order of priority will vary. We postulate that the functional elements of transport services could have greater relevance depending on the users' experience with the service, its sociodemographic characteristics, and the current levels of the attributes.

Fifth, in the PT literature, the analysis of the effects of Critical Incidents (CI) during a service on satisfaction have not been adequately developed. The satisfaction-loyalty relation is a well-established fact in the marketing literature (Oliver, 2010, Olsen, 2007, Paulssen and Birk, 2007). Loyalty behaviour refers to the intent to reuse the service and recommend it to friends and family (van Lierop et al., 2017). We found only three previous studies (Edwardsson, 1998, Friman et al., 1998, 2001) that modelled specific CI to explain attribute satisfaction levels and no studies considered the effect of CI on loyalty behaviour. We consider this to be a gap, since CI represents a more tangible policy variable.

Sixth, we found no previous study that addressed a major network reform in a PT system from the users' perceptions (i.e. satisfaction). We believe this to be a critical gap, since understanding how users perceive a major reform is vital for policy-making. Addressing a major reform from the satisfaction point-of-view can help bridging this gap.

Seventh, we found very few studies that dealt with how the real operational variables of a PT system affect users' satisfaction beyond the travel attributes reported by the user. We believe that this is another critical gap in the literature, since it is crucial for the PT authority to know how their operational input variables affect users' perception, for policy-development.

Eighth and last, we found no study in the literature addressing the satisfaction-evasion relationship. Fare evasion in PT is a major problem which hampers the PT authority's resources directly, especially in developing countries. Fare evasion is estimated to cost close to one billion euros per year to PT authorities' worldwide (Bonfanti and Wagenknecht, 2010). We postulate that a direct link exists between users' transit satisfaction and their fare evading behaviour. We will explore this problem with a specific survey design, and posterior modelling of the satisfaction-evasion link.

In summary, we have revealed eight different gaps in the literature. In this thesis, we make hypotheses and attempt to answer all of them with specific scientific evidence. In the next section, we present the hypotheses, objectives, and contributions of this research work.

1.3 Hypotheses

1.3.1 General hypothesis

It is possible to develop public transport users' satisfaction models, which replicate and predict satisfaction in different cities and mode-specific contexts, and which are consistent with Maslow's theory of human motivation.

1.3.2 Specific hypotheses

a) It is possible to develop satisfaction models that capture and correct for heterogeneity, regarding users' sociodemographic and travel characteristics.

b) It is possible to develop satisfaction models that allow for non-linear relationships between service attributes and global satisfaction.

c) It is possible to develop satisfaction models that allow for the inclusion of heterogeneous subpopulations through specific satisfaction models, which differ transversely in the subpopulations.

d) Maslow's theory of human motivation offers a plausible theoretical foundation for models of users' satisfaction with public transport.

e) Including critical incidents enhances public transport satisfaction models, offering policy-related knowledge.

f) Satisfaction models that include heterogeneity can aid in determining if and how major network reforms affect users' perceived satisfaction.

g) Satisfaction models that include heterogeneity can aid in determining if and how operational variables affect users' perceived satisfaction.

h) A satisfaction-evasion relation exists, and a behavioural fare-evader profile can be captured by modelling public transport satisfaction and evasion, accounting for heterogeneity.

1.4 Objectives

1.4.1 General objective

To develop a general model of users' satisfaction with public transport, which explains satisfaction of users in different contexts (modes of transport and cities), and is coherent with Maslow's theory of human motivation.

1.4.2 Specific objectives

a) To develop a satisfaction model which captures the heterogeneity associated with the sociodemographic characteristics of users and their types of travel.

b) To develop a satisfaction model which includes non-linear links between service attributes and overall satisfaction.

c) To formulate a model that allows capturing heterogeneous subpopulations through specific satisfaction models that differ transversally in the subpopulations.

d) To establish empirically whether Maslow's theory of human motivation offers a plausible theoretical foundation for models of users' satisfaction with public transport.

e) To determine empirically whether including critical incidents enhances public transport satisfaction models.

f) To determine if and how a major network reform affects the perceived satisfaction of public transport users.

g) To determine if (and how) including operational variables improves models of the perceived satisfaction of the public transport users.

h) To determine if a satisfaction-evasion relation exists and to develop a behavioural fare-evader profile, by modelling public transport satisfaction and evasion, accounting for heterogeneity.

1.5 Contributions

We discussed eight critical gaps in the literature regarding PT satisfaction models. The first three are covered in Chapters 2 to 4, where we present the methodological tools applied throughout the thesis. The latter five gaps are covered in specific applications presented in Chapters 2 and 5 to 8.

First, a specific methodology that allows modelling satisfaction is proposed, which corrects for heterogeneity of users according to sociodemographic and travel characteristics. Second, a methodological proposal is made that considers possible nonlinear relationships between service attributes and satisfaction. Third, a model that distinguishes between different subpopulations and incorporate them in the model is formulated and estimated. Fourth, we propose a theoretical foundation for PT satisfaction models based on Maslow's theory of human motivation, the hierarchy of needs, giving theoretical support to the proposed models. For this last point and the following, we use the methodologies derived from the previous three contributions. Fifth, we include the presence of critical incidents (CI) in satisfaction models and assess whether doing so improves their results. On this basis, we determine a policy variable that could be controlled by PT authorities. Sixth, we use the satisfaction models to assess if and how performing a major network reform improves or deteriorates the users' satisfaction with the PT system. Seventh, we include operational variables to assess if this improves the resulting models, over the user reported attributes. Following this, we obtain policy-related knowledge that aids decision-making. Finally, we explore the satisfaction-evasion link and develop a fare-evader profile, by modelling PT satisfaction and evasion, accounting for heterogeneity.

In a nutshell, one contribution of this research is to propose a methodological toolkit, by modelling satisfaction with PT allowing for flexibility to first, incorporate heterogeneity of users' perception and, next, allow for different subpopulations with different satisfaction models. A second key contribution consists in providing a theoretical foundation for the satisfaction models in order to draw more generalizable conclusions. For this, we compare satisfaction models in different cities for the same mode and assume that the priority of preference in different user populations is based on Maslow's theory of human motivation. The last contribution consists of four specific applications that allow us to develop additional public-policy knowledge for PT authorities. We investigate how critical incidents (CI), a major network reform, including operational variables, and users' fare evading behaviour, affect overall transit satisfaction. In the last case, we assess which satisfaction attributes induce the intention to fare-evade. All these results are generalizable to different settings and are critical for policy-making.

1.6 Methodology

The methodology consists of four phases: (i) literature review, (ii) data collection, (ii) data analyses, (iv) discussion and conclusions. Each chapter includes all these phases. However, overlapping will be minimised when the methods used are the same. In this section, we offer a brief overview of each phase of the methodology.

1.6.1 Literature review

For the specific objectives of the thesis, we studied four main themes: (i) the concept of satisfaction from the marketing point-of-view; (ii) the state-of-the-art and of the practice in modelling satisfaction with public transport; (iii) the capabilities of modelling latent constructs (i.e. satisfaction) with structural equations models (SEM), and (iv) Maslow's theory of human motivation. In particular, we studied the ability of SEM to model heterogeneity in the measurement of latent constructs, to model nonlinear relationships and to model heterogeneous subpopulations with different satisfaction approaches. We also review the literature on analysing critical incidents (CI) for modelling PT satisfaction. We review studies that analyse PT network reforms from the customer's point-of-view. Finally, we review the literature related with analysing fare evading behaviour from the passengers' point-of-view. All these are essential characteristics to correctly model the relationships between attributes and overall satisfaction in PT systems and, also, to gain policy-decision knowledge. Next, we offer some details on data collection in this research work.

1.6.2 Data availability

Seven different databanks (DB) were analysed, individually:

a) DBA-Santiago Bus: This survey was conducted in nine periods (three per year in 2013-2015). Each period had 5,000 repetitions. Individuals older than 13 years of age were surveyed at bus stops, obtaining their sociodemographic data and asked for their overall satisfaction with: (i) the PT system, (ii) with the specific bus line, and (iii) with the attributes of both the bus line and the entire system. The survey design was external to this investigation, and the data was provided by the *Dirección de Transporte Público Metropolitano* (DTPM) of the Chilean government.

b) DBB-Milano Rail: This survey was conducted in seven time periods (two per year in 2011-2013, and one in 2014). Each period had approximately 14,000 observations, onboard and at train stations. User's socioeconomic characteristics, travel habits, and trip characteristics were collected. Users were asked to express importance and satisfaction rates about 27 service quality attributes, and if they had suffered a CI on the specific attributes, during the last 30 days. Survey design was external to this investigation, and the data was provided by the Milan rail operator.

c) DBC-Madrid Metro: We analysed data collected in the fourth campaign of surveys conducted in 2015. Ten thousand passengers were interviewed in four subsamples, of approximately 2,500 interviews each. Travel characteristics and sociodemographic data of users were collected. Users were asked for satisfaction ratings of the overall Metro service and specific attributes of the service. There are four versions of the questionnaire. In each one, respondents expressed their perceptions about six (or seven) different service quality attributes. Then, all users were asked about inconveniences caused by any possible service interruptions (i.e. critical incidents). Finally, passengers were asked to express the intent to recommend Metro in a 1-10 scale, and the willingness to take part in future customer satisfaction studies conducted by the company. The survey design was external to this investigation, and the data was provided by the Madrid Metro operator.

d) DBD-Latin America Buses: This survey was conducted in one period (2014-2015). It considers between 800 and 2,000 observations, and users were surveyed at bus stops in different cities of Latin America: Santiago, Mexico City and four Brazilian cities (called *Cidade* A, B, C, D, since they are confidential data). Only individuals over 18 were surveyed, about their sociodemographic data, and their overall satisfaction with the PT system and with the attributes of both the line used and the complete system. The design was external to this research and the data was provided by SIMUS (Santiago, Mexico City) and EMBARQ (Brazil).

e) DBE-Barcelona Bus: This survey was conducted in three time periods (one per year between 2013 and 2015), and considered 5,000 observations on each period. Users were asked about their satisfaction with specific service attributes, and their sociodemographic and travel characteristics were also obtained, primarily if they used one of the new *Nova Xarxa de Bus* (NXB) lines, associated with the network reform. Once more, the design was external to this thesis and the data was facilitated by the TMB company.

f) DBF-Santiago Metro: This survey has been conducted monthly for several years, and we obtained data for 42 periods (36 between 2013-2015 and six in 2016). Each period comprises 1,000 observations of users between 18 and 60 years of age, who do not have a special discount pass, who are asked about their satisfaction with specific attributes of the service. The survey is conducted on the platforms, but users are accompanied onboard the train, if necessary. The survey design was external to this investigation and the data facilitated by Metro S.A. We also obtained operational variables for the Metro, and over 1,4 million records were analysed.

g) DBG-Santiago Metro and Bus: This satisfaction survey was based on an experimental design applied to Transantiago users in 2018, as part of this dissertation. First, two focus groups were used to inquiry about the PT attributes that users find relevant, both in Metro and buses. Then, an experimental design was generated based on a thorough literature review, the focus groups results, specific items detected in the above surveys, and other attributes (i.e. theoretical) that have not been considered before. Moreover, a set of items related with perception and (dis)satisfaction with evasion behaviour were included, plus two more items associated with users' fare evading behaviour, to allow exploring the satisfaction-evasion link. One thousand interviews were conducted in each mode.

For each databank specific analyses were performed. In the next subsection, we discuss briefly the analytical methodologies employed throughout the dissertation.

1.6.3 Data analyses

The data analyses were done using the R programming language and software (R Core Team, 2013). This has several package libraries that allow various statistical analyses, among which there are several SEM type models. In particular, the lavaan package (Rosseel et al., 2015) allows estimating SEM and MIMIC models, and the nlsem package (Umbach et al., 2015) allows modelling the two proposed finite mixture techniques (i.e. SEM with latent classes and non-linearities).

Throughout the dissertation, we applied the same specific analyses to different data banks. For example, the attribute-satisfaction items were analysed to form latent constructs. For the measurement system, principal components analysis (PCA) (Jolliffe, 2014) was done first on the set of items associated with the attributes, to obtain the principal components that are the basis of the latent constructs. Confirmatory factor analysis (CFA) was then carried out (Brown, 2015), to obtain a satisfactory measurement model that complies with proper global adjustment indices (Bentler, 1990, Hooper et al., 2008, Hu and Bentler, 1999). Also, the alpha value of Cronbach (1951) was used to assess internal validity, and all constructs were analysed for theoretical support (Churchill, 1979). Since satisfaction ratings are usually ordinal, there are two ways to model them: continuous/numeric and ordinal. We used both methods and, additionally, compared their effectiveness.

Our general modelling approach was as follows. We first used SEM-MIMIC models based on the hypothesis of differences in perception in all latent constructs regarding sociodemographic and travel differences (i.e., different subpopulations perceive the latent constructs differently). Binary variables were constructed for the socio-demographic and travel type categories. In the MIMIC models, these variables enter the latent constructs as regressors and those found significant remain in the model. To choose the best model, Z-statistics and global adjustment indices were used. As stated before, the satisfaction rates were first treated as numeric, and then the ratings were analysed with an ordinal-Probit model (McCullagh, 1980, McKelvey and Zavoina, 1975).

Second, we used SEM-MG (Multigroup) analyses to test for different satisfaction models across previously defined subpopulations. Third, we tested if there are non-linear relationships between attributes and satisfaction: the NL (non-linear) model. Fourth, we tested for the existence of heterogeneous subpopulations: the FM (finite mixture) model. Finally, we tested for a combination of the latter two, the NL-FM (non-linear-finite mixture) model (Klein and Muthén, 2007, Klein and Moosbrugger, 2000). In summary: descriptive statistics analyses, PCA, and structural equation models were performed using the R libraries.

1.6.4 Discussion and conclusions

Each chapter of the thesis includes its own discussion and conclusions. Notwithstanding, the last chapter considers our final conclusions regarding the completion of the dissertation's main objectives, the contributions, and future research.

1.7 Dissertation Structure

The thesis has nine chapters. After this Introduction, the next three chapters present the methodological toolkit used throughout the dissertation. Chapter 2 refers to the SEM-MIMIC model applied in the Santiago bus system, Chapter 3 presents a SEM-MIMIC ordinal Probit model applied in the Milan rail system, where the effect of critical incidents was analysed, and Chapter 4 presents the SEM-MG analysis models applied in the Madrid Metro system.

Chapter 5 presents one of our main results: the testing of the proposed Maslow's hierarchy of transit needs developed for the priority of relevance of the satisfaction attributes, using the Latin America's bus systems databank.

The next three chapters present specific policy-related applications regarding satisfaction with public transport. Chapter 6 deals with the effect of major network reform on transit satisfaction. Chapter 7 analyses if including operational variables improves the resulting satisfaction models, and Chapter 8 inquires whether users' fare evading behaviour affect overall transit satisfaction, and also explores the behavioural fare-evader profile.

Finally, Chapter 9 presents the conclusions of the dissertation: accomplishment of objectives, contributions and future lines of research.

2. MODELLING SERVICE-SPECIFIC AND GLOBAL TRANSIT SATISFACTION UNDER TRAVEL AND USER HETEROGENEITY

2.1 Introduction

A central objective of any public transport (PT) administrator should be to provide a better level of service for its customers, to retain them and to attract new users, mainly from modes such as the private car. In the PT literature, the heterogeneity of travel conditions and sociodemographic characteristics of PT users has not been adequately captured. Additionally, very few studies have considered the service-specific satisfaction and the global satisfaction independently.

This chapter focuses on an application in Santiago, Chile, taking advantage of recent interest by the authorities to improve the service quality delivered by Transantiago, the capital's integrated public transport system (Muñoz et al., 2009). Periodic surveys measuring users' satisfaction with its bus system component and information regarding users' type of travel and demographic data are analysed jointly here, to estimate the satisfaction with the specific service (bus-line) used by the respondents, and also their global satisfaction with the system. We use a structural equation multiple cause multiple indicator (SEM-MIMIC) modelling approach. When applied to PT systems, this novel methodology could help the authorities to prioritise resources for improving the system while considering users' travel heterogeneity.

Our work allows us to make the following contributions. First, we propose a complete structural equation model including two separate regressions for both Bus-line Satisfaction (service encounter) and System Satisfaction (global), for the bus component of an integrated public transport system, Transantiago. To the best of our knowledge, this is a novel approach that brings marketing concepts into the PT arena and allows for both bus-line and global PT system insights. The ability to correct for heterogeneity is fundamental in the model, which provides valuable information regarding how the different satisfaction latent constructs are perceived. The framework could be generalizable in any service setting where one can separate service transaction-specific and system attributes (Jones and Suh, 2000). In each case, the modeller would need to consider user's sociodemographic and specific service encounter characteristics, to construct a MIMIC model.

Second, we introduce a three-step MIMIC model, including a measurement model, a MIMIC structural model, and a Mediation Analysis (MacKinnon et al., 2007) for PT satisfaction. We introduce the MIMIC model into the PT satisfaction paradigm to consider various travel and user heterogeneities. A novelty, since current literature has not captured these elements adequately. We only found one previous study which used the MIMIC approach for modelling PSQ (Guirao et al., 2016). Our framework leads the way for future research studies to include heterogeneity into MIMIC models when utilising the SEM approach. The Mediation Analysis permits acquiring information on how primary policy variables affect System Satisfaction. Finally, we present the first SEM-MIMIC-Satisfaction application reported in the literature for a Latin American PT system.

The rest of the chapter is organised as follows. Section 2.2 summarises the results of several literature reviews. First, we present an analysis of the theoretical framework of PSQ and user satisfaction in the service setting. Then, we offer a summary of contemporary research approaches regarding the search for relevant factors explaining users' satisfaction with PT systems, focusing on previous attempts to consider heterogeneity. Last, we present a review of other essential aspects concerning satisfaction with PT, such as attitudes towards PT and growing mass transit markets. Section 2.3 offers a methodological review, briefly describing structural equation models, the MIMIC (Bollen, 1989) approach, and Mediation Analysis. In Section 2.4 we introduce our Case Study, we describe the Santiago case, specifically the Transantiago bus system, provide details about the sample, and the survey. In Section 2.5 we present the Model structure and results.

Finally, in Section 2.6 we present the discussion, and in Section 2.7 we conclude with our most important findings, alongside key policy recommendations for the Transantiago administration.

2.2 Literature Review

There are two different perspectives when modelling service quality and user satisfaction: perceived quality based on users' experience, and expected quality, which determines the users' expectations from a PT system (Bordagaray et al., 2014). In the marketing science literature, Miller (1977) found that when asked about expectations, customers elicited several different types of expectations, including expectations of ideal, minimum, predicted and normative performance. Therefore, depending on the type of expectation measured, different strengths of relationships with other constructs were found. Thus, in this research, we base our focus on perceived service quality (PSQ), which is also the most commonly used framework in the PT literature.

PSQ has been found to be one of the most important constructs in recent marketing literature (Laroche et al., 2004). It is a significant variable that correlates well both with customer satisfaction and value. PSQ indirectly measures how well the service delivery matches or exceeds customer's expectations. Hensher et al. (2003) note that as passengers may view specific aspects of service quality as positive or negative, the overall level of passenger satisfaction is best measured by how individuals evaluate the total package of services offered. The satisfaction level is an aggregate measure of perceived users' satisfaction with different aspects of the transport system (del Castillo and Benitez, 2012). Customer satisfaction is regularly measured using satisfaction surveys. In marketing, satisfaction ratings are considered means to strategic ends, such as repurchase behaviour and customer retention, that directly affect a firm's overall performance (Mittal and Kamakura, 2001). Customers' evaluations of service quality and their ratings of satisfaction offer critical inputs for developing marketing strategies (Ofir and Simonson, 2001). Regarding service quality, Parasuraman et al. (1985) state three essential features: intangibility, heterogeneity and inseparability. Services are intangible because they involve performance rather than object creation. They are heterogeneous because their performance often varies from producer to producer, from customer to customer, and from day to day. Finally, the production and consumption of many services are indeed inseparable. Additionally, in PT systems, users are subject to natural perturbations in passenger flows and congestion, which vary by the hour, causing additional variability in the services provided.

PSQ is the degree and direction of discrepancy between consumers' perceptions and expectations. In their exploratory work with focus groups for different scenarios, Parasuraman et al. (1988) produced a multiple-item scale instrument (SERVQUAL) to assess PSQ under different service settings. They reported five main dimensions: tangibles, reliability, responsiveness, assurance, and empathy. The objective of their instrument was to determine the relative importance of these dimensions in influencing customers' overall quality perceptions. When the original SERVQUAL experiment took place, *a striking result* reported by the authors was that reliability was consistently the most critical dimension and assurance the second.

We frame our application on the relationships between the concepts of sacrifice, PSQ, service value, and satisfaction. Cronin et al. (2000) conceptualised the effects of these four constructs on consumers' behavioural intentions. Their research empirically verified that service quality, service value and satisfaction were all directly related to behavioural intentions; their so-called "Research Model" is presented in Figure 2-1. Sacrifice is what is given up acquiring a service; we hypothesise that it affects Satisfaction negatively, as it enhances expectations. We also assume here that by improving PSQ and satisfaction, we are indirectly improving the user's intent to reuse the system and recommend it; a well-established fact in the marketing literature (Oliver, 2010, Olsen, 2007, Paulssen and Birk, 2007).

There is ample and relatively current research literature regarding how to obtain the relative importance factors of PSQ in PT systems based on customer satisfaction surveys. De Oña and de Oña (2015) offer a comprehensive review of the subject.

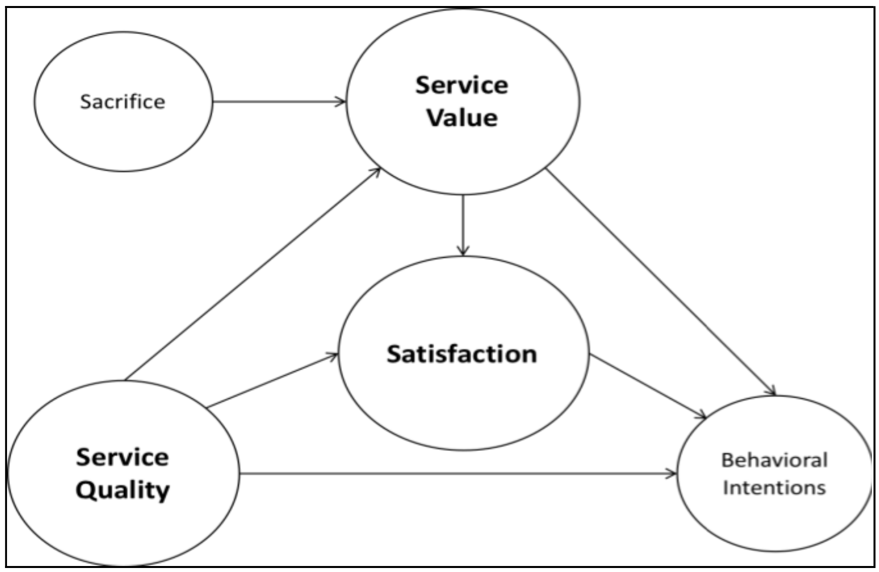

Figure 2-1: The Research Model

Adapted from: Cronin et al. (2000)

Herein we consider studies, which are relatively similar to our proposal, synthesising the particular approaches and main findings regarding satisfaction with PT systems; we include recent studies that have attempted to incorporate traveller's heterogeneity into the models. We report their methodologies and key results.

Eboli and Mazzulla (2007) formulated a structural equations model to assess the impact of service quality attributes on the global customer satisfaction. They analysed the bus service used by students to reach a university campus outside the urban area of Cosenza (Italy). Their key finding was that the latent variable with the most significant effect on global satisfaction was Service planning and reliability. De Oña et al. (2013) also used a structural equation approach to reveal specific latent aspects describing the perception of overall service quality using data from Granada (Spain). Their main result was that the latent construct obtaining the highest weight on overall PSQ was Service, a construct dependent on frequency, regularity, speed, proximity, and fares. Yaya et al. (2014) used exploratory and confirmatory factor analysis, and SEM, to assess causal paths and service quality perceptions among users with different demographic characteristics in the medium-sized city of Girona (Spain). Their findings showed that age and owning a driver's license were factors that directly and positively affected service quality, while education affected it negatively. The authors did not find any other factors to be significant.

Fellesson and Friman (2012) used factor analysis to compare the perceived satisfaction with the PT systems in nine European cities. Their study identified four satisfaction dimensions: system, comfort, staff, and safety. They found differences regarding how individual items loaded on attributes and in the general dimension structure between the cities. In other words, the satisfaction constructs were not cognitively perceived similarly in every city due to dissimilarities in culture and tradition.

Tyrinopoulos and Antoniou (2008) used two statistical methods, factor analysis and ordered logit modelling, to assess the policy implications of the variability of users' satisfaction across operators. They analysed five transit systems in Athens and Thessaloniki (Greece). The most relevant satisfaction attributes were service frequency, vehicle cleanliness, waiting conditions, transfer distance, and network coverage. Iseki and Taylor (2010) developed a methodology based on importance-satisfaction analyses and ordered logit regression models, to examine transit users' perceptions of services and the built environment at PT stops and stations in Los Angeles (USA). Surprisingly, their main result was that users cared more about personal safety and frequent, reliable services than for the physical conditions of the transit stops. Habib et al. (2011) estimated a hybrid multinomial logit model (with latent variables) to define perceptions and attitudes towards transit service quality in Calgary (Canada). Their main finding was that people valued reliability and convenience higher than ride comfort. Del Castillo and Benitez (2012) estimated three types of models to determine the relationship between global satisfaction ratings and the satisfaction evaluations from users of the public bus company of Bilbao (Spain): weighted means, multivariate discrete distribution, and a generalised linear model. Their main finding was that the most relevant items were reliability, adequacy of the bus-stop location, punctuality, connection to other lines and service frequency.

The following studies incorporate heterogeneity in their analysis. Bordagaray et al. (2014) modelled bus transit quality in the city of Santander (Spain) using travellers' perception data. PSQ was estimated with a random-ordered Probit model allowing for systematic taste variations (Ortúzar and Willumsen, 2011, p. 279). Reliability was found to be an essential factor for users; other aspects of high relative importance were journey time, available information and driver kindness. Tarigan et al. (2014) performed segmentation and ordered Probit analysis using paratransit users' opinions on service quality for their trip-making behaviour in Bandung (Indonesia). They divided users into six segments, confirming that paratransit users, especially women, perceived negative experiences related to failing to achieve on-time performance and security issues inside the vehicles to be the most critical attributes regarding PSQ. De Oña et al. (2015) considered perceptions by different groups of users regarding transit service quality for a suburban rail service in Milan (Italy). Using a regression and classification tree (CART) approach, they found that although the quality was perceived differently among the various groups, the most critical factors for the overall sample were regularity and punctuality.

Mouwen (2015) analysed a satisfaction survey with national PT services in the Netherlands, using moderating effects to evaluate whether user characteristics affected the way satisfaction was perceived. He found that people over 65, those that used regional trains and users living in dense urban areas perceived satisfaction differently. The most important attributes were punctuality, the speed of travel and frequency of service. Mahmoud and Hine (2016) investigated the influence of PSQ on active and potential users in Belfast (UK), using a binary logistic regression. They quantified the relationship between PSQ and overall user perception through 29 indicators. They found eleven attributes to be essential: frequency, reliability, transfer and waiting time, station/stop safety, comfort, discounted monthly ticket availability, stop information, fare, need for transfer, stop location, and availability of a park and ride service. Morton et al. (2016) used factor analysis to model the difference in perceptions of bus riders in Scotland. They found that PSQ varied across socioeconomic cohorts, females tended to exhibit relatively negative opinions regarding bus interior quality; also, that improvements to service frequency, availability, reliability, and stability would increase satisfaction. Finally, Guirao et al. (2016) performed the only study we found that included SEM-MIMIC models applied to PSQ in PT. Using data from Madrid (Spain), they derived direct estimates of the importance of service attributes: age, gender and trip purpose were significant characteristics that affected the latent constructs.

All studies reviewed in this section systematically concur that users value reliability over any other attribute. This result is consistent with the SERVQUAL experiment (Parasuraman et al., 1988), where reliability had the highest relative importance over other dimensions, across different types of services. We did not discover any literature on similar models applied to Latin American transit system data, as most studies come from Europe. With few exceptions, most authors do not consider user and travel heterogeneity in their econometric models sufficiently. None considered service specific and global satisfaction separately. Consequently, we propose A SEM-MIMIC framework that considers user perception heterogeneity, integrates it into a model that accounts for service specific and global satisfaction and does this for Latin American PT data. Next, we summarise some relevant studies regarding consumer attitudes towards transit and a literature review on growing PT markets.

Wachs (1976) examined the literature regarding users' attitudes towards transit services in the 1970s, and his key findings still hold. One of his original statements was that attitudinal and modal-choice research revealed that total travel-time reliability was more relevant than total elapsed travel-time. He also concluded that out-of-vehicle time was more onerous than in-vehicle time and stated that travel cost reductions were probably less important than improving elements of travel time as a strategy for attracting a larger user base. The only relevant conveniences that should be included in transit improvement programs were temperature control and reductions in the proportion of passengers who must stand, an increase in seat assurance. The difference in responses to alternative transportation modes was caused by disparities in users' experiences and the service levels impacting different sociodemographic groups. The author also stated that with favourable performance on the dimensions above, the services could attract riders from demographic groups traditionally considered automobile users.

Beirão and Sarsfield Cabral (2007) performed a qualitative study of both PT and car users to understand travellers' attitudes towards transit and perceptions of PT service quality based on a series of in-depth interviews in the Porto region (Portugal). As often found, car was considered the most attractive transport mode for its convenience, speed, comfort, and individual freedom; and PT would need to adjust its services to levels desired by consumers to induce a modal shift. They also found that different user segments might evaluate the same service quality construct differently and that different service attributes could influence their satisfaction. Another significant finding was that the choice of transport mode was influenced by factors such as individual characteristics and lifestyles, type of travel, perceived performance, and situational travel variables. These results advocate the need for segmenting the user base considering travel attitudes and trip context (e.g. purpose).

Currie and Wallis (2008) provide a summary of evidence about PT patronage growth based on bus improvement measures in urban settings. Their compendium includes experiences in Europe, North America and Australasia, and focus on service improvement measures such as network structure, service levels, bus priority measures, vehicle and stop infrastructure, fares and ticketing systems, passenger information and marketing, and personal safety and security. Their source is a mixture of a literature review, their own experience, and the results of an international Delphi survey of bus experts regarding improvement measures specially designed for patronage growth. Their main finding from the synthesis of evidence, is that elasticities from the three primary attributes of a bus service (fare, frequency, and in-vehicle time) are of a similar magnitude. Nonetheless, if funding were no problem, the greatest user base growth increase would most probably be the result of improved frequencies and service levels, followed by reduced fares, and finally reduced travel times. Additionally, they found that in places where reliability is low, improving it could provide significant patronage gains at a relatively low cost.

Additionally, the international bus experts' survey identified the following key features to achieve high levels of bus patronage growth: service frequency increases, bus reliability, and the speed features associated with a Bus Rapid Transit (BRT) system. Again, the reliability dimension appears as one of the main factors driving users' behavioural intentions from a market perspective.

2.3 Methodological Review

SEM has been applied to psychology, sociology, education and marketing research due to its easiness of modelling latent constructs. It has also seen applications in economics to measure abstract constructs such as value and socioeconomic status, and applications in travel behaviour research date from 1980 (Golob, 2003). Its use is growing due to increasing user-friendly software.

A SEM is a flexible linear-in-parameters multivariate statistical technique (Golob, 2003) that allows handling a large number of endogenous and exogenous variables and also latent (unobserved) variables. The latent constructs (factors) represent theoretical, abstract concepts or phenomena such as attitudes, behavioural patterns, cognition, social experiences, and emotions that cannot be observed or measured directly (Bowen and Guo, 2012). Factor or measurement models focus on how one or more latent constructs are measured, or represented, by a set of observed variables. Thus, for intangible concepts, a measurement model allows capturing the underlying construct metric via a finite number of indicators that correlate with the latent variable's real value.

A SEM is a confirmatory method since it requires the analyst to build a model structured in a system of unidirectional effects of one variable on another. Standard regression models test hypotheses about the strength and direction of relationships between predictor (regressor) variables and a specific outcome variable. SEM additionally, can accommodate regression relationships among latent variables and between observed and latent variables. Finally, SEM can also estimate in a single analysis system models where one or multiple variables are predicted and predictor variables at the same time (Bowen and Guo, 2012). A SEM is estimated using the covariance analysis method (Bollen, 1989, Hoyle, 2012). After estimation, various goodness-of-fit tests can be used to decide if the specified model is consistent with the pattern of variance-covariance in the data (see Hooper et al., 2008).

A SEM can be composed of up to three sets of simultaneous equations: (i) a measurement model for the endogenous (dependent) variables, (ii) a measurement model for the exogenous (independent) variables, and (iii) a structural model. The measurement model specifies latency (unobserved) variables as weighted averages of other variables in the system, which are called indicators of the latent constructs, and is similar -but not the same- to factor analysis (FA) and principal component analysis (PCA). PCA is a statistical procedure that uses a transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components (see Jolliffe, 2014). However, in PCA all elements of the matrix defining the latent variables expressed as linear combinations of observed variables take on non-zero values (Golob, 2003).

In a SEM and confirmatory factor analysis (CFA), the modeller decides which parameters may be restricted to be zero, freely estimated or even constrained. Since the measurement model allows for a large number of possible combinations, exploratory factor analysis (EFA) or PCA are often used to guide the construction of a structural equations measurement model. A SEM offers the additional capabilities of being able to estimate direct, indirect, and total effects; direct effects include all links between a productive variable and the variables that are the target of the effect. They represent the causal modelling aspect of a SEM. In synthesis, a SEM allows for endogenous and exogenous variables with measurement errors, latent variables with multiple indicators, and separation of measurement errors from specification errors. It also allows testing a complete model rather than just its coefficients, modelling mediating variables (indirect effects), accounting for missing data, and the analysis of non-normal data (Golob, 2003). Next, we will review the SEM-MIMIC model and the Mediation Analysis technique.

The Multiple Cause Multiple Indicator (MIMIC) model (Hauser and Goldberger, 1971, Joreskog and Goldberger, 1975), allows for the possibility of detecting heterogeneity in the measurement of latent variables between different groups of the population when using a SEM. The MIMIC approach allows considering the restriction of a group-invariant covariance matrix for the observed response variables (indicators), conditional on grouping variables represented by regressors (see Figure 2-2). In other words, we can estimate group differences on the perceptions of the latent variables through the specification of the MIMIC model

(Kline, 2015, p. 323), where the latent variables are regressed on one or more binary indicators that represent group membership.

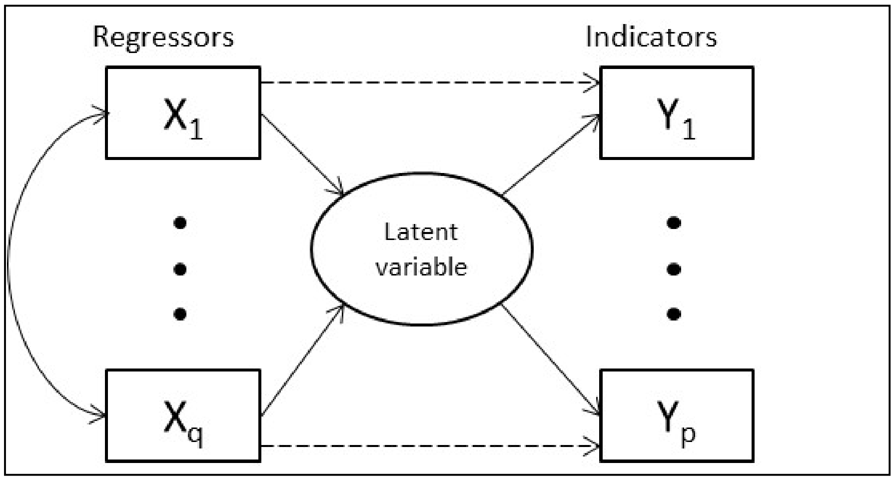

Figure 2-2: A one-factor MIMIC Model

Adapted from: Muthén (1989)

Advantages of the SEM-MIMIC approach are that there is no need to partition the population into subsamples at the modelling stage. Hence, there are no special identification requirements beyond the usual ones for single samples (see Kline, 2015, p. 323). Additionally, one can test several different grouping variables all at once, instead of performing a multigroup analysis with just one variable at a time.
In essence, the inclusion of a set of relevant explanatory variables provides MIMIC modelling with valuable extra information about the measurement model and enables the investigation of hypotheses of invariance across subpopulations (Muthén, 1989). For a highly heterogeneous population, such as bus users, one could model the latent constructs for PSQ and condition the latent variables for different groupings using the explanatory variables (regressors), and test for differences across the subpopulations. Fundamentally, the MIMIC model provides extra information about the possible causes of differences in the measurement of the latent variables, across groups. As an example, if we have a binary regressor

such as gender, where 0 means Male, and 1 represents Female; if we use this variable as a regressor and obtain a negative sign, this would mean that, in general, Female users perceive this latent construct more negatively compared to Male users.

With a highly heterogeneous population, we could use any categorical binary regressor to test for heterogeneity in the perceptions. Hence, this approach proved convenient in our case, since we expected differences in the perceptions across several types of travel and sociodemographic characteristics, the model may provide useful information for policy design. It is important to note that with this approach we assume that the structural coefficients are the same for all subpopulations. In our case, these are the parameters in the Bus-line and System Satisfaction structural regression models. The MIMIC approach has also been used for hybrid discrete-choice analysis in the travel and marketing behaviour literature (Ashok et al., 2002, Raveau et al., 2010).

Mediation Analysis (MA) focuses on relations between two variables, X and Y. In its purest form, mediation represents the addition of a third variable to this X ➔ Y link, whereby X causes the mediator, M, and M causes Y, resulting in X ➔ M ➔ Y. Mediation is used to determine the total effect that the variable X has on Y, considering that mediator M (see MacKinnon et al., 2007).

In this study, MA refers to the total effect of all travel and sociodemographic characteristics (regressors) on our variables of interest: Bus-line Satisfaction and System Satisfaction. We will build a model where the regressors affect the mediating variables (latent satisfaction constructs) and the dependent variables, Bus-line Satisfaction and System Satisfaction. The MA will consider all possible paths of the regressors to the System Satisfaction, including the effects through the mediating latent satisfaction constructs and Bus-line Satisfaction (see Figure 2-5, Appendix B). The regression coefficients and the mediating paths are both considered when computing the total effects. The standard errors of the mediated effects are computed using the multivariate delta method (MacKinnon et al., 2007).

This technique will prove valuable in determining total effects of significant regressors in our model.

2.4 Case Study, Sample and Survey

Transantiago is the mass transit system that serves Chile's capital. It is considered the most ambitious transport reform undertaken by any developing country. In Chile, some label it the worst public policy application in the nation's history. It was launched chaotically in February 2007 in a "Big Bang" fashion where services were fully modified, and fare integration and a touchless smart card were introduced in a single day. There were several problems with the design and implementation of the plan (Muñoz et al., 2008) causing significant inconveniences to users. They were forced to make more transfers, transport capacity was insufficient, bus reliability was inferior, and the new system was difficult to understand for its users. As a result, the Metro system, a highly efficient public company, became the preferred mode of the system by many, as it was its most reliable component.

Transantiago had the following primary objectives:

a) Encourage and promote the use of public transport.

b) Enhance the quality of public transport.

c) Diminish the city's air/noise pollution by reducing the number of buses and improving its standards.

d) Reduce public transport travel times.

The previous system had lengthy and inefficient bus-lines operated by private operators competing with the Metro without integration. Transantiago implemented an integrated fare structure, allowing for a minimal surcharge for transfers between buses and Metro, and introduced a new contactless smart card. The fare integration permitted many PT users in Santiago to afford Metro for their commuting trips, causing Metro demand to almost double. The new system also featured a set of trunk bus-lines complementing the Metro network as the backbone of the system and a set of feeder bus-lines handling shorter trips to feed the Metro and trunk bus-lines.

In 2010, a new governmental body called Metropolitan Transit Directory (DTPM) started managing Transantiago operations. DTPM has focused on continuous improvements to the system, both at service levels and regarding overall system quality. Since 2012, modifications to the contracts with the private bus operators were implemented to allow the creation of incentives for better performance of the scheme, which included penalties for non-compliance of various service parameters, and rewards for controlling fare evasion, which had rocketed.

In June 2012, DTPM discarded the notion that different companies should operate the feeder and trunk lines. Connections between lines serving different zones were encouraged to eliminate unnecessary transfers and thus provide better services. From that moment on, DTPM formed seven Operating Units (U1-U7), each assigned to a concessionaire that included both trunk and feeder lines. This configuration remains until today.

As of January 2013, DTPM has periodically measured the service delivered by Transantiago, regarding overall user satisfaction, through a study commissioned by a consulting company. The primary objective of this study is to monitor user satisfaction continuously. It also aims to determine which aspects have higher relative importance regarding the satisfaction provided by the system and by the different bus-lines operators.

One of the main problems that remain to date in the system is that the present contracts do not force operators to stop at every stop: "… *For example, an issue that has become more and more frequent is that buses go past bus-stops. An incentive to avoid this is needed*." (Beltrán et al., 2013). This issue may prove to be of high relevance in our present study since the frequencies reported and the compliance measures enforced by contract on the operators may not adequately reflect the frequencies perceived by the users. For example, Gómez-Lobo and

Briones (2013) state: "*However, some operators could meet these operational targets (compliance measures) by supplying the correct number of kilometres, buses, and capacity and still offer a bad service; for example, by not stopping at bus-stops when requested*."

The satisfaction survey mentioned above was conducted three times a year, from January 2013 until May 2014, and applied to a sample of approximately 5,000 Transantiago users (men and women over 13 years) in every wave. It is a descriptive quantitative analysis, based on face-to-face questionnaires filled at bus-stops. The semi-structured questionnaire is organised as follows. Part 1 includes questions about the users' socioeconomic characteristics (age, gender, and socioeconomic status, SES), time of day, frequency and motive of travel. The SES is obtained through a methodology used by marketing companies in Chile, based on a double entry classification of occupation and education level (this original information was not made available to us). The SES is divided into seven categories: A (high-high), B (high-low), C1 (middle-high), C2 (middle-middle), C3 (middle-low), D (low-high), and E (low-low). A survey requisite is that the current service (bus-line) is the most frequent service used by the respondent.

In the next parts of the survey, users were asked to assign a score related to their perception of the service. The rates are given on a 1-7 scale (where one is very poor, and seven is excellent), formulated as a Likert scale (Likert, 1932). As in Chile, school and university grades use this 1-7 range, with four (4) being the minimum passing grade; the population understands the survey's scoring system intuitively.

Part 2 includes the System Satisfaction Items. Specifically, users have to assign a satisfaction score to the whole system (System Satisfaction - P1). Additionally, there are 9 attribute-specific satisfaction items related to the Overall PT system (P2C-P2K).

Part 3 includes the Bus-line Satisfaction Items, users assign a satisfaction score to the specific bus-line being used (Bus-line Satisfaction - P3). Also, there are 17 attribute-specific bus-line items. These relate to the frequency of the buses (P5A-P5B), quality of bus-stops (P6D-P6F), convenience and accessibility (P7A-P7E), and quality of buses and driver's behaviour (P8A-P8G), all specific to the bus-line.

Part 4 includes three peripheral items to the bus PT system related to satisfaction with other user's behaviour (P9), the Metro (P11), and the information provided by the PT system (P19). These three items are not bus-line specific.

The sample in this research consists of the five complete surveys totalling 25,094 questionnaires, equally distributed in each of the five surveys performed in January 2013 (Summer), May 2013 (Autumn), November 2013 (Spring), January 2014 (Summer) and May 2014 (Autumn). The main characteristics of the sample are presented in Table 2-1.

Table 2-1: Sample characteristics

Characteristics	**Percentage**
Socioeconomic Status (SES)	ABC_1C_2 (high/middle): *16%*, C_3 (middle/low): *66%*, DE (low): *18%*.
AGE (years)	13-17: *7%*, 18-29: *38%*, 30-45: *30%*, 46-60: *20%*, 61+: *5%*
GENDER	Male: *42%*, Female: *58%*
TIME	AM-Peak: *35%*, PM-Peak: *38%*, Off-Peak: *27%*
DAY	Weekday: *74%*, Saturday: *18%*, Sunday: *8%*
Frequency of Bus Use (days/week)	0-4: *19%*, 5-7: *81%*
Frequency of Metro Use (days/week)	0-1: *32%*, 2-4: *21%*, 5-7: *47%*
Trip Purpose:	Work *66%*, Study: *16%*, Other: *18%*
Perceived Waiting Time (PWT, min)	0-6: *26%*, 6-10: *31%*, 10-20: *31%*, 20+: *12%*
Perceived Travel Time (PTT, min)	0-20: *28%*, 20-40: *33%*, 40-60: *24%*, 60+: *15%*

All these measures, except the SES and gender, are self-reported; therefore, the perceived waiting time (PWT) and perceived travel time (PTT) represent what users perceive to experience when travelling in their regular bus-lines. We note that most of the sample belongs to the C3 (middle-low) socioeconomic category (66%), the majority also belong to both the 18-29 and 30-45 age groups (68%). Most of the users are Female (58%). The survey was performed mainly in both peak periods (73%) and during weekdays (74%). Most respondents (81%) are frequent bus users, and half of them (47%) are also frequent Metro users. Additionally, most users (62%) have a PWT between 6 and 20 minutes, and most of them (85%) have a PTT between 0 and 60 minutes, while only 15% have PTT of more than an hour. It is safe to say that most users are commuting workers that frequently travel in the peak-hours and during weekdays, belonging to Chile's middle/low-class stratum.

Next, we present the average scores of all items from the survey in Table 2-2. All items are divided into their specific domains. From the overall satisfaction items, we note that users evaluate better their particular bus-line (P3, 4.61) than the overall system (P1, 4.39). From the system domain, the best score belongs to the room for improvement item (P2H, 4.91), and the worst one to the availability during weekends item (P2K, 4.07). From the frequency domain, the bus frequency item (P5A, 4.35) has a lower score than the reliability item (P5B, 4.56).

From the bus-stop domain, the comfort item (P6D, 4.84) has the highest score, and the protection from the weather item (P6E, 4.70), the lowest. Users rate the accessibility domain high; the best score goes to the mobility item (P7B, 5.22), the lowest scored is the safety item (P7E, 4.83). From the buses and drivers' domain, in general, drivers are better assessed than buses. The lowest scored item is the bus cleanliness item (P8B, 4.43), and the highest score goes to the respectful drivers' item (P8E, 5.01).

Last, from the peripheral domain, we note considerable differences amongst scores. The best goes to the Metro service item (P11, 4.97), the worst to the other users' behaviour item (P9, 4.30), and in the middle lies the information means item (P19, 4.70). We consider these three items, peripheral items as they do not involve a direct assessment of the service given by the bus system, nevertheless, are more image/tangible items. The overall evaluation of the bus-line (4.39), passes the minimum requirement (>4). However, there is room for improvement to reach an adequate (>5) evaluation regarding users' assessment.

Table 2-2: Survey results

Domain	Service quality attributes	Average satisfaction score (1-7 scale)
Overall	P1: *In general, with what grade do you evaluate the whole bus system in Transantiago*?	4.39
	P3: *In general, with what grade do you evaluate the service delivered by this bus-line?*	4.61
System	P2C: *It is an easy-to-use system*	4.80
	P2D: *They endeavour to inform and educate the users*	4.46
	P2E: *It is a reliable and transparent system*	4.41
	P2F: *It is a system that cares about the users*	4.31
	P2G: *It is a modern system*	4.73
	P2H: *It is a system with room for improvement*	4.91
	P2I: *I have bus-line alternatives to get to my destination*	4.76
	P2J: *I always find buses when I have to go out at night during weekdays*	4.20
	P2K: *I always find buses when I have to go out at night during weekends*	4.07
Frequency	P5A: *The buses run frequently, I don't have to wait too much*	4.35
	P5B: *I can always get on the bus, it is not full, and it stops at the bus-stop*	4.56
Bus-stop	P6D: *Is it comfortable to wait at the bus-stop?*	4.84
	P6E: *The bus-stop protects me from sun and rain*	4.70
	P6F: *I feel safe at the bus-stop*	4.74
Accessibility	P7A: *I can trust that the bus will come*	4.98
	P7B: *This bus-line lets me get where I need to go*	5.22
	P7C: *I don't have to make many transfers*	5.03
	P7D: *It's easy to make transfers*	5.02
	P7E: *I feel safe during this trip*	4.83
Buses and drivers	P8A: *The buses are safe*	4.72
	P8B: *The buses are spot on clean*	4.43
	P8C: *The buses are comfortable*	4.48
	P8D: *The buses are modern/new*	4.72
	P8E: *The drivers treat you with respect*	5.01
	P8F: *The drivers are responsible while driving*	4.85
	P8G: *The drivers always stop when asked, and it is permitted*	4.86
Peripheral	P9: *How would you evaluate users' behaviour in the Public Transport System?"*	4.30
	P11: *In general, how would you assess the service delivered by Metro?*	4.97
	P19: *In general, how would you evaluate the means used by Transantiago (PT) to inform users?*	4.70

2.5 Model Structure and Results

We present the model results in three-steps: (i) Measurement model, (ii) MIMIC structural model, and (iii) Mediation Analysis. All steps were computed using R statistical packages, specifically the Lavaan (Rosseel, 2012) package for the SEM-MIMIC models. The Likert-type items were treated as continuous/numeric, as we have seven response options (>5) and all items approximated a Normal distribution; following Hoyle (2012, p. 497).

2.5.1. Measurement model

Before estimating the Measurement model, PCA was performed on the set of perceived quality items to obtain adequate latent constructs for the bus-line and system's global satisfaction. Initially, the PCA recovered six factors from the dataset according to the Kaiser rule (Kaiser, 1960). However, after iterating several times, getting different alpha Cronbach's values (Cronbach, 1951) and considering satisfaction theory, we defined ten factors: six for the Bus-line Satisfaction and four for the System Satisfaction measurement models. We did not include individual questions that did not correlate well with other survey items, as no latent constructs could be formed.

The result is two separate measurement models, see Figure 2-3; ovals represent latent constructs and rectangles, individual survey items. The first, the Bus-line Items, are made up of six Bus-line Components, BL: Frequency (BL1), Bus-stop (BL2), Safety (BL3), Transfer (BL4), Buses (BL5), and Drivers (BL6). Items represent individual questions from the survey, see Table 2-2. The second, the System Items, are made up of four System Components, SC: Convenience (SC1), Transparency (SC2), Availability at night (SC3), and Peripheral-SPT (SC4). This last variable was chosen as the three peripheral variables loaded heavily into one factor in the PCA: satisfaction with users of the PT system, with Metro, and with information in the PT system. We named it peripheral-system of public transport (Peripheral-SPT), as it is not related directly to the bus system but appears to be an

Image component. The final model is represented by the ten latent constructs. The measurement model applied to the whole sample provides a statistical tool to discern if the entire sample perceives the different latent constructs homogeneously. The CFA provides statistics of the items loading onto individual latent variables: Z-scores, standard errors (S.E.), and standardised coefficients (St. Coef.). The standardised coefficients refer to how many standard deviations a dependent variable will change, per standard deviation increase in the predictor variable. A standard coefficient > 0.5 is considered acceptable, and > 0.6 good. We tested different measurement models provided that the individual items linked to the latent construct were consistent with satisfaction theory. The chosen model had the best fit.

The measurement system is presented in Figure 2-3, ovals represent latent constructs, and rectangles represent single items. With the measurement model, a confirmatory factor analysis (CFA) was performed to test for goodness-of-fit. The model yielded the following fit indices which are fully explained in Hooper et al. (2008), and in parenthesis are their recommended cut-off values (Hu and Bentler, 1999): comparative fit index, CFI: 0.973 (>0.95), Tucker-Lewis index, TLI: 0.967 (>0.95), goodness of fit index, GFI: 0.990 (>0.95), adjusted goodness of fit index, AGFI: 0.986 (>0.95), root mean squared error approximation, RMSEA: 0.040 (<0.08), and standardized root mean square residual, SRMR: 0.021 (<0.06).

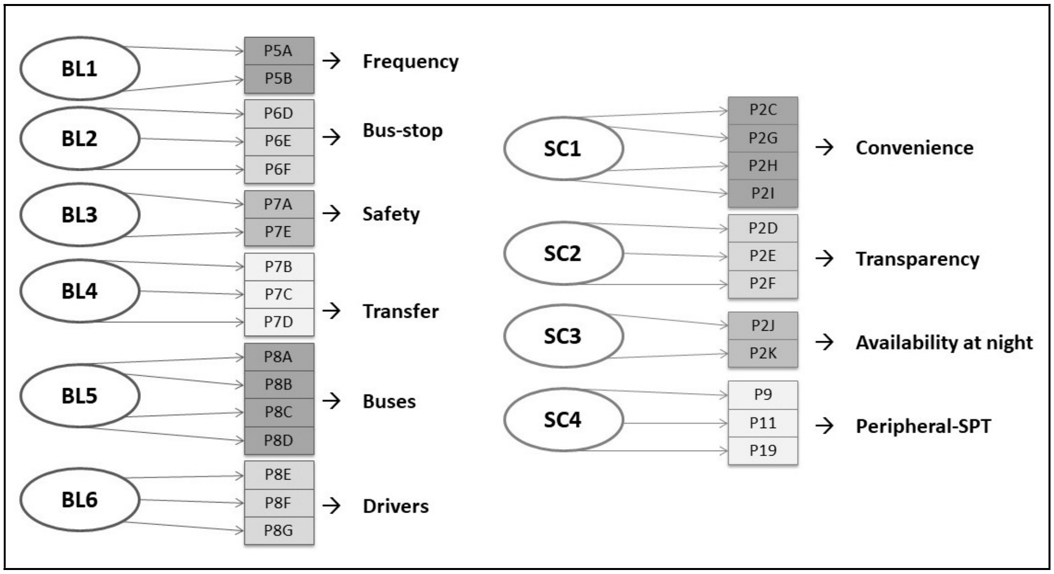

Figure 2-3: Bus-line Satisfaction and System Satisfaction measurement models

All values confirm an excellent fit to the measurement model proposed, validating it. The most important ones are the CFI and the AGFI, the closer they are to one the better, and they should be above 0.9 for an adequate fit. The RMSEA is frequently used as an indicator of fit; below 0.08, it is considered acceptable. Table 2-3 shows the results of the CFA, representing the final measurement model.

From the CFA model, we can confirm that the Z-scores are high, and the standardised coefficients are all between 0.75-0.95 (high reliability, see Table 2-3), except for one latent construct, Peripheral-SPT, as it is partially well measured according to the measurement model. The highest standard coefficient, for this construct, belongs to the users' behaviour (P9) item with 0.54, which is adequate but the other two items are a little below 0.4. For modelling purposes, we decided to retain this variable as it is. However, we will discuss this variable also in the concluding section and give recommendations for future studies.

Table 2-3: CFA Measurement model for the satisfaction latent constructs

	Estimate	S.E.	Z-value	St. Coef.
BL1				
P5A	1.221	0.010	124.86	0.803
P5B	1.233	0.010	123.44	0.852
BL2				
P6D	1.000			0.841
P6E	1.069	0.009	121.33	0.850
P6F	1.047	0.009	115.55	0.855
BL3				
P7A	1.014	0.009	110.31	0.796
P7E	1.062	0.009	116.29	0.833
BL4				
P7B	1.000			0.813
P7C	1.068	0.009	122.04	0.849
P7D	1.089	0.009	119.81	0.894
BL5				
P8A	1.000			0.838
P8B	1.067	0.008	131.82	0.826
P8C	1.103	0.008	139.48	0.879
P8D	1.009	0.008	122.99	0.849
BL6				
P8E	1.000			0.848
P8F	1.066	0.008	127.43	0.887
P8G	1.036	0.009	110.66	0.812

	Estimate	S.E.	Z-value	St. Coef.
SC1				
P2C	1.000			0.768
P2G	1.032	0.010	101.69	0.797
P2H	1.032	0.011	97.41	0.766
P2I	0.983	0.010	99.14	0.754
SC2				
P2D	1.000			0.821
P2E	1.063	0.008	130.57	0.872
P2F	1.061	0.008	124.93	0.852
SC3				
P2J	1.300	0.008	155.65	0.926
P2K	1.302	0.008	157.81	0.914
SC4				
P11	1.000			**0.362**
P9	1.431	0.046	31.24	**0.537**
P19	0.991	0.038	25.89	**0.368**

2.5.2. MIMIC structural model

Our methodological review revealed that many applications had not accounted for different travel conditions and sociodemographic heterogeneity despite the fact that service perception evidently varies from customer to customer, day to day, season to season, and by different producers. One of the primary objectives and contribution of this research was to try and capture these differences in service perception. In this sense, our model hypothesises differences among all possible travel conditions measured in the surveys, and we will test if these differences hold.

We built (n-1) dummy variables for each attribute category and introduced them as regressors in the MIMIC model, for every latent construct; the variables will be explained in the following paragraphs. We also considered two separate linear-in-parameters regression equations, one for the Bus-line Satisfaction model, and another for the System Satisfaction model; the latter has the Bus-line Satisfaction (P3) variable as an additional regressor. We are implicitly assuming that customers have a Bus-line Satisfaction that can approximate the service encounter satisfaction. A novel approach and a contribution in itself, according to our literature search only a few studies in the transport field report differences in the type of satisfaction due to differences in service encounter and global satisfaction, and none modelled it separately. We also assumed that travel conditions and the whole array of user characteristics (regressors) might influence both the specific Bus-line Satisfaction and System Satisfaction; therefore, we used all variables in the separate regression equations, allowing us to compute direct, indirect, and total effects, via the Mediation Analysis technique. Again, this is a novel approach in itself, according to our literature review.

Initially, we introduced the following sociodemographic variables into the MIMIC model (base case in parentheses): two regressors for SES (DE), four for AGE (13-17 years), one for GENDER (Male) and one for Trip Purpose (Other). We included trip characteristics' variables: two for TIME (Off-peak), two for DAY (Weekday), one for Frequency-Bus-Use (5-7 days/week), two for Frequency-Metro-Use (5-7 days/week), one for Transfer (No transfer), three for PWT (0-6 min), and three for PTT (0-20 min). Additionally, we included six dummy variables for the Operator (U1) category and four for Date (November 2013), for a total of 32 dummy variables.

Figure 2-4 represents the complete MIMIC Structural Model; ovals represent latent variables, the joined rectangles represent the categorical regressors, and the large rectangles represent the bus-line and system satisfaction items. Initially, we tested a specification including the 32 variables, but later left (for each latent construct and satisfaction variable) only those regressors that were significant at the 95% level and removed the rest. We kept those regressors that effectively affect the latent constructs to correct for heterogeneity without maintaining a larger model needlessly.

The MIMIC model allows correcting for possible bias due to heterogeneity in the users' perception given different trip conditions and sociodemographic characteristics. We introduced more parameters into the model to capture heterogeneity (Hooper et al., 2008), as this was one of the main objectives of our study. However, there is evidence that increasing the number of variables in the model, as we did, tends to worsen the CFI and the TLI (Kenny and McCoach, 2003). In fact, the final fit indices were: CFI: 0.848 (>0.95) and TLI: 0.812 (>0.95). Notwithstanding, all the rest with exception of SRMR, remained well within their acceptable values: GFI: 0.981 (>0.95), AGFI: 0.967 (>0.95), RMSEA: 0.051 (<0.08), SRMR: 0.118 (<0.06). Despite the abovementioned shortcomings, we consider that the model produced an acceptable fit to the data (RMSEA<0.08) and generated valuable policy-related information about heterogeneity in the perception of the satisfaction constructs.

We present the MIMIC results for the ten individual latent constructs regressions, six for the Bus-line Satisfaction constructs regressions (BL1-BL6), and four for the System Satisfaction construct regressions (SC1-SC4), in Tables 2-4 to 2-8.

All categories of regressors produced significant results for all latent constructs, which highlights the importance of correcting for heterogeneity; we note that the different R^2 vary between 0.13-0.29 (see Tables 2-4 to 2-8), implying a low to moderate fit for each latent construct regression. It is not an unexpected result, as the regressions only capture heterogeneity and do not necessarily include all the possible variables that cause each distinct satisfaction construct. Notice that the results are presented for two regressions in each table (for example, in Table 2-4 the BL1 and the SC2 regressions are presented), where BL represents a Bus-line component and SC a System component. We consider that it is key to note which categorising variables affect each individual construct, and also it is important to assess the signs. For easiness, we will compare the different tables accordingly, and subsequently we will discuss all results together.

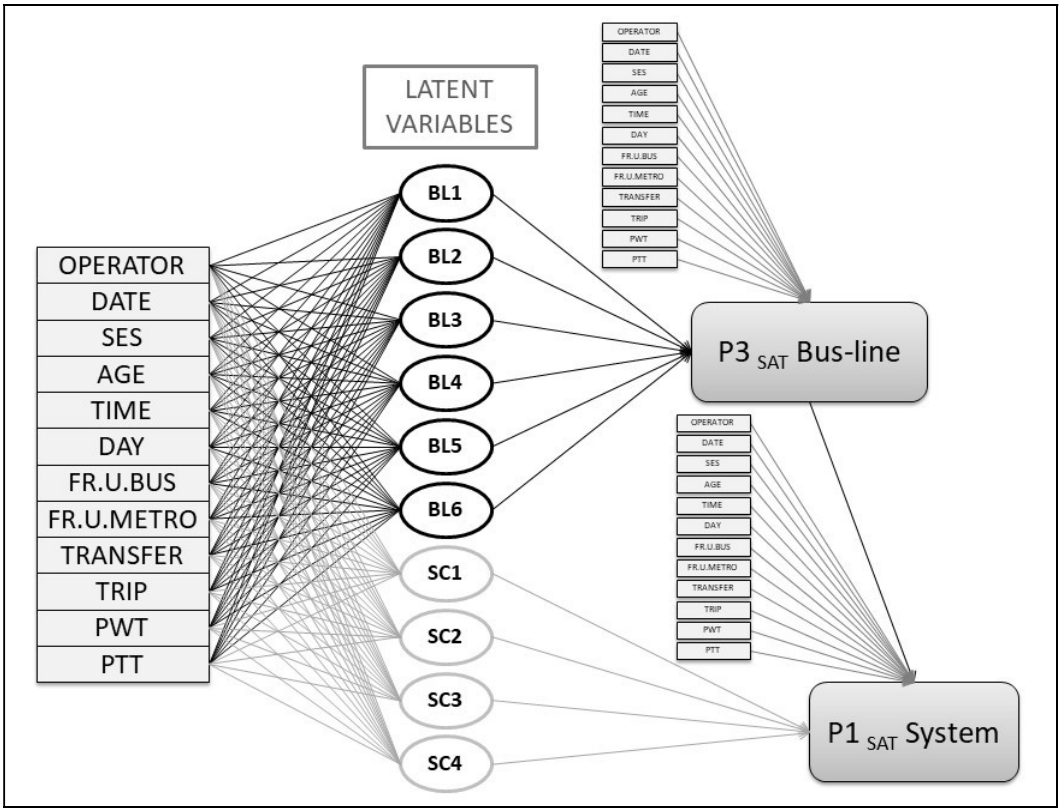

Figure 2-4: MIMIC Structural Model

From Table 2-4, we can confirm that both Frequency (BL1) and Transparency (SC2) present heterogeneity according to the following variables: Operator, Seasonal, SES, Time, PWT, and PTT. However, some differences appear, the Age variable is more significant for the System (SC2) variable, in all age groups. Also, the gender variable is only significant for the Frequency (BL1) construct, as is the Trip Purpose category. Additionally, Day, Frequency of use, and Transfer variables were significant for the BL1 construct only.

Table 2-4: MIMIC latent constructs regression results (BL1, SC2)

Satisfaction with Frequency **(R^2 = 0.283)**

BL1	Group	Estimate	S.E.	Z-value
OPER.	U2	-0.116	0.024	-4.88
	U3	0.142	0.022	6.51
	U5	0.268	0.023	11.44
	U6	0.196	0.022	8.88
	U7	-0.116	0.024	-4.83
SEASON	JAN13	-0.178	0.019	-9.31
	JAN14	-0.061	0.020	-3.10
	MAY14	-0.440	0.023	-19.29
SES	AB.C1.C2	-0.188	0.037	-5.07
	C3	-0.493	0.023	-21.17
AGE	AGE61+	-0.090	0.035	-2.50
GENDER	FEMALE	-0.053	0.013	-3.99
TIME	PM	-0.068	0.014	-4.94
TRIP	WORK	-0.062	0.016	-3.90
PWT	PWT6.10	-0.293	0.023	-12.93
	PWT10.20	-1.031	0.025	-41.90
	PWT20+	-1.865	0.031	-60.01
PTT	PTT20.40	-0.065	0.019	-3.48
	PTT40.60	-0.088	0.021	-4.29
	PTT60+	-0.178	0.024	-7.53

Satisfaction with Transparency **(R^2 = 0.145)**

SC2	Group	Estimate	S.E.	Z-value
OPER.	U2	-0.093	0.021	-4.35
	U3	0.084	0.017	4.94
	U4	0.036	0.017	2.08
	U6	0.059	0.059	3.45
	U7	-0.078	0.078	-3.60
SEASON	JAN13	-0.137	0.022	-6.17
	MAY13	-0.077	0.022	-3.55
	JAN14	-0.148	0.022	-6.83
	MAY14	-0.338	0.024	-14.02
SES	AB.C1.C2	-0.158	0.036	-4.44
	C3	-0.421	0.022	-19.35
AGE	AGE18.29	-0.079	0.020	-3.86
	AGE30.45	-0.065	0.022	-3.03
	AGE46.60	-0.125	0.025	-4.98
	AGE61+	-0.171	0.038	-4.54
TIME	AM	-0.075	0.014	-5.52
	PM	-0.068	0.013	-5.08
DAY	SATURDAY	-0.034	0.014	-2.45
FR.U.BUS	FR1234	-0.063	0.018	-3.58
TRANSF.	TRANSFER	-0.048	0.015	-3.12
PWT	PWT6.10	-0.161	0.022	-7.45
	PWT10.20	-0.649	0.023	-28.53
	PWT20+	-1.164	0.028	-42.27
PTT	PTT20.40	-0.040	0.018	-2.24
	PTT40.60	-0.068	0.020	-3.38
	PTT60+	-0.119	0.024	-4.95

Tables 2-5 and 2-6, allow us to confirm that these four Bus-line constructs (BL2-BL5) present heterogeneity according to the following variables: Operator, Seasonal, SES, PWT, and PTT. However, some differences are present. Age is significant only for the 18-29 age group, for Bus-stop (BL2) and Transfer (BL4); also, Time is significant for these two constructs (BL2, BL4), and the two Peak hour periods are perceived negatively. Frequency of Metro use is perceived positively (less Metro use) for the Safety (BL3) and Transfer (BL4) constructs.

Having to Transfer (Transfer) affects only the Safety (BL3) and Buses (BL5) constructs, and positively.

Table 2-5: MIMIC latent constructs regression results (continued) (BL2, BL3)

Satisfaction with Bus-stop ($R^2 = 0.193$)

BL2	Group	Estimate	S.E.	Z-value
OPER.	U2	-0.390	0.025	-15.63
	U3	-0.302	0.025	-12.31
	U4	-0.157	0.025	-6.34
	U6	-0.104	0.025	-4.19
	U7	-0.261	0.025	-10.43
SEASON	JAN13	-0.220	0.025	-8.80
	MAY13	-0.089	0.024	-3.71
	JAN14	-0.098	0.024	-4.07
	MAY14	-0.261	0.027	-9.65
SES	AB.C1.C2	-0.211	0.039	-5.42
	C3	-0.537	0.024	-22.48
AGE	AGE18.29	0.034	0.016	2.15
TIME	AM	-0.042	0.020	-2.12
	PM	-0.084	0.019	-4.31
PWT	PWT6.10	-0.247	0.025	-10.01
	PWT10.20	-0.712	0.026	-27.77
	PWT20+	-1.199	0.031	-38.76
PTT	PTT40.60	0.049	0.018	2.71

Satisfaction with Safety ($R^2 = 0.233$)

BL3	Group	Estimate	S.E.	Z-value
OPER.	U2	-0.388	0.023	-16.75
	U5	0.075	0.017	4.49
	U6	0.079	0.023	3.46
	U7	-0.173	0.023	-7.43
SEASON	JAN13	-0.363	0.025	-14.52
	MAY13	-0.170	0.022	-7.70
	JAN14	-0.101	0.017	-6.02
	MAY14	-0.583	0.025	-23.01
SES	AB.C1.C2	-0.355	0.041	-8.74
	C3	-0.726	0.025	-29.43
FR.U.ME.	FR.M01	0.041	0.017	2.47
TRANSF.	TRANSFER	0.096	0.012	7.72
PWT	PWT6.10	-0.249	0.024	-10.41
	PWT10.20	-0.824	0.026	-32.29
	PWT20+	-1.443	0.032	-45.60
PTT	PTT20.40	-0.120	0.020	-6.06
	PTT40.60	-0.076	0.022	-3.46
	PTT60+	-0.115	0.026	-4.45

Finally, the Trip purpose variable only affects the Transfer (BL4) construct, and positively for the Study category. We note that most differences are present depending on the characteristics of the trip and the type of construct, hence the importance of this type of model, to assess these differences.

Next, we compare the Drivers (BL6) and Convenience (SC1) regressions. From Table 2-7, we confirm that both Drivers (BL6) and Convenience (SC1) present heterogeneity according to the following variables: Operator, Seasonal, SES, PWT, and PTT. However again, some differences are present; the Age variable is positive for the Drivers construct, and negative for the Convenience construct.

Rather unexpectedly, having to make a Transfer is positive for both constructs. For the Convenience (SC1) construct, using the bus less is negative and using the Metro less is positive, this indicates inconveniences in having to transfer to Metro.

Table 2-6: MIMIC latent constructs regression results (continued) (BL4, BL5)

Satisfaction with Transfer (R^2 = 0.165)

BL4	Group	Estimate	S.E.	Z-value
OPER.	U2	-0.313	0.018	-17.44
	U6	0.101	0.018	5.70
	U7	-0.189	0.018	-10.47
SEASON	JAN13	-0.285	0.017	-16.79
	MAY13	-0.122	0.016	-7.44
	MAY14	-0.443	0.019	-22.96
SES	AB.C1.C2	-0.235	0.031	-7.59
	C3	-0.498	0.019	-26.24
AGE	AGE18.29	0.053	0.009	5.67
TIME	AM	-0.030	0.011	-2.67
	PM	-0.031	0.011	-2.82
FR.U.ME.	FR.M01	0.057	0.014	4.20
	FR.M234	0.037	0.011	3.25
TRIP	STUDY	0.052	0.012	4.29
PWT	PWT6.10	-0.196	0.020	-9.94
	PWT10.20	-0.591	0.020	-28.86
	PWT20+	-0.952	0.025	-38.64
PTT	PTT20.40	-0.103	0.015	-6.67
	PTT40.60	-0.056	0.017	-3.31
	PTT60+	-0.148	0.019	-7.62

Satisfaction with Buses (R^2 = 0.180)

BL5	Group	Estimate	S.E.	Z-value
OPER.	U2	-0.241	0.024	-10.20
	U3	0.057	0.020	2.81
	U4	0.048	0.020	2.35
	U5	0.154	0.024	6.52
	U6	0.376	0.024	15.91
	U7	-0.272	0.024	-11.45
SEASON	JAN13	-0.259	0.022	-11.94
	MAY13	-0.107	0.021	-5.05
	JAN14	-0.047	0.021	-2.25
	MAY14	-0.274	0.024	-11.60
SES	AB.C1.C2	-0.223	0.034	-6.65
	C3	-0.477	0.021	-23.13
TRANSF.	TRANSF.	0.051	0.015	3.43
PWT	PWT6.10	-0.216	0.022	-10.02
	PWT10.20	-0.699	0.023	-30.96
	PWT20+	-1.141	0.027	-41.81
PTT	PTT20.40	-0.108	0.017	-6.16
	PTT40.60	-0.108	0.020	-5.43
	PTT60+	-0.140	0.023	-5.99

Last, for the Drivers (BL6) construct the Work trip is perceived positive, and for the Convenience construct, the Study trip is perceived positive; in comparison to Other.

Next, we compare the last two constructs Availability at night (SC3) and Peripheral-SPT (SC4) regressions. From Table 2-8, we confirm that both Availability at night (SC3) and Peripheral-SPT (SC4) constructs present heterogeneity according to the following variables: Operator, Seasonal, SES, Trip, PWT, and PTT. Both Trip categories Work and Study are perceived negatively for both constructs.

Table 2-7: MIMIC latent constructs regression results (continued) (BL6, SC1)

Satisfaction with Drivers				(R^2=0.154)
BL6	Group	Estimate	S.E.	Z-value
OPER.	U2	-0.405	0.022	-18.01
	U5	0.070	0.022	3.12
	U6	0.181	0.022	8.13
	U7	-0.309	0.023	-13.69
SEASON	JAN13	-0.255	0.024	-10.81
	MAY13	-0.157	0.023	-6.77
	JAN14	0.076	0.023	3.30
	MAY14	-0.239	0.026	-9.27
SES	AB.C1.C2	-0.230	0.035	-6.54
	C3	-0.498	0.022	-22.31
AGE	AGE18.29	0.106	0.025	4.29
	AGE30.45	0.144	0.026	5.51
	AGE46.60	0.190	0.027	6.98
	AGE61+	0.281	0.036	7.84
TRANSF.	TRANSF.	0.120	0.016	7.38
TRIP	WORK	0.034	0.014	2.44
PWT	PWT6.10	-0.207	0.024	-8.79
	PWT10.20	-0.634	0.025	-25.82
	PWT20+	-1.088	0.030	-36.77
PTT	PTT20.40	-0.089	0.019	-4.68
	PTT40.60	-0.089	0.022	-4.11
	PTT60+	-0.105	0.026	-4.12

Satisfaction with Convenience				(R^2=0.150)
SC1	Group	Estimate	S.E.	Z-value
OPER.	U2	-0.184	0.019	-9.81
	U5	0.036	0.015	2.36
	U7	-0.112	0.019	-5.93
SEASON	JAN13	-0.070	0.021	-3.39
	MAY13	-0.060	0.020	-3.01
	JAN14	-0.071	0.020	-3.58
	MAY14	-0.367	0.022	-16.40
SES	AB.C1.C2	-0.252	0.033	-7.74
	C3	-0.477	0.020	-24.03
AGE	AGE46.60	-0.091	0.016	-5.59
	AGE61+	-0.138	0.031	-4.49
FR.U.BUS	FR1234	-0.038	0.016	-2.32
FR.U.ME.	FR.M01	0.059	0.010	5.77
TRANSF.	TRANSF.	0.036	0.014	2.56
TRIP	STUDY	0.077	0.014	5.54
PWT	PWT6.10	-0.194	0.020	-9.52
	PWT10.20	-0.615	0.021	-28.86
	PWT20+	-1.037	0.026	-40.18
PTT	PTT20.40	-0.085	0.016	-5.15
	PTT40.60	-0.062	0.019	-3.32
	PTT60+	-0.111	0.022	-5.02

Some differences are present once more: Age is negative for all categories in the Peripheral-SPT construct, and negative only for the Age 61+ category for the Availability at night construct. For Availability at night, using less bus and Metro is perceived negatively, while for Peripheral-SPT only using less Metro is perceived negatively. Availability at night is perceived negatively on the Saturday category, and for the Peripheral-SPT being a Female brings a negative perception.

Table 2-8: MIMIC latent constructs regression results (continued) (SC3, SC4)

Satisfaction with Availability at night (R^2 = 0.137)

SC3	Group	Estimate	S.E.	Z-value
OPER.	U2	-0.067	0.021	-3.19
	U4	-0.060	0.020	-3.04
	U5	0.053	0.021	2.51
	U7	-0.151	0.021	-7.14
SEASON	JAN14	-0.063	0.018	-3.52
	MAY14	-0.089	0.021	-4.31
SES	AB.C1.C2	-0.111	0.034	-3.24
	C3	-0.404	0.022	-18.53
AGE	AGE61+	-0.096	0.035	-2.73
DAY	SAT.	-0.037	0.016	-2.25
FR.U.BUS	FR1234	-0.115	0.017	-6.81
FR.U.ME.	FR.M01	-0.039	0.014	-2.78
TRANSF.	TRANSF.	-0.056	0.014	-3.97
TRIP	STUDY	-0.059	0.022	-2.65
	WORK	-0.089	0.019	-4.65
PWT	PWT6.10	-0.194	0.022	-8.89
	PWT10.20	-0.590	0.023	-25.62
	PWT20+	-1.095	0.028	-38.78
PTT	PTT20.40	-0.058	0.018	-3.25
	PTT40.60	-0.101	0.020	-4.99
	PTT60+	-0.157	0.024	-6.57

Satisfaction with Peripheral-SPT (R^2 = 0.199)

SC4	Group	Estimate	S.E.	Z-value
OPER.	U2	-0.123	0.017	-7.12
	U3	0.068	0.016	4.25
	U6	0.110	0.017	6.48
SEASON	JAN14	0.123	0.015	8.41
	MAY14	-0.184	0.017	-10.69
SES	C3	-0.212	0.014	-15.39
AGE	AGE18.29	-0.110	0.023	-4.85
	AGE30.45	-0.191	0.025	-7.61
	AGE46.60	-0.168	0.026	-6.47
	AGE61+	-0.164	0.034	-4.89
GENDER	FEMALE	-0.063	0.011	-5.69
TIME	PM	-0.027	0.011	-2.44
FR.U.ME.	FR.M01	-0.210	0.015	-14.32
	FR.M234	-0.074	0.014	-5.13
TRIP	STUDY	-0.047	0.021	-2.29
	WORK	-0.072	0.015	-4.75
PWT	PWT6.10	-0.138	0.018	-7.80
	PWT10.20	-0.386	0.019	-19.79
	PWT20+	-0.530	0.023	-22.60
PTT	PTT20.40	-0.088	0.015	-6.07
	PTT40.60	-0.073	0.015	-4.79
	PTT60+	-0.151	0.018	-8.41

For all the Bus-line specific latent constructs we obtained similar results; the significant variables included: Operator, Seasonal, SES, Time, PWT, and PTT. For the System components, other variables were also significant: Trip and Age. Overall, each latent construct produced consistent and systematic results; however, there are some specific differences on the variables Gender, Day, Bus-Use-Frequency, Metro-Use-Frequency, and Transfer. All the latent constructs for both Bus-line and System Satisfaction presented the same pattern relating to the critical policy variables, PWT and PTT: a decrease in satisfaction within each category as PWT and PTT increased, as expected, but had a more distinguishable effect in the PWT variable. Next, we present the structural model results in Table 2-9.

Table 2-9: MIMIC structural model results

Linear Regression 95% Sig. Factors **(R^2=0.531)**

P3 BL Satisfaction	Est.	S.E.	Z-value	St. Coef.
Frequency (BL1)	0.667	0.008	83.48	**0.613**
Safety (BL3)	0.251	0.008	30.11	**0.223**
Buses (BL5)	0.071	0.010	7.38	0.060
Drivers (BL6)	0.071	0.009	8.11	0.064

Linear Regression 95% Sig. Factors **(R^2 = 0.358)**

P1 Syst. Satisfaction	Est.	S.E.	Z-value	St. Coef.
BL Satisfaction (P3)	0.255	0.006	43.54	**0.282**
Convenience (SC1)	0.171	0.016	10.54	0.143
Transparency (SC2)	0.246	0.014	17.28	**0.227**
Avail. at night (SC3)	0.084	0.008	9.97	0.078
Periph.-SPT (SC4)	0.566	0.027	20.93	**0.277**

For the structural model, we only left out two latent constructs from the Bus-line Satisfaction model, which were not significant at the 70% level, and one of them had negative sign. The latent constructs Frequency (BL1), Safety (BL3), Buses (BL5), and Drivers (BL6) explain Bus-line Satisfaction (P3) best. In turn, the latent constructs Convenience (SC1), Transparency (SC2), Availability at night (SC3), Peripheral-SPT (SC4), and Bus-line Satisfaction (P3) explain the System Satisfaction (P1) best. We note that the R^2 for the Satisfaction regressions indicate adequate and moderate fits respectively: 0.53 for Bus-line Satisfaction and 0.36 for System Satisfaction. As expected, we obtained better fit for the Bus-line Satisfaction model than for the System Satisfaction, as the survey was performed in the bus component of the PT System. Nonetheless, it is evident that some other variables may be missing to explain the System Satisfaction.

It is important to note the standardised coefficients for both regressions. Values near 0.10 are considered weak, values near 0.30 are considered moderate, and values at or above 0.50 are considered large (Currie and Delbosc, 2017), for the structural coefficients. For the Bus-line Satisfaction, Frequency has a large effect (0.61), and Safety a moderate one (0.22). For the System Satisfaction, Bus-line Satisfaction (as a regressor) has a moderate effect (0.28) alongside Transparency (0.23), and Peripheral-SPT (0.28).

The SEM-MIMIC measurement model and the Full Model statistics are shown in Tables 2-10 and 2-11. We note from Table 2-10, that some of the coefficients changed somewhat, generating some changes into which the most relevant items are, in the case of each latent construct. For example, now the Frequency (P5A) item is the most important item in the Frequency (BL1) construct. Also, in the Peripheral-SPT construct, Metro valuation (P11) gained importance in explaining this construct, and Users' behaviour (P9) remained close to a 0.5 standardised coefficient, which is adequate. We will further discuss these results in Section 2.6.

Table 2-10: Measurement model for Full SEM-MIMIC Model

	Estimate	S.E.	Z-value	St. Coef.
BL1				
P5A	1.121	0.008	142.14	0.873
P5B	0.934	0.007	128.59	0.770
BL2				
P6D	1.000			0.851
P6E	1.093	0.007	154.07	0.875
P6F	1.001	0.007	137.46	0.824
BL3				
P7A	0.897	0.007	130.41	0.800
P7E	0.922	0.007	125.42	0.816
BL4				
P7B	1.000			0.793
P7C	1.103	0.007	153.30	0.847
P7D	1.120	0.007	158.43	0.886
BL5				
P8A	1.000			0.823
P8B	1.082	0.007	152.61	0.827
P8C	1.123	0.007	164.51	0.879
P8D	1.003	0.007	153.08	0.829
BL6				
P8E	1.000			0.842
P8F	1.069	0.006	165.02	0.887
P8G	1.027	0.007	145.89	0.800

	Estimate	S.E.	Z-value	St. Coef.
SC1				
P2C	1.000			0.764
P2G	1.042	0.008	126.43	0.807
P2H	1.048	0.009	121.13	0.775
P2I	0.965	0.008	115.97	0.736
SC2				
P2D	1.000			0.820
P2E	1.058	0.007	156.84	0.867
P2F	1.056	0.007	151.63	0.845
SC3				
P2J	1.189	0.007	159.31	0.917
P2K	1.211	0.008	158.65	0.923
SC4				
P11	1.000			**0.441**
P9	1.093	0.036	30.39	**0.498**
P19	0.722	0.028	25.90	**0.327**

Table 2-11: Statistics for Full SEM-MIMIC Model

Number of Parameters	Chi-Square	Degrees of Freedom	P-value
342	79374.9	1177	0.000

2.5.3. Mediation analysis (total effects)

We can estimate total effects by considering all the paths, direct and indirect, from the regressors through the mediating variables. Accounting for all the effects regarding a particular regressor, from the latent construct regressions, the Bus-line Satisfaction regression, and the System Satisfaction regressions, we can compute Total Effects. From our model results, we decided to test the significance of the total effects of PWT and PTT on the System Satisfaction (see Table 2-12), as these were the two policy variables that had the most significant effects. However, one can choose any particular regressor to examine the full effect on System Satisfaction.

Table 2-12: Mediation analysis for PWT and PTT total effects on system satisfaction

P1 System Satisfaction	**Total Effects System (PWT)**		
Reference Level (0-6 PWT)	Estimate	S.E.	Z-value
(6-10 PWT)	-0.149	0.022	-6.64
(10-20 PWT)	-0.608	0.023	-26.11
(20+ PWT)	**-1.085**	0.028	-38.80

P1 System Satisfaction	**Total Effects System (PTT)**		
Reference Level (0-20 PTT)	Estimate	S.E.	Z-value
(20-40 PTT)	-0.039	0.016	-2.39
(40-60 PTT)	**-0.101**	0.013	-7.93
(60+ PTT)	-0.189	0.015	-12.50

Results confirm a significant total effect for users when they moved up to every category of perceived waiting time (PWT) and perceived travel time (PTT). The effect is more significant for PWT than for PTT, which is consistent with our literature review. In relative terms, moving from 0-6 min to 20+ min in perceived waiting time causes ten times more dissatisfaction than moving from 0-20 min to 40-60 min in perceived travel time.

2.6 Discussion of Results

This research benefitted from the SEM capabilities to test a complete measurement model for the perceived quality latent constructs of a transit system. After estimating the measurement model, we tested a MIMIC structural model including two separate regressions for the Bus-line Satisfaction and the System Satisfaction. In itself, the SEM-MIMIC model allowed us to gain insights into which user sociodemographic and travel characteristics, travel patterns, travel attributes, trip purpose, season, and bus operator company affected the perceptions of the different satisfaction constructs. All categories of variables produced significant results for most latent constructs, summarising:

a) **SES**: Belonging to the ABC1C2 (high/middle income) group has a negative effect on perception and belonging to the C3 (middle/low income) group has an even higher significant adverse effect. We interpret this by hypothesising that ABC1C2 bus users (which are non-captive) most probably choose bus because they find its level of service acceptable; C3 users, on the other hand, have a high proportion of captive users, making them more prone to being dissatisfied. As the DE (low income) group is content with just being able to use buses, in an analogous manner they are less dissatisfied than the other two groups.

b) **AGE**: the age variables have a more consistent adverse effect on the System variables than in the Bus-line variables. This result agrees with the fact that older people tend to get uncomfortable more easily. Also, the effect being more significant with the System Satisfaction may reflect more on the Image and Satisfaction produced by the transit system as a whole; as we age, we tend to become more selective and have a higher expectation regarding the satisfaction received from a service. An exception appears for the Satisfaction with Drivers (BL6) construct (see Table 2-7). Apparently, as we age, we tend to be more satisfied in this category. This result may be related to the fact that drivers tend to be gentler and caring with older people in Chile.

c) **GENDER**: the Female group having a negative effect compared to men can be explained, theoretically, from three sides: (i) women tend to travel with children and are more prone to make shopping trips, and that can be more uncomfortable when boarding/alighting. Also when riding a bus, (ii) the Female group is more abundant in the population, indicating that there may also be a higher proportion of captive users among women, males probably have more access to cars. (iii) Women tend to perceive themselves less safe outside their homes than men, and this indirectly affects their satisfaction with PT (Delbosc and Currie, 2012). Note also that the gender affects only Frequency (BL1) and Peripheral-SPT (SC4); both variables relate with waiting at the bus-stop and with other users (P9), which are both perceived as unsafe situations.

d) **TIME**: AM and PM peaks produce less satisfaction than the Off-peak. The effect is most significant for the PM peak. This result is consistent with the PM peak extending further than the AM peak in Santiago.

e) **DAY**: The only significant result is on Saturdays, with an adverse effect on the Transparency (SC2) and Availability at night (SC3) constructs. This finding indicates that people using the service on Saturdays might be less satisfied with the system, this is consistent with the fact that during weekends the bus service frequency goes down for all bus services, especially at night.

f) **BUS-USE-FREQUENCY**: The only significant result is for Infrequent Users (FR1234) in the Convenience (SC1), Transparency (SC2), and Availability at night (SC3) System constructs. An expected effect, indicating that infrequent users are less satisfied with the system, in this sense, as they are non-captive users, they have alternatives. Also as they are not frequent users, it may be possible that they do not move through the system as efficiently as would a frequent (avid) user.

g) **METRO-USE-FREQUENCY**: The only significant result is for the Safety (BL3), Transfer (BL4) and Convenience (SC1) constructs. An expected effect, indicating that bus users are less satisfied when they have to make infrequent Metro transfers. In this sense, as they are not frequent users, it may be possible that they do not move through the system as efficiently as would a frequent user, creating this negative perception when transferring and moving through the system.

h) **TRIP PURPOSE**: Both Work and Study have negative effects in comparison with Other. This outcome is related to having to arrive at the destination at a particular time (compulsory trips), and satisfaction should decrease if the user arrives late due to the unreliability of the bus system. Hence, the unreliability effect is magnified due to the time constraint. The effect is negatively significant for the Work category in the Frequency construct. Some other constructs have positive effects for the Study category.

i) **PWT, PTT**: all negative effects. Perceived Waiting Time is more significant by one order of magnitude than Perceived Travel Time. In fact, PWT has the greatest of all effects and every single latent construct is affected by it. Also, we notice that the effect in PWT is nonlinear, as moving to the 10-20 min PWT impacts strongly when comparing the coefficients (See Table 2-12). This result reinforces the fact that waiting time is highly valued and that it is a key policy variable for the administration. PTT is also a strategic policy variable and behaves similar to PWT, but the effect is not so pronounced.

j) **OPERATOR**: some negative, some positive. The introduction of these variables allows the Administration to examine the performance of the different bus-line operators, on different perception constructs, to understand the reason behind these effects.

k) **DATE**: all negative, compared to November 2013. These variables allow the Administration to check for seasonal effects, or for systematic periodic changes in the perception of the satisfaction constructs.

For the structural part of the model, the two regression equations allowed us to detect the relative importance parameters for both Bus-line Satisfaction and System Satisfaction. We can test each equation separately. The standardised coefficients are shown in parentheses. For Bus-line Satisfaction, the two latent constructs with the most significant effect are Frequency Satisfaction (0.61) and Safety Satisfaction (0.22). Both variables are related to the reliability of the service. The first indicates that users value their waiting time highly and may be dissatisfied with this service variable. The Safety variable also shows that users appreciate whether the bus is going to stop, and the level of safety in both the process of waiting and inside the bus. In general, this result shows the importance of having an excellent reliability of service and safety (security) during the waiting process. Both variables relate with the Reliability dimension as in the SERVQUAL experiment (Parasuraman et al., 1988). This empirical result is consistent with the reviews and case studies analysed in the literature review.

For the System Satisfaction Model, the two latent constructs with a more substantial effect are Peripheral-SPT Satisfaction (0.28) and Transparency Satisfaction (0.23). The Peripheral-SPT variable effect demonstrates that users value highly several peripheral items as part of their overall System Satisfaction: the quality of the Metro service, the availability of information, and the behaviour of other users. The Transparency variable indicates that users feel that it is essential that the system provides a friendly service that cares for the customer. This could relate to the Assurance dimension found in the SERVQUAL experience. It is interesting to note that both Reliability and Assurance are highly significant in our full model, a result that is also consistent with the original SERVQUAL's findings. Additionally, Bus-line Satisfaction (0.28), as a variable, has a moderate significance in this model, as expected, confirming the validity of our complete MIMIC structural model. In particular, the Bus-line Satisfaction is relevant for the System Satisfaction; an intuitive result confirmed empirically.

2.7 Conclusions and Policy Recommendations

This study allowed us to reach several conclusions. The MIMIC approach allows distinguishing how the different travel conditions and sociodemographic variables affect how users perceive several Satisfaction Latent Constructs regarding the Transantiago bus component. The SEM-MIMIC paradigm permits the estimation of a full model including two separate Satisfaction regression equations for a PT system: Service Encounter (bus-line) satisfaction, and Global satisfaction with the system. Also, having used a large sample (n = 25,094) allows us to obtain a rich model that incorporates the trip conditions and sociodemographic characteristics of travellers. The main findings regarding these features are consistent with previous empirical results.

Our principal finding is that reliable/frequent services represent the latent constructs that weigh more heavily on Bus-line Satisfaction (service encounter). On the other hand, the Tangibles/Image (Peripheral-SPT) and Assurance (Transparency) latent constructs are essential for the global System Satisfaction. Both results are in-line with the original SERVQUAL results. From our work, we can conclude that peripheral aspects of the system such as satisfaction with Metro, with the users of the PT system, and with the information provided, are perceived as one latent construct, and are highly valued for the System Satisfaction. This attribute could be an Image component. Nonetheless, we must recommend that additional items are designed for future studies regarding these three different subdomains.

The perceived waiting time (PWT) is a key policy variable, even more so than the perceived travel time (PTT). We can conclude that all the satisfaction latent constructs are negatively impacted as PWT rises. It is also evident that the effect is nonlinear, as moving to the 10-20 min PWT range affects both the Bus-line and System Satisfaction very negatively (see Tables 2-4 to 2-8 and 2-12). We believe that the PWT is highly related to the variability of waiting time. Our result is consistent with findings from Batarce et al. (2016), in their study about crowding conditions from PT users in Santiago, Chile. They report that a focus group result indicated that one of the most critical variables was PT headway regularity. Additionally, their model results show that the coefficient of variation of waiting time is one of the most significant variables. These results are in-line with ours, it could well be, that users perceive their waiting time longer the more variable it is; this is intuitive as they have to adjust their departure time because of the waiting time variability. Our results are also in-line with those of Wachs (1976) and Currie and Wallis (2008), as the variability of waiting time impacts more than that of in-vehicle travel time.

Another significant finding is that two latent constructs Bus-stop Satisfaction and Transfer Satisfaction were not significant in explaining the Bus-line Satisfaction in the first regression. We interpret this, as previous empirical research has (Iseki and Taylor, 2010), in that the most important aspect for users is a reliable and frequent service. In the case of the Transfer Satisfaction (BL4), we hypothesise that the sample used in this research, mainly regular users, had this attribute fulfilled and, as such, did not weigh it as critical to the Bus-line Satisfaction. Another possibility may be that this variable is highly correlated with the Safety Satisfaction (BL3, see Appendix A, Table 2-13, BL3-BL4 covariance). Thus, we recommend commissioning a separate survey, which samples infrequent users and users of other transport modes, to test if these variables are not relevant there also, and check for additional indicators to capture each variable's effect conclusively.

We consider that our framework could be applied in any service setting that considers specific service encounter and global satisfaction planning and tactical strategies, not only for PT services. As our application proves, there may be specific operational variables that could impact both service encounter and global satisfaction, and there will be particular attributes that will affect the service encounter and the global satisfaction independently. It is interesting to note that with our model we can measure how much the specific service encounter satisfaction affects the system satisfaction. As it stands, the tool may help any service provider design specific strategic and operational policies, to help increase their users' PSQ and satisfaction, indirectly improving customer retention.

Another highlight of the model is that it allows decision makers to assess how users with different sociodemographic characteristics perceive specific satisfaction attributes. Thus, particular trends can be examined, and specific unsatisfied user groups can be analysed within any satisfaction attribute, for particular policy design within the service setting being studied.

From our results, we highlight some key policy recommendations: the PWT is the single most relevant policy variable in both Bus-line and System Satisfaction. This finding warrants the Administration to investigate a different type of contract entirely, to resolve the issues of frequency and waiting time (43% of users have a PWT of 10 min or more). Explicitly, the problem may be the variability of waiting time. This critical service variable (PWT) is obviously under-performing in Santiago. We recommend investments in improvements towards service frequency increases and reliability features. The current buses are not providing the reliability expected by users, especially regarding their waiting time.

Related to the previous point, the Frequency and Safety latent constructs are the most relevant variables in the Bus-line Satisfaction. Based on this result, we recommend coordinating an investment plan to tackle the most important issues causing the system to have low reliability. From the literature review there is evidence that operators and bus drivers, in Santiago, do not have incentives to stop at bus-stops, so sometimes they do not stop. The PTA should include key contract clauses and enforcement controls for operators to stop at bus-stops when requested. Also, a major investment plan should be coordinated to provide all the BRT type characteristics in the system (see Herrera et al., 2016), as these allow to improve reliability (i.e. designated lanes corridors and closed stations, amongst others). A schedule operation should be considered in the Transantiago system, especially in corridors with low frequency. This policy could help alleviate the high waiting times perceived by users.

The System Satisfaction is mostly affected by the Peripheral-SPT variable, which is in part made up of users' satisfaction with Metro, other users' behaviour, and information. The administration (Transantiago) should consider these findings in their policy and operations planning. We recommend commissioning another survey that can expand the original one, including additional indicator items for these components, in this way we can assess each subdomain individually. More specific items will help to assess better which specific subdomains are more relevant for the user. This study could include extra items regarding these variables; our model suggests that the two service variables: security in the PT system (users) and information provided are essential policy variables. It would be advisable to entertain the idea of improving the information system provided, and additional safety measures on the PT system, when appointing the survey. By adding items to the survey, we may create separate latent constructs for these variables and distinguish the effects of Metro, the users and information availability, for better policy recommendations.

3. EFFECT OF CRITICAL INCIDENTS ON PUBLIC TRANSPORT SATISFACTION AND LOYALTY: AN ORDINAL PROBIT SEM-MIMIC APPROACH

3.1 Introduction

Public Transport (PT) service quality is a driving element in passengers' travel choices and PT administrators sustainable transport policies. If passengers remain satisfied after experiencing a PT system, they would most probably choose to reuse the service and recommend it to others (de Oña et al., 2016a). As several aspects characterise a PT service, the quality of the whole service depends on the quality of its various elements (Eboli and Mazzulla, 2015). The most usual perception measure is the passengers' judgment of the different service attributes, collected by traditional attribute-based surveys (e.g. Fellesson and Friman, 2012, de Oña et al., 2015, 2016b, Morton et al., 2016).

In this study we consider a further element influencing user perceptions, which could be taken as an alternative measure, the so-called Critical Incidents (CI). CI are defined as encounters that are particularly satisfying or dissatisfying (Bitner et al., 1990). CI may be considered as more concrete than traditional attribute-based satisfaction ratings (Stauss and Hentschel, 1992). For example, they are not restricted to evaluations of predefined service attributes because customers who experience them, remember and can even describe the incident, they have a clearer recollection of the event. Our primary aim is understanding how CI affect passengers' satisfaction. We implement a framework that is innovative in the sense of the theoretical concepts measured and in the methodological sequence used to achieve the results. Nevertheless, our fundamental contribution is including and testing CI for their effect on PT satisfaction.

We introduce item-specific CI to explain attribute-specific (e.g. reliability, safety, cleanliness) satisfaction levels. In turn, the attribute-specific satisfaction explains overall service satisfaction, which in turn leads to loyalty (i.e. intent to recommend the service). We connect all links through a Structural Equation Modelling (SEM) approach. Our model shows how and how much CI affect users' overall satisfaction, through the attribute-specific satisfaction levels. Additionally, it allows complete mediation of the CI over attribute-specific satisfaction, overall satisfaction and loyalty, measured as the users' willingness to recommend the service. Thus, it allows assessing the full effect of CI on overall satisfaction and loyalty constructs for PT policy planning. We found only two previous studies (Edwardsson, 1998, Friman et al., 1998, 2001) that modelled attribute-specific CI to explain attribute satisfaction levels, and no studies considering the effect of CI on loyalty behaviour. The satisfaction-loyalty relation is a well-established fact in the marketing literature (Oliver, 2010, Olsen, 2007, Paulssen and Birk, 2007), and this is confirmed in our study.

Methodologically, we model the presence of CI as a binary variable and demonstrate that treating ordinal items as ordinal scales, produces better and more convincing results than recurring to a numerical scale. The latter is the framework employed in most PT satisfaction studies, but it may yield biased estimates. We prove that significant differences appear between both approaches, even for numerical scales up to 10 categories (i.e. present case study).

Last, we present a full Structural Equation Multiple Cause Multiple Indicator (SEM-MIMIC) modelling framework, which allows for heterogeneity in the satisfaction with the service attributes, overall service satisfaction and loyalty. No previous studies, to the best of our knowledge, have modelled heterogeneity using a MIMIC model for both overall satisfaction and loyalty in PT systems. We believe this can significantly aid PT administrators, as information about heterogeneity in the intent to recommend a service is affected by the users' socioeconomic characteristics and travel habits. PT services are highly heterogeneous, and as such, need to have the heterogeneity accounted for, to avoid biases. Our large sample size (96,763) allows accounting for heterogeneity and obtaining a rich model which correctly quantifies which service attributes are more relevant to users.

The rest of the chapter is structured as follows. Section 3.2 presents a literature review of studies investigating the relationship between CI and Customer Satisfaction (CS) and adds the rationale for our survey design regarding CI. Section 3.3 presents a brief methodological review concerning all models used in this work. In Section 3.4, we introduce the case study by describing the PT system under consideration: the railway services offered in the hinterland of the city of Milan, and also briefly explain the survey and dataset employed. In Section 3.5 we describe our methodology and present the model results in detail, which we discuss in Section 3.6. Finally, conclusions and future research paths are given in Section 3.7.

3.2 Critical Incidents on Customer Satisfaction: A Literature Review

Roos (2002) reports a comparative review of methods for investigating CI, from the traditional Critical Incident Technique (CIT) to alternative methods such as Olsen's Technique (1996), the Sequential Incident Technique (Stauss and Weinlich, 1997) and the Switching Path Analysis Technique (Roos, 1999). However, the majority of studies on the influence of CI on customers' perceived service quality, have been analysed by applying the CIT, which was initially developed by Flanagan (1954). CIT has become the most popular incident-based method in the context of CS studies (Backhaus and Bauer, 2001). It analyses the content of stories related with purchasing and selling interactions of companies with customers, by classifying incidents into specific categories. CIT has been used mainly in the consumer service industry (e.g., hotels, airlines and restaurants) where single transactions were the objects of evaluation. Bitner et al. (1990) report a review of several studies applying CIT to such services. The technique has also been extensively used in the services management and marketing literature (e.g., Edvardsson, 1992, Edvardsson and Strandvik, 2000, Wong and Sohal, 2003, Gremler, 2004, Gremler and Gwinner, 2008, van Doorn and Verhoef, 2008).

In the literature regarding PT, we found few studies based on the analysis of CI and even less on the influence of CI on CS. Some authors did analyse how CI influence other aspects related to transport such as travel behaviour, route-choice, and transport mode-choice. As an example, Sundling et al. (2016) investigated the influence of positive and negative CI on the travel behaviour of older people when using PT. Grison et al. (2013) tried to achieve a better understanding of the variables affecting the choice of route in PT by asking individuals to report on positive and negative real events experienced (i.e. the CI). Finally, van der Waerden and Timmermans (2003) did qualitative research on the impact of CI on transport mode-choice switching behaviour.

Concerning how CI influence customer satisfaction, a significant part of the literature about PT service quality focuses only on the traditional attribute-based surveys. In these studies, researchers propose methods and models that evaluate how passengers perceive the various SQ attributes, expressed through judgements, influence the overall satisfaction and service quality. Only one research group investigated the influence of CI on CS (Edwardsson, 1998, Friman et al., 1998, 2001). In a preliminary study, Edwardsson (1998) analyses written complaints and information from interviews with customers of PT services. The purpose of his work was to describe, examine and interpret passengers' experience of CI. A central part of his research regarding the analysis and interpretation of data led to the formation of categories. In two successive papers, Friman et al. (1998, 2001) analysed negative CI regarding a PT service based on the most frequent complaints. Differently from studies based on the CIT, in which the experienced CI are described in detail and then classified, Friman et al. (1998) obtained the negative CI from the archival customer-complaint data. This approach was motivated by previous findings by Cadotte and Turgeon (1988), according to whom complaints contain unsatisfactory experiences comparable to negative CI. Subsequently, Friman et al. (2001) found that attribute-specific satisfaction was directly and negatively affected by the frequency of negative CI.

From these studies, we may conclude that in the case of PT services negative CI (e.g., a delay, no seats available, lousy driving) may have more impact than positive CI, in contrast to many other services, such as restaurants or entertainment, for which achieving satisfaction is more important and convenient than avoiding dissatisfaction. As an example, the delay of a bus leads to perceiving the service as non-reliable, while a bus running on time is not considered a positive event, but just a regular experience (Friman et al., 1998). Moreover, Backhaus and Bauer (2001) interpreted that negative incidents should be more readily accessible to memory than positive incidents and should elicit a stronger effect. They also highlighted that the first CI exerts the most substantial impact on overall satisfaction, and the degree of influence on satisfaction decreases with each additional CI. Finally, they also found that the effect of a CI is more significant on satisfaction with an attribute that has a particular closeness to the CI than on overall satisfaction. This fundamental result helps to justify our framework.

Following the evidence arising from the literature review, a CS survey was designed explicitly aimed at collecting information about CI and passenger satisfaction for attribute-specific domains. With this methodology, which combines the passengers' judgements of whether they experienced CI and attribute-specific satisfaction, we were able to combine the traditional attribute-based CS survey with the more specific CI-based survey.

Considering our experience in CS survey design and methodological development for data analysis based on passengers' perceptions, a questionnaire was developed where:

a) The passengers interviewed were asked to give information about only negative CI (i.e., not positive).

b) Passengers were also asked to state whether they experienced a CI (or not) concerning each specific service attribute listed in the questionnaire. They were asked if they had found any problems (negative CI) on these attributes.

c) Passengers had to consider the most recent CI, or instead events that occurred during the last 30 days before the compilation of the survey.

We avoid requesting passengers for CI frequency, or a detailed description of the experienced event because we believe that these specific requests could be problematic if passengers do not remember earlier incidents or the details of the experience. Also, the literature suggests that the first episode has the most substantial impact on overall satisfaction.

We hypothesise that CI are highly salient for users when they rate their satisfaction with specific attributes of the PT system. By considering that CI will have more influence on the satisfaction with an attribute that is closely related to the incident, we retain that it is convenient to ask passengers to give information concerning attribute-specific satisfaction and CI, at the same time. Our methodology provides added value to the existing literature because it allows exploring how and how much CI indirectly affect users' overall satisfaction, through the satisfaction with every single attribute.

3.3 Methodological Review

3.3.1. What is SEM?

SEM is a general methodology that contains many commonly employed statistical models, such as analysis of variance, analysis of covariance, multiple regression, factor analysis, path analysis, simultaneous equations, amongst others (Bowen and Guo, 2012). In this sense, SEM is a methodology that comprises a set of multivariate statistical approaches to empirical data. Frequently SEM is used as a data analysis method that combines simultaneous equations and factor analysis. Factor analysis test hypotheses about how well sets of observed variables in an existing dataset measure latent constructs (i.e. factors). These latent constructs represent abstract and psychological concepts such as attitudes, behaviour, and emotions that cannot be directly observed or measured with single items (Bowen and Guo, 2012). For this reason, in the literature factor analysis is also known as a measurement model.

Regression models test the strength and direction of relationships between predictor variables (independent) and an outcome variable (dependent). SEM, unlike standard regression models, can include regression relationships among latent variables and between observed and latent variables. In other words, it can estimate, in a single analysis, models where one or more variables act as predicted and predictor variables concurrently (Bollen, 1989, Bowen and Guo, 2012). SEM offers advantages over other statistical methods. Amongst its capabilities are: (i) treating endogenous and exogenous variables as random variables with measurement error, (ii) explaining latent variables with multiple indicators, (iii) overall testing of model fit over coefficients only, and (iv) handling non-normal data (Golob, 2003). It also permits testing models where there are multiple dependent variables (Bowen and Guo, 2012).

The distinction between latent and observed variables represents a fundamental difference between SEM and traditional regression modelling. In the SEM framework, latent variables are of interest, but they cannot be directly measured. Observed variables are a function of model-specific latent constructs and latent measurement errors. In this framework, researchers can isolate “real” causes of scores and variations in scoring due to unrelated causes. Tests of relationships among the resulting latent variables –structural equations– are superior to tests among variables containing irrelevant variance (Bowen and Guo, 2012).

In general, factor analysis methods are used to analyse the relationships among measured variables. Initially, Principal Component Analysis (PCA) helps to explore how the factors are composed. PCA is a statistical procedure that uses a transformation to convert a set of observations of correlated variables into a set of values of linearly uncorrelated variables called principal components (see Jolliffe, 2014). Once the factor structure is known, the modeller can specify a confirmatory factor analysis (CFA).

CFA answers different questions than the exploratory approaches (PCA). CFA assesses the fit of the proposed model and can help to determine the adequacy of a measurement model for the available sample, before performing a substantive latent variable analysis (Bowen and Guo, 2012). Establishing the measurement model adequacy before testing the structural relationships in SEM is best practice (Anderson and Gerbing, 1988, Bollen, 1989). Many of the rules of interpretation regarding the assessment of model fit in structural equation modelling apply equally to CFA. The approach is distinguished from SEM by the fact that in CFA there are no directed arrows between latent factors. In other words, factors do not directly cause one another in CFA, whereas SEM often specifies particular factors and variables to be causal. In the context of SEM, the CFA is often called the measurement model, while the relations between the latent variables, with directed arrows, are called the structural model.

We will follow this sequence in our case study, (i) we will conduct PCA for the CI and Satisfaction items, (ii) then we will perform CFA to test the proposed measurement models and, (iii) subsequently, we will examine the full SEM models for the proposed relations between the latent variables. Finally, (iv) we will test the SEM-MIMIC models, to assess the perception of heterogeneity by PT users.

3.3.2. SEM ordinal-Probit formulation of categorical-ordered variables

To justify our methodological approach, we want to specify the following. Commonly, the answers to customer satisfaction surveys are expressed on a Likert-type scale (Likert, 1932), for example from 1 (Very dissatisfied) to 10 (Very satisfied). A problem with ordinal scales, is that a 4, say, does not necessarily mean being twice as satisfied as a 2. Also, the difference between 1 and 2 does not mean, necessarily, the same as the difference between 4 and 5, similar to any grading system (i.e. 1-5 or 1-7). For this reason, a numerical treatment may induce bias in the results. The ordinality of the survey responses is unavoidable, and there are possible biases inherited for reproducibility (i.e. different scales across individuals). As such, the key is to garner several items for each latent construct to obtain a better measurement, thus reducing the bias.

In our case, the scale used was identical to the Italian grading system (1-10). By using a well-known scale, we reduce the bias for reproducibility. Another bias that is often overlooked is the regional bias: people from different geographical regions, states, countries, urban-suburban or rural locations, tend to show systematic differences in their interpretations of point scales and their tendencies to give higher or lower scores. Nonetheless, in our case, as we treat the whole conurbation of the hinterland of Milan, we can assume that all users have a similar scale. Again, the fact that it is a well-known scale to the respondents reduces the possibility of this bias.

For SEM, the most common and basic estimation algorithm is the Maximum Likelihood (ML) estimator, which is appropriate for an interval, ratio level, or continuous data with normal distributions and large sample sizes. Bollen (1989) provides details of the robustness of estimators under various conditions. For non-normal data, Weighted Least Squares (WLS) is one of the most recommended and preferred estimators (Bollen, 1989). The WLS-estimator makes minimum assumptions about the distributions of the observed variables.

When variables in the analysis are ordinal, the recommended analysis matrix is a polychoric correlation matrix (Bowen and Guo, 2012). The main assumption is that behind the ordinal categories for measuring the latent constructs a continuous normally distributed phenomenon lies. Since we cannot directly observe the variable, we define it as a latent response variable, denoted by y*. The relationship between ordinal y (with K response categories) and y* is given by:

$$y = k \Leftrightarrow \tau_{k-1} < y^* < \tau_k \tag{3.1}$$

for the categories k = 1, 2, …, $K - 1$; furthermore, let $\tau_0 = -\infty$ and $\tau_K = +\infty$.

The K values are called cut points or thresholds. The typical assumption is that y* is normally distributed with mean zero, and unit variance:

$$y^* \sim N(0,1) \qquad (3.2)$$

Rosseel (2012) presents a proof that this model leads to a general ordinal-Probit formulation.

In our case study, we use the Diagonally Weighted Least-Squares (DWLS) estimator. DWLS estimators use only the diagonal of the weight matrix for inverting the matrix. However, DWLS estimators use the full weight matrix noted in the WLS equation, to estimate standard errors of parameter estimates and the overall model chi-square index (χ^2) accurately. DiStefano and Morgan (2014) provide further details about the DWLS estimator and its variations. For our case study, we use the DWLS estimator included in the Lavaan package for R (Rosseel, 2012), in this sense, we have the flexibility to consider all possible ordinal-ordinal pairwise set of variables in the SEM analysis.

We did not find any other application that considered ordinal-ordinal variable relationships in an SEM model in the PT satisfaction literature. However, there are a few studies that include ordered Probit approaches, yet not in the SEM framework, such as dell'Olio et al. (2010). We will apply both the numeric-numeric approach and the ordinal-ordinal approach to compare both models according to SEM model fit indices, demonstrating the benefits of implementing the DWLS estimator and treating the ordinal variables with a latent normally distributed scale. This approach contributes to the PT satisfaction literature.

3.3.3. The SEM-MIMIC model

The MIMIC model (Joreskog and Goldberger, 1975) allows for the possibility of detecting heterogeneity in the measurement of latent variables between different groups within the population when using SEM. The MIMIC approach considers the restriction of a group-invariant covariance matrix for the observed indicators, conditional on grouping variables represented by regressors, which can be categorical values (i.e. binary variables).

Including a relevant set of explanatory variables provides MIMIC modelling extra information about the measurement model and enables the investigation of hypotheses of invariance across subpopulations explicitly. Allen et al. (2018) applied the SEM-MIMIC framework recently for modelling service-specific and global transit satisfaction in a Santiago (Chile) case study, specifically for the bus component of an integrated PT system. They found the following characteristics to be significant for the satisfaction latent constructs: socioeconomic status, age, gender, time of day of the trip, day of the week, bus and Metro use frequency, trip purpose, perceived waiting and travel time, bus operator, and date of the survey. In fact, all characteristics tested were significant.

In the same spirit, we propose an SEM-MIMIC model that considers all the possible socioeconomic and travel characteristics of the present case study. We expect that for a heterogeneous population, such as the suburban and regional rail users in Milan, the latent constructs for attribute-specific satisfaction can be conditioned for different groupings using the explanatory variables (regressors), and then testing for differences across the subpopulations.

The SEM-MIMIC approach has been previously implemented for an attribute-specific satisfaction model by Guirao et al. (2016) using data from urban bus users in Madrid (Spain). They derived direct estimates of the importance of service attributes. The groupings age, gender and trip purpose, were significant characteristics that affected the latent constructs. However, as they used a smaller sample size (n = 520) than ours, we expect more variables to be significant in our model, similar to Allen et al. (2018). Likewise, we want to compare whether the results are close to both these case studies, to test for possible generalisations.

3.4 Case Study

3.4.1 PT system

The PT operator collected data supporting this research through Customer Satisfaction Surveys (CSS) addressed to passengers of railway services in the North of Italy, from May 2011 to June 2014, and with a frequency of two times per year. The analysed PT system covers all regional railways transport services operating in the Lombardy region, which is the most attractive region in Northern Italy from an economic and financial point-of-view. In the last 15 years, residents of this area have experienced an increase in PT use, combined with a significant amount of investments by central and local governments. Investments were oriented towards the regional railway network and the High-Speed national network connecting the main Italian cities with Milan. Specifically, from 2005 to 2009, the corridor Turin-Milan-Naples and Milan-Venice was implemented as part of the Trans-European Networks (TEN-T), sponsored by the European Commission.

Two different transit operators provided regional railways services in the Lombardy region: a transit operator working at a national level, which managed only the regional lines, and a transit operator working at a local level, which controlled all the suburban lines and a few other minor regional lines. To optimise rail services from both quantitative and qualitative points-of-view, in 2009 the two transit operators were joined in a single Transit Company. The new company manages now 1,900 kilometres of railway lines among the busiest in Europe: 42 regional lines connecting the city of Milan with the bordering regions such as Piedmont and Veneto, nine suburban lines connecting the hinterland of Milan with the city centre, and a unique line connecting Milan city with Malpensa Airport. There are over 2,000 daily departures, and over 570 thousand passengers travelling per day (see de Oña et al., 2014 for additional details). Despite the impressive numbers and a specific mobility demand captured by the High-Speed railway services, the Lombardy regional railway system is today still inadequate due to a continuously expanding demand level. This demand must find a practical alternative to the private car, a less sustainable transport mode, in the local transit service.

Suburban lines connect the hinterland with the city of Milan using railway services designed for satisfying urban mobility. These services have a high frequency (services operating 365 days per year, from 6:00 a.m. to 12:00 p.m., and with a scheduled departure every 30 min), high capacity, and trains stopping at all railway stations. On the other hand, the regional service meets the needs of moving from Milan to the central locations bordering the Lombardy capital and the region. Regional lines are characterised by less frequent services, with no schedule, and trains stopping only at the most important railway stations. Regional and suburban lines can be used conjointly with urban lines (Metro) through integrated tickets and city passes; there is an electronic ticketing system implemented.

For all the regional and suburban lines 1,321 trains are used every day, of which 90% have an electrical power supply, 62% have air conditioning, 100% permit bringing bicycles on board, 24% have facilities for passengers with reduced mobility (PRM). Additionally, there are 14 waiting rooms at the stations where assistance for travel can be organised for PRM. All the trains operating along Malpensa Express line have air conditioning, a display of information on board and facilities for PRM. About 88% of the regional trains arrive within 5 min of the scheduled time, and over 97% arrive within 15 min; for suburban lines, the number of trains arriving within 5 min of the scheduled time is over 90%. About 80% of checks satisfy the standard level of quality regarding cleanliness on board. Complaints are managed through a dedicated website and a 24-hour Contact Centre: in 2009 about 6,600 complaints were handled, with an average replying time of 18 days.

Regarding the services at railway stations, over 70% of the stations have aided accessibility for PRM and other services for the disabled (8% escalators, 15% lifts, and 38% toilets). Also, 95% of the stations have a waiting room, 95% video surveillance installations, 89% car parking, 65% bike parking, 67% a ticket office, and 75% automated ticket machines.

3.4.2 The survey

Since 1999, the transit operator has conducted periodic CSS for continuously monitoring passengers' perception of its service. The surveys were carried out in compliance with the Italian guidelines for customer satisfaction surveys (UNI 11098: 2003 standard) and consisted of face-to-face interviews on board and at railway stations with a representative sample of travellers. In the case of the Malpensa Express line, questionnaires were also translated into different languages to consider foreign travellers. The data used herein consist of seven complete CSS, conducted two times per year, from May 2011 until June 2014: four Spring surveys (May-June) and three Autumn surveys (November-December). The interviews were conducted during the whole week, even on holidays, and at any time of the day.

The questionnaire structure contains six parts, totalling 109 questions. Part 1 focuses on general information and includes items about train, line and railway station. Part 2 refers to user's socioeconomic characteristics, covering questions about gender, age, monthly income and so on. Part 3 is about travel habits and trip characteristics and includes trip origin and destination, access and egress transport mode, and trip purpose. Part 4 focuses on passengers' perceptions of the services. Users were asked to express importance and satisfaction rates (S1-S27) about several service quality attributes. The former was not used in this research. Ratings were requested according to an ordinal 1 to 10 scale, where 1 means "not very important"/"not very satisfied" and 10 means "very important"/"very satisfied". The population intuitively understands this scoring scale because in Italy school grades use a similar 1-10 range, where six is the minimum passing grade. In total, 27 service quality attributes were investigated, including safety, cleanliness, comfort, information, and personnel.

Additionally, for each of the 27 service quality items users had to indicate the occurrence of a problem (negative CI) in the last 30 days before the interview, and this had a “yes” or “no” answer (C1-C27). Part 5 has three questions concerning opinions about the overall service. Specifically, an overall level of satisfaction was requested separately for services managed by the national (S28) or local transit operator (S29). Also, users had to state if they noticed any improvements in the services managed by the local transit operator in the last year (S30), with three possible answers: “big improvements”, “some improvements”, and “no improvements”. Finally, extra specific questions requested a level of satisfaction on the same scale (E32-E35) and consist of perceptions about the information system in abnormal operating conditions (E32), the ticket-purchasing system (E34) and the interventions implemented for improving service quality (E35). Finally, passengers were asked if they would recommend (“yes”, “no” or “I do not know”) the train service to relatives and friends (E38).

3.4.3 The sample

The CSS totalized 96,763 questionnaires (approximately 14,000 records for each periodic survey). We present the main characteristics of the sample in Table 3-1. The sample is made up of more females (56%) than males and is mostly composed of people aged between 16 and 40 (75%). Most of the journeys are for work or study purposes. Despite the high level of education (79% of the sample has a High School or bachelor’s degree), the net monthly income is very low or No Income at all (59%). Respondents are prevalently commuters (54%), daily PT users (57%), PT users on the weekdays (81%), and at peak hours (58%). As frequent users, they use a pass in most cases (86%).

Table 3-1: Sample characteristics

Passenger's socioeconomic characteristics	
Gender	Male (44%), Female (56%)
Age (years)	16-25 (43%), 26-40 (32%), 41-65 (22%), >65 (3%)
Employment Status	Employee (38%), Professional (14%), Student (38%), Other (10%)
Monthly Income (Euros)	No Income (43%), < 1000 (16%), 1001-1500 (20%), 1501-3000 (17%), > 3000 (4%)
Education Level	Elementary School degree (8%), Middle School degree (13%), High School degree (51%), University degree (28%)
Passenger's travel habits	
Frequency of Use	Daily (57%), Weekly (17%), Occasional (26%)
Journey Purpose	Work (38%), Study (34%), Other (28%)
Ticket Type	One-Way ticket (6%), Book of tickets (8%), Pass (86%)
Access Mode	Walk (39%), Car (26%), Cycle (5%), Other public transport (PT) mode (30%)
Egress Mode	Walk (41%), Car (26%), Cycle (5%), Other PT (27%)
Other information	
Day	Workday (81%), Holiday and Pre-holiday (19%)
Time of Day	Off-Peak Hour (42%), AM Peak Hour (17%), PM Peak Hour (41%)
Service	Regional (56%), Suburban (40%), Malpensa express (4%)
Type of User	Commuter Worker (28%), Commuter Student (26%), Other (46%)
Line	Northern Regional (20%), Southern Regional (12%), Eastern Regional (19%), Western Regional (7%), Suburban and Minor Regional* (42%)

3.5 Model Results

Section 3.5 is divided into four parts: the PCA, the Measurement Model, the comparison between SEM Numeric and SEM ordinal, and finally the SEM-MIMIC ordinal Probit Model. In synthesis, (i) the PCA allows exploring the composition of the latent constructs perceived by users regarding CI and Satisfaction items; (ii) the CFA enables testing the measurement models; (iii) the SEMs allow to examine the structural relations between the CI and Satisfaction latent constructs, and the Overall Satisfaction construct and Loyalty item. Additionally, we compare the model fits between the numeric and ordinal SEMs.

Finally, (iv) the SEM-MIMIC model allows correcting for heterogeneity in the perception of the Satisfaction constructs, Overall Satisfaction, and Loyalty.

3.5.1 Principal component analyses (PCA)

Before designing the measurement model, it is best practice to run an exploratory PCA separately on the CI items and the CS items. Doing this helps determine which items are represented by a specific latent construct; in other words, which items are related and represent one construct as a whole. Therefore, we performed two separate PCA on the set of CI and CS items to obtain initial adequate latent constructs for the model. After iterating several times, according to the Kaiser rule (Kaiser, 1960), we defined eight attributes concerning the CI items and nine attributes for the CS items.

We did not include individual questions that did not correlate well with other survey items, as no latent constructs could be formed. When performing the PCA, this occurred for the item Ticket price concerning the service offered (C12, S12), which was consistently unrelated to any other construct. The reason is that this item has more meaning as a service value item, which is not attribute-specific, but it is more of an overall rating. Hence, we omitted this question in the subsequent PCA and left it for later analysis of overall satisfaction. The PCA results are reported in Appendix C (Tables 3-6 and 3-7).

The resulting eight CI components are Safety (CI1), Cleanliness (CI23), Comfort (CI4), Reliability (CI5A), Accessibility (CI5B), Additional Services (CI6), Information (CI7), and Personnel (CI8). The nine CS components are Safety (CS1), Cleanliness on Board (CS2), Cleanliness at Station (CS3), Comfort (CS4), Reliability and Accessibility (CS5), Additional Services (CS6), Information (CS7), Personnel (CS8) and Added-Value Services (CS9). The criterion used was a resulting absolute loading > 0.4, which was met for all but two items that loaded 0.37. We considered this satisfactory to continue with the CFA.

In turn, each item represents an individual question from the survey. There are differences between the CI and CS constructs. For the CI items, the five cleanliness items were grouped into one latent construct (CI23), but for the CS items, we obtained two separate constructs, CS2 (Cleanliness on Board) and CS3 (Cleanliness at Station). Additionally, for the CS items, the five reliability and accessibility items loaded into one latent construct (CS5), but for the CI items we found two separate constructs, CI5A (Reliability) and CI5B (Accessibility).

3.5.2 Measurement model: numeric and ordinal

Once the different latent constructs were defined, we set up the Measurement Model. In our case, we intended to compare a CFA-numeric with a CFA-ordinal. Hence, we computed both types of CFA. The model is the same for both and is presented in Figure 3-1: the ovals represent the latent constructs being measured, and the rectangles represent the respective CI (C01-C27) and CS items (S01-S30, E32-E35).

With the measurement model, a confirmatory factor analysis (CFA) was performed to test for goodness-of-fit. The CFA-numeric model yielded the following fit indices (fully explained in Hooper et al., 2008), with their recommended cut-off values in parentheses (Hu and Bentler, 1999): comparative fit index, CFI: 0.931 (>0.95), adjusted goodness of fit index, AGFI: 0.896 (>0.95), root mean squared error approximation, RMSEA: 0.040 (<0.08), and standardized root mean square residual, SRMR: 0.028 (<0.06). The most commonly reported fit indices are the CFI and the RMSEA; strictly, a good model fit should have a RMSEA of less than 0.8. In recent years, the RMSEA has become regarded as one of the most informative fit indices (Diamantopoulos and Siguaw, 2000, p.85) due to its sensitivity to the number of estimated parameters in the model, as it favours more parsimonious models.

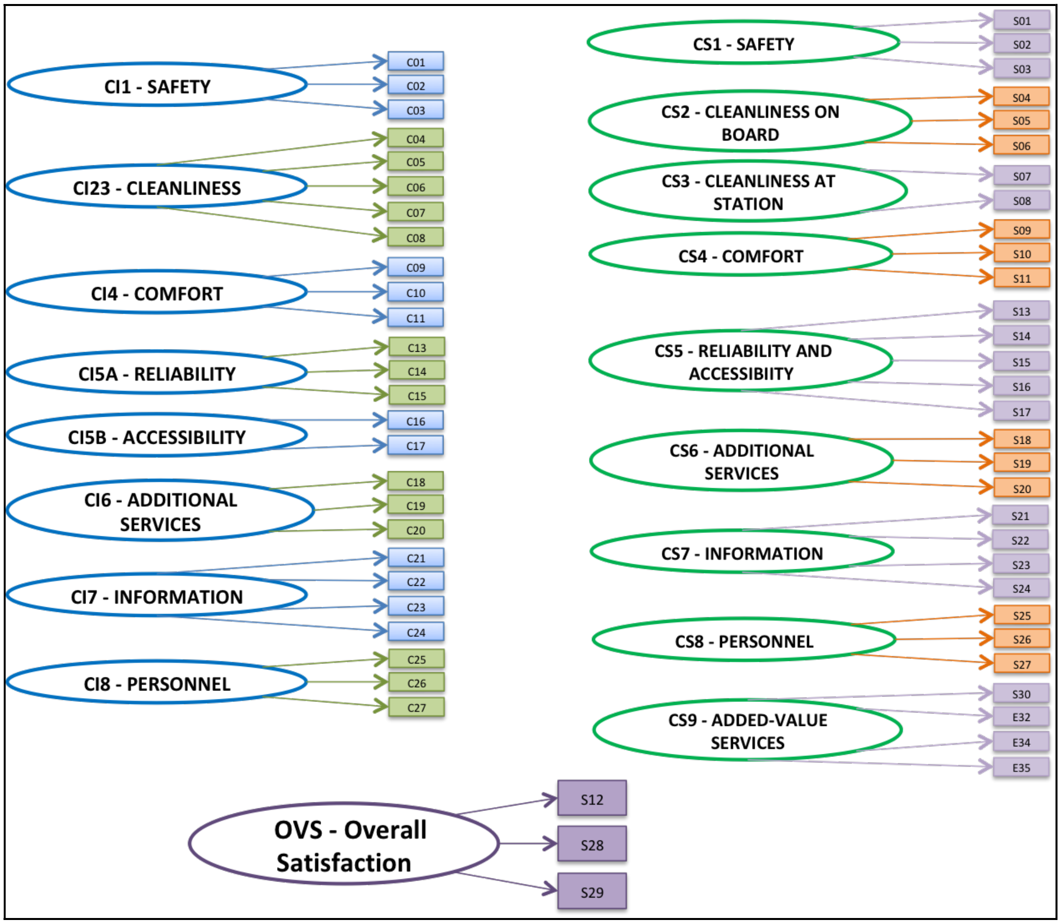

Figure 3-1: Measurement Model (Confirmatory Factor Analysis)

On the other hand, the CFA-ordinal model yielded the following fit indices: CFI: 0.997 (>0.95), AGFI: 0.996 (>0.95), RMSEA: 0.032 (<0.08), and SRMR: 0.032 (<0.06). We can confirm that the goodness-of-fit indices under the CFA-ordinal model are significantly better than under the CFA-numeric model. In other words, the CFA-ordinal model provided a much more adequate fit. We discuss these results in section 3.6.

3.5.3 Comparison of SEM-numeric with SEM-ordinal models

Once we have satisfactory measurement models, we can set up the SEM. Again, we computed two types of SEM: an SEM-numeric where all variables, specifically the satisfaction scores, are treated with a numeric range, and an SEM-ordinal where the satisfaction scores are treated as ordered categorical variables. The model schema is the same for both SEM, and it is presented in Figure 3-2: the ovals represent the latent constructs being measured, and the rectangle represents the respective Loyalty item (E38): Intent to recommend to friends and relatives.

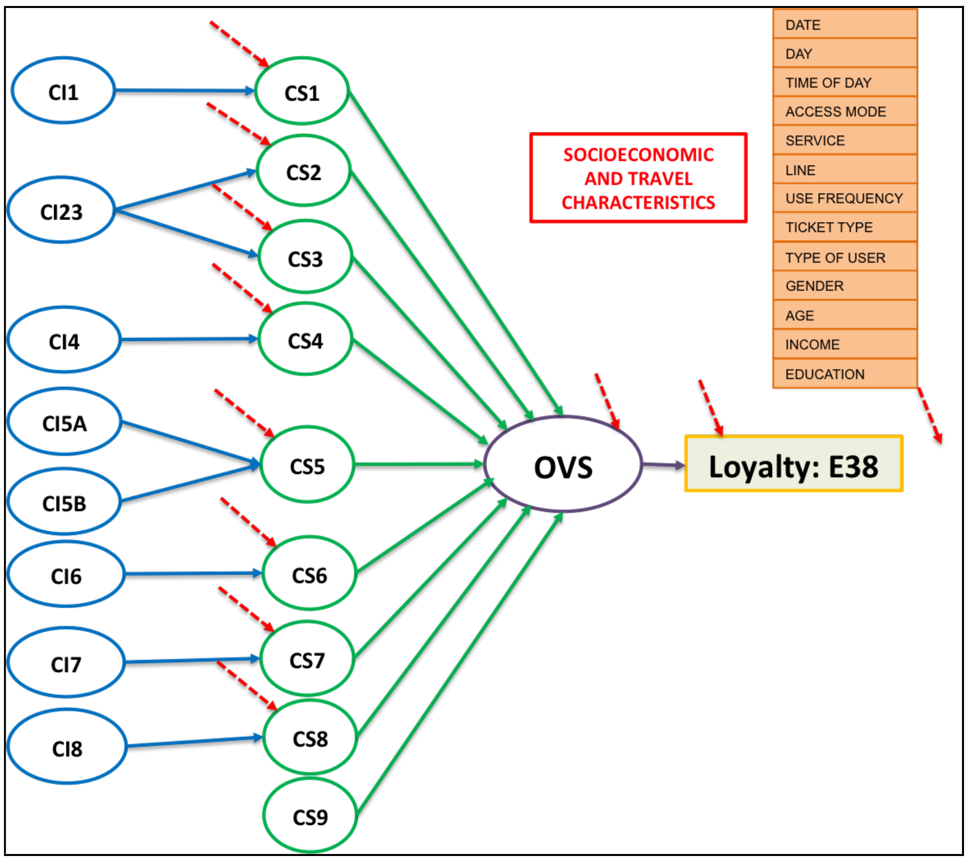

Figure 3-2: Schema for SEM and SEM-MIMIC models

We assume that the attribute-specific CI affect their attribute-specific CS constructs directly. We also assume that attribute-specific CS constructs affect and comprise Overall Satisfaction. Afterwards, Overall Satisfaction leads directly to Loyalty. The latent constructs are measured with the Measurement Model presented previously (Figure 3-1). Additionally, socioeconomic characteristics and travel habits are included in the SEM-MIMIC models (next section). The only difference between both models is that the item Perception of Improvements (S30) was defined as an antecedent for Added-Value Services in the Numeric model, but later was categorised as part of the Added-Value Services, via the measurement model, in the ordinal model.

Thus, both numeric and ordinal SEM models were tested to assess goodness-of-fit. The SEM-numeric model yielded the following fit indices: CFI: 0.803 (>0.95), AGFI: 0.699 (>0.95), RMSEA: 0.064 (<0.08), and SRMR: 0.183 (<0.06). On the other hand, the SEM-ordinal model yielded the following fit indices: CFI: 0.962 (>0.95), AGFI: 0.962 (>0.95), RMSEA: 0.101 (<0.08), and SRMR: 0.147 (<0.06). Thus, the fit indices for the SEM-numeric model imply a poor fit and those of the SEM-ordinal model are significantly better.

We present the estimates, standard errors (S.E.), Z-values, and standardised coefficients (Std.Coeff.) of the SEM parameters in Table 3-2. The standardised coefficients refer to how many standard deviations a dependent variable will change, per standard deviation increase in the predictor variable.

Comparing the standardised coefficients for the CS constructs regressions, it is clear that the SEM-ordinal regressions provide more substantial effects when the CI regress on the CS. Precisely, the standardised coefficients lie between 0.35-0.49 for the SEM-numeric, and between 0.79-0.94 for the SEM-ordinal model. For the Overall Satisfaction regression, the results are similar in magnitude, but different constructs are significant in each case. The most relevant constructs have a high significance in both models: Reliability and Accessibility, and Added-Value Services. For the Loyalty regression, the SEM-ordinal regression provides, once more, a more significant effect than the SEM-numeric regression. Specifically, the standardised coefficient for the SEM-ordinal regression is 0.62, and for the SEM-numeric regression is only 0.40.

Table 3-2: SEM parameter estimates

	SEM-numeric				SEM-ordinal			
	Estimate	S.E.	Z-value	Std.Coeff.	Estimate	S.E.	Z-value	Std.Coeff.
CS1 Safety								
CI1 CI Safety	-3.867	0.049	-78.62	**-0.337**	-1.007	0.011	-94.12	**-0.814**
CS2 Cleanliness on Board								
CI23 CI Cleanliness	-2.972	0.023	-130.99	**-0.444**	-1.067	0.003	-322.26	**-0.903**
CS3 Cleanliness-Station								
CI23 CI Cleanliness	-1.135	0.011	-100.51	**-0.351**	-1.592	0.009	-178.83	**-0.790**
CS4 Comfort								
CI4 CI Comfort	-2.933	0.026	-111.90	**-0.490**	-1.196	0.008	-156.84	**-0.939**
CS5 Reliab.-Accessibility								
CI5A CI Reliability	-3.456	0.034	-100.49	**-0.430**	-1.058	0.008	-140.03	**-0.801**
CS6 Additional Services								
CI6 CI Addit. Services	-3.211	0.047	-67.65	**-0.309**	-0.982	0.010	-93.83	**-0.805**
CS7 Information								
CI7 CI Information	-3.407	0.033	-104.74	**-0.406**	-1.167	0.007	-166.05	**-0.894**
CS8 Personnel								
CI8 CI Personnel	-4.163	0.044	-93.68	**-0.394**	-1.168	0.010	-120.15	**-0.854**
CS9 Add. Value Services								
S30 Improvements	0.889	0.009	99.22	**0.384**				
OVS Overall Satisfaction								
CS1 Safety	0.083	0.003	29.80	**0.084**				
CS2 Cleanliness on Board	0.135	0.002	72.48	**0.211**	0.148	0.002	83.32	**0.173**
CS3 Cleanliness at Station								
CS4 Comfort	0.163	0.003	59.97	**0.189**	0.124	0.003	40.01	**0.112**
CS5 Reliab.-Accessibility	0.314	0.003	103.71	**0.354**	0.357	0.003	136.73	**0.322**
CS6 Additional Services								
CS7 Information	0.085	0.002	37.84	**0.110**				
CS8 Personnel	0.147	0.003	57.46	**0.175**	0.082	0.002	39.36	**0.088**
CS9 Add. Value Services	0.311	0.004	83.77	**0.303**	0.606	0.006	93.71	**0.352**
S30 Improvements	0.287	0.007	40.51	**0.121**				
E38 Loyalty								
OVS Overall Satisfaction	0.131	0.001	118.28	**0.398**	0.755	0.004	200.06	**0.615**

The results are confirmed by the R^2 explained variance statistics, presented in Appendix D (Table 3-8). The SEM-ordinal regression provides much better-explained variance, ranging from 0.62-0.88 for the Satisfaction regressions and

0.77 for the Overall Satisfaction regression. For Loyalty, the SEM-ordinal model has 0.38 against 0.16 in the SEM-numeric model. We can confidently state that the SEM-ordinal regression provides a much better representation of the SEM model proposed. For the SEM-MIMIC formulation, we will continue the analysis with only the ordinal framework.

3.5.4 SEM-MIMIC ordinal

Among the objectives of this research, we wanted to capture the heterogeneity of users' perception of service quality. Our model hypothesises differences in users' perceptions as a function of their socioeconomic characteristics and travel habits and tests them within the model. The SEM-MIMIC model allows correcting for possible biases due to heterogeneity in users' perceptions. To capture heterogeneity, we introduce more parameters into the model (Hooper et al., 2008), supported by the significant sample available (96,763 records).

First, the proposed SEM model contains eight CI latent variables, nine CS latent variables, one OVS latent variable, and one indicator measuring Loyalty (E38). The items regarding the occurrence of CI measure the CI latent variables. These affect the CS latent variables, which are related to eight different service aspects, each measured by indicators representing satisfaction judgements. In turn, the CS latent variables affect the OVS latent variable representing the satisfaction with the overall service. Finally, the OVS latent variable affects the indicator of Loyalty. In synthesis, we start with the SEM-ordinal model shown in Figure 3-2.

Additionally, (n-1) dummy variables regarding users' socioeconomic and travel conditions were defined for each user category and introduced as regressors for specific latent constructs. We assume that these user characteristics (regressors) affect satisfaction with the particular service attributes (CS1-CS8 latent variables), the satisfaction with the overall service (OVS latent variable), and Loyalty. We used all variables in the separate regression equations. We assumed that the CI latent constructs are linked directly to the CS latent constructs, which in turn are directly related to overall satisfaction and indirectly to loyalty (see Figure 3-2).

Initially, we introduced dummy variables into the MIMIC model (base category case in parentheses), by distinguishing between: (i) travel characteristics: six for Survey Date (Autumn.2012), one for Day (Workday), two for Time of Day (Off.Peak), and three for Access Mode (Access.Walk); (ii) travel habits: two for Service (Regional), four for Line (Suburban), two for Frequency of Use (Daily), two for Ticket Type (Ticket.Pass), and two for Type of User (Other); and (iii) socioeconomic characteristics: one for Gender (Male), three for Age (Age.16-25), four for Net Monthly Income (No.Income), and three for Education Level (High.School.Degree).

Figure 3-2 represents the complete MIMIC Structural Equation Model. Note that all dummy variables act as regressors onto the CS constructs, the Overall Satisfaction, and Loyalty indicator, totalling nine separate regressions. Firstly, we tested a specification including the 35 variables, but later left for each latent construct, only those regressors that were significant at the 95% level and removed the rest. We kept those regressors that effectively affected the latent constructs, avoiding a large model needlessly. The full model has 612 parameters, 2,864 degrees of freedom, and the Chi-Square statistics is significant at the 95% level.

We discuss the full model fit indices; the results indicate a CFI: 0.965 (>0.95). This index shows a slight improvement when compared to the original SEM-ordinal model discussed in Section 3.5.3. Additionally, the rest of the fit indices, except SRMR, remained well within their acceptable values: AGFI: 0.959 (>0.95), RMSEA: 0.070 (<0.08), and SRMR: 0.133 (<0.06). These values confirm an excellent fit to the measurement model proposed, it generates valuable information about heterogeneity, validating it. CFI should be above 0.9 for an acceptable fit and above 0.95 for an excellent fit. Regarding RMSEA (< 0.08), frequently used as an indicator of fit, the model produced an adequate fit, outperforming the SEM-ordinal model.

Table 3-3 presents the estimates and statistics for the full SEM-MIMIC ordinal measurement model.

Table 3-3: SEM-MIMIC ordinal: CI and CS constructs measurement models

	Estimate	S.E.	Z-value	Std.Coeff
CI1				
C01	1.000			0.618
C02	1.014	0.011	89.72	0.626
C03	1.006	0.011	88.57	0.622
CI23				
C04	1.000			0.777
C05	0.994	0.004	273.56	0.772
C06	0.863	0.004	207.52	0.671
C07	0.956	0.004	227.91	0.743
C08	0.942	0.005	202.82	0.732
CI4				
C09	1.000			0.541
C10	1.097	0.008	136.47	0.593
C11	1.142	0.009	130.33	0.618
CI5A				
C13	1.000			0.516
C14	0.986	0.009	115.06	0.508
C15	0.956	0.009	105.50	**0.493**
CI6				
C18	1.000			0.518
C19	1.160	0.016	72.89	0.601
C20	1.177	0.015	77.74	0.610
CI7				
C21	1.000			0.630
C22	1.020	0.007	146.52	0.643
C23	1.017	0.008	121.23	0.641
C24	1.034	0.009	116.16	0.652
CI8				
C25	1.000			0.626
C26	0.851	0.009	94.59	0.533
C27	0.945	0.009	107.60	0.592

	Estimate	S.E.	Z-value	Std.Coeff
CS1				
S01	1.000			0.800
S02	1.138	0.003	400.97	0.908
S03	1.097	0.003	386.33	0.875
CS2				
S04	1.000			0.952
S05	1.001	0.001	1277.78	0.953
S06	0.927	0.001	961.20	0.886
CS3				
S07	0.571	0.002	309.61	0.943
S08	0.556	0.002	330.90	0.919
CS4				
S09	1.000			0.740
S10	1.127	0.003	374.21	0.830
S11	1.181	0.003	367.57	0.869
CS5				
S13	1.000			0.737
S14	1.174	0.003	422.35	0.862
S15	1.164	0.003	418.93	0.855
S16	1.034	0.003	367.72	0.762
S17	0.972	0.003	349.49	0.717
CS6				
S18	1.000			0.669
S19	1.082	0.005	237.01	0.723
S20	1.185	0.005	224.88	0.790
CS7				
S21	1.000			0.865
S22	1.008	0.001	684.24	0.872
S23	0.981	0.002	611.63	0.849
S24	0.943	0.002	573.30	0.817
CS8				
S25	1.000			0.878
S26	0.802	0.002	333.00	0.706
S27	0.970	0.002	481.89	0.852
CS9				
S30	1.000			**0.466**
E32	1.532	0.014	111.86	0.714
E34	1.214	0.012	102.85	0.566
E35	1.736	0.015	119.05	0.809
OVS				
S12	1.000			0.819
S28	1.132	0.002	480.88	0.922
S29	1.100	0.002	471.67	0.897

The measurement model allows assessing the reliability of measurement of the separate latent constructs. We consider items that have a standardised coefficient of 0.6 or more to have good reliability, and 0.5, acceptable reliability. All items except two (shown in bold in Table 3-3) have a standardised coefficient higher than 0.5; we consider this to be acceptable since both are reasonably close to 0.5. It is worth noting that for the CI measurement model the standardised coefficients range from 0.49-0.78, implying acceptable to good reliability. For the CS measurement model, they range from 0.47-0.95, suggesting acceptable to high reliability. The slight difference may be because the CI items are binary, while the CS items have a more appealing 1-10 scale. In any case, our methodology allows capturing the inherent error from the CI questions, by considering multiple indicators for an attribute-specific CI. We believe this to be a positive contribution to the PT satisfaction literature. For the enthusiastic reader, we include results about the threshold parameters for all CI items, some CS items and the Loyalty item (E38) on Table 3-9, Appendix E (see Section 3.3.2 and Rosseel (2012) for interpretation).

Next, we present two tables showing the results obtained from two CS latent construct regressions, CS Safety and CS Cleanliness on Board (Table 3-4), and for Overall Satisfaction and loyalty regressions (Table 3-5). The rest of the CS constructs regressions are presented in Appendix F (Tables 3-10 to 3-12). Nevertheless, the discussion below considers all regressions.

Most categories of regressors produced significant results for all the CS latent constructs, demonstrating the importance of correcting for heterogeneity. The values of the R^2 statistics, the explained variance, vary between 0.64-0.91, implying moderate-high to excellent fits for each CS latent construct regression. We can also note that in some service quality respects (see Table 3-4 and tables 3-10 to 3-12 in Appendix B) some variables concerning users' socioeconomic characteristics and travel conditions are not present.

Table 3-4: SEM-MIMIC ordinal: Satisfaction latent constructs regressions (CS1-CS2)

CS1 SAFETY	**(R^2 = 0.722)**	Estimate	S.E.	Z-value	Std.Coeff.
CRITICAL INCIDENT	CI1 CI Safety	-1.086	0.012	-89.36	**-0.829**
SURVEY DATE	SPRING.2011	0.079	0.010	7.90	0.037
	AUTUMN.2011	-0.121	0.010	-12.75	-0.057
	AUTUMN.2013	-0.084	0.011	-7.45	-0.030
	SPRING.2014	-0.122	0.012	-9.90	-0.048
TIME OF DAY	AM.PEAK	0.030	0.008	3.95	0.014
ACCESS MODE	ACCESS.CAR	-0.051	0.007	-7.09	-0.027
	ACCESS.CYCLE	-0.054	0.011	-4.94	-0.021
	ACCESS.PT	-0.039	0.007	-5.72	-0.022
LINE	NORTHERN.REGION	-0.066	0.007	-9.54	-0.033
	WESTERN.REGION	-0.132	0.011	-11.64	-0.040
	SOUTHERN.REGION	-0.082	0.008	-9.71	-0.033
TICKET TYPE	TICKET.ONE-WAY	0.133	0.008	17.66	0.067
GENDER	FEMALE	-0.196	0.006	-35.70	-0.121
AGE	AGE>65	0.051	0.015	3.54	0.012
MONTHLY NET INCOME	1001.1500	0.025	0.008	3.03	0.013
	1501.3000	0.071	0.009	7.86	0.032
	>3000	0.070	0.014	5.09	0.017
EDUCATION LEVEL	UNIVERSITY.DEGREE	0.027	0.007	4.19	0.015

CS2 CLEANLINESS-BOARD	**(R^2 = 0.828)**	Estimate	S.E.	Z-value	Std.Coeff.
CRITICAL INCIDENT	CI23 CI Cleanliness	-1.105	0.004	-282.86	**-0.879**
SURVEY DATE	SPRING.2011	-0.106	0.013	-8.40	-0.041
	AUTUMN.2011	-0.348	0.011	-31.52	-0.135
	SPRING.2012	-0.108	0.012	-9.19	-0.042
	SPRING.2013	-0.188	0.013	-14.72	-0.060
	AUTUMN.2013	-0.186	0.013	-14.01	-0.054
	SPRING.2014	-0.174	0.014	-12.09	-0.057
TIME OF DAY	PM.PEAK	-0.072	0.009	-8.41	-0.028
ACCESS MODE	ACCESS.CAR	-0.027	0.008	-3.16	-0.012
	ACCESS.CYCLE	-0.055	0.013	-4.13	-0.018
SERVICE	SUBURBAN	0.144	0.007	21.25	0.072
LINE	NORTHERN.REGION	-0.209	0.008	-26.00	-0.086
	WESTERN.REGION	-0.326	0.014	-23.78	-0.082
	SOUTHERN.REGION	-0.179	0.010	-17.97	-0.060
TICKET TYPE	TICKET.ONE-WAY	0.196	0.009	22.20	0.082
TYPE OF USER	COMMUTER.STUDENT	-0.175	0.010	-18.40	-0.079
GENDER	FEMALE	-0.208	0.006	-32.10	-0.106
AGE	AGE>65	0.080	0.017	4.69	0.015
MONTHLY NET INCOME	<1000	-0.047	0.011	-4.40	-0.018
	1001.1500	-0.087	0.010	-8.79	-0.036
	1501.3000	-0.070	0.011	-6.62	-0.027
EDUCATION LEVEL	ELEM.SCHOOL.DEGREE	0.048	0.010	4.92	0.017
	UNIVERSITY.DEGREE	-0.035	0.008	-4.54	-0.016

For example, variables regarding the Type of User were not significant for the service aspect concerning Safety and Income was not significant for the service aspect linked to Cleanliness at Station. Furthermore, it is important to note the values of the standardised coefficients. Values near 0.1 are considered weak, values near 0.30 are considered moderate, and values at or above 0.50 are considered large (Currie and Delbosc, 2017). By observing the coefficients of the variables affecting the satisfaction with the eight attribute-specific service aspects (CS latent variables), we can state that CI is the variable having the most critical weight on the CS constructs. In fact, the standardised coefficients of the CI latent variables are consistently higher than 0.8 (see Table 3-4, and tables 3-10 to 3-12 in Appendix B). This result indicates that the CI are not only relevant for users but also the most critical aspects regarding satisfaction.

The above results are highly encouraging since our research hypothesis stated that the attribute-specific CI would significantly and negatively affect the attribute-specific CS constructs, and this effect holds for all CS constructs regressions. CI is the most critical variable in every case.

On the contrary, the standardised coefficients of the variables concerning passengers' socioeconomic characteristics and travel conditions present values that are in most cases lower than 0.1, and sometimes only slightly higher than 0.1, indicating in both cases that these characteristics have weak effects. However, they are significant and as such, if not included may induce in bias on the other estimates. As these variables give a contribution to satisfaction and correct for heterogeneity, it is important to consider them, especially when considering PT policy, because they aid in the search for differences among users' perceptions, findings that can be useful for PT operators.

In Table 3-5 the value of the R^2 statistic, for the OVS latent constructs shows a good fit, demonstrating a high amount of estimated variance (0.78). On the other hand, the value of R^2 relating to Loyalty shows only a moderate fit (0.41), meaning that the explained variance is not as considerable. Next, we comment on the parameter estimates.

Table 3-5: SEM-MIMIC ordinal: Overall Satisfaction and loyalty regressions (OVS-E38)

OVERALL SATISFACTION	**(R^2 = 0.778)**	Estimate	S.E.	Z-value	Std.Coeff.
SATISF. ATTRIBUTE	CS2 Cleanliness on Board	0.143	0.002	78.74	**0.167**
	CS4 Comfort	0.154	0.003	49.87	**0.138**
	CS5 Reliability and Accessibility	0.357	0.003	132.33	**0.319**
	CS8 Personnel	0.079	0.002	37.38	**0.084**
	CS9 Added-Value Services	0.576	0.006	88.83	**0.321**
SURVEY DATE	AUTUMN.2011	-0.096	0.006	-16.29	-0.044
	AUTUMN.2013	-0.060	0.007	-8.58	-0.020
	SPRING.2014	0.035	0.008	4.49	0.013
TIME OF DAY	PM.PEAK	-0.038	0.005	-8.25	-0.017
ACCESS MODE	ACCESS.CAR	-0.020	0.004	-4.48	-0.010
SERVICE	SUBURBAN	0.031	0.004	6.95	0.018
LINE	NORTHERN.REGION	-0.046	0.004	-10.45	-0.022
	WESTERN.REGION	-0.090	0.007	-12.69	-0.027
	SOUTHERN.REGION	-0.050	0.005	-9.62	-0.020
TICKET TYPE	TICKET.ONE-WAY	0.072	0.005	15.04	0.035
TYPE OF USER	COMMUTER.STUDENT	-0.106	0.005	-20.80	-0.056
GENDER	FEMALE	-0.030	0.004	-8.59	-0.018
AGE	AGE>65	0.102	0.009	11.75	0.022
MONTHLY NET INCOME	1001.1500	0.027	0.007	3.96	0.013
	1501.3000	0.067	0.006	11.93	0.030
	>3000	0.087	0.009	9.52	0.021
EDUCATION LEVEL	ELEM.SCHOOL.DEGREE	0.035	0.005	6.60	0.014
	UNIVERSITY.DEGREE	-0.031	0.005	-5.74	-0.017

E38 LOYALTY	**(R^2 = 0.408)**	Estimate	S.E.	Z-value	Std.Coeff.
OVERALL SATISFACTION	OVS Overall Satisfaction	0.745	0.004	190.15	**0.601**
SURVEY DATE	AUTUMN.2011	-0.065	0.013	-5.14	-0.024
	SPRING.2012	-0.090	0.013	-6.69	-0.033
	AUTUMN.2013	-0.103	0.014	-7.11	-0.028
	SPRING.2014	-0.246	0.017	-14.84	-0.076
TIME OF DAY	PM.PEAK	-0.050	0.010	-5.16	-0.019
ACCESS MODE	ACCESS.CAR	-0.046	0.010	-4.74	-0.019
SERVICE	SUBURBAN	0.093	0.008	12.03	0.044
LINE	NORTHERN.REGION	-0.113	0.009	-12.08	-0.044
	WESTERN.REGION	-0.110	0.015	-7.25	-0.026
	SOUTHERN.REGION	-0.128	0.011	-11.13	-0.040
TYPE OF USER	COMMUTER.STUDENT	-0.125	0.011	-11.42	-0.053
AGE	AGE>65	0.129	0.022	5.87	0.023
MONTHLY NET INCOME	<1000	0.069	0.013	5.53	0.024
	1001.1500	0.117	0.011	10.31	0.046
	1501.3000	0.133	0.012	10.70	0.047
	>3000	0.096	0.020	4.84	0.019
EDUCATION LEVEL	ELEM.SCHOOL.DEGREE	0.039	0.012	3.33	0.013

For the structural model, we included only five of nine CS latent constructs. These variables describe the service aspects linked to Cleanliness on Board, Comfort, Reliability and Accessibility, Personnel, and Added-Value Services. Two of the four variables left out presented negative signs contradicting satisfaction theory, and all were not significant at the 85% level. Thus, we decided to leave them out of the final models. Note the positive signs of the coefficients of the CS latent constructs on the Overall Satisfaction regressions.

Concerning the weight (standardised coefficient) of the CS constructs on Overall Satisfaction, we can say that Reliability and Accessibility (CS5) and Added-Value Services (CS9) have a moderate effect (0.319 and 0.321, respectively); nevertheless, compared to the other attributes these two have the highest impact. Next, Cleanliness on Board (CS2) and Comfort (CS4) have an effect that can be classified as intermediate, between low and moderate (0.167 and 0.138, respectively). Finally, Personnel (CS8) has a weak effect (0.084). Considering that Overall Satisfaction is a weighted sum of all the attribute-specific satisfaction constructs, it is not unexpected that no single component dominates (i.e., has a high standardised coefficient). Hence, the results are intuitive and appealing, since they suggest two high importance attributes, two of average importance, and one of low importance, but important still. For a final assessment, the R^2 statistic is 0.78, suggesting a good fit. We may argue that the CS constructs provide most of the explained variance; the model still does not explain 22% of it. Nevertheless, it provides a high explained variance.

Additionally, service attributes mostly influencing Overall Satisfaction (OVS) belong to the service aspect concerning Added-Value Services, which includes specific items such as information under abnormal conditions, the purchasing ticket mode, and specific operator interventions to improve the service.

The other service aspect considered of high importance to passengers is Reliability and Accessibility, which includes timetable (concerning frequency and daily distribution), punctuality, and regularity.

In turn, Overall Satisfaction has a significant effect on Loyalty; the standardised coefficient is 0.60. This result agrees with Satisfaction theory, which states that a satisfied customer is more likely to recommend the service to his friends and relatives (Oliver, 2010). The value of the R^2 statistic for Loyalty is 0.41; this indicates a low to moderate fit (see Table 3-5). The reason for this may be because only one item was collected for Loyalty in the survey: Intent to recommend. We think that if other items like Intent to reuse and Intent to use more frequently, or others, were used alongside Intent to recommend, then the Loyalty construct could have been better measured and would provide a higher level of explained variance. Nevertheless, as it stands, we can attest that Overall Satisfaction dramatically enhances Loyalty behaviour, specifically, the intent to recommend.

3.6 Discussion of Results

First, we will further discuss the Measurement Model and the structural relations of the SEM-MIMIC ordinal results. Next, we will detail the results obtained from the MIMIC regressions concerning the travel and socioeconomic characteristics. Finally, further analyses will be provided by comparing our results with similar previous studies.

3.6.1 SEM-MIMIC measurement and structural models

We found that the occurrence of an attribute-specific CI causes a decrease in CS with that element, which in turn causes a reduction in OVS. Subsequently, OVS influences Loyalty. That is, an increase in passengers' overall service satisfaction leads to an increase in the probability of the passenger recommending the service to friends and relatives. Concerning the relationship between attribute-specific satisfaction and overall service satisfaction, we can state that passengers retain as most important the service characteristics that are fundamental for its existence (i.e. Reliability and Accessibility) and also Added-Value Services. Aspects that would improve the quality of service but are not necessary for its existence are Cleanliness, Comfort, and qualities of the Personnel.

Each latent construct is explained by several items or observed variables. By analysing how the CI latent variables are defined by the indicators representing CI occurrences, and also how the attribute-specific CS constructs are formed, we can conclude that in general, all observed items explain the corresponding latent variables representing the particular CI and CS variables in a similar way. In other words, the CI latent constructs and the CS latent constructs have a similar item-to-item composition. This result is expected, as one would suspect items with identical description to have similar weights, even though the concepts of CI and CS are not entirely the same.

Still, there are some differences. For example, the latent variable CI Cleanliness is mostly explained by the occurrence of CI concerning the cleanliness of the carriage and seats, and less by CI involving cleanliness of the toilets. This result could indicate that maybe only a small fraction of passengers use toilets on the trains and, for this reason, there is a low probability of an incident. Another example emerges from the relationship regarding the latent variable describing CI Comfort, which is mostly explained by the occurrence of a CI in the workings of windows and doors rather than one regarding temperature on board. We assume that a critical incident associated with the opening of a door or a window can be more irritating (and maybe also dangerous) than a malfunction of the air-conditioning system.

By observing the latent variable of CI Additional Services, we can state that a critical incident associated with the facilities for people with restricted mobility (PRM) is most important than another concerning parking at the departure station. Both service factors concern a restricted portion of people, PRM and people who access the station with their private car. Also, an incident involving a PRM is more ethically inadmissible than one concerning parking a vehicle. Finally, the CI Personnel has more relevance if it concerns their courtesy on board than their function of checking tickets on board. Yet again, this result may stem from the fact that CI regarding staff behaviour is more salient than the specific ticket-checking task. These results derive from comparing the standardised coefficients reported in Table 3-3.

Now, we briefly comment on the relationship between the attribute-specific satisfaction latent constructs, and the indicators explaining them. As in the CI case, the CS Cleanliness on Board is mostly defined by the cleanliness of the carriage and seats, instead of the cleanliness of the toilets. Similarly, CS Comfort is explained chiefly by the satisfaction with the working of windows and doors rather than with satisfaction with temperature on board or overcrowding. For the CS Reliability and Accessibility, as expected, fundamental service characteristics such as train punctuality and regularity are considered more important than the distribution of stations in the region. Another similar result, confirming the one regarding CI, is given by CS Personnel. In this case, CS Personnel is mostly explained by the courtesy rather than the ticket-checking function of the staff. Interventions of the operator for improving the service are the item primarily influencing CS Added-Value Services. Finally, the latent variable representing Overall Satisfaction (OVS) is mainly explained by the indicators regarding satisfaction with the regional lines and with suburban lines. This result is intuitive as the other two items are indirect measures of overall satisfaction.

The CS constructs regressions allow to understand which variables affect mostly each latent satisfaction construct. For the eight latent constructs representing the main aspects of the service, we can conclude that the variables relating to CI have the highest coefficient values. They are higher and more significant than all the remaining coefficients; in most cases, the standardised coefficient is approximately 0.8. In other words, a critical incident has a substantial influence on CS; this is the most critical and compelling result of our research. It is fundamental in identifying policy recommendations to PT operators and administrators. Intuitively, minimising CI in the most relevant service attributes ought to be the strategy to prioritise.

3.6.2 SEM-MIMIC socioeconomic and travel characteristics discussion

We also investigated how the socioeconomic characteristics of passengers and their travel habits influence the CS constructs. Although we found interesting relationships, these traits have a minor influence compared to CI.

Concerning gender and age, for example, we found similar evidence for all the satisfaction latent constructs. More specifically, being female causes a decrease in satisfaction with the eight service aspects. Being young, as opposed to older than 65, also causes a decrease in satisfaction. The same analysis can be made for Overall Satisfaction and Loyalty (see Table 3-5). We observed negative values for the variable Female and positive values for Age>65 in both cases, implying that men and people older than 65 are more satisfied with the Overall Service and more inclined to recommend it to their peers. We hypothesise that women may have higher expectations than men when travelling by PT; these could be due to having more requirements concerning attributes such as security, comfort, cleanliness, among others. On the other hand, we also hypothesise that older users may be more in need of using PT than younger people, because of less availability of driving, and therefore they could be more inclined to travel by train, hence more satisfied than bus users.

Next, we analysed the effect of educational level. Again, it has a similar impact on the satisfaction with the various service aspects and with overall satisfaction and loyalty. In particular, people having a low educational level (such as Elementary School degree) are more satisfied with the service than people having a High School degree. Furthermore, people having a University degree are less satisfied than people having a High School degree. This result suggests that PT satisfaction decreases with the educational level. This fact is supported by the coefficient values of the variables Elementary.School.Degree (with positive values) and University.Degree (with negative values). We hypothesise that higher educated users are more demanding, and require a higher standard for the PT level of service.

On the contrary, for the variable Net Monthly Income, we can observe different results among the eight service aspects. As an example, income has a complicated relationship with CS Safety. Satisfaction increases if income increases until 3,000 euros, but for people with higher income (more than 3,000 euros) we register a decrease in satisfaction. A similar trend can be observed for the service characteristics: CS Reliability and Accessibility, and CS Personnel. For CS Cleanliness on Board, we found the opposite situation; that is, satisfaction decreases if income increases. The same result is present in CS Information. In other words, wealthier people are more demanding regarding having clean trains and a reasonable level of information when they travel. Conversely, when assessing Overall Satisfaction, it increases if income increases to 3,000 euros (i.e., poor people are less satisfied with the service). However, the satisfaction of higher income passengers (more than 3,000 euros) decreased, and the same tendency is observed for Loyalty (see Table 3-5). This result indicates a non-linear and diminishing effect of income on Overall Satisfaction and Loyalty. One hypothesis is that low-income people have to make a substantial effort to use the service. Hence, they face a relatively higher price (train is more expensive) than middle-income users. Meanwhile, the higher income (more than 3000 euros) passengers may have higher standards than the rest, and for this reason, have a decrease in overall satisfaction and intention to recommend the service. Still, more insight is warranted to determine the causes of differences in perceptions between the different income groups.

Among the passengers' travel habits, ticket type presents a similar behaviour for all satisfaction constructs and Overall Satisfaction. A positive coefficient sign is present for passengers using the one-way ticket, meaning that their satisfaction is higher than for passengers using passes. Thus, habitual users are less satisfied than occasional users; this might be related to being captive train users. This result especially concerns the satisfaction with the CS Comfort attribute, to which habitual users are probably more sensitive as they tend to spend more time on the trains.

Another travel habit that shows a similar behaviour for the CS constructs and Overall Satisfaction is the access mode to reach the train. By analysing the coefficient of the dummy variables Access.Mode, we can conclude that accessing by car, or bike/motorcycle (or other public transport modes) to train, as opposed to accessing by foot implies a reduction in satisfaction. This outcome means that organising the journey by combining more modes with the train is less satisfying, suggesting that walking access is less complicated, or hassle-free.

Analysing the travel conditions, we offer interesting findings. For example, concerning the time of day when users travel, we see that travelling during the morning peak implies an increase in satisfaction compared to travelling during off-peak hours. Instead, travelling during the afternoon-peak and evening-peak implies a decrease in satisfaction. This result could be caused by the fact that people travelling in the last part of the day are tired and thus, tend to be less satisfied with the service. Meanwhile, people travelling in the morning are more favourable; a similar consideration can be made regarding Loyalty. Being a commuter for studying purposes implies a decrease in Overall Satisfaction and Loyalty compared to being a user travelling for other reasons.

Passengers are more satisfied when travelling on suburban trains than regional trains. By observing the coefficients linked to the variable Service, we conclude that the suburban service is better than the regional one. The reason for this may be that people travelling by suburban trains spend less time on board than regional passengers, and also because regional lines are characterised by less frequent services without a schedule and not many stops. This result is confirmed by the behaviour of the variable Service regarding Overall Satisfaction and Loyalty, a decrease in both compared to travelling by suburban lines. Finally, concerning the survey date, we can state that satisfaction with the various service aspects decreases in general when comparing all surveys to the base survey done in the autumn of 2012, this may be fortuitous, but it could bring insight to the administrator. Also, satisfaction with the Overall Service receives a positive contribution from the variable associated with the survey date in the case of Spring 2014, which is the last survey analysed.

3.6.3 Comparison with previous SEM-MIMIC satisfaction models for PT

Allen et al. (2018), who also proposed an SEM-MIMIC approach to investigate PT service quality, found that reliability and frequency-waiting time satisfaction were the most relevant variables for bus satisfaction. Their result is in-line with ours; in fact, we found that specific service characteristics, such as reliability concerning frequency and punctuality, were among the most important for users. The same finding was obtained by Guirao et al. (2016), who also proposed an SEM-MIMIC approach. Thus, we can positively conclude that independent from the type of transport mode (bus or train), the fundamental attributes for users are related to the service reliability.

Another similar result with the work of Allen et al. (2018), is that the service aspect linked to bus stops (in our case railway stations) was not significant in the model. Guirao et al. (2016) found that this attribute was not in the top of the list either. Nonetheless, rail passengers spend more time in stations than bus passengers at bus stops. We suspect that in our case, this aspect was not relevant for users because the services analysed are used similarly than bus services, this is, for commuting purposes. Therefore, passengers spend a short time at the station, as they are avid users, similar to bus passengers.

Concerning passengers' characteristics, while the effect of gender on satisfaction was the same as observed by Allen et al. (2018), they found that older people tended to get uncomfortable more easily, and we found the contrary. This difference could be due to the difference in the characteristics of the transport modes analysed. In other words, older people tend to be more satisfied when travelling by train than by bus. Regarding passengers' travel conditions, we found that travelling in peak hours in the morning produced more satisfaction than during off-peak hours and that travelling in the afternoon and evening peak hours produced less satisfaction, while Allen et al. (2018) found that both AM, and PM peaks delivered less satisfaction than the off-peak. Travelling by train in the most crowded hours of the morning is more tolerable than travelling by bus. We hypothesise that this is related to the buses sharing space with cars, while the trains have separated corridors. Finally, we found similar results to those of Allen et al. (2018) concerning the type of user. Specifically, we found that commuter students are less satisfied than people travelling for other purposes.

3.7 Conclusions and Future Research

Our most relevant result is that, effectively, CI is highly relevant for users when appraising their satisfaction with service-specific (domain) attributes. In other words, we show that for every attribute-specific latent construct the CI latent variable produced a substantial negative effect on the CS constructs. Thus, CI is crucial to policy-making for PT services. Since the satisfaction constructs are more subjective, less tangible, and less controllable by the administration, we contend that the elimination of CI on specific service attributes may be the most logical policy that an administration can undertake in their planning and operational strategies.

We conducted a four-step methodology, where we estimated (i) PCA, (ii) CFA, (iii) SEM numeric and ordinal, and finally (iv) SEM-MIMIC ordinal Probit models. First, this sequence allows to correctly determine how the CI and CS are formed, and subsequently allowed to test the structural relations between these latent constructs, and Overall Satisfaction and Loyalty. In other words, our method permits to confirm which are the essential items comprising CI and CS constructs, and this would not have been possible unless an SEM framework was utilised. This is a crucial result for policymaking. Furthermore, the CI is measured allowing correction for measurement error. This is relevant since users may inherently introduce errors when recalling CI for some attributes. If a standard regression procedure had been used, a less adequate model would have been obtained, as no such correction would have been possible. Again, this methodological contribution is not present elsewhere in the literature.

We estimated a full SEM-MIMIC ordinal Probit model, which allows distinguishing between differences in perceptions amongst different types of travellers, different travel habits, and different socioeconomic characteristics. This is key towards achieving accurate results since not accounting for heterogeneity would have resulted in biased estimates. We measure this difference in the perception of Loyalty behaviour. We did not find previous works that measured Loyalty using an SEM-MIMIC framework in the PT Satisfaction literature. Additionally, we also captured heterogeneity for the CS constructs and Overall Satisfaction. Consequently, we were able to compare our results with those of two previous studies that used this framework providing critical policy-related insight. We conclude that Reliability and Accessibility are a critical component when measuring Overall Satisfaction.

Finally, we compared two different SEM models: a numeric and an ordinal one. With these results, we were able to show that the SEM ordinal model produces more significant and convincing results, tested via the explained variance and full model fit indices; shedding light on the biases of the numerical framework. This particular result contributes to the PT literature by stating that the ordinal framework is more accurate than the numerical counterpart. In this case, a model that has more parameters provides better fit indices than the simple numeric SEM.

For future research, we recommend using the SEM ordinal framework and, if possible by having a large enough sample size, to employ the SEM-MIMIC approach also. In our case study, the SEM-MIMIC ordinal model delivered an almost identical fit than the SEM ordinal; however, it improved regarding the RMSEA index, proving to have a better fit. Furthermore, it offers more information regarding differences in perceptions which can be useful policy-making information for PT operators and administrators. Again, we insist that not taking account this heterogeneity, may lead to biased results. As such, a more straightforward model may lead to incorrect policy-related conclusions.

Our final recommendation is to include the CI items in future Customer Satisfaction Surveys since this element can provide a more tangible and policy-related variable to PT operators and administrators. We showed that all the CI are highly significant in the regression results, implying that by reducing them we will gain in the attribute-specific satisfaction constructs, improving overall satisfaction. Also, it allows building a more complex model structure that suggests a direct and causal relationship between the CI, a more policy-related variable, and the CS constructs; without significantly extending the size and time for performing the surveys.

4. THE ROLE OF CRITICAL INCIDENTS AND INVOLVEMENT IN TRANSIT SATISFACTION AND LOYALTY

4.1. Introduction

Service quality is an abstract concept, difficult to define and often interchangeably used with satisfaction (Lien and Yu, 2001, Lai and Chen, 2011, Sumaedi et al., 2012). However, the differences between both variables have been clarified in the literature (de Oña and de Oña, 2015). As an example, Berry et al. (1990) point out that … "*customers are the sole judges of service quality*". Therefore, if service quality is measured from the customer's perspective, transit quality depends on the passengers' perceptions about each attribute characterising the service (de Oña et al., 2013). In itself, customer satisfaction is commonly used as a measure of service quality.

Past investigations have suggested that satisfaction is an excellent predictor of repurchase intentions (Petrick, 2004). Therefore, we can consider that satisfaction is a predictor of transit passengers' intentions to reuse the service in the future. More recently, the concepts of service quality and customer satisfaction have been analysed together with the concept of behavioural intentions. According to Zeithaml et al. (1996), behavioural intentions are signals showing whether a customer is willing to continue using a company's service or switch to a different provider. When we talk about customer behavioural intentions, we are talking about customer loyalty, since a loyal customer will continue to use the service. Loyal customers are more likely to reuse the services and also to recommend them to potential new users (Imaz, 2015). Although many studies attempt to identify the factors affecting customer attraction in other fields, transit user loyalty has been far less investigated in the literature.

We use a Structural Equation Modelling (SEM) approach with the objective to investigate the relationship of specific transit service quality attributes and customer satisfaction on loyalty. Understanding passengers' behavioural intentions after experiencing transit services is an essential task for transit managers. Doing so, they can define effective strategies to meet passengers' needs, retain existing passengers and also attract new ones (Lai and Chen, 2011), expressly by enhancing the loyalty component.

We introduce three distinct concepts that allow enriching the satisfaction models. Firstly, a Loyalty construct considered as the intention to recommend the service. We hypothesise that Overall Satisfaction does not only directly affect Loyalty, but that Loyalty is also affected by other attribute-specific satisfaction constructs. Unveiling which specific satisfaction constructs affect Loyalty directly is an essential finding for PT policy development. Secondly, we introduce Critical Incidents (CI) regarding the closure of a line in the last three months or an anomaly in the service in the past month. We hypothesise that this construct negatively impacts all the attribute-specific satisfaction constructs. Lastly, we introduce Involvement, as the intent to participate in future PT marketing studies. We hypothesise that both Overall Satisfaction and Loyalty may affect this variable. The three independent results contribute to the literature, and also embody a key finding for PT policy design.

We have available four satisfaction surveys applied to samples from the same population, each with an approximate sample size of 2,500. In essence, the surveys are the same, except that the attribute-specific satisfaction items, which total 26, are divided into the four samples (i.e. each survey only has six or seven of these items). Thanks to this peculiarity, we can assess which type of attributes are more relevant for users, from each separate survey. We can also verify which model better explains users' Overall Satisfaction with the PT system. Thus, we gain the possibility to assess which are the best attribute-specific items to include in future surveys. We believe this is an essential contribution to the literature, not presented elsewhere.

We will additionally conduct an SEM Multi-Group (SEM-MG) analysis using the survey dataset that yielded the best model. Our objective with this analysis is to determine the potential heterogeneity in the satisfaction model when considering different travel characteristics of the users and demographic characteristics. Explicitly, we will assess if gender, age, nationality, time of day, travel frequency, and trip purpose affect the resulting satisfaction model. One advantage of our dataset is that we have a relatively large sample (2,500) for each of the four cases. Consequently, this allows us to obtain different models from the SEM-MG analysis. We consider our findings may be essential for PT policy development since we found significantly different models for different user types. As we uncover these differences, we give specific recommendations for policy-making.

The organization of the chapter is as follows. Section 4.2 provides a literature review of the studies investigating loyalty in PT. We also mention previous studies that have analysed CI in the PT Satisfaction literature and review the Involvement concept and its use in this literature. Section 4.3 provides a succinct methodological review of both SEM models and SEM-MG analysis. Section 4.4 deals with the description of the case study, and specifically the depiction of the transit service, the summary of the survey and the sample characteristics. Sections 4.5 and 4.6 present the results and discussion, respectively, and Section 4.7 our main conclusions.

4.2. Literature Review

Customer loyalty has been extensively analysed in the field of marketing and other related industries such as air transportation services. Loyal customers can indirectly attract new customers, generally through personal recommendations or even by making them follow social trends (Webb, 2010). These general benefits can be in most cases directly translated to the PT context.

In the last few decades, the PT literature has started to consider the analysis of customer loyalty. However, there is no consensus yet about how loyalty should be measured. Many authors claim that loyalty should be based only on intended future usage and willingness to recommend (Lai and Chen, 2011, Minser and Webb, 2010, Zhao et al., 2014). Lai and Chen (2011) suggest that intention to use is often considered a proxy for actual future usage, and according to Reichheld (2004), users who are willing to recommend the service to others are commonly loyal customers. A fundamental assumption reported in the literature is that users will continue to use PT and recommend it to their peers only if they are satisfied with the service performance (van Lierop and El-Geneidy, 2016). For example, captive users of PT will not necessarily recommend its use to friends and family.

Several researchers have analysed the relationship between concepts such as service quality, customer satisfaction, and behavioural intentions, by adopting structural equation modelling. Some authors talk about loyalty, while others about behavioural intentions. We consider that both elements represent the same concept. Park et al. (2004) explored the relationship among service expectation, perceived performance, perceived value, satisfaction, and behavioural intentions. Chou and Kim (2009) adopted structural equation modelling to measure the impact of service quality, corporate image, satisfaction, and customer complaints on loyalty. They assessed loyalty through repeat business, willingness to recommend, and price tolerance. Webb (2010) applied structural equation models to obtain a better understanding of customer loyalty in the PT context and found that service quality, service value and customer satisfaction affected behavioural loyalty (i.e. the willingness to continue using transit in the future). Minser and Webb (2010) defined loyalty based on likeliness to continue using the service and willingness to recommend. They found that service quality, service value, and customer satisfaction directly influenced loyalty, as well as that problems and agency image indirectly affected it.

Lai and Chen (2011) explored the roles of service quality, perceived value, satisfaction, and involvement in behavioural intentions of public transit passengers. Chen (2016) measured intentions to repurchase and to recommend to others, through questioning “the possibility to ride again”. Van Lierop and El-Geneidy (2016) developed a series of structural equation models that reflected the different groups using transit. They classified users into captive riders (dependent on transit), choice riders (car owners who choose transit), and captive-by-choice riders (they rely on transit but could own a car). They defined a loyal customer based on his/her Overall Satisfaction with the service, likeliness to continue using the service in the future and likeliness to recommend it to others. De Oña et al. (2016) hypothesised that transit users’ behavioural intentions depended on several constructs. They considered that the user’s levels of satisfaction with the service, perceptions about the costs and benefits from using transit service, their judgments about the different characteristics of the service, their opinions about alternative transport modes, and about the transit system, in general, affected their behavioural intentions.

The studies mentioned above coincide by considering that general constructs such as service quality and customer satisfaction influence loyalty or behavioural intentions of transit passengers. In this study, we also decided to investigate the effects of customer satisfaction on loyalty. However, we have an additional and more specific objective, which is to demonstrate that not only passengers’ satisfaction with the overall service directly affects loyalty, but that there are direct effects of specific service factors also affecting loyalty. We pose that it is fundamental to understand which service factors mostly affect loyalty, to identify the best strategies that agencies have to undertake for customising the service for the passengers and, at the same time, attract new ones. We believe that that the factors affecting customers’ satisfaction the most are not necessarily the same that affect customers’ loyalty or have the same magnitude. As an example, as stated in Imaz (2015), a significant but occasional delay during a subway ride is unlikely to drive a loyal customer away from transit, but it may make him less satisfied. Understanding the factors that affect user loyalty may be crucial for transit agencies as they can use that information to incorporate operational and marketing instruments to maximise customer retention (Imaz, 2015).

Few researchers have investigated our objective. As an example, Figler et al. (2011) found that service reliability and safety, both while waiting for and riding the bus, were good predictors of bus loyalty. The study proposed by Imaz (2015) used data provided by a mixed Stated Preference/Revealed Preference survey to identify the factors affecting customer loyalty in the context of PT. They found that service quality attributes such as vehicle crowding, and travel time played a critical role in transit user loyalty, while initiatives such as the provision of real-time information panels or availability of park and ride facilities had a less determinant effect on the customers' mode shifting decisions.

Concerning the introduction of CI in the analysis of transit service quality and customer satisfaction, there are very few studies in the area of PT service quality that introduce this concept. Friman et al. (1998, 2001) analysed negative CI regarding a PT service based on the most frequent complaints. They obtained the negative CI from the archival customer-complaint data. Allen et al. (2018) implemented a framework including CI and introduced attribute-specific CI to explain attribute-specific satisfaction levels. Their respondents were asked to give information about negative CI, in the last thirty days and to state whether they experienced the CI or not concerning a list of service-specific attributes. They demonstrate that CI strongly and negatively influence users' perceptions.

Research about the concept of Involvement is scarce. Authors such as Lai and Chen (2011) have shown that involvement influence behavioural intention. They measured involvement by the item "I always pay attention to the information about public transit". Machado et al. (2016) found that the involvement of transit users could positively affect their evaluations of service quality and enhance their intentions to reuse the service and recommend it to others. Involvement was measured by collecting users' level of agreement with three statements. These are: "I feel that using transit is consistent with my lifestyle", "I feel that by using transit I help to protect the environment", "I like others to know the fact that I use transit" and "I like people who use transit". These correspond to an attitude of involvement towards public transport. However, in our study, we measure involvement as an intention and not as an attitude. Hence, we will test two separate hypotheses: (i) a satisfied customer presents more intentions of involvement towards the transit system, and (ii) a loyal customer presents more intentions towards involvement with the system.

4.3. Methodological Review

4.3.1. Structural equations modelling

SEM is a statistical analysis tool frequently used by social workers and researchers to develop multi-item measures with the objective of analysing complex phenomena including attitudes, cognitions, behavioural patterns, social experiences, and emotions (Bowen and Guo, 2012). SEM encompasses a set of multivariate statistical approaches to| treat empirical data, both conventional (e.g. analysis of variance and covariance, factor and path analysis) and more innovative approaches, such as methods for analysing latent classes cross-sectionally and over time (mixture modelling) and latent growth curve modelling. Often SEM is used as a data analysis method that combines simultaneous regression equations and factor analysis.

Factor analysis tests hypotheses about how well specific observed variables in a dataset measure latent constructs. These theoretical and abstract concepts or phenomena cannot be observed or measured directly or with single items (Bowen and Guo, 2012), due to possible measurement error. Factor analysis, also called measurement models, focus on how one or more latent constructs are measured, or represented, by a set of observed variables. The observed variables can be responses to questionnaire items and census figures, amongst others (Bollen, 1989). Latent variables with adequate statistical properties can then be used in cross-sectional and longitudinal regression analyses, for testing hypotheses about the strength and direction of relationships between predictor variables and outcome variables (structural model). Unlike standard regression models, SEM permits regression relationships among latent variables and between observed and latent variables. Therefore, in a single analysis, it can estimate models where one or more variables act as predicted and predictor variables simultaneously as is the case in our present study.

SEM has several advantages over other commonly used statistical methods. Among them are: (i) treating endogenous and exogenous variables as random variables with measurement error, and (ii) measuring latent variables with multiple indicators. SEM can also: (iii) test coefficients and overall model fit, (iv) test models where there are multiple dependent variables, and (v) handle non-normal data (Bowen and Guo, 2012, Golob, 2003). The observed variables are a function of model-specific latent constructs and latent measurement errors. As such, researchers can beneficially isolate real causes of scores from variations in scoring due to unrelated causes. The statistical tests of the relationships among the resulting latent variables are superior to tests among variables containing irrelevant variance (Bowen and Guo, 2012).

Principal Component Analysis (PCA) or Exploratory Factor Analysis (EFA), can aid to unveil how many dimensions of the phenomenon are represented by the measured items, and which items are associated with each latent construct (factors). PCA is a statistical procedure that uses rotation techniques to transform a set of possibly correlated variables into a set of linearly uncorrelated factors called principal components (Jolliffe, 2014). Once the modeller knows the factor structure, he can specify a Confirmatory Factor Analysis (CFA), and assess fit indices and parameter estimates of the proposed model. Establishing the measurement model adequacy before testing the structural relationships in SEM is best practice (Anderson and Gerbing, 1988, Bollen, 1989).

4.3.2. Measurement invariance and multi-group analysis

SEM offers the flexibility for analysing moderation effects when the moderator is a categorical variable representing subgroups of a sample. Testing moderating effects means that the modeller hypothesises that the impact of an exogenous variable (x) on the endogenous variable (y) varies with the level of another exogenous variable called the moderator (M). In our case study, we are interested in testing whether SEM models differ by travel and sociodemographic characteristics of the users. Thus, in our models, we will evaluate whether the effect of the Satisfaction Attributes (x) on the Overall Satisfaction (y) differs across different subgroups (M) in the population. We categorise the latter by gender, age, nationality, time of travel, ticket type, trip purpose and travel frequency.

We aim to find out if and how much the patterns of effects of the Satisfaction Attributes on Overall Satisfaction differ by group. For this purpose, the Multi-Group Analysis (MGA) technique offers beneficial capabilities. The process involves testing out a series of SEM analyses for the groups (SEM-MG). Before doing this, Measurement Invariance (MI) needs to be examined. An MI analysis is performed to test "*…whether the items comprising a particular measuring instrument operate equivalently across different populations (e.g. gender, ability, cultural groups). In other words, is the measurement model group-invariant?*" (Byrne, 2009, p.197, cited by Bowen and Guo, 2012). Testing MI means testing the CFA model with several groups and assessing whether the factor structure and parameter estimates are statistically the same for the groups (Bowen and Guo, 2012). If not, the model is not invariant across groups. Thus, there are two options: if the model is invariant, we can assess and compare between groups with the same model. If not, different models are required for each group, and the latent constructs will be measured differently. Henceforth, in the latter case, no direct comparisons could be made about measuring the same construct across groups. However, if this is the case, it may well be that the different groups perceive the latent constructs differently. In such a heterogeneous population as users of PT, this may be the case, and as such may be modelled separately across groups.

Once we test for MI, we can pose additional research questions regarding implications for policy, research, and practice. Specifically, we may want to compare whether structural path coefficients are different among the groups. In our case study, we want to assess whether the relevance, or priorities, of specific Satisfaction Attributes, remain the same across groups, for different categorising variables. In this work, we consider this fundamental question for policy design and operation of PT. Since we do find different attributes being more relevant in different groups, we consider this a novel contribution to the PT Satisfaction literature.

Multi-group comparison involves a sequential assessment. First, a baseline model specifies identical form (that is same factor composition) for all groups, generating estimates for each group, and obtaining a χ^2, and other fit indices that apply to the entire multi-group model. Subsequently, the modeller constraints one part of the model at a time, for example, the structural paths to have equal coefficients across groups (Bowen and Guo, 2012). We expect that a model with additional constraints will have a worse fit. Henceforth, an analysis considering the worsening of fit indices will need to be addressed.

Both MI and MGA will be assessed using a modelling approach (Beaujean, 2014). As with the statistical method (using the χ^2 statistic), the assessment involves evaluating the increasingly constrained models against each other and determining whether the worsening of the model fit is significant within a numeric threshold. In the SEM literature, the most popular recommended fit to assess for measurement invariance is the difference (Δ) between nested models of the comparative fit index (CFI). When the ΔCFI for two models of different levels of measurement invariance is greater than 0.01, then invariance is probably untenable (Cheung and Rensvold, 2002). We will use this threshold for testing for MI, and we will use a threshold of 0.02 for testing for MGA.

In the next paragraphs, we will explain five of the types of MI, with their respective restrictions. There are degrees of invariance, usually described as a hierarchical sequence, from weakest to the strictest form (Beaujean, 2014).

i) *Configural invariance*: The most basic level indicates that the latent variables have the same factor structure in all groups. There is no evidence that the latent variables are measuring the same construct in each group. No direct comparison between groups can be made.

ii) *Weak invariance:* This level adds the restriction of the loadings being the same across groups. Latent variance across groups vary. It allows a comparison of the latent variance and covariances between groups.

iii) *Strong invariance*: This level adds the restriction of the intercepts to be the same for any given indicator. Thus, individuals at the same level of a given latent variable have the same expected value on the indicators. It allows for the comparison between groups of latent means, variances, and covariances.

iv) *Strict invariance*: This level adds the restriction of the error variances to be equal across groups. This form of invariance is usually tested in conjunction with the evaluation of invariance of the latent variances (v).

v) *Homogeneity of latent variable variances*: If there is homogeneity, this indicates that the groups used an equivalent range of the construct for the indicator variables' values. If it does not hold, it suggests that the group with the lesser amount of latent variance is using a smaller range of the construct than the group(s) with the higher amount.

In our case study, we will focus on the first four types as they deal with the aspects of MI. However, if there is a case where strict invariance does not hold, we will still test for MGA. As our objective is to assess for differences in the relevance, priority in preference of the Satisfaction Attributes, we will still be able to understand which attributes are more relevant to the users, even if the constructs are not measured identically. We will present those cases that do have a difference in the MGA. This result warrants that the groups in the population have different SEM models, showing differences in the preferences of specific-satisfaction attributes in the population.

We found previous work in the PT Satisfaction literature that followed the SEM-MG approach. Joewono and Kubota (2007) analysed one variable, being a student or not, with a sample size of 980, in the case of Paratransit service in Bandung, Indonesia. Their analysis concluded that both subpopulations had a similar model. Antonucci et al. (2012) analysed six variables: gender, age group, educational level, employment status, residence area, time of day, the frequency of use and reason to use. They used a sample of 3,000 users waiting at bus stops in Bari, Italy. They found that two variables resulted in different structural satisfaction models: residence area and frequency of use. Nonetheless, their study did not report the resulting SEM-MG models for the different subpopulations. Finally, de Oña et al. (2018) analysed one variable, year, with a sample size of 4,633, in the case of a Metropolitan bus transit system in Granada, Spain. First by running individual models, they found differences in the structural coefficients. However, when they ran the SEM-MG models, the structural coefficients proved to be statistically invariant. Our case study has the most similarities with the research by Antonucci et al. (2012). However, additionally, we will identify the resulting SEM-MG models, to assess which structural factors are different for the subpopulations. We consider this to be a contribution to the existing PT Satisfaction literature. For all computations, we will use the R statistical code, and specifically the Lavaan package for R (Rosseel, 2012) in all SEM models.

4.4. Case Study

Madrid's Metro system opened in 1919, with the launch of Line 1. At that moment, it consisted of 3.48 kilometres and eight stations, between "*Sol*" and "*Cuatro Caminos*". Since then, the Metro network has been continuously expanding. Currently, the Metro network has 12 lines with 294 kilometres and 301 stations, it is an interurban system that reaches several towns in the metropolitan area. Also, the Metro system is interconnected with the light rail system (3 lines), introduced in 2007, and with the suburban railways servicing short distance trips to and across the city.

Madrid's Metro system is the third most extensive Metro in Europe and the ninth in the world (Rohde, 2017). Because of its extension, about 80% of the inhabitants in the twelve municipalities reached by Metro have an available station less than 600 meters from their home or their usual destination. In average, the Metro system transports two million passengers daily, and in 2017 it accommodated 626 million passengers, an increase of 7.1% over the number in 2016 (Metro de Madrid, 2017b). This growing trend started in 2013, after some years of decrease in the demand for Metro (Metro de Madrid, 2016).

The company that operates the Metro service carries four campaigns of surveys per year to identify the users' level of satisfaction about several attributes of transit service quality and the level of importance that users give to each element. The surveys are conducted at stations using face-to-face interviews. The questionnaire contains five parts. The first collects the main aspects of the trip and the socio-demographic characteristics of the respondents. Part 2 registers user's perception of the service quality of the overall Metro system. In Part 3, respondents are asked to declare the levels of satisfaction and importance that they associate with specific attributes of the transit service. This part is directed to the characteristics of the particular line that the respondent is using while answering the questionnaire. Part 4 asks about inconveniences caused by any possible service interruptions. Finally, in Part 5, passengers express the intent to recommend or not the use of Metro to other people, the level of recommendation, and the willingness to take part in future customer satisfaction studies conducted by the company.

The service quality factors investigated in Part 2 are 26 and cover all aspects of the Metro service: operation, comfort, cleanliness and customer service, amongst others. However, the survey has a particular and unique structure. It includes four versions of the questionnaire that differ among them in Part 2. In each version, respondents express their perceptions about six (or seven) different service quality attributes. Consequently, the overall sample can be divided into four sub-samples that answer various questions about the quality of the Metro service. We will call these, Surveys A, B, C and D.

Our study was developed with the data collected through the fourth campaign of surveys conducted in 2015. The sample consists of 10,159 passengers, and each sub-sample counts approximately 2,500 interviews, as reported in Table 4-1. The surveys were addressed to passengers of all subway lines and those of line ML1 of the Light Rail (LRT) system. Most interviews were conducted at stations in two different time periods of the day: 9.30-13.00 and 18.00-21.00.

Table 4-1: Survey characteristics

	Category (Percentage)
Questionnaire	Survey A: 2569 (25.3%), Survey B: 2550 (25.1%), Survey C: 2527 (24.9%), Survey D: 2513 (24.7%)
Time period	7.00-9.30: 2206 (21.7%), 9.30-13.00: 2264 (22.3%), 13.00-16.30: 2112 (20.8%), 16.30-18.00: 839 (8.2%), 18.00-21.00: 2264 (22.3%), 21.00-22.00: 474 (4.7%)

We summarise the main characteristics of the sample in Table 4-2. Respondents were mainly female (59.6%) and aged between 16 and 24 (25.1%). About 84% of users are national residents, and about 13% are foreign residents. Few tourists were interviewed. The ticket type mostly used is the travel pass (71.4%). The most common trip purpose is work (57.7%), followed by study (20.8%) and leisure (9.7%). More than one-third of respondents declared to make about 4-10 trips per week (37.4%). About equally numerous are the passengers who make between 11 to 20 weekly trips (34.5%). Therefore, the typical respondent is a commuter habitually travelling for work or study with a high frequency of trips in a week.

Table 4-2: Sample characteristics

Characteristics	Category (Percentage)
Gender	Male: 4100 (40.4%), Female: 6059 (59.6%)
Age	16-24: 2550 (25.1%), 25-34: 2547 (25.1%), 35-44: 2248 (22.1%), 45-54: 1693 (16.7%), >55: 1121(11%)
Nationality	National resident: 8519 (83.8%), National tourist: 186 (1.8%), Foreigner resident: 1381 (13.6%), Foreigner tourist: 73 (0.7%)
Ticket type	One-way ticket: 625 (6.1%), Carnet: 2,272 (22.4%), Travel pass: 7250 (71.4%), Tourist ticket: 12 (0.1%)
Trip purpose	Work: 5860 (57.7%), Study: 2109 (20.8%), Leisure: 991 (9.7%), Medical visit: 394 (3.9%), Shopping: 147 (1.4%), Other: 658 (6.5%)
Travel frequency	Weekly trips <4: 1172 (11.5%), weekly trips 4-10: 3795 (37.4%), weekly trips 11-20: 3505 (34.5%), weekly trips >20: 1687 (16.6%)

Table 4-3 shows the users' satisfaction average rates and reports the service quality attributes evaluated in the Surveys A, B, C, and D. All respondents were asked to evaluate the overall Metro service on an 11-level scale, from 0 - the lowest level - to 10 - the highest level - (same as the academic grading system in Spain) and the average score results equal to 7.30. About 76% of the respondents evaluated the overall Metro service with scores higher than 7. Most of the sample (50.9%) declared that the service had remained the same in the last year, and 23.3% declared that the service had improved. On the other hand, 25.8% of the respondents gave a negative judgement declaring that the service has worsened. Almost all the respondents (97%) recommend the use of Metro of Madrid as a means of transport, and the average recommendation level is 8.19 measured on the same 11-level scale (0-10).

Table 4-3: Survey results

	Service quality attributes	Average satisfaction score
Survey A	Operation of automatic ticket vending machines (S01)	7.60
	Platform waiting time (S03)	6.40
	Personal security from assault and theft (S06)	7.30
	Attention to complaints and suggestions (S08)	6.58
	Temperature and ventilation system at stations (S13)	6.87
	Noise level on the vehicles (S20)	6.58
	Operation of elevators (S24)	6.94
Survey B	Operation of turnstiles (S02)	7.62
	Speed of train travel (S04)	7.44
	Kindness of the vigilantes (S07)	7.09
	Cleanliness at stations (S15)	7.40
	Cleanliness of vehicles (S16)	7.44
	Information about incidents of the service (S19)	7.17
	Onboard ticket validators (S26)	7.03
Survey C	Escalator operation (S09)	7.20
	Abnormal stops of trains (S11)	6.85
	Temperature and ventilation system onboard (S12)	6.98
	Maintenance of station (S17)	7.69
	Signage in a station (S18)	8.16
	Maintenance of vehicles (S21)	7.59
Survey D	Safety against accidents (S05)	7.88
	Available space onboard (S10)	7.26
	Attention and kindness of employees (S14)	7.36
	Lighting of trains (S22)	8.22
	Lighting at stations (S23)	8.26
	Accessibility for disabled (S25)	7.53

The respondents of the type A questionnaire are more satisfied with "Operation of automatic ticket vending machines (S01)" to which they attribute an average score of 7.60. The less satisfactory attribute is "Platform waiting time (S03)" (6.40). All the evaluated attributes reached the 6 score, which corresponds to "good" according to Spanish academic grading and therefore represents a satisfactory judgment.

The service quality attributes evaluated by respondents to Survey B show an average score higher than 7. In particular, the best-evaluated attribute is "Operation of turnstiles (S02)", whereas respondents gave the lower scores to the factors "Kindness of the vigilantes (S07)" and "Onboard ticket validators (S26)" (evaluated only for line ML1 of the LRT).

For the respondents of Survey C, the attribute showing the highest satisfaction rate is "Signage in a station (S18)", exceeding the score of 8. The lowest satisfaction rate was registered for "Abnormal stops of trains (S11)", the average score of which still exceeds the satisfactory level equal to 6.

The service quality attributes included in Survey D were well evaluated by the users, resulting in average scores higher than 7. Users were more satisfied with "Lighting at stations (S23)" (8.26) and "Lighting of trains (S22)" (8.22). "Available space onboard (S10)" obtained the lowest satisfaction rate (7.26).

Looking at the overall results, we can say that users seem more than satisfied with all the service quality attributes. The best attribute of service quality refers to lighting at stations, and the least satisfying attribute was "Platform waiting time (S03)".

4.5. Models

In this section, we present the model results. These include: (i) PCA for Satisfaction items for the four surveys, (ii) Measurement Models for the four surveys, and (iii) SEM models for the four surveys. Also, included are the (iv)

Measurement Invariance for Survey A (MIA), and finally (v) Multi-Group Analysis for Survey A (MGA), which includes three variables. In this Section, we briefly comment the results. The PCA allowed us to build the Satisfaction constructs, and the measurement models allowed us to test the proposed measurement systems. The SEM enabled us to unveil the significance of the models presented, and permits obtaining the best satisfaction model: SEM A. With this model, we tested for measurement invariance, and finally, we derived the three different SEM-MG models for the categorising variables: Age, Time, and Travel Frequency.

4.5.1. Principal component analyses (PCA)

It is best practice to run exploratory PCA separately in the Satisfaction items of each of the four Surveys, before designing the Measurement models. Doing this helps us to determine which items are better represented by specific latent constructs. In this sense, we can decide which items are related and serve one construct as a whole. When an item loaded heavily on several components, it was eliminated from the PCA, but kept as a single indicator or variable of a specific construct, for subsequent analyses. Therefore, we performed four PCA to obtain initial adequate latent constructs for the models. After several iterations for each Survey and according to the Kaiser (1960) rule, we defined two or three different attributes. We note that in the case of Survey B, the item "Onboard ticket validators (S26)" was removed from the analysis, as it was only asked for the specific ML1 line and not for the rest of the system. The PCA results are reported in Table 4-4. With bold letter the scores corresponding to that specific principal component.

Table 4-3: PCA for Satisfaction Items

PCA		PC1	PC2	PC3
Survey A	Operation of automatic ticket vending machines (S01)	-0.20	**0.56**	-0.36
	Personal security from assault and theft (S06)	0.17	-0.16	**-0.63**
	Attention to complaints and suggestions (S08)	-0.04	0.03	**-0.68**
	Temperature and ventilation system at stations (S13)	**0.64**	0.00	-0.12
	Noise level on the vehicles (S20)	**0.71**	0.07	0.02
	Operation of elevators (S24)	0.11	**0.81**	0.12
Survey B	Operation of turnstiles (S02)	-0.10	**0.70**	
	Kindness of the vigilantes (S07)	0.08	**0.56**	
	Cleanliness at stations (S15)	**0.68**	-0.02	
	Cleanliness of vehicles (S16)	**0.68**	-0.01	
	Information about incidents of the service (S19)	0.25	**0.44**	
Survey C	Escalator operation (S09)	0.12	**0.57**	
	Abnormal stops of trains (S11)	-0.04	**0.82**	
	Temperature and ventilation system onboard (S12)	**0.54**	0.01	
	Maintenance of station (S17)	**0.58**	-0.01	
	Maintenance of vehicles (S21)	**0.60**	-0.00	
Survey D	Safety against accidents (S05)	-0.27	**0.47**	
	Available space onboard (S10)	**-0.53**	-0.07	
	Attention and kindness of employees (S14)	0.04	**0.88**	
	Lighting of trains (S22)	**-0.60**	-0.04	
	Lighting at stations (S23)	**-0.54**	0.06	

The resulting components are explained in the following paragraphs. The criterion used was a resulting absolute loading >0.4, precisely in just one component. For Survey A, the three constructs formed are Safety (S06, S08), Comfort (S13, S20), and System (S01, S24). The letter A, B, C, or D next to the name indicates that the constructs belong to that survey. As item "Platform waiting time (S03)" loaded on several components it was left out of the final PCA and will be modelled subsequently as the Service.A variable. For Survey B, the two components formed are Comfort (S15, S16) and System (S02, S07, S19). Again, as item “Speed of train travel (S04)” loaded on several components it was left out of the final PCA and will be modelled subsequently as the Service.B variable. For Survey C, the two components formed are Comfort (S12, S17, S21) and System (S09, S11). The item “Signage in a station (S18)” loaded on several components. Hence it was left out of the final PCA and will be modelled subsequently as the System.C2 variable. For Survey D, the two components formed are Safety (S05, S14) and Comfort (S10, S22, S23). Finally, as item “Accessibility for disabled (S25)” loaded on several components it was left out of the final PCA and will be modelled subsequently as the Service.D variable.

4.5.2. Measurement models

Having defined the Satisfaction latent constructs for each Survey, we set up the Measurement Models or CFA. Additionally, we included specific latent constructs for Critical Incidents (CI) and Overall Satisfaction, and these are measured with the same items for all Surveys. Notably, the CI latent construct is composed of “Affection by a closure of a Line this summer (CI1)” and “Affection by an anomaly in the last month” (CI4). The Overall Satisfaction latent construct is composed of “Overall service satisfaction (OS1)”, “Overall service improvement this past year (OS2)”, and “Line service satisfaction (OS3)”. Figure 4-1 presents the schema for the Measurement Models, ovals represent latent constructs, and rectangles represent single items.

The results of the Measurements Models are presented in Tables 4-5 to 4-7. The first shows the estimates for measurement models A and B, the second presents the estimates for measurement models C and D, and the third the goodness-of-fit tests. The standard coefficients represent the effect of a change in one standard deviation of the respective variable on the dependent variable. For the measurement models, values above 0.5 are considered acceptable and above 0.6 good. The R^2 represents the explained variance. Values have been rounded to two decimal places for readability.

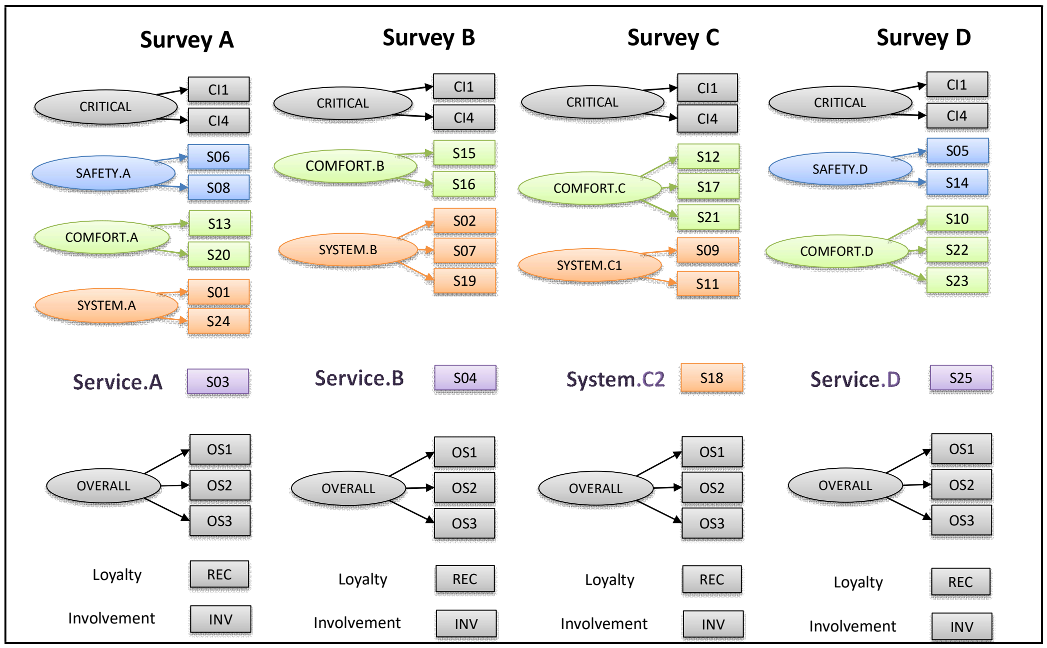

Figure 4-1: Measurement models (CFA) for Surveys A, B, C, and D

Most standard coefficients are over 0.5 in Tables 4-5 and 4-6, but in general the latent construct CI has the lowest sat. We expect this to happen since one of the items asked for events during the summer (CI1), and the other one asked for events during the last 30 days (CI4).

Table 4-4: Measurement models A and B

		Estimate	S.E.	Z-value	Std.Coeff.	R^2
Measurement Model A	**CRITICAL INCIDENTS**					
	Events during summer (CI1)	0.15	0.02	10.08	0.33	0.11
	Events last month (CI4)	0.21	0.02	11.42	0.54	0.30
	SAFETY.A					
	Security from assault and theft (S06)	1.55	0.04	37.65	0.75	0.56
	Attention: complaints/suggestions (S08)	1.65	0.05	36.20	0.72	0.51
	COMFORT.A					
	Temperature/ventilation at stations (S13)	1.74	0.04	41.75	0.81	0.65
	Noise level on the vehicles (S20)	1.55	0.04	37.80	0.73	0.54
	SYSTEM.A					
	Operation of ticket vending machines (S01)	1.29	0.05	28.73	0.68	0.46
	Operation of elevators (S24)	1.37	0.05	25.72	0.58	0.34
	OVERALL					
	Overall satisfaction (OS1)	1.00			0.84	0.71
	Service improvement (OS2)	0.24	0.01	20.46	0.46	0.21
	Line satisfaction (OS3)	0.94	0.03	28.72	0.70	0.48
Measurement Model B	**CRITICAL INCIDENTS**					
	Events during summer (CI1)	0.13	0.02	8.63	0.29	0.09
	Events last month (CI4)	0.26	0.03	10.13	0.66	0.43
	COMFORT.B					
	Cleanliness at stations (S15)	1.74	0.03	58.44	0.93	0.87
	Cleanliness of vehicles (S16)	1.74	0.03	59.59	0.94	0.89
	SYSTEM.B					
	Operation of turnstiles (S02)	1.00			0.52	0.27
	Kindness of the vigilantes (S07)	1.55	0.07	21.61	0.65	0.43
	Information about service incidents (S19)	1.78	0.08	23.12	0.79	0.62
	OVERALL					
	Overall satisfaction (OS1)	1.00			0.85	0.73
	Service improvement (OS2)	0.23	0.01	20.18	0.44	0.20
	Line satisfaction (OS3)	0.90	0.03	30.31	0.72	0.52

Table 4-5: Measurement models C and D

		Estimate	S.E.	Z-value	Std.Coeff.	R^2
Measurement Model C	**CRITICAL INCIDENTS**					
	Events during summer (CI1)	0.10	0.02	6.77	0.22	0.05
	Events last month (CI4)	0.20	0.03	8.01	0.51	0.26
	COMFORT.C					
	Temperature and ventilation onboard (S12)	1.00			0.74	0.55
	Maintenance of station (S17)	0.85	0.02	39.53	0.81	0.66
	Maintenance of vehicles (S21)	0.99	0.02	42.04	0.90	0.80
	SYSTEM.C					
	Escalator operation (S09)	1.28	0.05	27.01	0.62	0.38
	Abnormal stops of trains (S11)	1.30	0.05	25.92	0.58	0.34
	OVERALL					
	Overall satisfaction (OS1)	1.00			0.83	0.69
	Service improvement (OS2)	0.25	0.01	20.90	0.47	0.22
	Line satisfaction (OS3)	0.95	0.03	28.97	0.72	0.52
Measurement Model D	**CRITICAL INCIDENTS**					
	Events during summer (CI1)	0.13	0.02	8.78	0.28	0.08
	Events last month (CI4)	0.23	0.02	10.51	0.58	0.34
	COMFORT.D					
	Available space onboard (S10)	1.00			0.62	0.38
	Lighting of trains (S22)	1.01	0.03	32.74	0.88	0.77
	Lighting at stations (S23)	0.97	0.03	32.42	0.85	0.73
	SAFETY.D					
	Safety against accidents (S05)	1.41	0.04	38.07	0.80	0.64
	Attention and kindness of employees (S14)	1.24	0.04	29.36	0.60	0.37
	OVERALL					
	Overall satisfaction (OS1)	1.00			0.83	0.70
	Service improvement (OS2)	0.25	0.01	20.93	0.48	0.23
	Line satisfaction (OS3)	0.92	0.03	28.14	0.71	0.50

From the estimates presented in Tables 4-5 and 4-6, we consider that, in general, all the measurement models adjust to the data. Table 4-7 shows the goodness-of-fit indices. We compare these against the recommended cut-off values presented by Hu and Bentler (1999): comparative fit index (CFI), Tucker-Lewis index (TLI), adjusted goodness of fit index, and root mean squared error approximation (RMSEA). All four Measurement Models (MM) comply with the cut-off values, just the MM from Survey D does not abide by the TLI criterion. However, it is very close, so we conclude that all four MM are well suited for the datasets.

Table 4-6: Measurement models fit indices

MM-Fit Indices	CFI	TLI	AGFI	RMSEA
A	0.984	0.975	0.981	0.035
B	0.994	0.991	0.987	0.027
C	0.986	0.978	0.981	0.038
D	0.965	0.946	0.958	0.061
Cut-offs	>0.95	>0.95	>0.95	<0.08

4.5.3. SEM models

In the SEM models we tested several relationships, and at the end, we kept those that were significant at a 90% level in the models. We included (i) a direct effect from the CI latent construct to all the attribute-specific satisfaction constructs (e.g. Safety.A). Also, we added (ii) direct effects between the attribute-specific satisfaction constructs and Overall Satisfaction, and (iii) direct effects between these attribute-specific satisfaction constructs and Loyalty. The evaluation of this effect was one of the primary objectives of this work, and we found three direct effects to be significant, to be further explained in Section 6. We include (iv) a direct effect of Overall Satisfaction to Loyalty (REC), knowing that this effect has been well proven in the literature. Next, we tested (v) a direct effect between Overall Satisfaction and Involvement (INV), which did not result significant in any of the Surveys and lastly, we examined (vi) a direct effect between Loyalty and Involvement, which was significant in all Surveys. Figure 4-2 presents the schema for the SEM models: ovals represent latent constructs (measured with the MM from Figure 1), and rectangles represent single items.

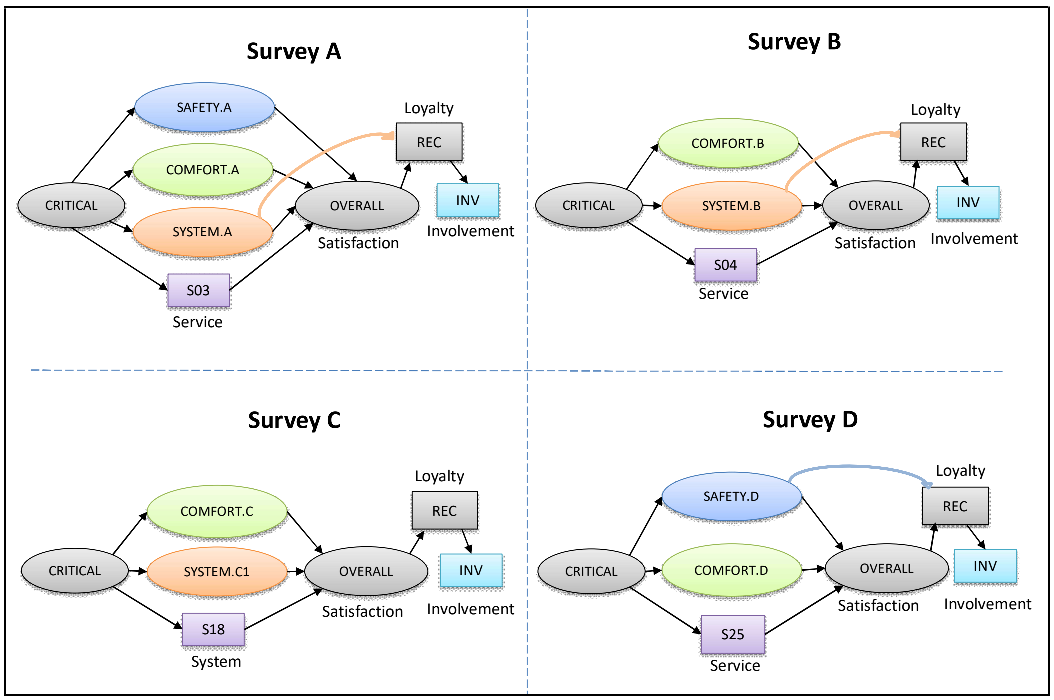

Figure 4-2: SEM models for Surveys A, B, C, and D

The results of the SEM models are presented in Tables 4-8 to 4-10. The first shows the estimates for SEM models A and B, the second presents the estimates for SEM models C and D, and the third the goodness-of-fit tests. For the structural coefficients, values below 0.1 are considered very low, between 0.1 and 0.3, low, between 0.3 and 0.5 moderate, and above 0.5 high.

Table 4-7: SEM A and B

		Estimate	S.E.	Z-value	Std.Coeff.	R^2
SEM A	**SAFETY.A**					0.83
	CRITICAL INCIDENTS	-2.18	0.21	-10.45	-0.91	
	COMFORT.A					0.72
	CRITICAL INCIDENTS	-1.59	0.10	-16.21	-0.85	
	SYSTEM.A					0.60
	CRITICAL INCIDENTS	-1.23	0.09	-14.41	-0.78	
	Service.A (S03)					0.46
	CRITICAL INCIDENTS	-1.58	0.05	-34.64	-0.68	
	OVERALL					0.53
	COMFORT.A	0.10	0.03	3.32	0.14	
	SYSTEM.A	0.18	0.04	5.02	0.22	
	SAFETY.A	0.05	0.03	1.70	0.09	
	Service.A (S03)	0.23	0.01	16.65	0.41	
	Loyalty (REC)					0.44
	SYSTEM.A	0.13	0.03	4.52	0.13	
	OVERALL	0.69	0.03	20.00	0.58	
	Involvement (INV)					0.01
	Loyalty (REC)	0.02	0.01	4.05	0.08	
SEM B	**COMFORT.B**					0.61
	CRITICAL INCIDENTS	-1.25	0.06	-22.66	-0.78	
	SYSTEM.B					0.87
	CRITICAL INCIDENTS	-0.90	0.04	-24.00	-0.94	
	Service.B (S04)					0.47
	CRITICAL INCIDENTS	-1.28	0.04	-34.76	-0.68	
	OVERALL					0.53
	COMFORT.B	0.09	0.03	3.16	0.10	
	SYSTEM.B	0.45	0.06	7.59	0.32	
	Service.B (S04)	0.29	0.02	15.77	0.40	
	Loyalty (REC)					0.42
	SYSTEM.B	0.14	0.05	3.12	0.09	
	OVERALL	0.67	0.03	19.33	0.59	
	Involvement (INV)					0.01
	Loyalty (REC)	0.02	0.01	4.15	0.08	

Table 4-8: SEM C and D

		Estimate	S.E.	Z-value	Std.Coeff.	R^2
SEM C	**COMFORT.C**					0.73
	CRITICAL INCIDENTS	-1.34	0.04	-32.44	-0.85	
	SYSTEM.C1					0.94
	CRITICAL INCIDENTS	-3.91	1.37	-2.86	-0.97	
	System.C2 (S18)					0.52
	CRITICAL INCIDENTS	-1.15	0.03	-36.51	-0.72	
	OVERALL					0.43
	COMFORT.C	0.29	0.05	5.93	0.35	
	SYSTEM.C	0.11	0.04	2.96	0.34	
	Loyalty (REC)					0.40
	OVERALL	0.71	0.02	29.77	0.64	
	Involvement (INV)					0.00
	Loyalty (REC)	0.01	0.01	2.53	0.05	
SEM D	**COMFORT.D**					0.71
	CRITICAL INCIDENTS	-1.06	0.04	-27.26	-0.84	
	SAFETY.D					0.87
	CRITICAL INCIDENTS	-2.56	0.39	-6.63	-0.93	
	Service.D (S25)					0.39
	CRITICAL INCIDENTS	-1.23	0.04	-30.61	-0.62	
	OVERALL					0.37
	COMFORT.D	0.16	0.05	3.00	0.14	
	SAFETY.D	0.20	0.03	6.28	0.40	
	Service.D (S25)	0.09	0.02	4.88	0.13	
	Loyalty (REC)					0.42
	COMFORT.D	0.18	0.03	6.59	0.15	
	OVERALL	0.62	0.03	21.19	0.55	
	Involvement (INV)					0.01
	Loyalty (REC)	0.02	0.01	4.26	0.09	

The standardised coefficients for the CI construct over the attribute-specific satisfaction constructs are all high and negative, in Tables 4-8 and 4-9. In general, the effect from the attribute-specific satisfaction constructs to Overall Satisfaction is positive and between low and moderate, as expected, since we do not expect a single attribute to dominate. Next, we found significant effects between some attribute-specific satisfaction constructs and Loyalty. Also, a low weight between Loyalty and Involvement. We will discuss and explain the results in more detail in Section 6.

From the estimates presented in Tables 4-8 and 4-9, we consider that in general, all SEM models adjust to the data, and the theoretical hypotheses presented. We found some attribute-specific satisfaction constructs affecting Loyalty, but not all, this is a fundamental result of PT policy design. Also, another notable result is that Loyalty only affects the intent of Involvement with a low effect, yet significant, for all four Surveys. Its key to note that for the Overall Satisfaction the explained variance was 0.53 for SEM A and B, 0.43 for SEM C, and 0.37 for SEM D, indicating that the attribute-specific constructs better-explain Overall Satisfaction in the first two cases. We will further develop on this in Section 4.6.

In Table 4-10 the goodness-of-fit indices are presented, again, we compare against the recommended cut-off values presented by Hu and Bentler (1999). All four SEMs comply with the cut-off values, except for SEM models C and D, but barely. However, we can reasonably conclude that all four SEMs are well suited for the datasets, since the essential fit indices are CFI and RMSEA, and all models comply with the cut-off values proposed.

Table 4-9: SEM Fit indices

SEM	CFI	TLI	AGFI	RMSEA
A	0.967	0.956	0.967	0.043
B	0.978	0.970	0.969	0.042
C	0.958	0.945	0.956	0.052
D	0.951	0.935	0.950	0.057
Cut-offs	>0.95	>0.95	>0.95	<0.08

4.5.4. Measurement invariance and multi-group analyses

For the MIA and MGA, we will only use Survey A, since it provided the most explained variance (alongside Survey B) for the Overall Service (OS), and additionally, included the four attribute-specific domains for the Overall Service regression. We follow three steps: (i) perform an MIA on all the possible travel characteristics and sociodemographic categorical variables. Once we establish measurement invariance, (ii) we perform an MGA with the variables and test for different SEM models according to the same variables. If any of the models present a statistical worsening of the CFI index, then the MGA is warranted and will be presented (see Section 4.3 for details). Also, (iii) if MIA is not guaranteed we will still perform the MGA. If any of the models present a statistical worsening of the CFI index, the MGA is warranted and will be presented.

The categorical variables available to test the MGA are Time of day, Gender, Age, Nationality, Ticket type, Trip purpose, and Travel frequency. The objective here is to determine whether different groups in the population have different models, according to the SEM models presented in Section 4.5.3. Firstly, we show the measurement invariance for the variables that do not guarantee it. The variables that do not present measurement invariance are Time, Age, and Travel frequency (Table 4-11). For these three variables, there is evidence that the measurement models should be different.

We run the MGA for all the variables. Once again, we present the results for only the variables providing evidence for a worsening of the fit according to the MGA. The results are shown in Table 4-12, and we can state that the variables Time of day, Age and Travel frequency warrant an SEM-MG model, since their fit index CFI worsens significantly when the parameters are fixed to be the same for all the groups when performing the MGA. In other words, PT users in the different groups have different satisfaction models, for these three variables.

Table 4-10: Measurement invariance

Measurement invariance	TIME		AGE		TRAVEL FREQ	
	CFI	**ΔCFI**	**CFI**	**ΔCFI**	**CFI**	**ΔCFI**
Fit configural	0.985	-	0.983	-	0.986	-
Fit loadings	0.985	0.000	0.976	0.007	0.987	0.001
Fit intercepts	0.984	0.002	0.957	**0.019**	0.984	0.003
Fit residuals	0.974	**0.010**	0.950	0.008	0.967	**0.017**
Fit means	0.969	0.005	0.940	0.010	0.960	0.007
Thresholds		**<0.01**		**<0.01**		**<0.01**

Table 4-11: MG analysis

	CFI	**ΔCFI**	**TLI**	**AGFI**	**RMSEA**
TIME					
All loose multi-group	0.965	-	0.954	0.992	0.044
All fixing parameters	0.941	**0.024**	0.945	0.991	0.048
AGE					
All loose multi-group	0.955	-	0.942	0.991	0.050
All fixing parameters	0.914	**0.041**	0.922	0.989	0.057
TRAVEL FREQUENCY					
All loose multi-group	0.965	-	0.953	0.992	0.044
All fixing parameters	0.924	**0.041**	0.930	0.989	0.054
Cut-offs	>0.95	**<0.02**	>0.95	>0.95	<0.08

Next, we show the SEM regression models for the Overall Satisfaction, for these three variables. We want to test if the relevance or priorities differ according to the different groups. In Tables 4-13 to 4-15, we show the results for the MGA-Time of day, MGA-Age, and MGA-Travel frequency. We will discuss the results in detail in Section 4.6.

Table 4-12: MGA – Time

MGA – TIME	Latent variables	Estimate	S.E.	Z-value	Std.Coeff.	R^2
Group 1 [7-9.30]	**OVERALL**					0.52
	COMFORT.A	0.10	0.07	1.54	0.16	
	SYSTEM.A	0.16	0.07	2.21	0.19	
	SAFETY.A	0.06	0.07	0.83	0.10	
	Service.A (S03)	0.22	0.04	5.96	0.41	
Group 2 [9.30-13]	**OVERALL**					0.55
	COMFORT.A	0.13	0.07	1.85	0.21	
	SYSTEM.A	0.00	0.10	0.01	0.00	
	SAFETY.A	0.08	0.07	1.20	0.18	
	Service.A (S03)	0.26	0.04	7.04	0.44	
Group 3 [13-18]	**OVERALL**					0.58
	COMFORT.A	0.04	0.07	0.49	0.05	
	SYSTEM.A	0.33	0.08	4.19	0.40	
	SAFETY.A	0.02	0.06	0.29	0.03	
	Service.A (S03)	0.22	0.03	6.78	0.40	
Group 4 [18-22]	**OVERALL**					0.54
	COMFORT.A	0.11	0.08	1.51	0.16	
	SYSTEM.A	0.17	0.08	1.99	0.22	
	SAFETY.A	0.05	0.06	0.77	0.11	
	Service.A (S03)	0.21	0.03	7.66	0.38	

Table 4-13: MGA – Age

MGA - AGE	Latent variables	Estimate	S.E.	Z-value	Std.Coeff.	R^2
Group 1 [16-24]	OVERALL					0.52
	COMFORT.A	0.15	0.06	2.65	0.24	
	SYSTEM.A	0.22	0.06	3.90	0.27	
	SAFETY.A	0.02	0.05	0.35	0.03	
	Service.A (S03)	0.16	0.03	5.56	0.36	
Group 2 [25-34]	OVERALL					0.51
	COMFORT.A	0.05	0.07	0.74	0.07	
	SYSTEM.A	0.16	0.09	1.83	0.19	
	SAFETY.A	0.18	0.06	2.87	0.21	
	Service.A (S03)	0.22	0.03	6.35	0.38	

Table 4-14: MGA – Age (continued)

Group 3 [35-44]	OVERALL					0.66
	COMFORT.A	0.04	0.07	0.57	0.07	
	SYSTEM.A	0.30	0.08	3.75	0.38	
	SAFETY.A	0.01	0.06	0.21	0.02	
	Service.A (S03)	0.28	0.03	9.35	0.47	
Group 4 [45-54]	OVERALL					0.54
	COMFORT.A	0.03	0.07	0.38	0.05	
	SYSTEM.A	0.12	0.11	1.13	0.14	
	SAFETY.A	0.19	0.06	3.00	0.27	
	Service.A (S03)	0.22	0.04	5.15	0.40	
Group 5 [55+]	OVERALL					0.53
	COMFORT.A	0.23	0.13	1.85	0.41	
	SYSTEM.A	0.04	0.15	0.26	0.05	
	SAFETY.A	-0.09	0.13	-0.73	-0.11	
	Service.A (S03)	0.27	0.05	5.66	0.44	

Table 4-14: MGA – Travel frequency

MGA - TRAVEL FREQ	Latent variables	Estimate	S.E.	Z-value	Std.Coeff.	R^2
Group 1 [0-3]	OVERALL					0.39
	COMFORT.A	-0.03	0.11	-0.26	-0.09	
	SYSTEM.A	0.05	0.13	0.37	0.06	
	SAFETY.A	0.17	0.15	1.14	0.36	
	Service.A (S03)	0.23	0.05	5.08	0.38	
Group 2 [4-10]	OVERALL					0.53
	COMFORT.A	0.09	0.06	1.47	0.12	
	SYSTEM.A	0.17	0.06	2.65	0.21	
	SAFETY.A	0.06	0.05	1.24	0.14	
	Service.A (S03)	0.21	0.03	7.76	0.39	
Group 3 [11-20]	OVERALL					0.56
	COMFORT.A	0.16	0.06	2.56	0.21	
	SYSTEM.A	0.19	0.07	2.58	0.25	
	SAFETY.A	-0.01	0.05	-0.17	-0.02	
	Service.A (S03)	0.23	0.03	8.22	0.41	

Table 4-15: MGA – Travel frequency (continued)

Group 4 [21+]	OVERALL					0.61
	COMFORT.A	0.06	0.09	0.65	0.08	
	SYSTEM.A	0.27	0.11	2.48	0.26	
	SAFETY.A	0.09	0.08	1.10	0.13	
	Service.A (S03)	0.28	0.04	6.84	0.46	

4.6. Discussion of Results

From the PCA, we can observe several interesting findings. Before discussing the results, recall that the constructs regarding safety are not only represented by specific safety components, but also by the attention given by the staff of the Metro system. Consequently, we can conclude that effectively the staff presence at stations and other installations combines with the actual real safety and security perception to integrate two distinct safety constructs, in Surveys A and D. Concerning the comfort constructs and the system constructs, note that they are differently formed in the various surveys, as it can be observed in the following discussion. Finally, the service components did not have more than one item in Surveys A, B, and D, so no latent constructs were formed. There was no specific service construct in Survey C.

SEM A results allow to see that the most influential variable for Overall Satisfaction is waiting time at the platform (S03). Also, the latent construct concerning more technical aspects, such as "Operation of automatic ticket vending machines (S01)" and "Operation of elevators", has considerable relevance, while the elements concerning comfort and safety have a less prominent role in Overall Satisfaction.

SEM B results are similar. Concretely, the attribute influencing Overall Satisfaction the most is the speed of train travel (S04). However, the latent construct measured with "Information before incidents of the service (S19)", "Operation of turnstiles (S02)", and "Kindness of the vigilantes (S07)", has a similar weight. As in SEM A, comfort has a less significant influence on the Overall Satisfaction.

SEM C shows some differences with the previous two models. Overall Satisfaction is influenced similarly by comfort and system latent variables. This may happen because there were no service variables included in Survey C. Comfort was represented by "Temperature and ventilation system onboard (S12)", "Maintenance of station (S17)", and "Maintenance of vehicles (S21)". The system was represented by "Escalator operation (S09)" and "Abnormal stops of trains (S11)".

Finally, the service aspects that influence Overall Satisfaction the most in SEM D are the variables concerning safety. This result is different from SEM A, where safety could not be considered as a relevant aspect. In the case of SEM D, safety is represented by "Safety against accidents (S05)" and "Attention and kindness of employees (S14)". Survey D included a service variable, Accessibility for disabled (S25). However, it is not an essential service variable as those involved in Surveys A and B, thus, it did not turn out as relevant as in the case of the first two surveys.

Also, by comparing all four models by their explained variance, R^2, the former two are better than the latter two. We may conclude that some relevant variables may have been missing in the latter two, specifically some item of service speed, frequency or reliability; which were included in the first two. Another interesting finding from the comparison of the four models is that the influence of the various service attributes on the Overall Satisfaction depends on which attributes are selected for representing the service aspects linked to service, system, comfort, and safety. As an example, when comfort is expressed by characteristics related to cleanliness and lighting on trains and at stations, the temperature at stations, noise level onboard, or available space onboard, it is not considered as an essential aspect by the passengers. On the contrary, passengers find comfort the most relevant service aspect when it is represented by temperature onboard and maintenance of station and vehicles. Another difference among the models concerns the element linked to safety. Safety is not considered as a fundamental aspect when it is expressed by attributes such as personal security from assault or theft, or attention to complaints and suggestions. On the contrary, passengers view safety as the most critical aspect when it is evaluated regarding safety against accidents and attentiveness and kindness of employees. These results suggest that for passengers of Metro systems travelling in a comfortable manner means travelling with an agreeable temperature onboard and having trains and stations well maintained. At the same time, travelling without the occurrence of accidents and having assistance from employees make them feel safe, that is, the presence of personnel who pay attention to passengers remains reassuring.

Observing the results of all the four models, we note that although the highest weight of the Loyalty regression is due to Overall Satisfaction, there are also specific attributes which additionally influence it directly, such as some service aspects concerning the system, and others concerning comfort. As an example, system functioning attributes such as the operation of ticket machines and elevators, and comfort attributes like available space onboard, and lighting of trains and stations have a specific direct influence on Loyalty. These aspects are considered relevant if passengers have to recommend the service to other people, even if, they have not ranked these elements to be among the most influential towards their Overall Satisfaction.

From all four models we can observe that the CI latent construct negatively affects all latent constructs representing the various service aspects, and in a significant manner (i.e. all weights are over 0.6). This result indicates that increasing the occurrence of critical incidents causes a decrease in satisfaction with all service aspects, which will directly impact the Overall Satisfaction, in a negative fashion, and hence the Loyalty. Another similar result present in all four models is the effect of Loyalty towards Involvement, as there is a slight positive standardised coefficient, in all cases.

Another relevant result is that for all four surveys, Overall Satisfaction is mostly represented by the overall service satisfaction (OS1) and also by the level of satisfaction with the line used at the time of the interview (OS3). On the contrary, the indicator representing passenger's assessment of the service's improvement in the past year (OS2) has a lower weight, even if the value of the coefficient is significant.

Exciting findings can also be observed from the results of the MGA models. We found that passengers' perceptions depended on travel characteristics such as the time of travelling and travel frequency, and on the sociodemographic characteristic, age. Investigating the differences regarding the time of day (Table 4-13), we verified that waiting time at the platform (S03) is the service quality attribute that mostly influences Overall Satisfaction, independently of the time of the day when passengers travel. The primary evidence regarding differences between user groups can be found by observing the coefficients of the latent variables associated with comfort and safety, which have low values for the group of passengers travelling from 1 pm to 6 pm. On the contrary, this attribute is more relevant for passengers travelling in the morning and also in the evening. These latter passengers are more demanding regarding comfort (measured in this case through temperature in the stations and noise level on the vehicle) and safety (measured as security from assault and theft, and attention to complaints and suggestions). The afternoon is a period where passengers are more relaxed and do not perceive as essential the above aspects linked to comfort and safety. On the contrary, in the afternoon they are more demanding concerning the functioning of the system (measured through operation of automatic ticket vending machines and operation of elevators). These same aspects are not considered too valuable in the other periods of the day.

About the differences regarding Age (Table 4-14), similar results are obtained for the service attribute linked to the waiting time at the platform, which is the most relevant also in this case, independently of the group of passengers. Concerning comfort, we found that older passengers (over 55 years old) are the most exigent.

Surprisingly, young people (16 to 24 years old) give particular importance to the temperature at the station and the noise level on the vehicle, while for passengers from 25 to 54 years old these aspects are not relevant. Older passengers (over 55 years old) do not give importance to the aspects concerning system and safety. On the contrary, safety is very relevant for passengers between 45 and 54 years old, maybe because they are more subject to assault or theft and to make complaints, belonging to a middle-aged class characterised by maturity and experience but at the same time activity and business. Finally, passengers between 35 and 44 years old are the most exigent concerning the functioning of the ticket machines and elevators. A reason for this result could be that people belonging to this age class may be more demanding of a hassle-free operation and often travel with children. Thus, they demand a better functioning system. Also, people in this age-class are probably workers requiring PT systems with high levels of efficiency because they pay much more attention to the time spent in travel activities for reaching their workplaces.

The last MGA model investigated passengers' differences concerning travel frequency (Table 4-15). Also, in this case waiting time at the platform was the most relevant service attribute for all groups of passengers. Interesting contrast of perception concerns the attributes linked to the functioning of the system, which is essential for all groups, except the occasional users who travel less than four times per week. They do not pay attention to this aspect because they travel only occasionally. On the contrary, they give more importance to the elements linked to safety, maybe because a reason why they travel by PT only occasionally is their perception of security. Finally, comfort is more relevant for passengers travelling from 4 to 20 times per week, than for the occasional passengers and the most regular ones (travelling more than 21 times per week). Occasional passengers may not have a perception of this attribute because they travel too few times, and probably they are more inclined to travel in a not very comfortable manner because of their occasional travel. On the contrary, habitual passengers do not pay much attention to comfort levels because they travel too many times, and therefore are accustomed to the comfort offered when daily travelling. They are most probably captive travellers.

4.7. Conclusions

Our results show that the CI variable was the worst measured construct. This result was expected since one of the items asks for events during the summer, and the other asks for events during the last 30 days. Also, in this case, we considered only two indicators and not a separate indicator for each service quality item, as performed by Allen et al. (2018b). Nevertheless, we verified the hypothesis that the CI latent construct negatively affected all latent constructs representing the various service aspects. Thus, we confirm the findings from Allen et al. (2018b) and argue that this should be considered in PT satisfaction studies, to gain a better knowledge of what is causing dissatisfaction in specific characteristics of the system, and hence obtain a natural policy variable for PT operators. We also recommend creating attribute-specific CI items.

Concerning the Loyalty variable, it has a positive, yet small, weight on the variable Involvement. The Involvement of passengers in the evaluation and monitoring processes needs to be improved by transit operators, even though our findings show a weak relationship between Loyalty and Involvement, and no correlation between Overall Satisfaction and Involvement. Still, there appears to be a link that should be addressed, with future studies focusing on Overall Satisfaction, Loyalty, and Involvement, with additional items that should be carefully obtained from the marketing literature.

Regarding our primary objective of finding key attributes that affect Loyalty besides Overall Satisfaction, we showed that the most critical attributes affecting Overall Satisfaction were the service variables: speed of PT and waiting time at the platform. Instead, the variables that mostly affect Loyalty are other types of variables; specifically, some system and safety variables. This result indicates that for PT policy, these variables cannot be ignored and should be enhanced, to increase the loyalty of the PT patrons. For the Metro of Madrid case, we recommend using all items at the same time or conducting a small pilot survey with all of them to determine which are the best ones. We believe our results give insight of which could be the best indicators to use. We recommend using at least two indicators for each separate latent construct, in the initial survey design.

From the MGA, we were able to unveil differences in passenger's perceptions among user with different characteristics, that is occasional versus habitual users, young versus old people, passengers travelling in the morning and evening versus passengers travelling in the afternoon. These differences should be taken into account as key policy recommendations for PT operators. We recommend applying the same methodology in future studies on different cities to address the generalizability of our findings. One key outcome, worth mentioning again from our research, is that infrequent passengers perceive higher importance in the security variables regarding thefts, a policy variable to be addressed directly. Comparing with Antonucci et al. (2012), they obtained residence area and frequency of use as variables that supported different satisfaction models, and we obtained the time of day, age, and travel frequency. We conclude that the categorical variable, travel frequency, is supported to indicate different satisfaction models amongst the subpopulations, in two different studies. In fact, in our case study, this variable provided the highest statistical proof of difference by having a higher ΔCFI in the MG Analysis (see Table 4-12).

Waiting time at the platform and speed of the Metro are the service quality attributes that mostly influence passenger's Overall Satisfaction. In fact, we showed that waiting time is the most relevant attribute for all different types of users addressed in the MGA. We believe this to be a crucial result since we can conclude robustly that this is the most crucial attribute, the "maintain at all costs" service attribute.

Regarding the different SEM obtained in all four surveys, we recommend caution in making conclusions with the models SEM C and SEM D. As shown with the explained variance, it appears that these two surveys are lacking relevant variables: speed or frequency of the service. As such, due to possible correlation, other variables may be gaining importance. From these findings, we recommend to always have at least two items for all possible service domains, and include them all, in our case: service, comfort, safety, and system. We recommend, if possible, to separate these domains further depending on each specific PT case study, and its unique characteristics.

5. UNDERSTANDING PUBLIC TRANSPORT SATISFACTION: USING MASLOW'S HIERARCHY OF (TRANSIT) NEEDS

5.1. Introduction

Citizens are increasingly demanding cleaner or more sustainable modes of transport, such as walking and cycling, and more efficient public transport modes/vehicles regarding energy and the environment (see Litman, 2007 for a detailed summary). In addition to this, governments are interested in encouraging people to transfer from private modes to public transport (Beirão and Sarsfield Cabral, 2007) to mitigate congestion (Anable, 2005, Steg, 2005), among other externalities. To encourage the use of PT, the level of satisfaction perceived by their customers must be high.

Satisfaction surveys allow detecting global satisfaction levels with a PT system, in addition to satisfaction with its different attributes. However, although the scientific literature registers various econometric models to determine which attributes are more relevant (Allen et al., 2018a, Eboli and Mazzulla, 2007, Fellesson and Friman, 2012). Moreover, the literature reports differences in the importance of the different attributes of a PT system depending on the modes of transport associated with it and the city where the system operates (Tyrinopoulos and Antoniou, 2008). Few models correctly capture the sociodemographic heterogeneity of users and the variations in the characteristics of their trips (Allen et al., 2018a, 2018b). In fact, most studies treat users as a single homogenous mass; possibly introducing biases in the models, severely limiting them for use in practical applications.

More importantly still, we found little evidence of any psychological theory of behaviour to justify the satisfaction models proposed in the transport literature. This critical gap led us to propose and test a plausible justification for PT satisfaction models based on Maslow's theory of a hierarchy of needs (Maslow, 1943, 1954). Thus, we propose three types of attributes in a hierarchy of transit needs: functional (utilitarian), security (protection) and hedonic (excitement). We detail the contributions of the study in the next paragraphs.

Our primary contribution is to test the hypothesis of the existence of a hierarchy of PT attributes, based on Maslow's hierarchy of needs (Maslow, 1943). For this, we focus our analysis on a benchmark customer satisfaction survey performed in the Bus Rapid Transit (BRT)-type systems of four major cities in Latin America: Santiago de Chile, Mexico City, and two cities in Brazil. We focus on the following specific domains: reliability (functional), safety (security), customer services (hedonic), and comfort (hedonic), and use a structural equation (SEM) approach to test our hypothesis. We are interested in assessing differences in the satisfaction models for each city, in other words, differences in the order of priority of the various attributes. Within this framework, we provide a general mechanism about the order of relevance of attributes in PT satisfaction models. Secondly, to reinforce this contribution, we apply two robustness checks. Initially, we use two different types of SEM models for the MGA: a numerical one, using the satisfaction rates as natural numbers, and an ordinal one, using the satisfaction rates as ordered levels. Our results demonstrate that with both approaches we obtain very similar results.

Thirdly, we use multigroup analyses (MGA) to assess differences in the satisfaction models according to differences in travel and sociodemographic characteristics. We test for differences between the pooled data (all four cities combined), and also test for two of them separately, Santiago and Mexico City. In this way we are able to test our initial hypothesis between different sociodemographic and travel groups, providing a deeper generalisation of our results. We have found no previous studies performing a similar analysis.

Finally, we consider a structural equation model with finite mixtures (SEMM), yielding latent class SEM models. These allow for grouping the subpopulations according to users' responses rather than predefined categories. Again, we have not found studies in the PT satisfaction literature using this type of models. We compare the results from Santiago and Mexico City against the MGA results and, again, obtain similar results. All these checks reinforce our confidence in the results obtained and strengthen the evidence supporting our hypothesis: the existence of the hierarchy of transit needs.

The rest of the chapter is organised as follows. Section 5.2 presents a literature review of three topics: (i) individual satisfaction attributes and Maslow’s hierarchy of needs, justifying its use for PT satisfaction models; (ii) studies that cover PT satisfaction in multiple cities, and (iii) a brief mention of the SEMM model introduced here in the PT satisfaction literature. Section 5.3 presents the case studies, samples and survey, regarding BRT-type systems in four cities in Latin America. Section 5.4 presents the model results detailing the methodology. Section 5.5 discusses all the results and interpretations in some detail, whilst in Section 5.6 we summarise the most important conclusions from our study and offer key policy recommendations for PT administrators from our findings. In particular, we comment on the possible generalisation of our results and suggest future research avenues to further examine our hypothesis.

5.2. Literature Review

5.2.1. Satisfaction attributes and Maslow’s hierarchy of needs

We first address satisfaction attributes according to the quality and marketing literature. We refer to service and product indistinctly, according to the dominant logic in marketing (Vargo and Lusch, 2007a, 2007b). An important aspect is to distinguish the relationship between the level of compliance of an attribute and the overall service satisfaction. Accordingly, four types of attributes may be defined (Kano et al., 1984, Matzler et al., 2004):

a) Basic or dissatisfaction attributes. These must reach minimum compliance for the user to be satisfied. That is, if they are not met, the user will be dissatisfied, but they do not provide a high degree of satisfaction once the minimum expectation of the user is met. A non-linear relationship between these attributes and global satisfaction is expected. In the case of public transport, we suspect that basic attributes could be accessibility to the service (e.g. a maximum walking and access time) and safety (e.g. a minimum level of safety when travelling standing).

b) Performance or critical attributes. These are essential both to provide satisfaction or dissatisfaction on the whole metric of the attribute. There will always exist an opportunity to improve or diminish overall satisfaction depending on these attributes, regardless of their current level. A linear relationship between global satisfaction and these attributes is expected. In the case of public transport, we suspect that critical attributes could be service reliability and travel time.

c) Excitement and delight, or satisfaction attributes. Once a customer is in the positive satisfaction range, these attributes can provide additional satisfaction. A non-linear relationship is expected between these attributes and global satisfaction. For public transport, we suspect that excitement attributes could be having real-time information of the vehicle location and arrival times, availability of air conditioning, and the provision of wireless connectivity inside the vehicle.

d) Neutral attributes. Regardless of the user's level of dissatisfaction or satisfaction, these attributes do not produce a difference in the levels of overall satisfaction with the service, as they have no direct relationship with it. No significant correlation between these attributes and overall satisfaction is expected. In the case of public transport, a neutral attribute could be the service schedule since in most cases it is known at what time the service operates so the attribute would only be relevant if the schedule is modified.

Note that as some attributes will only be valued if others are adequately met, valuations may significantly depend on the city-mode context and the current levels of satisfaction. Figure 5-1 presents a conceptual scheme with the four types of attributes and their relationships with global satisfaction. The relationship between an attribute and global satisfaction may vary in time if the context changes (Falk et al., 2009, Mittal et al., 2001). For example, we expect some attributes to evolve in the same market and go from being excitement attributes to performance attributes, after some time on offer. Simple examples of this dynamic concept are power windows and cupholders in cars. When they entered the market, they were attributes of excitement because few vehicles offered them; but over time, they became basic or dissatisfaction attributes, since consumers expected any car to have them.

The types of attributes may also vary depending on market consolidation and local quality standards. Thus, we expect a same set of attributes to present different levels of service depending on the individual context of PT systems. Some of these will depend on the available technologies, and the levels of demand and penetration they enjoy in their market (e.g. location of buses in real time, mobile phone applications for finding routes). In this way, by modelling non-linear relationships, the relative importance of an attribute will depend on its current levels. Finally, there may be heterogeneity of preferences among users (Huse and Evangelho, 2007) because, for a given service, an attribute can be perceived differently depending on the user's level of expectation and experience.

We consider of interest searching for a theoretical justification for these relationships in the theory of human motivation of Maslow (1943). His Hierarchy of Needs establishes that the basic needs of humans (functional or utilitarian) must be met before seeking to satisfy higher-level (hedonic) needs. For each service, there will be functional attributes that must be satisfied before seeking to satisfy hedonic attributes, and we expect a sequential order of preferences. Interestingly, in the literature related with social housing, there is evidence of this phenomenon.

Greene and Ortúzar (2002) found that creditors of social housing valued more those attributes that they lacked, and those that they received along with the housing bond, than other attributes. The authors confirmed this behaviour with three groups of creditors, in differentiated circumstances of need, finding in all cases that creditors valued more those attributes that they currently did not possess.

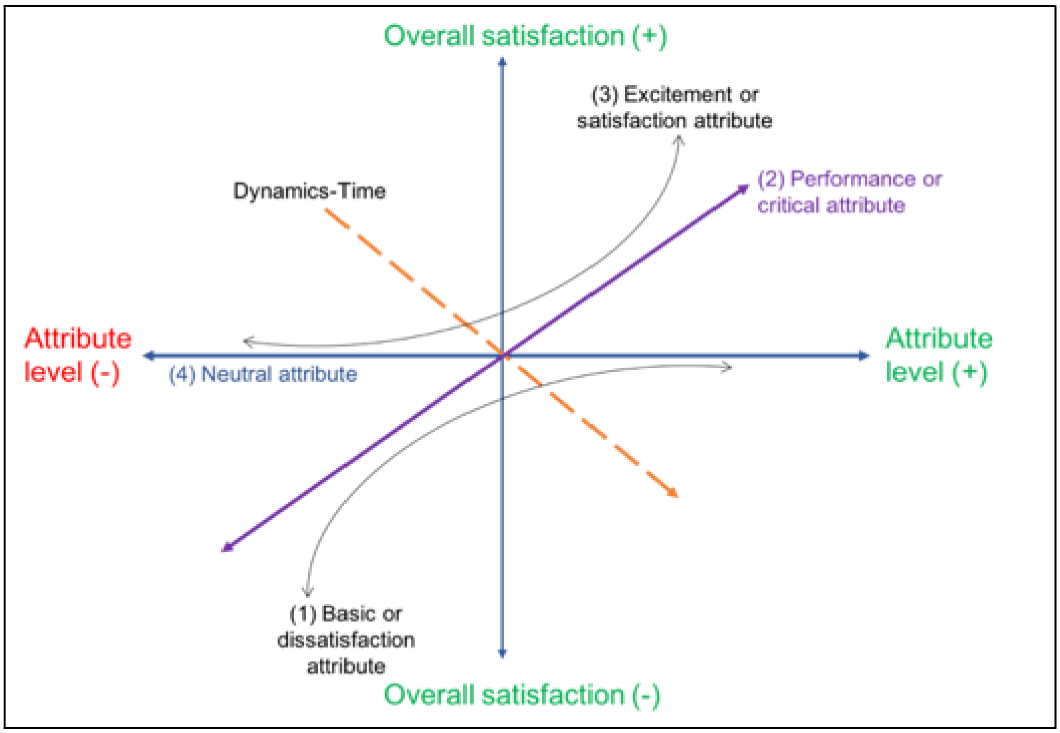

Figure 5-1: The relationship between attributes and satisfaction: Four attributes theory

Adapted from Kano et al. (1984) and Matzler et al. (2004)

In the transport literature, Donoso et al. (2013) report differences in the relevance of urban bus service attributes depending on the type of service: collector or trunk, showing that they also vary according to the level of service already being provided. Bordagaray et al. (2014) report that access (or walking) time, is not a significant attribute in Cantabria, given that there is a local policy to guarantee access to the entire population, so there are bus stops within 300 m of all households. Thus, given that implicitly the population has this need covered, it is not perceived as relevant, and it could be a dissatisfaction attribute. Consequently, there is empirical evidence that the relevance of an attribute depends on its current level of service. This phenomenon is similarly present in a study by Fellesson and Friman (2012), who report differences in the relevant attributes defining satisfaction in different cities; we suspect that differences in the service levels offered in those cities could cause these. In the following paragraphs, we introduce Maslow's theory of human motivation.

The hierarchy of human needs is a psychological theory presented by Maslow (1943, 1954), which has produced a substantial impact in the development of psychology and related sciences. Recent studies validate the theory and build on it with slight adaptations in the structure of needs (Kenrick et al., 2010). Maslow formulates a hierarchy of human needs and stipulates that as basic needs are met, humans develop higher needs and desires. All these needs are what motivate the human being to exercise activities in search of satisfying them. In itself, the theory provides a way of understanding the priority mechanisms that humans give to their needs and, therefore, how they seek to satisfy them. These needs govern the way human beings make decisions to act. The hierarchy has five levels: the first four are deficit needs, and the upper level is called self-realisation. The fundamental idea is that superior needs are searched upon only when the lower needs have been met. The aim is to satisfy the higher needs, and this motivates behaviour. Maslow classified the needs into:

a) Basic needs. They are basic physiological needs to maintain homeostasis and survival. They refer to: breathe, drink water, feed, rest, eliminate body wastes, avoid pain, maintain body temperature and have clothing.

b) Security and protection needs. They refer to the need to feel safe and protected. They are needs of physical and health security, security of resources (e.g. home, work) and housing (protection).

c) Social needs (affiliation). They are related to the social nature of the human being. They are a function of social relationships: friendship, partners, colleagues, family, and social acceptance.

d) Estimation needs (recognition). Maslow poses two types of esteem: high and low. The former corresponds to the need of respect for oneself (self-respect). Low esteem concerns respect received from other people. Together they are called self-esteem: an individual's overall subjective emotional evaluation of his/her worth.

e) Self-realization. It is the highest psychological need of human beings, and it is through its satisfaction that justification for life through the potential development of activities is found.

Maslow's theory expresses that only unmet needs influence people's behaviour; the satisfied needs do not generate any behaviour. While basic or physiological needs develop when we are born, others evolve as we grow. Higher needs do not arise if the lower ones have not been partially met, both could be related (interrelated) but the basic ones will predominate over the higher ones. Maslow's theory has supported the classification of service attributes in various marketing strategies. Both Kano et al. (1984) and Falk et al. (2009) based their studies on this theory to categorise the types of attributes concerning their relationship with service satisfaction. Perone et al. (2005) and Winters et al. (2001) are the only studies we are aware of aiming at applying Maslow's hierarchy of needs for PT service levels. Peek and van Hagen (2002) also applied it for analysing waiting conditions at train stations.

In this study, based on the theory of Maslow and the interpretation of Falk et al. (2009), we propose a classification of the different attributes of a public transport service. We consider that the hierarchy of transit needs is composed of three types of attributes: functional (utilitarian), security (protection) and hedonic (excitement), as detailed below:

a) Functional (utilitarian) attributes. These attributes refer to the reliability, accessibility and mobility offered by a PT service. These are all the characteristics related to the basic production of the service. They are classified as basic since they provide the most elemental transportation service: to move the person from one place to another in a reliable way and with a basic standard of comfort (e.g., comfortably travelling standing). Among these attributes we should have frequency, reliability, accessibility and speed.

b) Security (protection) attributes. These attributes refer to security and protection. Among them will be the perception of security against traffic accidents and assaults, either waiting or in the vehicle. Also, the perception of safety in the event of falls, disruption of the services or travelling too crowded during the journey. They also include protection against weather conditions and environmental hazards inside the vehicles and at the stops or subway stations.

c) Hedonic (excitement) attributes. These attributes refer to accessory aspects of the system that the user may perceive positively. Among them is the perception of comfort (extra comfort) in the system, including all its components. Specific attributes are the perception of the number of seats available, space available within the vehicle, additional services such as availability of a wireless connection and air conditioning. Other attributes may include integrated services regarding payment (single card), availability of payment places, vehicle cleanliness and physical condition, noise, smell, aesthetics, convenience and ease of use, user information, customer services, driver's courtesy, and image (status).

Therefore, in this study we classify the attributes of a PT system and formulate a hypothesis that is coherent with Maslow's theory of human motivation. The attributes should have the following order of preference: (i) functional attributes, (ii) safety and protection attributes, and (iii) hedonic attributes. In essence, the functional attributes will only be relevant when they are not at a satisfactory level. If they are, the security attributes will become relevant; and if they reach a threshold, the hedonic attributes will gain importance. Maslow stipulates that the emergence of higher needs to be gradual: as most needs of a certain level are met, the next level gains relevance according to the hierarchy.

In practical terms, we expect that, for example, reliability will be a very relevant attribute if it is below a minimum standard. If reliability exceeds this standard, but there is a severe crime problem (e.g. New York Metro 1980s), safety will become critical; given that case users will assign greater relevance to security and protection attributes, and less to reliability. We expect that if reliability and security are satisfactorily fulfilled, the hedonic aspects will gain relevance. In a nutshell, this is the working hypothesis to be tested in this study. Next, we summarise previous studies that analysed satisfaction models in different cities and modes of transport with a single customer satisfaction survey.

5.2.2. Public transport satisfaction models in multicity surveys

A limited number of studies have analysed multiple cities with a single customer satisfaction benchmark survey. Fellesson and Friman (2012) used factor analysis and structural equation models (SEM) to compare perceived satisfaction with public transport in nine European cities: Barcelona, Berlin, Copenhagen, Geneva, Helsinki, Manchester, Oslo, Stockholm, and Vienna. The study identified four dimensions of satisfaction: system (operation), comfort, company personnel, and security. The primary result was that there were differences in perception of the different attributes and the general structure between cities; the latent constructs were perceived differently indicating heterogeneity in how the dimensions were structured cognitively. The authors state that this result could be due to cultural dissimilarities and tradition between different cities. We argue that this may be, in fact, caused by different mean levels of service among the cities.

Tyrinopoulos and Antoniou (2008) used factorial analysis and ordinal-logit models to analyse the implication of the variability in the perception of user satisfaction for different operators; they analysed five public transport systems in two cities, Athens and Thessaloniki, Greece. The essential satisfaction attributes cross-cutting the operators were: frequency of service, cleaning of the vehicles, waiting conditions, the distance of transfer and coverage of the network. Another research effort by Susilo et al. (2010), explored the impact of attitudes and past experiences on various public transport policies in Indonesia. Interviews were carried out in three metropolitan areas: Jakarta, Bandung, and Jogjakarta. By performing cluster analysis in one specific city, Bandung, they showed that sociodemographic aspects and learning processes substantially determined the acceptability of individuals towards different policies over time. In the scientific-academic literature, we found no other studies comparing satisfaction models between cities with the same benchmark survey.

One advantage of our own survey is that it was standardised to serve as a benchmark between the different cities, and a minimum of 2,000 questionnaires was filled in each one, allowing for complex analysis such as SEM-MGA and SEMM. Additionally, the survey was intended for specific BRT-type systems, not for PT in general. Therefore, specific insight may be obtained about the satisfaction models, and specific operative and policy recommendations for BRT systems can be determined from the analysis. The next subsection introduces the SEMM model. For additional details on the SEM, SEM-MIMIC, and SEM-MIMIC ordinal Probit models applied to PT satisfaction, please refer to Allen et al. (2018a, 2018b).

5.2.3. Structural equation models with finite mixtures (SEMM)

The SEM-MIMIC model (Hauser and Goldberger, 1971, Joreskog and Goldberger, 1975) considers heterogeneity in the measurement but does not consider heterogeneity in the cross-sectional preferences of different segments of the population. However, it is common that in economics, psychology and finance, for example, heterogeneity exists in different segments of the population (Jedidi et al., 1997a). Likewise, satisfaction research suggests that decision processes vary transversally for segments of the population (Oliver, 2010).

Feasible solutions include the SEM-MGA, which allows for different models in individual categorised subpopulations: we use these models in the present study, see Allen et al. (2018c) for details. Another solution is to use models that allow for a higher degree of complexity, such as structural equations with finite mixtures (SEMM). These treat heterogeneity by forming subpopulations in the context of a specified structural model (Jedidi et al., 1997b). Once the structural model is specified, the modelling allows generating different subpopulations to be segmented, depending on the totality of the responses or items. The approach is considerably more general than the conglomerate analysis methods, confirmatory multigroup factor analysis and SEM-MGA. These last two techniques only apply when the subpopulations have been identified before the analysis.

Advantages of the SEMM include: first, it allows for a segmentation based on the responses to a consumer decision model (e.g. satisfaction model), while simultaneously considering measurement errors (latent variables). Second, it allows detecting unobserved moderating factors (latent classes) that take into account heterogeneity (Jedidi et al., 1997a) in the responses and preferences. Third, once the moderating factors are identified, the membership of segments can be linked to observable characteristics at the individual level (e.g. sociodemographic variables and characteristics of the type of trip), and thus improve the operational policies and marketing strategies of the service. Finally, for these models, there are selection criteria that allow defining the correct number of latent classes to be modelled (Henson et al., 2007).

5.3. Case Studies, Samples and Survey Forms

5.3.1. BRT-type systems in Latin America

Data supporting this research were collected by the PT operators of four different cities through Customer Satisfaction Surveys for specific bus rapid transit (BRT)-type systems. The data from Santiago and Mexico City were commissioned by a private entity (SIMUS) through the local PT operators in both cities. The original questionnaire was translated from Portuguese into Spanish and edited to fit local language in both cities; thus, in essence, the questionnaire is the same. The data from the two Brazilian cities were commissioned by a different private entity (WRI Brazil) through the local PT operators in both cities, this entity designed the original questionnaire focusing on generating a benchmark between Brazilian cities (Lindau et al., 2017). Next, we present the BRT-type systems analysed in this study.

The Santiago system is an integrated bus-Metro system (Transantiago) implemented in 2007. It currently operates through an integrated fare structure (one-way trip costs approximately 1 USD) that allows for minimal surcharges for transfers between buses and Metro, through a contactless smart card (Muñoz et al., 2008). Metro is the backbone of the PT system; however, a set of trunk bus-lines complement the Metro trunk network skeleton, and a set of feeder bus-lines handles shorter trips to feed both Metro and the trunk bus-lines. Santiago hosts close to 7 million people, and approximately 4 million use Transantiago each working day, while 60% of these trips use Metro in one of its legs. The speed of Metro ranges between 25 and 40 km/h depending on the line. The speed of the bus reaches 25 km/h in segregated corridors and between 8 and 15 km/h for regular bus services. To date, Transantiago has less than 80 km of designated bus-lines of the 300 km planned initially, and 118 km of Metro lines. As the segregated corridors lack off-level payment stations, they cannot be considered BRT corridors strictly. Studies by Batarce et al. (2016) and Allen et al. (2018a), focusing on the

Transantiago bus component, found that the reliability of waiting time is of significant concern for users and the reliability and waiting time is the most critical attribute regarding users' satisfaction. In this study we cover only the bus system, the least favoured by most PT users in Santiago.

Mexico City's Metrobus, is a BRT system that operates since 2006. By 2018, Metrobus was carrying 1.8 million passengers each working day, with a fleet of 720 buses on seven lines. As an example, Line 1 replaced 370 standard buses and microbuses with 210 articulated buses that travel at an average speed of 20 km/h. A one-way trip costs approximately 0.4 USD. Ticketing is done by prepaid smart cards; which travellers have to validate at turnstiles at the entry points to the separated bus platforms. Currently, Metrobus has a total extension of 125 km and 239 stations and is a successful implementation of a full BRT system in one of the biggest cities in the world.

In Brazil, BRT-systems are quite popular. Curitiba's Rede Integrada de Transporte was a pioneering system in 1974 and became the first BRT implemented in the world. The maximum peak-load capacity here is 22,500 passengers per hour. Since Curitiba, more than twelve BRT-type systems have been implemented in Brazil, making it one of the countries with more kilometres of BRT worldwide (BRT Centre of Excellence, 2018). For our case study, due to confidentiality agreements, we cannot disclose the names of the cities selected for the study; thus, we call them Brazil B1 and Brazil B2. These cities were chosen from a four-city sample because they contained the largest samples. Fortunately, we are allowed to share essential information about the cities that describe the two systems when the Surveys were collected (see Table 5-1).

Table 5-1: Brazilian cities BRT-system characteristics

City	Scope of survey	Travel time/day (min)	Income (USD)	Fare (USD)	Segregated lanes (km)	Passenger-km (PKM)	Fleet age (years)	Speed (km/h)
B1 (2015)	All city including BRT corridors	94.1	950	0.95	83.9	4730	6.05	N/A
B2 (2015)	BRT corridor only	113.7	650	1.05	39.0	9820	1.89	30

Source: WRI-Brazil

The survey was applied with the same methodology in all cities. The minimum number of questionnaires applied through simple random sampling, for a 95% confidence level, for systems carrying more than 20,000 daily passengers was approximately 400. Nevertheless, the four cities had larger samples, reducing sampling error and allowing for stratification. Samples were divided by bus line and time periods (AM Peak, PM Peak, Off-Peak) to ensure a random distribution of questionnaires proportional to the line's demand and ensuring that the selected lines covered 90% of the total system demand. Likewise, the time distribution was proportional to the time-periods' demand, ensuring that the samples covered 90% of the system demand (Lindau et al., 2017).

5.3.2. The survey

The benchmark survey was designed to compare satisfaction ratings between BRT-type systems in different cities in Latin America (Lindau et al., 2017). It measures the perception of bus system users. WRI-Brazil designed the survey based on an extensive literature review of practices and surveys applied in different cities and bus systems. The questionnaire was standardised and included additional modules to suit the needs of a particular city or transit system. However, in this study we focus only on the basic modules, that is, those that were required and completed in all four cities.

The survey is composed of the following four basic modules (Lindau et al., 2017):

a) Customer profile. These are sociodemographic characteristics: gender, age, education level, occupation status, driver's license, ownership of car, motorcycle and bike, and gross household income.

b) Usage profile. This includes frequency of bus use and, for the most frequent trip: the purpose (e.g. work, study, other), time of day, use of bus-only lanes or corridors, number of buses used on a one-way trip, and two-way total travel time.

c) Satisfaction. This includes 16 quality factors about the bus system. Five operational factors (S1-S5): access to transport, availability, speed, reliability, easiness to transfer; four comfort factors (S6-S9): at bus-stops, at stations, at integration terminals, and inside buses; three customer services factors (S10, S11, S15): customer services, information, and easiness to pay and reload travel card; three safety items (S12-S14): security, road safety, and exposure to noise and pollution; the last two factors refer to satisfaction with expenses with public transport (S16), and general satisfaction with the PT system (S17). All items were rated with a Likert (1932)-type scale from 1 (Very dissatisfied) to 5 (Very satisfied).

d) General perception. This includes eight additional attitudinal or perception components. In this study, we used only three of these: "taking the bus contributes to increasing my quality of life" (C1), "I can rely on the bus system for mobility" (C2), and "I would recommend the public transport system" (C8). All were rated with another 1-5 Likert type scale.

The sample consists of one application, performed in 2015, in the four different cities. It considers between 2,000 and 4,387 observations in each city, collected at bus stops. Only individuals of legal age (>18) were interviewed, obtaining their sociodemographic data and asked about their overall satisfaction with the BRT-type public transport system. In the next subsection, we present the sample travel and sociodemographic characteristics, and the average survey results for all cities, and for each city separately.

5.3.3. Sample and survey results

The data collected in 2015 totalled 10,688 observations, divided among Santiago (2387), Mexico City (4289), Brazil B1 (2012), and Brazil B2 (2000). Table 5-2 summarises the sample's main sociodemographic characteristics.

In the pooled sample there are more Females (54%), the biggest age group is under 30 (43%), and the smallest age group is over 45 (25%). Most PT users have a Primary or Secondary education (55%). However, this changes drastically in Mexico City, as most PT users there have University studies or a University degree (64%). In the pooled data most users do not possess a driver's license (67%). However, in Brazil B1, approximately half do not possess a driver's license (54%). The minority of respondents own an automobile (41%), yet again, Brazil B1 has opposite numbers, and the majority own a car (59%). Most respondents do not own a motorcycle (90%), and most do not own a bicycle (62%).

For the gross household income (GHI) categories, we defined the following ranges: a Low stratum representing a GHI of less than US$320; a Mid-Low stratum between US$350-750; a Mid-High stratum between US$750-1500 and a High stratum range for GHI over US$1500.

From the pooled data, the most prominent group is the Mid-Low stratum (45%), next, the Mid-High (21%), then, the High (18%), and finally the Low (16%). Mexico City and Brazil B1 resemble the pooled data's GHI distribution. However, there are differences in the other two cities. Santiago has a sizeable High stratum (34%) and a smaller Low stratum (8%). Brazil B1 has a different distribution, both Low (21%) and Mid-Low (55%) strata are larger than in the pooled data, and the High stratum is the smallest (6%) of all the cities. In a nutshell, according to the respondents of the survey, Santiago is the richest, Mexico and Brazil B2 are middle-income, and Brazil B1 is poor.

Table 5-2: Sample sociodemographic characteristics

CATEGORY	GROUP	All	Santiago	Mexico City	Brazil B1	Brazil B2
GENDER	Male	46.1	44.1	49.8	40.3	46.5
%	Female	53.9	55.9	50.2	59.7	53.5
AGE	<29	43.3	48.8	37.9	53.4	38.1
%	30-45	32.2	29.1	34.5	24.7	38.4
	>46	24.5	22.1	27.6	21.9	23.5
EDUCATION	Primary	12.8	9.6	7.9	13.9	25.7
%	Secondary	42.6	56.0	28.4	46.6	53.0
	Univ. Studies	21.0	19.4	25.8	22.3	11.5
	Univ. Degree	23.6	15.0	37.9	17.2	9.8
OCCUPATION	Other	10.7	7.5	15.4	9.6	5.4
%	Student	20.6	23.9	19.4	29.6	10.4
	Employed	57.6	63.9	52.0	47.8	71.5
	Professional	11.1	4.7	13.2	13.0	12.7
DRIVER'S LICENSE	Yes	32.9	23.0	35.0	45.8	27.2
%	No	67.1	77.0	65.0	54.2	72.8
AUTO AT HOME	Yes	41.3	31.9	42.6	58.9	32.4
%	No	58.7	68.1	57.4	41.1	67.6
MOTO AT HOME	Yes	10.4	5.6	13.1	11.7	8.9
%	No	89.6	94.4	86.9	88.3	91.1
BIKE AT HOME	Yes	37.6	46.4	32.5	41.0	34.8
%	No	62.4	53.6	67.5	59.0	65.2
INCOME	Low	15.8	8.4	17.7	15.6	20.5
%	Mid-Low	44.9	40.0	44.4	41.7	55.3
	Mid-High	20.9	17.4	22.2	25.0	18.0
	High	18.4	34.2	15.7	17.7	6.2

Table 5-3 summarises the sample’s main travel characteristics in the BRT-type systems. Pooling all the data, most respondents are frequent PT users (73%) and most travel to their workplaces (67%). Half of the respondents have a designated lane in their daily commute (48%), but this varies by city: for Santiago, it is 57%, for Mexico City, 19%, for Brazil B1, 59%, and for Brazil B2, 88%. Both Santiago and Brazil B1-B2 overrepresent respondents that use a designated lane corridor, while Mexico City’s respondents are more representative of the regular PT users. The design of this study was intended to survey more users of the BRT-corridors. For the pooled data, most respondents have to use two or three different bus-lines for their commute (63%). However, in Santiago, most respondents take a single bus-line for their commute (61%), and this, again, may have to do with the overrepresentation of designated lane corridor users.

In the travel time category, the most prominent group is the 60-120-minute two-way round trip (43%) and this percentage is similar across cities. However, Brazil B2 also has a significant >120-minute group (42%) and the highest travel time average of all. Finally, from the time-of-day category, the pooled data strata are well spread out, with the 9-13 being the smallest group (23%) and the 13-17 the most abundant (29%). However, there are some differences. Santiago was sampled mostly in peak periods, 5-9 (37%) and >17 (38%), and Mexico City was mostly sampled in off-peak periods, 9-17 (90%). This difference arises from the difference in time in which the bus system is mostly used in each city. For our purposes, all these differences will be considered when performing the SEM-MGA. When we treat the pooled data, it is crucial to account for these differences.

Table 5-3: Sample travel characteristics

CATEGORY	GROUP	All	Santiago	Mexico City	Brazil B1	Brazil B2
SAMPLE (n)	n	10688	2387	4289	2012	2000
%	%	100.0	22.4	40.1	18.8	18.7
FREQUENCY OF USE	5-7	73.1	82.8	69.6	63.7	78.0
%	3-4	11.3	9.8	12.1	13.0	9.7
	0-2	15.6	7.4	18.3	23.3	12.3
MOTIVE	Work	67.1	67.1	61.6	61.9	84.1
%	Study	18.2	22.3	18.3	21.0	10.3
	Other	14.7	10.6	20.1	17.1	5.6
DESIGNATED LANES	Yes	47.7	57.0	18.8	58.6	87.8
%	No	52.3	43.0	81.2	41.4	12.2
NUMBER OF BUSES	1	36.7	61.2	32.1	24.4	29.6
%	2	47.6	32.3	53.6	49.8	50.7
	>3	15.7	6.5	14.3	25.8	19.7
TRAVEL TIME	<60	27.3	29.8	31.9	25.9	15.6
%	60-120	43.0	47.2	38.8	47.3	42.6
	>120	29.7	23.0	29.3	26.8	41.8
TIME OF DAY	05-09	23.5	36.7	0.0	30.6	36.2
%	09-13	22.7	12.8	42.8	14.0	15.7
	13-17	29.0	12.6	47.6	26.9	25.5
	>17	24.8	37.9	9.7	28.5	22.6

The satisfaction average survey results for the pooled data, and for each city independently are shown in Table 5-4. We will comment on the results from the pooled data.

Table 5-4: Survey satisfaction results: mean values (1-5 Likert scale)

DOMAIN		Service quality attribute	All	Santiago	Mexico City	Brazil B1	Brazil B2
Reliability	S1:	Access to transport	3.64	3.39	3.84	3.44	3.71
	S2:	Availability: time interval	3.33	3.24	3.49	2.96	3.48
	S3:	Speed	3.49	3.31	3.70	3.02	3.73
	S4:	Reliability: arrival on time	3.42	3.27	3.59	3.08	3.54
	S5:	Easiness to transfer	3.55	3.34	3.74	3.42	3.50
Comfort	S6:	Comfort at bus stops	3.12	3.11	3.63	2.43	2.76
	S7:	Comfort at stations	3.25	3.20	3.63	2.73	3.03
	S8:	Comfort at terminals	3.26	3.19	3.59	2.85	3.05
	S9:	Comfort inside buses	3.20	3.13	3.60	2.76	2.88
Customer services	S10:	Customer services	3.43	3.27	3.55	3.57	3.22
	S11:	Customer information	3.31	3.21	3.39	3.41	3.13
	S15:	Easiness to pay and reload	3.54	3.34	3.80	3.52	3.24
Safety	S12:	Security: thefts and assaults	2.84	2.78	3.54	2.09	2.14
	S13:	Road safety	3.25	3.02	3.62	2.86	3.13
	S14:	Exposure: noise and pollution	3.15	2.81	3.56	2.68	3.15
Overall satisfaction	S16:	Expenses: public bus system	3.10	2.70	3.75	2.34	2.92
	S17:	Satisfaction: bus system	3.39	3.01	3.79	3.05	3.33
Loyalty	C1:	Bus increases quality of life	3.48	3.18	3.83	3.09	3.47
	C2:	I rely on the bus for mobility	3.56	3.35	3.76	3.36	3.56
	C8:	I recommend the bus system	3.40	2.97	3.74	3.25	3.31

All results are divided into specific satisfaction domains: reliability (S1-S5), comfort (S6-S9), customer services (S10, S11, S15), safety (S12-S14), overall satisfaction (S16-S17), and loyalty (C1, C2, C8). From the satisfaction domain values, the best-rated is reliability (3.49), and the worst is safety (3.08), on a 1-5 Likert scale.

From the reliability domain, the best-rated item is access to transport (S1, 3.64) and the worst is the availability-time interval (S2, 3.33). From the comfort domain, the best-rated is comfort at integration terminals (S8, 3.26) and the worst is comfort at bus-stops (S6, 3.12). In the customer services domain, the best-rated is easiness to pay fares and reload travel card (S15, 3.54) while the worst is the customer services item (S11, 3.31). In the safety domain, road safety (S13, 3.25) obtains the best score while security against thefts (S12, 2.84), the worst. In fact, this is the only item that scores under three as an average, and this indicates that, on average, the sample perceives the existence of a crime problem, in all cities.

In the overall satisfaction domain, the satisfaction with expense in public transport is treated as a value item, following Allen et al. (2018b). As such it is considered an overall satisfaction item, its average (S16, 3.10) is lower than the general satisfaction with the PT bus system (S17, 3.39). Finally, from the Loyalty domain, the best-rated item is I can rely on the bus system for mobility (C2, 3.56) while the worst is I would recommend the PT bus system (C8, 3.40). All Loyalty items have similar scores. Additionally, taking the averages of the overall satisfaction domain items (S16, S17), we can see considerable differences among the cities. The overall average is 3.24, while for the individual cities only Mexico City is above the average (3.77), while Santiago (2.85), Brazil B1 (2.70), and Brazil B2 (3.12) are below average. From Table 5-4, we notice differences between the cities in all the domains. We highlight that, notably, for Brazil B1, both the comfort and safety domains score rather low, as all average values for the individual items are below 2.86. We can conclude that this city presents different conditions in these two domains, compared to the other three.

5.4. Model Results

In this section, we present our modelling results. These include: (i) PCA for the satisfaction items, (ii) SEM continuous and SEM ordinal models, (iii) SEM-MGA and SEM-MG models, and finally (iv) the SEMM (latent class) models. From the PCA we will obtain the general structure of the satisfaction domains. We will test both types of SEM models, continuous and ordinal, to gain robustness in our results. We also test whether we obtain similar results from both types of models. Next, we will perform two MGA, again, with a continuous and an ordinal approach. From these analyses, we will test for different satisfaction models in the following conditions: (i) between the cities with the pooled data, (ii) among the different travel conditions and sociodemographic characteristics with the pooled data, and finally for (iii) Santiago and Mexico City, for both travel and sociodemographic conditions. All model results were obtained using the R software (R Core Team, 2013) and its associated library packages; in particular, we used the Lavaan package for R (Rosseel, 2012) in all SEM models.

In essence, we want to unveil which categorical variables generate heterogeneous satisfaction models (i.e. significative differences in the coefficients). Once we know which categorising variables yield different models, we run the SEM-MG and obtain regression results to assess such differences. Subsequently, we compare the mean values in the domains from the different groups from the categorical variables. Therefore, we test the hierarchy of transit needs, by assessing whether the different satisfaction models coincide with differences in the mean levels of the satisfaction domains.

We present six different cases from the pooled data, and the cities analysed where there are differences in the models, allowing to test our working hypothesis. Last, we present the two SEMM models, one for Santiago and another for Mexico City. With two latent classes for each city, we can compare the resulting models with the SEM-MG models. We present the results in this Section and discuss them further in Section 5-5.

5.4.1. Principal component analysis: satisfaction domains

Initially, we want to reveal how users perceive the different satisfaction constructs. For this, best practice is to run an exploratory PCA (Hoyle, 2012, Jolliffe, 2014) on the satisfaction items, and the additional perceptions items, separately. By doing this, we can determine which items represent specific latent constructs; in other words, which items represent one construct as a whole. Henceforth, we performed two separate PCA on the set of satisfaction items and the perception items. For the first, according to the Kaiser (1960) rule, we obtained four components: Reliability (RELI), Comfort (COMF), Customer services (CSER) and Safety (SAFE). The criterion used was a resulting absolute loading >0.4 (see Table 5-5).

Table 5-5: PCA: Satisfaction items

	Service quality attribute	**RELI**	**COMF**	**CSER**	**SAFE**
S1	Access to transport	**-0.43**	0.02	0.06	0.04
S2	Availability: time interval between buses: time and places I need	**-0.51**	-0.03	-0.07	0.00
S3	Speed	**-0.48**	-0.03	-0.06	-0.09
S4	Reliability: arrival on time	**-0.46**	0.02	0.01	-0.01
S5	Easiness to transfer between bus lines and other means of transport	**-0.32**	0.14	0.17	0.17
S6	Comfort at bus stops	-0.01	**0.47**	-0.03	-0.07
S7	Comfort at stations	0.01	**0.54**	0.00	0.02
S8	Comfort at integration terminals	0.00	**0.53**	0.02	0.04
S9	Comfort inside buses	0.00	**0.39**	0.01	-0.12
S10	Customer services	0.03	0.05	**0.60**	0.02
S11	Customer information	0.00	-0.01	**0.62**	0.02
S15	Easiness to pay fares and reload travel card	-0.01	-0.08	**0.45**	-0.17
S12	Security: thefts and assaults on the way to bus stops and on the bus	0.05	0.13	-0.08	**-0.53**
S13	Road safety	-0.03	-0.05	0.09	**-0.55**
S14	Exposure to noise and pollution produced by buses	-0.04	-0.04	0.01	**-0.57**

One item, Easiness to transfer (S5), only loaded 0.32 but we decided to keep it for theoretical reasons, as it was the only item that measured an indicator of transferability between PT modes.

Initially, there were eight different perception items. However, when performing the PCA, we unveiled only one theoretically justified construct, Loyalty. As we intend to comprehend differences in relevance towards satisfaction, we felt unnecessary to consider the other items. We formed the Loyalty (REC) with three of the perception (C1, C2, C8) items. We deem interesting to assess which satisfaction construct also affects Loyalty, following Allen et al. (2018c), for possible insights and policy recommendations. In the following subsection, we will present the SEM models.

5.4.2. Structural equation models: overall satisfaction and loyalty

Once defined the composition of the satisfaction constructs, we can construct a working SEM model. Best practice is to run a Measurement model first, and then the SEM model. We will present only the resulting SEM models, which contain both the measurement model and the regression results. The Measurement model structure and the SEM model structure are shown in Figure 5-2; the ovals represent the latent constructs, and the rectangles represent the single survey items. Left is the measurement system, and to the right the SEM.

We will use this structure for both the SEM numeric and SEM ordinal models. The SEM numeric model implies that all satisfaction items are treated with a numeric range; the SEM ordinal model, instead, implies that the satisfaction items are treated with an ordinal range. For both SEM models, we hypothesise first that all satisfaction constructs regress on the Overall Satisfaction (OSV). Additionally, we hypothesise that the satisfaction constructs may also affect the Loyalty construct; as such, we construct direct effects from the satisfaction constructs to Loyalty, as well as from Overall Satisfaction to Loyalty, following Allen et al. (2018c).

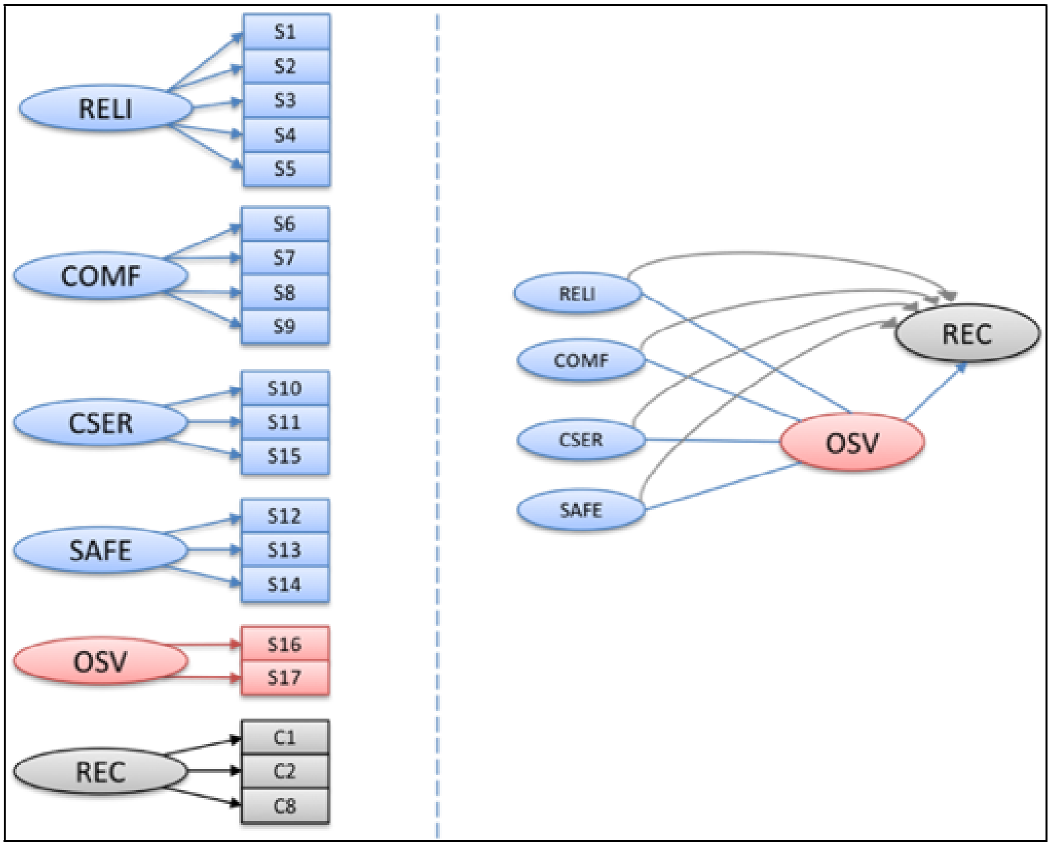

Figure 5-2: Schema for Measurement and SEM models' structure

Table 5-6 presents the results of the SEM continuous model using the pooled data: estimates, standard errors (S.E.), Z-values, and standardised coefficients (Std.Coeff.). The latter refer to how many standard deviations a dependent variable will change, per standard deviation increase in the predictor variable. From the measurement model, we can conclude that all latent variables have high reliability (Std.Coeff. >0.6) in the measurement, as all values lie in the 0.63-0.87 range. Note also the specific items that load more on the latent constructs: for RELI, it is Reliability (S4), for COMF, it is Comfort at stations (S7), for CSER, it is Customer services (S10), and for SAFE, it is Road Safety (S13). For OSV, it is general satisfaction with PT (S17), and for REC, it is the intention to recommend (C8).

Table 5-6: SEM continuous results for pooled data: All cities

Measurement model	Estimate	S.E.	Z-value	Std.Coeff.	R^2
RELI					
S1	1.00			0.69	0.48
S2	1.09	0.02	61.46	0.67	0.45
S3	1.17	0.02	67.98	0.75	0.56
S4	1.21	0.02	69.22	0.77	0.59
S5	0.96	0.02	58.42	0.63	0.40
COMF					
S6	1.00			0.81	0.65
S7	1.04	0.01	101.31	0.87	0.76
S8	0.96	0.01	96.22	0.84	0.70
S9	0.86	0.01	78.73	0.71	0.51
CSER					
S10	1.00			0.74	0.55
S11	0.98	0.02	61.86	0.69	0.48
S15	0.84	0.02	55.03	0.61	0.37
SAFE					
S12	1.00			0.70	0.49
S13	0.97	0.01	71.06	0.80	0.64
S14	0.98	0.01	66.80	0.74	0.55
OSV					
S16	0.42	0.01	50.74	0.73	0.53
S17	0.42	0.01	50.82	0.87	0.75
REC					
C1	1.00			0.63	0.39
C2	0.98	0.02	54.02	0.68	0.46
C8	1.16	0.02	56.53	0.73	0.54

Table 5-6: SEM continuous results for pooled data: All cities (continued)

Regressions	Estimate	S.E	Z-value	Std.Coeff.
OSV $R^2 = 0.77$				
RELI	0.83	0.04	19.28	0.27
COMF	0.32	0.03	11.19	0.14
CSER	0.64	0.05	13.20	0.24
SAFE	0.82	0.04	19.41	0.36
REC $R^2 = 0.68$				
OSV	0.18	0.01	23.11	0.55
RELI	0.27	0.02	15.25	0.26
CSER	0.07	0.02	3.74	0.08

For both regressions the explained variance (R2) is relatively high indicating reasonable models; explicitly, it is 0.77 for the Overall Satisfaction regression, and 0.68 for the Loyalty regression. For the structural coefficients, values below 0.1 are considered very low, between 0.1 and 0.3, low, between 0.3 and 0.5 moderate, and above 0.5 high (Allen et al., 2018c; Currie and Delbosc, 2017). We will further discuss these results in Section 5. Next, we ran separate models for all the cities and obtained fit indices and present them in Table 5-7.

Table 5-7: SEM continuous fit indices: Pooled data and cities

CITY	CFI	TLI	GFI	AGFI	RMSEA
Pooled data	0.943	0.932	0.945	0.926	0.059
Santiago	0.937	0.923	0.919	0.892	0.073
Mexico City	0.937	0.924	0.930	0.906	0.066
Brazil B1	0.936	0.923	0.965	0.953	0.043
Brazil B2	0.845	0.812	0.889	0.851	0.080
Cut-offs	>0.95	>0.95	>0,95	>0.95	<0.08

The objective is to assess whether the separate cities have a similar SEM model structure. From Table 5-7, we notice that for the first three cities, Santiago, Mexico City and Brazil B1, the fit indices indicate adequate models, as the CFI, TLI, AGFI and RMSEA present well-adjusted values. The cut-off values used were recommended by Hu and Bentler (1999).

We notice, explicitly, that the CFI is close to 0.94 for these three cities, being 0.95, the threshold value for an acceptable fit. The RMSEA is less than 0.08 for all cities, so even though Brazil B1 presents relatively worst fit indices, it still complies with the RMSEA which the strictest cut-off value for determining model fit. We conclude that the SEM structures adjust satisfactorily.

Table 5-8 presents the regression results for each city. The explained variance for the OSV regression is above 0.74 for all cities, indicating a reasonable explanation of the dependent OSV variable. The Loyalty regression has an explained variance over 0.66 for all cities except Brazil B1, an indication that for this city Loyalty is not so well explained. Keep in mind that this is the most under satisfied city of the four.

Notice also differences among the standardised coefficients between the different cities. Again, we will further discuss these results in the next section. We hypothesise that the differences are due to differences in the mean levels of perceptions. Next, we present the same SEM structure using an ordinal approach; this will allow comparing both approaches.

Table 5-8: SEM continuous results: Cities

Regressions	**Estimate**	**S.E.**	**Z-value**	**Std.Coeff.**
Santiago OSV $R^2 = 0.74$				
RELI	0.95	0.10	9.86	0.35
COMF	0.22	0.06	3.86	0.10
CSER	0.51	0.10	4.86	0.20
SAFE	0.70	0.07	9.83	0.31
Santiago REC $R^2 = 0.78$				
OSV	0.14	0.01	11.49	0.40
RELI	0.44	0.03	12.82	0.46
COMF	0.07	0.02	3.58	0.09
Mexico City OSV $R^2 = 0.78$				
RELI	0.71	0.08	8.89	0.23
CSER	1.55	0.15	10.57	0.61
SAFE	0.25	0.09	2.85	0.10
Mexico City REC $R^2 = 0.66$				
OSV	0.15	0.02	9.72	0.48
RELI	0.15	0.03	5.33	0.16
CSER	0.19	0.04	4.70	0.23
Brazil B1 OSV $R^2 = 0.76$				
RELI	1.96	0.25	7.82	0.49
COMF	0.92	0.12	7.88	0.32
CSER	1.00	0.18	5.48	0.27
Brazil B1 REC $R^2 = 0.23$				
OSV	0.07	0.02	3.40	0.25
RELI	0.18	0.07	2.66	0.17
CSER	0.11	0.06	2.05	0.12

Table 5-8: SEM continuous results: Cities (continued)

Brazil B2 OSV $R^2 = 0.74$				
RELI	0.68	0.12	5.90	0.16
COMF	0.15	0.08	2.00	0.06
CSER	1.13	0.14	8.08	0.45
SAFE	1.70	0.29	5.82	0.38
Brazil B2 REC $R^2 = 0.88$				
OSV	0.07	0.01	6.54	0.76
RELI	0.04	0.01	2.78	0.09
SAFE	0.08	0.03	2.36	0.18

Table 5-9 presents the SEM ordinal Probit results for the pooled data. From the ordinal measurement model, we can conclude that all latent variables have high reliability (Std.Coeff. >0.6) in the measurement, as all values lie in the 0.68-0.89 range. The measurement in this case is slightly better than in the numeric counterpart.

It is important to notice the specific items that load more on the latent constructs. For RELI, the highest loading item is Reliability (S4), for COMF, it is Comfort at stations (S7), for CSER, it is Customer services (S10), and for SAFE, it is Road Safety (S13). For OSV, it is general satisfaction with PT (S17), and for REC, it is the intention to recommend (C8). In this sense then, there are no differences between the ordinal and numeric models. For both regressions the explained variance (R^2) is high indicating an adequate model; explicitly it is 0.81 for the Overall Satisfaction regression, and 0.70 for the Loyalty regression; so, there is an improvement in relation with the numeric counterpart. Next, we ran separate ordinal models for all the cities (see Table 5-10). We will assess whether the separate cities have a similar SEM model structure.

Table 5-9: SEM ordinal Probit results Pooled data: All cities

Measurement model	Estimate	S.E.	Z-value	Std.Coeff.	R^2
RELI					
S1	1.00			0.77	0.59
S2	0.90	0.01	108.70	0.69	0.48
S3	1.02	0.01	129.05	0.79	0.62
S4	1.05	0.01	129.69	0.81	0.65
S5	0.95	0.01	112.57	0.73	0.54
COMF					
S6	1.00			0.86	0.74
S7	1.03	0.00	231.22	0.89	0.79
S8	1.00	0.00	227.05	0.86	0.74
S9	0.94	0.01	171.22	0.80	0.65
CSER					
S10	1.00			0.77	0.60
S11	0.93	0.01	106.46	0.72	0.52
S15	0.93	0.01	99.10	0.72	0.51
SAFE					
S12	1.00			0.76	0.57
S13	1.10	0.01	121.55	0.83	0.69
S14	1.04	0.01	115.26	0.78	0.61
OSV					
S16	0.34	0.01	65.11	0.78	0.61
S17	0.39	0.01	63.96	0.90	0.81
REC					
C1	1.00			0.68	0.46
C2	1.08	0.01	92.47	0.73	0.54
C8	1.19	0.01	91.85	0.81	0.65

Table 5-9: SEM ordinal Probit results Pooled data: All cities (continued)

Regressions	Estimate	S.E.	Z-value	Std.Coeff.
OSV $R^2 = 0.81$				
RELI	0.70	0.03	23.09	0.23
COMF	0.36	0.03	12.62	0.13
CSER	0.88	0.04	21.33	0.30
SAFE	1.07	0.05	22.99	0.35
REC $R^2 = 0.70$				
OSV	0.18	0.00	43.35	0.61
RELI	0.24	0.01	22.00	0.27

Table 5-10: SEM ordinal Probit fit indices: Pooled data and cities

CITY	CFI	TLI	GFI	AGFI	RMSEA
Pooled data	0.995	0.994	0.996	0.993	0.058
Santiago	0.996	0.995	0.997	0.994	0.058
Mexico City	0.995	0.994	0.995	0.992	0.069
Brazil B1	0.984	0.981	0.992	0.987	0.043
Brazil B2	0.936	0.923	0.966	0.942	0.111
Cut-offs	>0.95	>0.95	>0.95	>0.95	<0.08

From Table 5-10, we notice that for the first three cities, Santiago, Mexico City and Brazil B1, the fit indices confirm very reasonable models, as the CFI, TLI, AGFI and RMSEA present excellent fit values. We notice, explicitly, that the three cities comply with all the cut-off values. Specifically, the RMSEA is less than 0.08 in all cases. However, Brazil B2 does not comply with the RMSEA, and its other fit indices are also lower. Still, our primary objective is to compare the regression results, taking caution on the specific SEM model for Brazil B2. To perform the robustness checks, we performed the MGA for both continuous and ordinal model frameworks.

Thus, we were able to determine which variables were candidates for presenting heterogeneous satisfaction models within the subpopulations. Next, we compare the mean satisfaction scores within the domains, and test for the hierarchy. We present the MGA results and the testing of our hypothesis in the next subsection.

5.4.3. Multi-group analyses: testing Maslow's hierarchy of transit needs

We ran the MGA analysis with the pooled data, and then we analysed two cities, Santiago and Mexico City. There are three reasons for choosing these cities. First, they have the larger sample sizes, second, their mean values of satisfaction (OSV) are entirely different (Santiago, 2.85; Mexico City, 3.77), and last, we know precisely, which city they are, unlike the Brazilian cities. First, we present the two SEM-MGA: continuous in Table 5-11 and ordinal in Table 5-12. In bold are those variables that cross over the thresholds proposed and, for this reason, require a SEM-MG model.

We follow the methodology proposed by Allen et al. (2018c) to discover which moderating variables bring heterogeneity to the SEM satisfaction-loyalty models. ΔCFI is the difference between the CFI of a SEM-MG model with parameters allowed to be different for all groups, and a model where they are restricted to be the same.

If ΔCFI is large, it means that the SEM-MG with different parameters provides a better representation of the phenomenon, in other words, different parameters for each group are warranted). For the SEM-MGA continuous, we used a threshold of >0.01, which is the recommended value in the literature for a continuous type of analysis (Cheung and Rensvold, 2002). However, for the SEM-MGA ordinal, we did not find any recommended values in the literature and tested several possibilities, finding that we needed a stricter threshold (>0.003). We chose 0.003

as the other fit indexes started to deteriorate at this value (i.e. TLI, AGFI, RMSEA), also. We ran both frameworks, continuous and ordinal, to provide robustness to our results.

Table 5-11: SEM-MGA continuous results: Pooled data and two cities (ΔCFI)

CATEGORY	All	MGA	Santiago	MGA	Mexico City	MGA
CITY	**0.141**	**Yes**	----	----	----	----
FREQ OF USE	0.003	No	0.003	No	0.004	No
MOTIVE	0.004	No	0.005	No	0.006	No
DD LANE	**0.020**	Yes	**0.014**	Yes	**0.011**	Yes
No. OF BUSES	0.007	No	0.002	No	**0.012**	Yes
TRAVEL TIME	**0.011**	Yes	**0.011**	Yes	**0.014**	Yes
TIME OF DAY	**0.025**	Yes	0.004	No	0.004	No
GENDER	0.002	No	0.000	No	0.001	No
AGE	0.005	No	0.002	No	0.007	No
EDUCATION	**0.020**	Yes	0.005	No	0.005	No
OCCUPATION	0.004	No	0.008	No	0.006	No
LICENSE	0.003	No	0.001	No	0.005	No
INCOME	**0.010**	Yes	0.008	No	**0.013**	Yes
Cut-offs (ΔCFI)	>0.01	----	>0.01	----	>0.01	----

Once we determine which variables moderated the models, we computed the SEM-MG models to assess the differences in the regression parameters, among subgroups. We present six specific cases in Tables 5-13 to 5-18.

From Tables 5-11 and 5-12 we notice that the categorical variables that moderate the satisfaction models for the numeric and ordinal cases are similar. For example, for the SEM-MGA pooled data numeric, the categories that resulted needing an

MGA are CITY, DD (designated) LANE, TRAVEL TIME, TIME OF DAY, EDUCATION, and INCOME; in the SEM-MGA pooled data ordered, the only missing category is INCOME. For Santiago, only DD LANE and TRAVEL TIME appear in the numeric case. For the ordinal case, the same results are obtained, except for DD LANE.

Table 5-12: SEM-MGA ordinal Probit results: Pooled data and two cities (ΔCFI)

CATEGORY	**All**	**MGA**	**Santiago**	**MGA**	**Mexico City**	**MGA**
CITY	**0.047**	**Yes**	----	----	----	----
FREQ OF USE	0.001	No	0.001	No	0.001	No
MOTIVE	0.002	No	0.002	No	0.001	No
DD LANE	**0.008**	Yes	0.002	No	0.001	No
No. OF BUSES	0.002	No	0.002	No	**0.003**	Yes
TRAVEL TIME	**0.004**	Yes	**0.003**	Yes	0.002	No
TIME OF DAY	**0.015**	Yes	0.000	No	0.000	No
GENDER	0.001	No	0.000	No	0.000	No
AGE	0.002	No	0.001	No	0.001	No
EDUCATION	**0.008**	Yes	0.002	No	0.001	No
OCCUPATION	0.002	No	0.002	No	0.001	No
LICENSE	0.001	No	0.000	No	0.001	No
INCOME	0.002	No	0.002	No	**0.003**	Yes
Cut-offs (ΔCFI)	>0.003	----	>0.003	----	>0.003	----

Likewise, in Mexico City, for the ordinal model NO. OF BUSES and INCOME appear, while for the numeric model we also obtain DD LANE and TRAVEL TIME. We consider these results robust as we notice that the same categorising variables appear when we compare both frameworks, with the exception of one or two variables. This slight difference is not unexpected, as the analyses are done using different scales (i.e. continuous/numeric vs ordinal).

Next, we analyse the resulting MGA models for three data sets: pooled data, Santiago, and Mexico City. In the next six tables, we consider each specific case, and present the mean values of satisfaction, overall satisfaction and loyalty domains, among subgroups. Then, we present the regression coefficients (Std.Coeff.). The objective is to assess whether patterns of priority over specific domains exist according to the mean satisfaction levels across domains. We will comment briefly the results obtained for each table (Tables 5-13 to 5-18). However, we will further discuss these results in Section 5.5.

In Table 5-13, we present the regression coefficients for both the numeric and ordinal MGA. We notice that the two cities with the highest scores in the reliability domain, Mexico City (3.67) and Brazil B2 (3.59), also have the lowest regression coefficients (Std.Coeff.) for this latent construct (0.23 and 0.16 in the numeric regression). This fact initially supports the hierarchy of transit needs. We stated that if reliability was fulfilled, it would not be as relevant; this is the case in Mexico City and Brazil B1, which have the highest scores in reliability. Conversely the other two cities, which are lower scoring in reliability, Santiago (3.31) and Brazil B1 (3.19), show higher relevance for reliability, 0.35 and 0.49, in the numeric regression.

For the safety domain, only Mexico City has a satisfactory mean value (3.57) and we notice that only Santiago and Brazil B2 have a high regression coefficient. We conclude that for these three cities, Santiago, Brazil B2 and Mexico City, there is support for the hierarchy of transit needs, i.e. safety is relevant in Santiago and Brazil B2 (low levels of safety), and not in Mexico City (high levels of safety). From Table 5-13, we notice that Brazil B1 scores low in the comfort and safety domains but is only relevant in comfort. This counterintuitive result led us first to postulate that comfort may have some association with the safety appraisal for this city; we further comment on this in the next paragraphs.

Table 5-13: SEM-MG case analysis 1: City specific pooled data

Mean score (Likert 1-5)	**All**	**Santiago**	**Mexico City**	**Brazil B1**	**Brazil B2**
RELI	3.49	3.31	3.67	3.19	3.59
COMF	3.21	3.16	3.61	2.69	2.93
CSER	3.42	3.27	3.58	3.50	3.20
SAFE	3.08	2.87	3.57	2.54	2.81
OSV	3.24	2.85	3.77	2.70	3.12
REC	3.48	3.17	3.78	3.23	3.44
Continuous OSV (Std.Coeff.)					
RELI	0.27	0.35	0.23	0.49	0.16
COMF	0.14	0.10	----	0.32	0.06
CSER	0.24	0.20	0.61	0.27	0.45
SAFE	0.36	0.31	0.10	----	0.38
Ordinal OSV (Std.Coeff.)					
RELI	0.23	0.33	0.16	0.45	0.13
COMF	0.13	0.09	----	0.32	0.05
CSER	0.30	0.18	0.81	0.24	0.50
SAFE	0.35	0.37	----	0.07	0.34

We expected the next attributes in order of relevance to be customer services and comfort. For the customer services domain, the two cities with the highest scores are Mexico City and Brazil B2. However, in relative terms, as Brazil B2 has the reliability domain covered, it is expected to have a higher coefficient in customer services. We should note that both customer services and comfort are excitement attributes, and as such we expect them to play an important role only when the essential attributes are fulfilled. We found support for this assertion, as both Mexico City (0.61) and Brazil B2 (0.45) value customer services the highest; and, as we stated previously, they had the reliability domain mostly fulfilled.

In the comfort domain, only Brazil B1 has a high coefficient. Given this result and the previous one about safety, we decided to analyse the composition of this variable in this city. The worst rated item was comfort at bus stops: lighting, *protection*, cleanliness, number of people (S6, 2.43). As Brazil B1 has the lowest mean satisfaction score in Comfort (2.69), it is not unreasonable to hypothesise that users of the PT system may not feel safe at bus stops. Somehow this correlates with safety due to the *protection* part of the item; in fact, the first three comfort items (S6-S9) included the word protection in their statement, see Lindau et al. (2017). We are highly suspect that the safety domain is scored very low in Brazil B1 and it does not correlate well with the OSV (Overall satisfaction-value). We hypothesise that the comfort domain gained this relevance due to the protection aspect of the comfort items in the survey, for this particular city, Brazil B1.

In summary, we found support for the hierarchy of transit needs in all cities, excepting Brazil B1. This mixed result led us to believe that the safety domain may be correlated with the comfort domain in this city. In fact, the safety domain was not significant there; however, users scored this variable the lowest from all the domains. We conclude that there may be some association between both domains, and this was enhanced due to a general unsafe feeling in the city. All the previously mentioned results were supported by both types of regression: numeric and ordinal, giving robustness to the results obtained.

In Tables 5-14 to 5-18, we only present the regression coefficients for the numeric SEM-MG for conciseness; nevertheless, results from the ordinal regressions are entirely comparable. In Table 5-14, we analyse the Income grouping variable in the pooled data. We notice that the two strata with the highest regression coefficients for the reliability domain are the Mid-High and the High strata. In fact, these two groups behave similarly. Both have high coefficients for reliability and safety.

Table 5-14: SEM-MG case analysis 2: Income specific pooled data

Mean score (Likert 1-5)	**All**	**Low**	**Mid-Low**	**Mid-High**	**High**
RELI	3.49	3.55	3.48	3.49	3.45
COMF	3.21	3.24	3.20	3.23	3.19
CSER	3.42	3.44	3.40	3.43	3.47
SAFE	3.08	3.11	3.07	3.05	3.09
OSV	3.24	3.26	3.24	3.29	3.20
REC	3.48	3.58	3.50	3.46	3.35
Continuous OSV (Std.Coeff.)					
RELI	0.27	0.22	0.24	0.34	0.35
COMF	0.14	0.14	0.14	0.13	0.12
CSER	0.24	0.40	0.26	0.11	0.17
SAFE	0.36	0.28	0.37	0.39	0.36

If, for a moment, we think about the hierarchy of transit needs, it is intuitive that higher income users will have higher needs (i.e. higher expectations), and thus, will require more from the essential/basic attributes. Moreover, given the same conditions they should give more relevance to the reliability domain, first, and next to the safety domain, according to the proposed hierarchy, this result is confirmed in Table 5-14. If we move to the Mid-Low strata, we notice that now reliability loses some relevance, safety stays the same, and customer services gains relevance. In the same spirit as before, this group has fewer requirements for reliability; thus, they assign relevance to customer services. Now, as we move to the Low strata, safety loses some relevance, and customer services gains some more. Once again, this group has fewer requirements for reliability and safety, and as such gain it in customer services. In a nutshell, higher income groups have differentiated additional expectations in the essential/basic attributes than the lower income groups.

It is fascinating to notice that, across the board, comfort has an almost identical regression weight, which is relatively low. According to our working hierarchy, we postulate that as users do not have the essential/basic attributes fulfilled (i.e. reliability and safety); they do not assign importance to comfort. Notice that customer services is valued highly by the low-income group, and conversely have lower values for reliability and safety than the other groups. As we are working with the pooled data, the differences are subtle but present, as they involve users from different cities. In summary, we find evidence for the hierarchy of transit needs when grouping by income and find specific differences due to the differences in the level of requirements among the users of different income strata.

In Table 5-15, we analyse the Travel Time grouping variable in the pooled data.

Table 5-15: SEM-MG case analysis 3: Travel time specific pooled data

Mean score (Likert 1-5)	**All**	**<60**	**60-120**	**>120**
RELI	3.49	3.63	3.49	3.35
COMF	3.21	3.34	3.16	3.16
CSER	3.42	3.52	3.36	3.43
SAFE	3.08	3.24	3.01	3.03
OSV	3.24	3.42	3.16	3.19
REC	3.48	3.58	3.45	3.42
Continuous OSV (Std.Coeff.)				
RELI	0.27	0.26	0.22	0.37
COMF	0.14	0.16	0.19	0.06
CSER	0.24	0.25	0.24	0.22
SAFE	0.36	0.35	0.37	0.35

Notice that the >120 minutes group has the highest regression coefficients for reliability. This group may have the most complex journeys and probably value reliability higher because they experience longer trips, which are prone to be unreliable. This group has a higher requirement and give even higher relevance to the most basic reliability attribute, similar to what we discussed for the Income grouping above. Next, in Table 5-16, we analyse the Time of Day grouping variable in the pooled data.

Table 5-16: SEM-MG case analysis 4: Time of day specific pooled data

Mean score (Likert 1-5)	**All**	**5-9**	**9-13**	**13-17**	**>17**
RELI	3.49	3.34	3.59	3.59	3.41
COMF	3.21	2.99	3.41	3.34	3.09
CSER	3.42	3.31	3.49	3.51	3.37
SAFE	3.08	2.78	3.31	3.30	2.89
OSV	3.24	2.88	3.54	3.49	3.03
REC	3.48	3.28	3.66	3.64	3.31
Continuous OSV (Std.Coeff.)					
RELI	0.27	0.37	0.20	0.27	0.24
COMF	0.14	0.09	0.16	0.18	0.14
CSER	0.24	0.27	0.37	0.20	0.31
SAFE	0.36	0.27	0.28	0.36	0.34

One group values the reliability domain the highest, the 5-9 (AM-Peak) group. It is intuitive to think that this group gives the highest relevance to reliability as most user travel to work and study in the morning, and quite often need to arrive at a specific time. Because of this, they assign more importance to reliability and less to comfort, than the other time slot groups. Again, the AM-Peak has a higher requirement, and thus they give even higher relevance to the most basic reliability attribute, similar to what we discussed for the Income and Travel Time groupings, before.

In Table 5-17 we analyse the Designated Lanes grouping variable for Santiago.

Table 5-17: SEM-MG case analysis 5: Designated lanes in Santiago

Mean score (Likert 1-5)	**Santiago**	**Yes**	**No**
RELI	3.31	3.33	3.29
COMF	3.16	3.15	3.16
CSER	3.27	3.36	3.16
SAFE	2.87	2.92	2.81
OSV	2.85	2.87	2.83
REC	3.17	3.10	3.25
Continuous OSV (Std.Coeff.)			
RELI	0.35	0.46	0.26
COMF	0.10	0.05	0.13
CSER	0.20	0.10	0.33
SAFE	0.31	0.32	0.30

PT users of segregated bus lanes value the reliability domain higher. From the regression weights, we notice a similar dynamic than before. Designated Lane users value reliability more, and they value customer services or comfort, less. Again, we find support for the hierarchy of transit needs when grouping by use of Designated Lanes. As the PT users of the Designated Lanes are not fully satisfied with the overall service (keep in mind that the designated lanes in Santiago are not full BRT corridors), they value the most fundamental attribute, reliability, the highest; while assigning less relevance to hedonic attributes such as customer services and comfort. Interestingly, the safety domain has a similar coefficient for both groups (a relatively high one, 0.32, for the Designated Lanes group and 0.30 for the other one). Accordingly, the safety attribute is unfulfilled also.

We present our most exciting results in Table 5-18, where we analyse the Travel Time grouping variable in Mexico City. Here, as in general, Mexico City is the city with the highest satisfaction scores across all items; most users are satisfied, and so they assign most of the relevance to the customer services domain. However, by separating by travel time, we find that the >120-minute group behaves differently; they assign more weight to the reliability domain, less to customer services, and more to safety. We hypothesise that this group has to make more transfers or have long waiting times in their commutes, making them more prone to precarious waiting situations. Therefore, they assign more relevance to safety and reliability. Again, we find support for the hierarchy of transit needs, as the reliability and safety domains (basic attributes) are less fulfilled for the >120-minute group than to the other users, they assign more weight to them.

Table 5-18: SEM-MG case analysis 6: Travel time specific Mexico City

Mean score (Likert 1-5)	**Mexico City**	**<60**	**60-120**	**>120**
RELI	3.67	3.82	3.68	3.51
COMF	3.61	3.70	3.60	3.53
CSER	3.58	3.63	3.53	3.58
SAFE	3.57	3.66	3.56	3.50
OSV	3.77	3.90	3.74	3.68
REC	3.78	3.85	3.74	3.75
Continuous OSV (Std.Coeff.)				
RELI	0.23	0.20	0.15	0.28
COMF	----	----	----	----
CSER	0.61	0.71	0.73	0.48
SAFE	0.10	0.04	0.00	0.20

From the six cases in Tables 5-13 to 5-18, we found evidence for the hierarchy of transit needs. Additionally, from Tables 5-6 and 5-9, we can observe that for the Loyalty regression, reliability has an additional significant and high effect on this variable. In general, BRT-type users that value reliability highly, have the intention to recommend the service. As such, we believe that the existence of the hierarchy of transit needs holds. Direct and indirect pieces of evidence were presented from the six specific SEM-MG Case Analyses and the SEM regressions from the pooled data. We will further discuss these in Section 5.5. In the next subsection, we present the SEM latent class models for Santiago and Mexico City.

5.4.4. SEM with finite mixture models: two-class models

In this subsection, we provide additional insight by introducing the finite mixture SEM models. The objective of running these models was to obtain a two-class division in each city, Santiago and Mexico City, that had nothing to do with any preconceived classification. Our hypothesis, basing ourselves in the hierarchy of transit needs, is that two structurally different satisfied groups should appear in each city. In essence, we should be able to obtain an ever more extreme separation between the latent classes, concerning the relative weights they assign to the satisfaction domains. For this analysis, we used the nlsem (Umbach et al., 2017) library package from the R statistical programming language (R Core Team, 2013). Notice that we only include a regression for OVS (Overall satisfaction).

Table 5-19 shows, again, some fascinating results. In the case of Mexico City, the most prominent group (class 2, 54%) attach very high relevance to the reliability domain (0.44) and not so high to safety (0.23). According to the hierarchy, this indicates that these users should not assign much weight to customer services and comfort; and this statement holds. For class 1, in Mexico City (46%), reliability loses relevance (0.17). Conversely, safety (0.40), customer services first (0.27), and comfort (0.11) all gain relevance.

Table 5-19: SEMM-2 Class Analysis: Santiago and Mexico City

Santiago	**St.Coeff.**	**S.E.**	**Z-value**	**Prob. Class**
class1.RELI	0.17	0.06	2.92	0.46
class1.COMF	0.11	0.04	2.93	
class1.CSER	0.27	0.08	3.39	
class1.SAFE	0.40	0.05	7.98	
class2.RELI	0.44	0.05	9.54	0.54
class2.COMF	0.07	0.04	2.01	
class2.CSER	0.07	0.10	0.63	
class2.SAFE	0.23	0.06	4.09	
Mexico City	**St.Coeff.**	**S.E.**	**Z-value**	**Prob. Class**
class1.RELI	0.23	0.04	5.42	0.52
class1.CSER	0.55	0.07	8.16	
class1.SAFE	0.03	0.03	1.11	
class2.RELI	0.42	0.05	9.24	0.48
class2.CSER	0.48	0.05	9.96	

We conclude that the hierarchy of transit needs holds for these two classes of differentially satisfied users. Hedonic attributes are not valued (Class 1 users) until reliability is partially fulfilled (Class 2 users), as our hypothesis states.

As in the SEM model for Mexico City (see Table 5-8), the comfort domain was not relevant to any of the classes when we ran the latent class model. Users are not exigent about this domain in Mexico City, so we left it out of the model and regressed only the other three domains. Subsequently, we also obtain exciting results here. Class 2 (48%) values both reliability (0.42) and customer services (0.48) highly. The safety domain was very irrelevant, only Class 1 assigned it a very low relevance (0.03); in a sense, we hypothesise that this domain is fulfilled for most users. In fact, when revisiting Table 5-18, only the >120-minute group gave some relevance to safety (0.20), and this was the smallest group (29%). The other groups did not assign relevance to this domain. As such, according to the hierarchy, users are partially fulfilled in reliability, but not entirely. Class 1 gives higher relevance to the customer services domain (0.55) and less to reliability (0.23). As in the MGA cases, as users perceive the reliability domain to be fulfilled, they assign more weight to hedonic attributes, in this case, customer services. Again, we believe this gives clear evidence that the hierarchy of transit needs holds for these two classes of differentially satisfied users.

5.5. Discussion of Results

In this section, we discuss and synthesise our general results. At the end of the section, we offer policy recommendations. Initially, we set out to test our hypothesis of the existence of a hierarchy of transit needs among users of public transport. For this, we presented a framework that included several statistical analyses: (i) PCA, (ii) SEM, (iii) SEM-MGA, and (iv) SEMM models. In the end, we presented six specific SEM-MG Case Analyses and two SEMM models. All helped constructing a story to assess whether the hierarchy holds.

With the PCA (Table 5-5) we tested the satisfaction domains, explicitly, finding four: reliability, comfort, customer services, and safety. From our original hierarchy of transit needs, we classified them into functional (reliability), safety (safety) and hedonic (customer services and comfort).

From both the numeric and ordinal SEM models (Tables 5-6 and 5-9), we additionally confirmed that the measurement models presented high reliability (i.e. validity) for both frameworks. From the regressions, when using the pooled data, three domains presented high relevance: reliability, customer services, and safety. We confirmed that comfort was not highly valued across the cities: a counterintuitive result at first, but nevertheless coherent with our hypothesis. Additionally, as we also tested for possible satisfaction domains affecting the loyalty construct, we unveiled that both reliability and customer services had an extra effect on the intent to recommend. However, this effect was more pronounced for the reliability domain (see Tables 5-6 and 5-9). In fact, from the ordinal regression, only reliability presented significance; this fact gives further support for reliability being of generous relevance for the BRT-type users across cities. Empirically, we also confirmed that overall satisfaction had the highest effect on loyalty, consistent with satisfaction theory (Oliver, 2010).

The MGA numeric and ordinal models (Tables 5-11 and 5-12), allowed us to test which categorical variables produced heterogeneity in the satisfaction models. Comparable and similar results were obtained from both frameworks. As we tested with the pooled data, and for two different cities, we were able to discover subtle but significant differences. From the pooled data, the resulting categorical variables were CITY, DD (designated) LANE, TRAVEL TIME, TIME OF DAY, EDUCATION, and INCOME. For the specific cities, fewer variables produced heterogeneous models. For Santiago, the resulting variables were DD LANE and TRAVEL TIME, whilst in Mexico City they were DD LANE, No. OF BUSES, TRAVEL TIME and INCOME. Note that the variable that presented the highest difference was CITY, as expected, since we expect very different levels of service and travelling conditions among highly heterogeneous BRT-type systems.

Also, we can conclude that the variables resulting in heterogeneous satisfaction models are either policy related (designated lanes, travel time, time of day) or socioeconomic (education and income). We can argue that users have different expectations and levels of need depending on all of these variables. In itself, the MGA models help unveil these heterogeneities in the BRT-type users across cities. As we have not found a similar study in the PT literature, we believe this to be a contribution.

From the specific SEM-MG Case Analyses (Tables 5-13 to 5-18), we were able to test for the existence of a hierarchy of transit needs. From the city-specific analysis (Table 5-13), we confirmed its existence in three cities: Santiago, Mexico City and Brazil B2. In synthesis, Mexico City and Brazil B2 (BRT only corridor) have the reliability domain highly fulfilled and assign more relevance to customer services;

this is a clear indication that once an essential transport need is fulfilled, users start assigning more weight to hedonic attributes. Brazil B1 did not behave similarly; however, this city is particularly dissatisfied in both the comfort and safety domains. Even further, from these domains, the items ranking the lowest were comfort at bus stops: lighting, *protection*, cleanliness, number of people (S6, 2.43) and security from thefts and assaults on the way to bus stops and on the bus (S12, 2.09). We pose that there may have been a confusion with the word “protection”, as this is not a comfort concept (it is more of a safety item), however, it was included in three of the comfort items. In any case, users weighted heavily comfort towards their overall satisfaction, and even though safety also scored low, comfort was the domain with more relevance. Still, this result does not deter from the general results obtained for the other cities.

From Tables 5-14 to 5-16, we unveiled differences in levels of satisfaction by INCOME, TRAVEL TIME, and TIME OF DAY, across cities. Notably, from Table 5-14, we discovered that the higher the income, the higher the requirement for the reliability domain, and as such less requirement for customer services. Also, from Table 5-15, we noticed that users with the most extended travel times are particularly dissatisfied and, as such, assign more relevance to the reliability domain. Likewise, from Table 5-16, AM-Peak (5-9) users have higher requirements for reliability and assign a high weight to it. These three analyses unveiled which user groups are under satisfied in the reliability domain, the most basic need.

From Tables 5-17 to 5-18, we gained robustness in our results by being able to confirm the same behaviour obtained for the pooled data in city-specific contexts. Particularly for Santiago (Table 5-17), users travelling in designated lanes assign a high weight to the reliability domain, indicating a high demand for this domain, and less for the others. Regular-lane users assign more weight to customer services and slightly more to comfort and, in a sense, coherent with the hierarchy, they give less relevance to reliability. From Mexico City, users in the >120-min travel time group assign a high weight to reliability, meaning that they are still under satisfied with this domain and, as such, give less importance to customer services than the <120-min users. Finally, from Table 5-19 we also gain robustness in our results. By assessing for latent classes with different satisfaction models within Santiago and Mexico City, we confirmed the existence of users that assign a high relevance to reliability, and as such give less weight to customer services, from both cities.

In summary, all our results confirm evidence about the existence of a hierarchy of transit needs. The usefulness of this theory is that it allows for direct policy recommendations. From Figure 5-3 (left), based on the results obtained in this research, we were able to construct a hierarchy with four steps (domains), the pyramid of transit needs: reliability, safety, customer services and, finally, comfort. A PT administrator would need to establish a level of satisfaction in all domains and specify a working model (i.e. estimate a SEM model). Depending on the model's relative importance for each domain, the PTA can determine at what level of transit development the system is. See dynamic transit development, Figure 5-3 (right). A key aspect to keep in mind is that reliability will be by far the most important attribute to maintain at all costs. Exemplifying with our case study, even though Mexico City, in general, assigns lower importance to the reliability domain, this does not mean that we can forget about it. What this means, according to the hierarchy, is that Mexico City has fulfilled the reliability domain, and as such users now can develop higher order needs for hedonic domains, such as customer services and comfort.

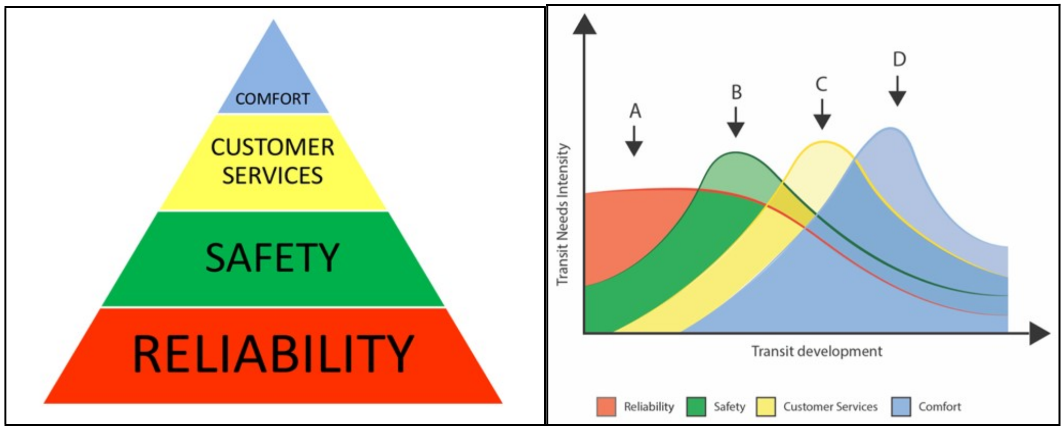

Figure 5-3: Pyramid of transit needs (left) and dynamic transit development (right)

Own design, based on Maslow's Hierarchy of Needs (1943, 1954)

In a nutshell, any system accounting for all users, should be situated across the transit development scale. Explicitly, according to Figure 5-3, in section A reliability would be the most important (Santiago and Brazil B1), in section B safety, in C customer services (Mexico City and Brazil B2), and finally in section D comfort. Also, our research helped verifying that by accounting for heterogeneity, with SEM-MG and SEMM models, we were able to discover differentiated levels of need across different type of users, inter and intracity-wise. These levels depend on the travel conditions (designated lanes, travel time, time of day), and their inherent requirements (income, education). We consider this to be a motivating contribution to public transport policy research.

5.6. Conclusions and Future Research

In this section, we present some conclusions from our study and future research questions that evolve from our findings.

The main conclusion is that we found strong evidence confirming the existence of a hierarchy of transit needs. We started this study with a working hypothesis, that there should be a psychological theory providing insight into more generalizable results regarding how PT users assign their relevance in relation with different satisfaction domains, when assessing their satisfaction with a PT system. We chose Maslow's hierarchy of needs (Maslow, 1943, 1954), as it provides a clear framework about how human beings generate their needs or necessities in life. We posed that this theory would help answer our questions and found evidence from different transport-related and other studies, suggesting that this theory could work for our purposes. Satisfaction is the compliance of a need or necessity (i.e. expectations), and as such, it is suitable to fit the hierarchy of needs framework.

We confirmed the existence of such a hierarchy using a benchmark survey taken in four different cities, with four different BRT-type systems. We gained abundant data about the systems and how they were in different stages of development. From the SEM-MG Case Analyses, we confirmed evidence for the hierarchy's existence in all cases, plus two more cases using SEMM models, although for the latter we did not have access to the mean satisfaction scores. Only in one city, Brazil B1, the hierarchy does not seem to hold. We analysed the questionnaire again, and found that a single word, *protection*, may be the culprit, and precisely in this city, as it presents very low scores in safety. Three comfort items included the word protection, and as this city scores low in both safety and comfort, we believe that some association or even misunderstanding might have occurred. Said that, we believe that this case does not deter from our general findings.

For future research, we recommend that whenever doing customer satisfaction studies, to separate the macro-domains in three: basic (functional), safety (security, protection), and hedonic (customer services, comfort). From there, more sub-domains can also be generated. We believe that it is difficult to separate some conditions; for example, unreliability causes users to have to wait more, and this may make waiting at bus-stops unsafe. Accordingly, it is crucial to perform focus groups and pilot studies to assess these particular nuisances in an effort to generate domain-specific items that do not conflict with each other. We understand the difficulties of constructing a questionnaire that works well in multiple-cities.

One limitation of our study is that we did not include non-linear relationships among the attributes. We believe this to be an essential next line of research, as from our results and the literature revised, it is evident that some domains may have a non-linear relation with overall satisfaction. For this we recommend testing the SEMM and NL-SEMM (non-linear SEM with finite mixtures) models.

One last comment. Most people use public transport to travel to work (i.e. to make money) or to study (i.e. to have a good job in the future). For this reason, when they move through PT, they are doing it to fulfil one of their basic needs in life, earn money to eat and to satisfy their basic needs. This is why reliability should always be at the bottom of the hierarchy. If they do not go to work, their survival will become threatened; henceforth, even in the most unsafe of travelling conditions, they will choose to travel. Allen et al. (2018a) present an extensive review where, unquestionably, reliability was the most critical aspect in many satisfaction PT studies. Nevertheless, herein we presented a specific case (Mexico City), where reliability was *not* the most important item, and we also presented the justification for this: the hierarchy of transit needs.

6. EFFECT OF A MAJOR NETWORK REFORM ON TRANSIT SATISFACTION

6.1. Introduction

Public transport authorities (PTA) are interested in providing high-quality services that users find attractive. Satisfied customers will most probably reuse the services and also recommend them to their relatives and friends (Lierop et al., 2018), encouraging users of private modes to transfer to public transport (Beirão and Sarsfield Cabral, 2007), resulting in a higher sustainable transport modal share (Redman et al., 2013), and mitigating congestion (Anable, 2005; Steg, 2005).

Customer satisfaction surveys measure the level of satisfaction perceived by the users. These surveys allow PTAs to obtain satisfaction ratings ranging from the overall PT service, the service-specific (bus/Metro-line), and the attribute-specific items, which generally consider attributes (i.e. domains) such as reliability, comfort, customer service, and safety. By statistically analysing the relations between the attribute domains and the overall service satisfaction, PTAs can determine which domains or areas are more relevant for users, resulting in direct policy-related information. Extensive literature exists regarding how to reveal which specific attributes users find more important when determining their satisfaction (Allen et al., 2018a; de Oña and de Oña, 2015; Eboli and Mazzulla, 2007).

More concretely, when governments and PTAs endure a substantial network reform in their system, it is of utmost importance to determine *if*, *how*, and *why* the user's perception of the system changes (i.e. their satisfaction levels). First, they want to determine if satisfaction changed, then they want to know how it changed (i.e. improves or deteriorates); and finally, they want to know why it changed. This "*why*" is the crucial piece of knowledge needed to assess which network reform element directly impacts users' satisfaction. In other words, it gives a crucial policy-related network-reforming variable that could aid in improving current PT policy knowledge and could provide insight for other cities and PT systems to use.

We found few studies that address this problem, from users' perceived satisfaction. Most studies focus on nationwide surveys (Cats et al., 2015), while some of them analyse specific macroeconomic changes (de Oña et al., 2018; Efthymiou and Antoniou, 2017). The latter studies analyse whether the satisfaction models changed structurally during a macroeconomic downturn. Still, few focus on specific dense cities with multiannual surveys. Additionally, we found no other study performed while a major network reform was underway in one specific PT system. We consider this a central gap in the literature since authorities require insights regarding which specific elements in the reforms are the ones that impact users' satisfaction the most. In our present study, we focus on this question. Subsequently, we provide details on the contributions of our work.

Our primary contribution is the detailed analysis of a three-year customer satisfaction survey conducted for the bus system of Barcelona while a major network reform was underway, the new bus network: Nova Xarxa de Bus. The analysis is also fostered with sociodemographic characteristics of the users of the new and old bus systems, and operational information about their trips. Using a structural equation (SEM) approach, we analyse whether critical variables from the bus network reform affect users' satisfaction significantly. This allowed us to infer whether the implementation impacted users' satisfaction. Moreover, our analysis incorporates the dynamic transition of the system regarding the number of km and lines of the new network and the number of passengers it attracted. We assess whether the overall satisfaction model, prioritising different satisfaction attributes, differs between users of the new improved network and users of the regular old one. We contribute to the literature by analysing how the users' satisfaction and their relative priorities are affected when a major network reform in a transit system is underway in a major dense city.

Our primary objective is to unveil which operative, travel, and sociodemographic characteristics affect the users' satisfaction and their priorities: we implement a novel two-step SEM approach. Initially, we test a Full SEM-MIMIC model for the whole population. From the results, we identify which categorical variables are candidates for a more specific analysis. Then, we run a Multi-Group Analysis (MGA) with those variables and evaluate for different satisfaction models across subpopulations. With this approach, we assess whether the users using the newly implemented network have a different satisfaction model than those using the regular old lines. We analyse dynamic satisfaction models within the first three years of implementation. This work also contributes to the literature by implementing a two-step approach for efficiently finding subpopulations with different satisfaction models; i.e. choosing out of the extensive set of candidate variables, which ones should enter the model.

Furthermore, we propose an innovative approach to capture the socioeconomic status (SES) of the transport users and correct for heterogeneous perceptions according to this variable, in the model. Satisfaction depends on users' expectations; it is often described as the difference between the perceived received service and the expected service (Oliver, 2010, 1980). Henceforth, we hypothesise that users with high expectations will perceive less satisfaction from the same service received than users with low expectations. Thus, we expect that the higher the SES of a user, the lower the perceived satisfaction that he/she will declare for the same service.

In the PT literature differences have been documented for users from different socioeconomic statuses (SES). Most studies associate the differences in SES utilising net family income, which may be a proxy. However, it does not represent SES accurately. SES is a multifactored construct that accounts for an individual's economic and social position concerning others, based on income, education, and occupation. Additionally, identifying income is quite problematic since some people prefer to omit this information, or under or overreport it in a survey (Moore and Welniak, 2000). To address this, we introduce an SES reflective latent variable, that attempts to correct for socioeconomic status in the SEM framework. In our model, SES is modelled through six personal attributes: education level, work status, ownership of a mobile phone, smartphone, and tablet, and possession of a debit/credit card. Thus, we provide a broader depiction of the SES.

In the PT literature, we found no study that captures the SES using a similar approach. In the economic science literature, few studies have captured this variable using such a framework. We think that a reflective latent variable method might be more efficient and could be used in different settings in which income information could be lacking or biased. With this construct, we can correct for SES in the satisfaction models. We consider this contribution essential for PT policy, allowing us to disentangle the net SES effect, since in many PT satisfaction studies SES may be omitted in the models.

The chapter is organised in the following manner. Section 6.2 presents the literature that is relevant to our case study. Its first part presents the few multiannual customer satisfaction surveys performed for PT systems. The second part addresses studies analysing the impact in user satisfaction of income and socioeconomic status variables. Finally, we briefly address some concerns regarding how the existing literature often considers SES. Section 6.3 presents the Barcelona mobility case study and provides details about the network reform. Section 6.4 presents the sample and the survey and Section 6.5, the model results. Section 6.6 discusses all the results and interpretations in detail. Finally, in Section 6.7, we summarise the most important conclusions and offer policy recommendations for PT authorities from our findings.

6.2. Literature Review

6.2.1 Multiannual PT customer satisfaction surveys and network reforms

Susilo and Cats (2014) investigate how the determinants of overall satisfaction with public transport evolve in time. They analysed data from all of Sweden's major PT systems, from 2001 to 2013. Their results showed that users' satisfaction had been deteriorating and a major cause for this was a decrease in satisfaction with customer interface and the length of the trip duration. Also, they claim that overall dissatisfaction with the PT operation was a major cause for this drop. This study is insightful; however, it does not analyse one particular PT system in depth. Nevertheless, it provides valuable general recommendations on policies to improve users' satisfaction.

De Oña et al. (2018) focus their attention on the effect of an economic downturn on public transport users' satisfaction. They analyse data from the bus transit service of Granada, Spain. They formulate three SEM models for 2008, 2011, 2014. Specifically, they show that, counterintuitively, the fare and service factors had a lesser influence on users' satisfaction when the economic downturn was more intense, in 2011. Other factors such as comfort and convenience remained constant in importance during the periods analysed. Although insightful, this study does not analyse any specific network reform taking place during the analysed timeframe. Their study focuses solely on the effect of the macroeconomic conditions, due to an economic downturn.

Efthymiou and Antoniou (2017) likewise analyse the effect of an economic downturn in PT users' satisfaction and demand, utilising factor analyses and hybrid discrete choice models (latent variables). They analyse data from two set points in Athens, Greece: 2008 and 2013; the 2013 point representing the aftermath of the economic downturn. Their main conclusions are that higher satisfaction with the quality of service of PT and higher car use and maintenance costs led to an increase in the PT demand. They also found that increased fares led some users away from PT. This study adds evidence of the behaviour of PT users during an economic downturn. Still, it does not analyse any effect of a specific network reform during the analysed timeframe.

As there is scarce-to-none literature in PT satisfaction during a major network reform, we refer to studies reporting how network reforms have impacted patronage. Mulley and Ho (2013) analysed the bus supply and patronage data for 15 metropolitan contract regions in Sydney (Australia) and tested whether bus reforms triggered patronage increments. Their results show that changes in km supplied drive positive changes in patronage, and that network reforms had a significant demand impact in different contract regions above the additional-km-supplied effect. We found no study that analysed the network reform effect from the passengers' point-of-view. We find this to be a critical gap in the available literature.

6.2.2 Socioeconomic status or income in PT satisfaction?

In their study covering eight European cities, Susilo and Cats (2014) identified that travel time reliability and station environment are determinants for low-income travellers' satisfaction. While frequency, onboard comfort and safety during waiting, do not influence their overall satisfaction much. Their results also indicate that specific traveller groups, such as women, young, and unemployed travellers have different determinants of satisfaction for different travel modes.

Dell'Olio et al. (2011) introduce the concept of desired quality, which is different from the perceived quality; it does not represent the daily experiences of users, but rather what they desire, hope for or expect. They find that potential users (i.e., not regular PT users) hardly consider cleanliness, employee attitude, and comfort to assess their desired quality. Instead, they only consider those related to the journey time, waiting time, and vehicle-occupancy level. They consider these variables much more critical than regular users. Despite this, the contribution that each variable makes to the utility functions estimated for the different income categories, in general, does not differ much. The only exception being waiting time, which is more valuable in high economic status groups. This result is expected as high-income users have a higher value of time. Low-income users give less importance to the vehicle environment (cleanliness, comfort, safeness) than high-income respondents (Morton et al., 2016). Moreover, respondents in the top income segment have a higher likelihood of displaying relatively adverse attitudes towards the perceived easiness of using buses compared to respondents in the low-income segments.

Income is one of the critical socio-demographic variables explaining heterogeneity across PT users. Nevertheless, it is also one of the most difficult attributes to capture accurately. Non-response, measurement error or omission in income variables are common in economic and transport studies. Measuring income in PT users' satisfaction studies is problematic when there is significant non-response. For example, Susilo and Cats (2014) report missing data on income status at 30%, due to respondents refusing to provide this information, or because they do not know their actual total income. Omitting observations due to the lack of information can bias the results since non-responsiveness is not evenly distributed across the population. As an example, Morton et al. (2016) report only respondents with full data on their socio-economic characteristics for urban bus services in Scotland. This issue would potentially limit the representativeness of the results obtained from such study, especially if there is a high non-response rate. Besides, discrepancies between respondents' reported earnings and their true earnings can bias the coefficient estimates and lead to imprecise standard error estimates (Angrist and Krueger, 1999). We argue that in some cases it might be preferable to indirectly measure the SES instead of asking to self-report the income.

There has been an interest in the use of principal components analysis (PCA) to measure SES (Vyas and Kumaranayake, 2006). Additionally, Kolenikov and Angeles (2009) propose utilising ordinal data instead of running a PCA with dummy variables. The authors argue that the ordinal framework yields better SES measurement. In our case study, we will utilise both the ordinal and numeric frameworks to assess which one provides a better representation of the phenomenon. In management science, there is no consensus about the convenience of using formative or reflective latent variables in SEM models. Edwards (2011) argues that using formative structures is problematic, precisely because one has to assume a causal relation, which may or may not be justified. The author states that using reflective latent variables, which are adequate in many different contexts, should be recommended. Eboli et al. (2018) compare both approaches in a PT satisfaction case study, and suggest that the reflective model is more reliable for describing the phenomenon of PT passenger satisfaction.

In our case study, we propose a reflective latent variable for the SES. We argue that the education levels, work statuses, and possession of mobile phones, tablets, and debit/credit cards are all indicators of SES. We do not expect any of them to be a perfect indicator of SES. Nonetheless, utilising all of them can provide a more accurate measurement of this critical policy-related variable. Note that the expectations of a particular user should grow with his/her SES. Hence, we presume a negative and significant coefficient for SES when regressed against satisfaction.

6.3. Barcelona Mobility and Policy Reform: *Nova Xarxa de Bus*

6.3.1 Barcelona mobility

Barcelona is a continuous and compact urban area with mixed land uses. There are commercial zones in most of the city area, and it faces mild weather conditions. With 1,620,809 inhabitants in a 102.15 km^2 area, it is the third densest city in Europe after Paris and Athens. In Barcelona, 1.83 million trips per weekday take place between its metropolitan area and the central city, while the number of intracity trips reaches 4.99 million. In 2015, half of the trips in the city of Barcelona were made by non-motorized modes -walking and cycling-, around 30%

by private modes and the remaining 20% by public transport (IERMB, 2015). The bus and metro modal shares are 11% and 14% respectively.

Despite the relatively low number of private mode trips, Barcelona has a high density of cars in its central city. While the number of cars per inhabitant is similar to other large European conurbations, Barcelona has one of the highest rates of cars per surface unit. One of the factors that allow a significant daily volume of motor-vehicle trips in such a dense space is that motorbikes make up a quarter of motorised internal trips within the city. The average number of trips in European cities is 2.8-2.9 trips per day and per person, whereas in the case of Barcelona, it was 3.3 in 2005 and it increased continuously up to 3.9 trips per day and person in 2015 (IERMB, 2015). Conversely, the total number of trips made by public transport has remained constant.

The Integrated Fare System is the zone-based ticketing system run by the regional regulator ATM (Autoritat del Transport Metropolità). This system provides a unified public transport fare integration throughout Barcelona's metropolitan area and beyond since 2001. In Barcelona, the Metro and bus networks share the same ticket. Users can combine different means of transport for 75 minutes paying a single fare. In the central city area, the public-owned firm TMB provides local bus transit. For the metropolitan area, TMB and several private firms supply different services in different areas (see Bel and Rosell, 2016). Supply has also remained quite stable over time. TMB offered 3,355 million seats-kilometres in 2016, while it supplied just over 3,050 million seats-kilometres in 1989. TMB behaves as a corporate company, upholding more flexibility than most public organisations.

The ratio of bus and tram network length per surface area ranks third amongst twenty-three European cities, and drops to sixth when corrected by population (EMTA, 2017). Still, regarding bus-kilometres supplied, it ranks in nineteenth place. Therefore, Barcelona has an extensive bus network while its vehicle-supply is low. Barcelona ranks last in the most vital bus quality indicator, its average bus speed reaches only 12 km/h. Its poor performance impacts the demand side, each inhabitant makes only 68 trips by bus each year, with only one-third of the European cities presenting a bus system attracting fewer per-capita demand.

6.3.2 Barcelona's new bus network: *Nova Xarxa de Bus* (NXB)

Until 2012, the Barcelona bus network was a clear example of a traditional direct trip-based network design. As the city grew, routes were prolonged, and line overlapping was common. The network had 63 direct-service lines, most of them centripetal. The proportion of transferring trips was only 11% (TMB, 2018), even though direct service was not offered between any two parts of the city. For the new user, the network was difficult to understand and to read on a map. On September 2012, a new bus network, *Nova Xarxa de Bus*, started its gradual implementation that finished in 2018.

In the mid-1800s, Barcelona expanded under a strongly gridded street plan known as l'Eixample. This design in part of the city allows Barcelona to possess an easy-to-navigate configuration. Furthermore, most of these streets run in a single direction. The new bus network has been designed to adapt to this street configuration. The design of the system is based on a model presented in Estrada et al. (2011) and Daganzo (2010) in which lines operate parallel services in this orthogonal grid. The *Nova Xarxa de Bus* design serves the whole city with 28 lines and has progressively replaced most of the redundant old bus network. Currently, only 20 secondary routes of the old network remain. From the 28 grid lines, 17 can be considered vertical lines, eight horizontal lines and three diagonals (see Figure 6-1).

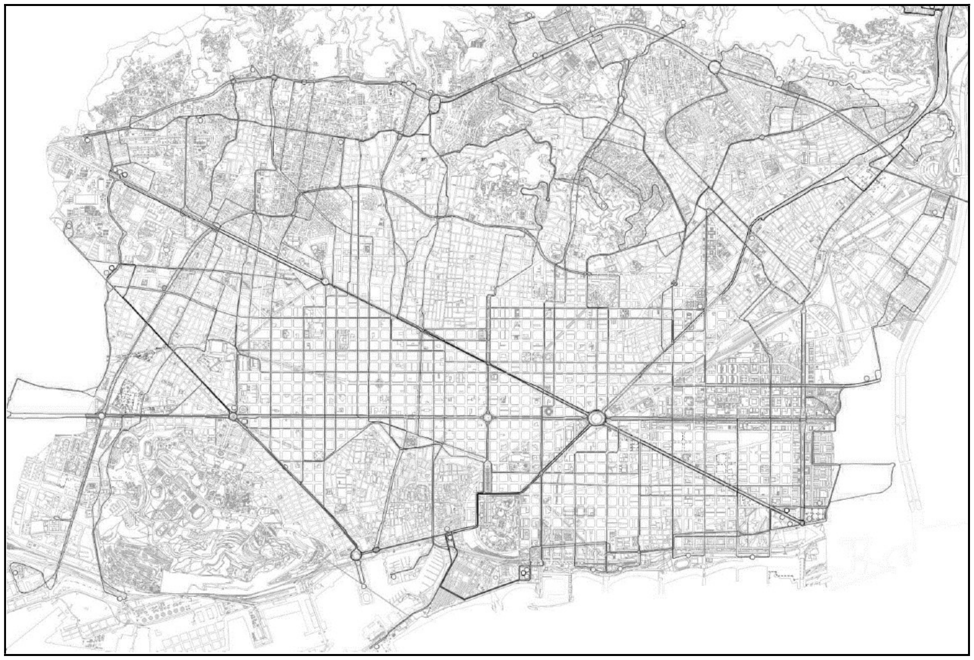

Figure 6-1: The *Nova Xarxa de Bus* design

Source: http://www.tmb.cat/.

This new design shifts the travelling experience from a direct-service network into a transfer-based network. The old spaghetti network has been replaced by a new one that should be significantly simpler to read by its users. As of mid-2018, 23 of these lines are open to the public.

Badia et al. (2017) strongly suggest that a transit system can attract more demand by providing a high-capacity integrated network with transfer-friendly conditions, than by providing a network of lines expected to provide transfer-free trips. They hypothesised that transit passengers are much less averse to transfers than often assumed in planning practice, reasoning that adequately designed transfer-based networks can be very appealing and even attract more demand than their conventional counterparts. While conventional bus networks in big cities exhibit transfer percentages ranging from 1.5% to 16%, the Nova Xarxa transfer percentage was 26% in 2015 and is expected to rise to 44% when the project

reaches completion in 2018. Still, NXB ensures that 90% of the trips in Barcelona can be completed with none or only one transfer. Between 2012 and 2014 TMB reduced its lines from 106 lines to 98 increasing its supply from 3,320 million seat-kilometres to 3,355. Demand grew from 179.97 to 195.80 million passengers in the same period: a 9% growth in demand against only a 1% supply increment.
Table 6-1 displays operational statistics from both bus networks.

Table 6-1: Regular and NXB operative statistics from 2012-2016

	2012	**2013**	**2014**	**2015**	**2016**
Total network (km)	927.7	916.6	908.3	873.2	871.5
NXB network (km)	44.8	95.2	131.1	131.5	160.7
Proportion NXB network (%)	4.8	10.4	14.4	15.1	18.4
Number total routes	107	107	105	100	99
Number NXB routes	5	10	13	13	16
Proportion NXB routes (%)	4.7	9.4	12.4	13.0	16.2
Passengers (millions)	180.0	183.0	184.3	187.8	195.8
Proportion passengers NXB (%)	13.3	27.4	34.1	34.9	40.9
Vehicles-net km (millions)	40.3	40.8	**39.7**	40.1	40.6
Seats-net km (millions)	3,320.9	3,357.1	**3,314.3**	3,338.3	3,355.0
Proportion vehicles-net km NXB (%)	1.9	9.8	21.8	26.4	31.1
Proportion seats-net km NXB (%)	2.4	12.4	27.2	32.5	37.6
Passengers per veh. and net km: regular lines	4.2	3.8	4.2	4.4	4.5
Passengers per vehicle and net km: NXB	31.8	12.6	7.3	6.2	6.4

Since 2012, NXB implementation has continuously grown, except for 2015 due to a change in the Barcelona local government. In 2016, the NXB proportion in the number of routes and kilometres was 16% and 18% respectively, while the number of passengers carried by NXB lines had reached 41%. In mid-2018, with 23 lines, more than half of the total passengers are carried by NXB (TMB, 2018). The demand captured by NXB routes outperform those of regular lines when correcting per vehicle or seat kilometre being offered. In 2015 and 2016, NXB carried 40% more passengers per vehicle and kilometre than regular lines. This reform shows that the demand captured by NXB goes beyond the supply it provides. It seems that the network reform is positively affecting the demand, which may seem coherent with the findings presented by Mulley and Ho (2013) for the Sydney's PT contracts.

Thus, the Barcelona case provides a perfect opportunity to identify which elements cause this demand surge and how are they affecting user satisfaction.

6.3.3 *Nova Xarxa de Bus*: perception questionnaire and participation process

Every year, TMB asks a sample of its users to answer a general questionnaire to understand customer satisfaction. In 2014, a complementary survey was conducted in order to assess users' perception about the new network (i.e. NXB). This perception questionnaire was answered by 2614 users representing more than half of the users surveyed in the general questionnaire. After two years of the launch of the new NXB lines, only two-thirds of the respondents were aware of its existence. The average grade obtained by all users surveyed was 7.46/10. 40% of the users stated that the network changes did not have an apparent effect. Ten percent claimed that the new network had an adverse effect because it increased travel time, there were fewer buses, or the new routes were inadequate. More than half responded positively to this reform due to better levels of service. When asked about the effect they expected once the new network would be fully operational, 80% considered that it would be beneficial. When users were asked if they used the bus more often since 2012, 80% of them claimed no major changes, while 18% reported a higher use.

During the first quarter of 2017, the Barcelona city council promoted a participation and debate process regarding the new bus network, attracting 1750 participants. The sessions were mostly attended by the elderly and people with reduced mobility. These users usually face transferring difficulties. The elderly claimed that the new network censors their travel patterns, as it forces them to make more transfers. This group saw little or no recognition of the advantages of the new bus network. Therefore, the outcome of the process was that NXB had not yet adapted and adjusted to the users' characteristics, needs, and priorities. In conclusion, the participation process and the NXB perception questionnaire show two realities: a group of people campaigning against NXB due to a suppression line policy and a majority of people that perceive this new service as better, and who are willing to continue using it in the future. Next, we present the results of the general questionnaire.

6.4. Case Study, Sample and Survey

6.4.1 The NXB bus survey: a satisfaction questionnaire

Data supporting this research was collected by TMB through Customer Satisfaction Surveys for the bus system. The data was collected with a standardised questionnaire, including eight modules, from 2013 until 2015. Since the NXB implementation started in 2012, we focus our analysis on the surveys from 2013-2015. In 2016, the questionnaire was changed. The survey measures bus transit system users' perceptions about bus-line service and its attributes. The main objective of the survey is to understand which factors affect service quality. We focus our work on the following six modules that characterise the user's travel patterns and socioeconomic attributes, and collect the satisfaction ratings:

a) *Starting data.* Whether the survey includes an importance module, time of the survey, the bus-parking lot zone (*Horta*, *Ponent*, *Triangle* and *Zona Franca*) -this is the zone where the buses are parked-, bus-line average headway (in minutes), place of the interview (at the bus stop or on board the bus), and whether it is an *NXB* line or a regular line. The surveyor fills this data before the interview begins.

b) *Ticket.* The user is asked which ticket he/she uses (i.e. full ticket or discount card).

c) *Satisfaction ratings*. Inquires the user to rate 19 bus-satisfaction (SB) items. Items are grouped in domains: four operational, four comfort, three information, two safety items, one about cleanliness inside buses, and two about customer service. Three items refer to TMB image perception items, and the final items is an overall assessment of the bus-line (SB.LINE). All ratings respond to a Likert scale from 0 to 10 (i.e. Spain's grade system).

d) *Current trip patterns*. Asks for the travel motive, if the trip involves transfers (and to/from which PT system), and if the respondent had a private vehicle available for this trip.

e) Travel habits. Bus weekly-use frequency and Metro weekly-use frequency.

f) *Users' profile*. This module collects users' socioeconomic information: gender, education level, work status, place of birth (i.e. Spain or foreign), and residence municipality. Non-traditional items collected are whether the user possesses a: mobile phone, smartphone with an internet connection, tablet, and debit/credit card.

In each of the three years (2013-2015) over 4000 questionnaires were answered. Only individuals older than 15 were interviewed. In the next section, we present the sample travel and socioeconomic characteristics, and the mean survey results.

6.4.2 Sample and survey results

In the three years the survey was answered by 12,511 users, 4089 in 2013, 4215 in 2014 and 4207 In 2015. Table 6-2 summarises the sample's travel characteristics subdivided by year and by type of bus-line. In the pooled sample, most are frequent bus users (70%), while frequent Metro users are a minority (36%). For the bus-garage zone distribution, the biggest stratum belongs to Zona Franca (35%) and the smallest to Ponent (16%). This holds for most years and across bus-line type except for NXB in 2013, in which most buses belonged to Zona Franca bus-garage (68%) and none to the Triangle bus-garage (0%). This information will be important for the bus authority, TMB.

Table 6-2: Sample travel characteristics: NXB and regular bus in each year's sample

CATEGORY	UNIT	ALL	2013.RB	2014.RB	2015.RB	2013.NX	2014.NX	2015.NX
SAMPLE	n	12511	3600	2663	2771	489	1552	1436
	%	100	29	21	22	4	12	12
FREQ.USE	5-7	70	70	71	64	76	74	70
BUS (D/W)	0-4	30	30	29	36	24	26	30
FREQ.USE	5-7	36	37	36	34	38	36	35
METRO (D/W)	0-4	64	63	64	66	62	64	65
BUS-GARAGE	*Horta*	24	24	26	27	16	20	20
ZONE	*Ponent*	16	17	18	17	16	13	15
	Triangle	25	25	28	28	0	22	27
	Zona Franca	35	34	28	28	68	45	38
HEADWAY	0-14	86	86	75	80	100	100	100
(MIN)	>15	14	14	**25**	20	0	0	0
TIME	06-10	23	23	22	22	25	24	24
OF DAY	10-14	25	24	25	26	24	27	25
(HRS)	14-18	28	28	29	30	28	26	27
	18-23	24	25	24	22	23	23	24
TICKET	Full	78	81	78	75	82	78	75
	Discount	22	19	22	25	18	22	25
TRANSFER	Yes	37	37	35	35	38	41	35
	No	63	63	65	65	62	59	65
TR.METRO	Yes	15	15	14	14	16	15	86
	No	85	85	86	86	84	85	14
TR.BUS	Yes	19	19	18	18	18	22	16
	No	81	81	82	82	82	78	84
TR.RENFE	Yes	2	2	3	2	2	3	2
	No	98	98	97	98	98	97	98
TR.FGC	Yes	2	2	1	3	2	3	3
	No	98	98	98	97	98	97	97

Regarding bus headways, most of the users were travelling in lines with an average headway lower than 15 min (86%). Since all *NXB* lines operate with headways lower than 15 min, we turn our attention to the regular lines. For 2014, 75% of them had headways lower than 15 minutes, while in 2013 and 2015 over 80% did. This fact matches our previous findings from the statistics in Table 1, that in 2014 supply was effectively lower than average. We notice that this affected the frequencies for the regular lines. Regarding time of the day, the distribution is quite even: the smallest stratum is 6 to 10 hrs (23%), and the largest, 14-18 hrs (28%).

Most users pay a full ticket (78%) while a minority possess a discount card (22%). Around 37% of users needed a transfer, specifically, 15% transferred to/from Metro, 19% to/from other buses, 2% to/from Renfe trains, and 2% to/from FGC trains. The transfers to/from Tram were negligible (<1%). The distribution for time of day and transfer categories remained fairly similar for all years and both types of bus lines.

Table 6-3 presents users' socioeconomic characteristics disaggregated by year and type of bus-line. Most users are Female (68%). However, the proportion of Male users increases from 30% in regular bus-line to 34% in *NXB* lines. Most users have High school and University studies (80%). Most users are employed (67%), while retired people and students correspond to approximately 13% in each group. Regarding private-vehicle availability, the lowest value corresponds to regular bus-line users in 2014 (20%).

Notice that for regular bus-line users the proportion was 29% in 2013 and 26% in 2015, while the overall average is 26%. This seems consistent with Barcelona statistics (IDESCAT, 2018) which state that in 2014, the vehicle ownership per 1000 residents reached its lowest value, following an economic downturn.

Most users live in the Barcelona municipality (86%), and most were born in Spain (81%). 97% of the users own a mobile phone, 81% own a smartphone, just 38% own a tablet, and 86% possess a debit or credit card. For these non-traditional items, the *NXB* users have slightly larger percentages than the regular line users, except for the tablets which show similar possession rates for both types of bus-line users.

Table 6-3: Sample socioeconomic characteristics: NXB and regular bus users by year

CATEGORY	UNIT	ALL	2013.RB	2014.RB	2015.RB	2013.NX	2014.NX	2015.NX
GENDER	Male	32	30	29	32	35	35	33
	Female	68	70	71	68	65	65	67
AGE	<30	28	26	28	28	37	27	31
(YR)	31-45	32	33	31	30	31	35	31
	>46	40	41	41	42	32	38	38
EDUCATION	Incomplete	2	1	1	2	1	1	3
	Primary	18	19	20	16	13	16	16
	Secondary	43	40	43	44	47	47	42
	University	37	40	36	38	39	36	39
WORK.STATUS	House-work	2	3	2	2	2	2	2
	Unemployed	6	6	7	5	6	5	4
	Retired	12	11	12	14	6	10	13
	Student	13	12	14	13	20	14	13
	Employed	67	68	65	66	66	69	68
PRIVATE.VEH	Yes	26	29	**20**	26	32	**26**	30
AVAILAB.	No	74	71	80	74	68	74	70
RESIDENCE	Barcelona	86	87	84	85	88	85	88
	Other	14	13	16	15	12	15	12
BIRTH.PLACE	Spain	81	82	82	79	81	82	79
	Foreign	19	18	18	21	19	18	21
MOB.PHONE	Yes	97	96	97	96	98	98	97
	No	3	4	3	4	2	2	3
SMART.PHONE	Yes	81	72	81	88	76	85	89
	No	19	28	19	12	24	15	11
TABLET	Yes	38	30	40	45	28	41	43
	No	62	70	60	55	72	59	57
DC.CARD	Yes	86	87	85	84	87	89	86
	No	14	13	15	16	13	11	14

In Table 6-4, we present the average satisfaction ratings, for each year, and for each type of bus-line. The last column presents the difference between the average *NXB* and the regular bus-lines ratings (Δ*NX*-RB).

Table 6-4: Survey results: averages by year and type of bus-line (Likert scale 0-10)

DOMAIN	Service quality attributes	ALL	2013	2014	2015	RB	NX	ΔNX-RB
Reliability	SB1: Waiting time	6.62	6.61	6.45	6.81	6.34	7.36	**+1.02**
	SB2: Frequency	7.09	7.25	6.78	7.23	6.95	7.44	**+0.49**
	SB3: Speed-reliability	7.32	7.44	7.19	7.33	7.21	7.60	**+0.39**
Comfort	SB5: Temperature	6.77	6.76	6.62	6.93	6.71	6.93	+0.22
	SB6: Accessibility to buses	7.87	7.97	7.79	7.84	7.79	8.07	+0.28
	SB7: Crowding conditions	6.45	6.58	6.36	6.42	6.37	6.68	+0.31
	SB8: Bus-stop comfort	7.46	7.51	7.31	7.55	7.39	7.64	+0.25
Information	SB9: Info at bus-stops	7.22	7.28	7.08	7.29	7.15	7.40	+0.25
	SB10: Of changes/incidents	6.49	6.48	6.30	6.68	6.39	6.74	+0.35
	SB11: Inside buses	7.38	7.48	7.24	7.42	7.34	7.48	+0.14
Safety	SB12: Accident safety	7.23	7.21	7.18	7.29	7.19	7.32	+0.13
	SB13: Security inside buses	7.58	7.60	7.53	7.62	7.52	7.75	+0.23
Customer	SB15: Staff courtesy	7.60	7.65	7.52	7.62	7.57	7.67	+0.10
Service	SB16: CS system	7.22	7.17	7.19	7.29	7.16	7.37	+0.21
System	SB18: TMB promotes civility	7.21	7.21	7.00	7.42	7.15	7.35	+0.20
	SB19: TMB's effort-improve	6.96	6.91	6.84	7.12	6.86	7.22	+0.36
SB.LINE:	Overall bus-line satisfaction	7.52	7.53	7.50	7.52	7.39	7.85	**+0.46**
		ALL	**2013.RB**	**2014.RB**	**2015.RB**	**2013.NX**	**2014.NX**	**2015.NX**
SB.LINE:	Overall bus-line satisfaction	7.52	7.49	7.31	7.34	7.89	7.83	7.85

We group the results into specific domains, or clusters: reliability, comfort, information, safety, customer service, system (the TMB image items), and finally present the overall bus line satisfaction (SB.LINE). The scoring on all items is on a 0-10 Likert scale. In Spain's (Barcelona) grading system a 5-7 is sufficient to pass, 7-9 is notable, and a 9-10 excellent.

From the reliability domain, the best-rated item is the speed-reliability item (SB3, 7.32) and the worst, waiting time (SB1, 6.62). From the comfort domain, the best-rated item is accessibility (SB6, 7.87) and the worst, crowding conditions (SB7, 6.45). In the information domain, the best-rated is information inside the bus (SB11, 7.38) and the worst, information about incidents and changes (SB10, 6.49). In safety, security inside the buses (SB13, 7.58) is better rated than accident safety (SB12, 7.23). In customer service, staff courtesy (SB15, 7.60) is better rated than the customer service system item (SB16, 7.22). Finally, the TMB civility promotion item (SB18, 7.21) is better assessed than the TMB's effort to improve item (SB19, 6.96).

The overall bus-line satisfaction mean value is 7.52. Still, there are substantial differences between the rating from the regular bus-lines and the *NXB* lines. In average, regular lines score 7.39, while *NXB* lines score 7.85, for a +0.46 differential. Among the 19 satisfaction items, the three highest differentials correspond to, not that surprisingly, reliability items: +1.02 for waiting time, +0.49 for frequency, and +0.39 for speed-reliability. Regarding the other items, the differentials are always positive and fall within the +(0.10-0.36) range. It is worth mentioning that the fourth to sixth places go to +0.36 for TMB efforts in improvements, +0.35 for information about incidents and changes, and +0.31 for crowding in the bus. Three relevant facts can be drawn from this table. The *NXB* lines outscore the regular lines in all the satisfaction items. The satisfaction in waiting time and all reliability items for the *NXB* lines is significantly higher than for the regular lines, and the crowding conditions and TMB's efforts to improve were also significantly better assessed in the *NXB* lines than in the regular lines.

From the SB.LINE mean values across year and type of line, the satisfaction for regular lines drops between 2013 and 2014. Although in a smaller fashion, this drop also happens for the *NXB* lines. In 2015, the satisfaction in both cases increased, but not up to 2013 levels. Thus, there may exist some systematic issue affecting the satisfaction ratings in 2014, with some impact carrying over for 2015 as well. We will return to this issue after analysing the model results.

6.5. Model Results

In this section, we present the model results. We work with the pooled data in all cases (i.e. all years). These include (i) PCA for the satisfaction items, (ii) SEM numeric and ordinal Probit models, (iii) SEM-MIMIC ordinal Probit model, (iv)

SEM-MGA, and SEM-MG case analysis models, again with the ordinal Probit framework.

From the PCA we obtain the general structure of the satisfaction domains. We test both types of SEM models, numeric and ordinal Probit, to assess for better model fit. As we obtain satisfactory results from the ordinal Probit model generating a better fit than for the numeric model, we continue the subsequent analyses with the ordinal Probit framework. Next, we perform an SEM-MIMIC model, from which we choose candidate categorical variables to test in the next step. Subsequently, we perform an SEM-MGA ordinal model, testing for different satisfaction models across subpopulations according to different travel conditions and sociodemographic characteristics. Explicitly, we test for year and type of bus-line. For all models, we use the software R (R Core Team, 2013) and its associated library packages, specifically, the Lavaan package for R (Rosseel, 2012) in all SEM models. In most tables, values were rounded to two decimal places for readability.

We want to discover which categorical variables produce different models across subpopulations. For this purpose, we ran the SEM-MGA to obtain the categories that effectively produce different models. From them, we obtain the regression results: six different case analysis that produce heterogeneous models. Integrating the results, we utilise both the SEM-MIMIC and the SEM-MGA case analysis results to conclude on the *if*, *how*, and *why* there are differences in users' satisfaction perceptions during the implementation of the NXB reform in its initial years. Herein, we comment briefly on the results obtained in each step. In Section 6.6 we discuss the results and their implications in detail.

6.5.1 Principal component analysis: satisfaction domains

As the first step, we want to uncover how a user perceives different satisfaction constructs. To reveal this, we run an exploratory PCA (Hoyle, 2012; Jolliffe, 2014) on the satisfaction items. This allows us to determine which items better represent specific latent constructs. In other words, which items are correlated, and are candidates to form a construct as a whole. Following the Kaiser (1960) rule we obtained seven components. In the process we left out three items which loaded into components with an absolute loading lower than 0.4. Hence, they could not be assigned to any components. Next, with the seven components, we computed Cronbach's (1951) alpha to assess internal validity. We present the results of the final PCA and Cronbach's alpha in Table 6-5.

Table 6-5: PCA: Satisfaction items

	RELI	COMF1	COMF2	INFO	SAFE	CSER	SYST
SB1: Waiting time	**0.61**	0.00	-0.02	-0.01	-0.02	-0.04	-0.01
SB2: Frequency	**0.60**	-0.05	0.01	-0.00	0.02	0.03	0.03
SB3: Speed-reliability	**0.49**	0.04	0.09	0.02	0.00	0.06	0.03
SB5: Temperature	-0.07	**0.71**	0.07	-0.03	-0.03	0.07	0.08
SB6: Accessibility to buses	0.01	-0.03	**0.80**	0.03	0.01	-0.01	-0.01
SB7: Crowding conditions	0.07	**0.68**	-0.07	0.04	0.04	-0.08	-0.08
SB8: Bus-stop comfort	0.02	0.11	**0.56**	-0.08	-0.03	-0.02	-0.05
SB9: Info at bus-stops	0.04	-0.05	0.04	**-0.61**	-0.03	-0.00	0.05
SB10: Info of changes/incidents	0.02	0.08	-0.15	**-0.55**	0.02	-0.01	-0.08
SB11: Info inside buses	-0.06	-0.02	0.08	**-0.56**	0.05	0.01	0.03
SB12: Accident safety	0.00	0.03	-0.03	0.01	**0.72**	0.02	0.02
SB13: Security inside buses	-0.00	-0.03	0.04	-0.02	**0.68**	-0.01	-0.03
SB15: Staff courtesy	0.01	0.03	0.02	0.04	0.06	**0.71**	0.07
SB16: Cust. Service system	0.01	-0.03	-0.04	-0.04	-0.06	**0.69**	-0.10
SB18: TMB promotes civility	-0.06	-0.05	0.04	0.02	0.03	0.03	**-0.74**
SB19: TMB's effort-improve	0.07	0.07	-0.02	-0.01	-0.03	-0.01	**-0.64**
Cronbach's Alpha (α)	**0.81**	**0.72**		**0.79**	**0.79**	**0.69**	**0.69**

Even though the PCA had divided the four comfort items into two components, we realised that regrouping them into a single component produced a more reliable Cronbach's alpha. Since both components dealt theoretically with the same comfort domain, we decided to model them as a single latent construct. The resulting final six components are reliability (RELI), comfort (COMF), information (INFO), safety (SAFE), customer service (CSER), and system (SYST).

This last item, system, joins the TMB's civility promotion (SB18) and efforts to improve (SB19) items, representing a TMB system/image construct.

6.5.2 Structural equation models: satisfaction with the bus-line and SES

With the resulting components of the satisfaction items, we created an SEM model. We followed the recommended practice of running a measurement model first and then an SEM model. In this section, we only present the final SEM models, which contain both the measurement models and the regression results. We are interested in forming an SES latent construct, composed by the following six socioeconomic attributes: education level, work status, mobile phone, smartphone and tablet ownerships, and debit or credit card possession. We ranked the education level progressively from incomplete to university, and the remunerated work status from non-employed to employed: housework, unemployed, retired, student, and employed. For the numeric SEM, we converted all the categories to numeric/integer form (e.g. incomplete education is 0 and university level is 3; housework is 0 and employed 4). For the ordinal Probit we utilised a more appropriate ordinal categorisation.

As we stated before, we present both frameworks. We expect to obtain a correctly measured construct by using six indicators of the users' SES. We additionally combined the mobile phone and smartphone to form a new dummy variable, MOB.PHONE, which equals to one if the user had either a mobile phone or a smartphone, or equals zero, otherwise. This new variable alone performed better than using variables associated with both attributes separately, as such, we continued to use only this newly formed variable.

Our SEM model additionally contains the six latent satisfaction constructs obtained from the PCA (see Table 6-5). We present both the Measurement model structure and the SEM model structure in Figure 6-2. The ovals represent the latent construct, and the rectangles represent the single survey items. The Measurement system is at the left, and the SEM at the right.

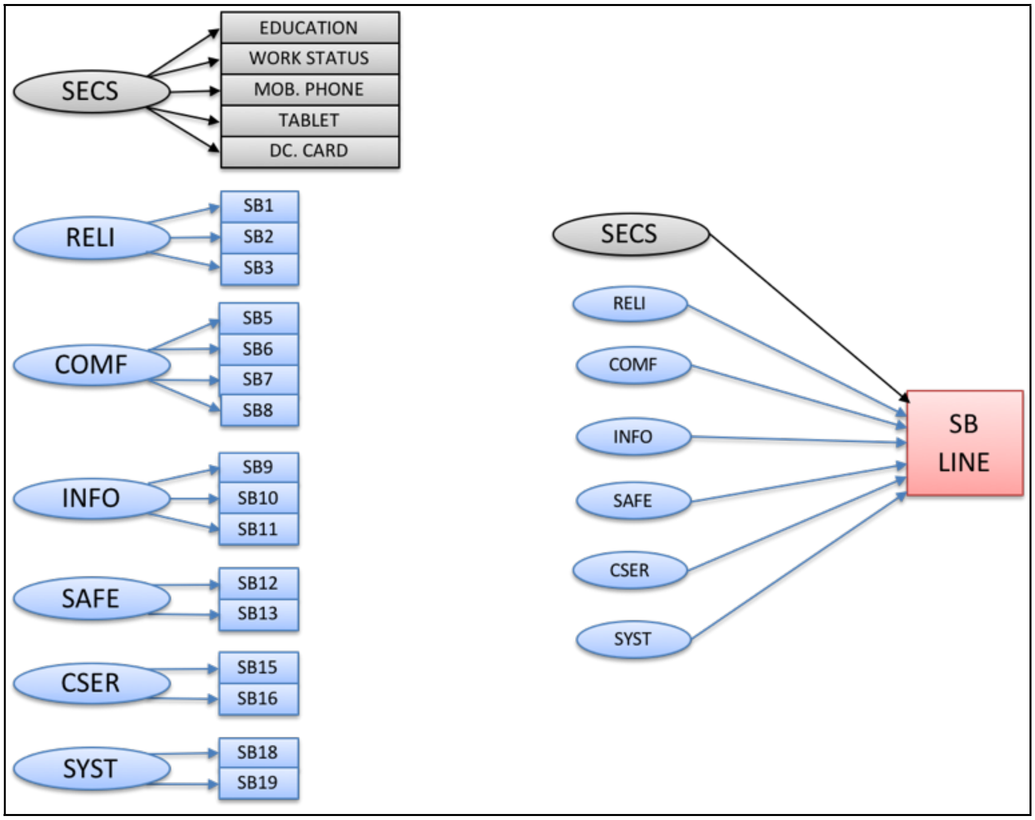

Figure 6-2: The schema for the Measurement and SEM model structures

We use this structure for both the SEM numeric and SEM ordinal Probit models. The SEM numeric model treats the satisfaction items with a numeric range (i.e. 0 to 10), while the SEM ordinal Probit treats the satisfaction items with an ordinal range (for details, see Allen et al., 2018b). For both SEM models, we hypothesise that all the satisfaction constructs and the SES have a statistically direct impact on bus-line satisfaction (SB.LINE).

Table 6-6 presents the results of the SEM numeric model: estimates, standard errors (S.E.), Z-values, and standardised coefficients (Std.Coeff.) associated with each item. The standardised coefficients refer to the standardised change of the dependent variable of interest when the predictor variable changes one standard deviation.

Table 6-6: SEM continuous results for pooled data

Measurement	Estimate	S.E.	Z-value	St.Coeff.	R^2
SES					
EDUCATION	1.00			0.65	0.42
WORK.STATUS	0.91	0.03	26.98	0.44	0.19
MOB.PHONE	0.09	0.00	22.11	0.30	0.09
TABLET	0.31	0.01	22.52	0.31	0.10
DC.CARD	0.29	0.01	26.42	0.41	0.17
RELI					
SB1	1.00			0.78	0.61
SB2	0.94	0.01	86.19	0.80	0.64
SB3	0.73	0.01	78.55	0.72	0.52
COMF					
SB5	1.00			0.60	0.36
SB6	0.85	0.02	55.01	0.63	0.39
SB7	1.08	0.02	53.32	0.60	0.36
SB8	1.00	0.02	59.69	0.71	0.50
INFO					
SB9	1.00			0.77	0.59
SB10	1.17	0.02	79.18	0.75	0.57
SB11	0.94	0.01	78.82	0.75	0.56
SAFE					
SB12	1.43	0.02	93.86	0.79	0.63
SB13	1.43	0.01	99.46	0.83	0.69
CSER					
SB15	1.14	0.01	80.36	0.73	0.53
SB16	1.27	0.02	79.81	0.72	0.52
SYST					
SB18	1.22	0.02	77.33	0.69	0.47
SB19	1.39	0.02	87.32	0.77	0.60
Regression	**Estimate**	**S.E.**	**Z-value**	**St.Coeff.**	R^2
SB.LINE					0.56
RELI	0.36	0.01	33.79	0.43	
COMF	0.13	0.03	4.93	0.10	
SAFE	0.13	0.02	7.01	0.09	
CSER	0.09	0.02	4.06	0.07	
SYST	0.22	0.02	9.17	0.16	
SES	-0.21	0.02	-8.53	-0.07	
CFI	**TLI**	**GFI**	**AGFI**	**RMSEA**	**SRMR**
0.971	0.964	0.978	0.970	0.034	0.023

For the SES latent construct, the items present moderate to high reliability (i.e. Std.Coeff.) in the Measurement model. The values lie within the 0.30-0.65 range. We consider them acceptable since the goal is to get a good measurement of SES, a usually intangible construct. Notice that all the coefficients are positive indicating that all the items behave as expected (i.e. the higher education level, the higher SES). The highest coefficients are for education and work status, an expected result since they are often considered critical components of the SES. Socioeconomic status (SES) is defined by material wealth, occupation, and participation in educational and social institutions (Oakes and Rossi, 2003).

We also conclude that the measurement for each satisfaction construct has high reliability since all their Std.Coeff lie within the 0.60-0.83 range. The following specific items load more on the latent constructs: for RELI it is *frequency* (SB2), for COMF it is *the bus-stop comfort* (SB8), for INFO it is *information at bus-stops* (SB9), for SAFE it is *security inside the bus* (SB13), for CSER it is *staff courtesy* (SB15), and for SYST it is *TMB's efforts to improve* (SB19).

The regression part of the SEM numeric model presents a high R^2 (>0.5). However, the overall satisfaction with the bus-line is measured with only one item (SB.LINE). Usually, this means that there is some inherited error in the measurement, as is it not performed with multiple items. We consider an R^2 >0.5 acceptable, and >0.6 adequate. For the structural coefficients (St.Coeff), values below 0.1 are considered very low, between 0.1 and 0.3, low, between 0.3 and 0.5, moderate, and above 0.5 high (Allen et al., 2018a; Currie and Delbosc, 2017). The most relevant construct is reliability, while the SES presents a negative sign, as expected. Regarding fit indices, the CFI, TLI, GFI, and AGFI are all above the recommended cut-off (>0.95) values by Hu and Bentler (1999). The RMSEA, the most important index, complies with the cut-off value as well (<0.08).

Next, we present the same SEM structure using an ordinal Probit approach, allowing us to compare both frameworks. Table 6-7 presents the SEM ordinal Probit results for the pooled data.

Table 6-7: SEM ordinal results for pooled data

Measurement	Estimate	S.E.	Z-value	St.Coeff.	R^2
SES					
EDUCATION	1.00			0.75	0.57
WORK.STATUS	0.69	0.03	21.68	0.52	0.27
MOB.PHONE	0.94	0.05	19.12	0.70	0.50
TABLET	0.52	0.03	19.19	0.39	0.15
DC.CARD	0.79	0.04	22.41	0.59	0.35
RELI					
SB1	1.00			0.77	0.59
SB2	1.05	0.01	133.22	0.81	0.65
SB3	1.03	0.01	131.76	0.79	0.63
COMF					
SB5	1.00			0.63	0.39
SB6	1.09	0.01	90.67	0.68	0.47
SB7	0.99	0.01	85.22	0.62	0.38
SB8	1.19	0.01	97.75	0.75	0.56
INFO					
SB9	1.00			0.79	0.62
SB10	0.98	0.01	144.85	0.78	0.60
SB11	1.00	0.01	140.68	0.79	0.62
SAFE					
SB12	0.80	0.00	196.52	0.80	0.64
SB13	0.86	0.00	220.52	0.86	0.74
CSER					
SB15	0.77	0.01	162.58	0.77	0.59
SB16	0.76	0.01	155.59	0.76	0.58
SYST					
SB18	0.71	0.01	136.81	0.71	0.51
SB19	0.79	0.01	161.08	0.79	0.62
Regression	**Estimate**	**S.E.**	**Z-value**	**St.Coeff.**	**R^2**
SB.LINE					0.59
RELI	0.54	0.01	41.49	0.41	
COMF	0.17	0.03	6.40	0.11	
SAFE	0.08	0.01	7.64	0.08	
CSER	0.08	0.01	6.07	0.08	
SYST	0.19	0.01	14.30	0.19	
SES	-0.11	0.01	-9.54	-0.09	

CFI	TLI	GFI	AGFI	RMSEA	SRMR
0.997	0.996	0.998	0.997	0.032	0.031

From the ordinal measurement model, the SES St.Coeff. values lie within the 0.39-0.75 range. Education and mobile phone present the highest coefficients, which differ from the numeric counterpart where the second highest was work status. We also conclude that all the latent satisfaction variables have high reliability (i.e. validity) in the measurement, since all their standardised coefficient values lie within the 0.62-0.86 range. The ordinal Probit Measurement model presents higher reliability than the numeric counterpart. The specific items that load more on the latent constructs are almost the same as in the numeric model.

For the SB.LINE regression the explained variance (R^2) is higher (0.59) than the numeric (0.56) indicating a better fitting model. Again reliability is the most critical construct, and the SES has a negative sign, but larger in magnitude than for the numeric regression, according to the standardised coefficients. From the fit indices, the CFI, TLI, GFI, and AGFI are all above the recommended cut-off (>0.95) values by Hu and Bentler (1999). In fact, all values are above 0.99, indicating an excellent fit to the data. The RMSEA, 0.032, also complies with the cut-off value (<0.08), even complying with the strictest <0.05 cut-off value for a good fit.

6.5.3 SEM-MIMIC: ordinal Probit

In this subsection, we present the SEM-MIMIC model (Joreskog and Goldberger, 1975). We base this model on the SEM ordinal Probit presented before, since it showed better measurement properties than the numerical counterpart. We follow the methodology presented in Allen et al. (2018b).

We introduce dummy variables into the SB.LINE regression that attempt to capture the heterogeneity of the perceptions depending on the users' travel and socioeconomic characteristics. The SB.LINE is treated as an ordinal variable. For all categories, we add n-1 dummy variables, where n is the number of groups in each category. We introduced (base in parentheses): two regressors for year (2013), three for bus-garage zone (*Zona Franca*), one for headway (<15 min), one for gender (Male), three for time of day (6-10), one for location of the survey (Bus), one for bus-line type (Regular), two for age (15-30), and one for ticket type (Full Ticket). Moreover, we introduced four dummies for mode-specific transfers (No), one for vehicle availability (No), one for bus-use frequency (0-4 days/week), and one for Metro-use frequency (0-4 days/week). Lastly, we introduced one dummy for birthplace (Spain), one for residence Municipality (Barcelona), and one for if the survey had the importance module (No). This variable was included as having the importance module may lengthen the questionnaire, causing dissatisfaction. Hence, we want to test whether this occurs.

In total, we included 25 dummy variables, and we kept those that showed statistical significance over the 85% level. Still, we decided to keep all the transfer variables as these are policy-related. We present the results in Table 6-8. We kept 20 dummy variables in the model, the ones that did not resulted significant were: the time of day (3), the location of the survey (1), and vehicle availability (1). From the significant variables, the ones with the highest coefficients were the NXB lines (0.16) and the Birthplace Foreign (0.20). The year dummies present negative and similar values, indicating that there was a systematic deterioration of the satisfaction in the years 2014 and 2015, which is not explained by the structural model or other variables. This result lines up adequately with the mean satisfaction rating from summary statistics (Table 6-4): while in 2013 it was 7.58, in 2014 and 2015, it decreased to 7.50 and 7.52, respectively.

Surprisingly, the transfer variables present very low values, and the only significant one is for the transfer to Renfe trains (-0.02), indicating that Transfers are not a major dissatisfactory issue in this bus network. This is a promising result indicating that bus-users do not penalise transferring highly in their bus-line satisfaction. Again, the model complies with the recommended cut-off values (Hu and Bentler, 1999).

Table 6-8: SEM-MIMIC ordinal Probit results for pooled data

Regression		Estimate	S.E.	Z-value	St.Coeff.	R^2
SB.LINE						0.60
RELI		0.53	0.01	39.91	0.38	
COMF		0.17	0.03	6.12	0.10	
SAFE		0.08	0.01	7.51	0.08	
CSER		0.09	0.01	6.66	0.08	
SYST		0.19	0.01	13.96	0.18	
2014		-0.12	0.02	-4.86	-0.05	
2015		-0.11	0.02	-4.46	-0.05	
ZN.HORTA		0.07	0.03	2.71	0.03	
ZN.PONENT		0.13	0.03	4.62	0.05	
ZN.TRIANGLE		0.09	0.03	3.40	0.04	
NXB.BUS		0.38	0.02	16.54	**0.16**	
HWAY>15MIN		-0.11	0.03	-4.33	-0.04	
SES		-0.12	0.01	-8.76	-0.09	
FEMALE		-0.06	0.02	-3.00	-0.03	
AGE31.45		-0.07	0.03	-2.92	-0.03	
AGE46		-0.06	0.03	-2.46	-0.03	
DISCOUNT.CARD		0.17	0.02	7.05	0.07	
TR.METRO		-0.01	0.03	-0.19	-0.00	
TR.BUS		0.01	0.02	0.28	0.00	
TR.RENFE		-0.14	0.06	-2.32	-0.02	
TR.FGC		-0.01	0.06	-0.20	-0.00	
FR.BUS.5W		-0.03	0.02	-1.52	-0.02	
FR.METRO.5W		-0.03	0.02	-1.39	-0.01	
BP.FOREIGN		0.53	0.02	22.48	0.20	
MUNI.OTHER		0.13	0.03	4.78	0.04	
SURVEY.IMP		-0.03	0.02	-1.49	-0.01	
	CFI	**TLI**	**GFI**	**AGFI**	**RMSEA**	**SRMR**
	0.977	0.975	0.983	0.976	0.044	0.029

To be precise, we obtain an excellent fit to the data as the CFI, TLI, GFI, and AGFI are all above 0.97. Also, the RMSEA is less than 0.05, indicating a superb fit. Adding all the dummy variables did not deteriorate the model fit. The explained variance improved slightly to 0.60, compared to the SEM ordinal Probit model (i.e. without the MIMIC variables). We conclude that the SEM-MIMIC has an excellent fit to the data. More importantly, the model allows the heterogeneous conditions of the bus users to be incorporated. We obtain significant results for the *NXB* variable, which was one of our original research objectives. This effect adds over the effect of the satisfaction constructs within the model (i.e. RELI, COMF, SAFE, CSER).

6.5.4 SEM-MGA ordinal Probit and case analyses

In this subsection, we present the SEM-MGA using the ordinal Probit framework. From the SEM-MIMIC we obtained several significant categories; meaning that users have a different perception (positive or negative) of bus-line satisfaction according to those particular categorical variables. Here we unveil whether different subpopulations exist according to those variables (i.e. heterogeneous satisfaction models). In other words, we assess whether there are significantly different structural models (i.e. standardised coefficients) among these subpopulations. Explicitly, we want to discover whether there are different satisfaction models across the user population in the different bus-line types (NXB) and years of the survey.

We follow the methodology of Allen et al. (2018c) to discover which moderating variables bring heterogeneity to the SEM bus-line satisfaction models. For this purpose, we compute the ΔCFI. This value represents the result of subtracting the CFI of an SEM-MG model were all the parameters are restricted to be the same, from the CFI of an all-loose SEM-MG model (see Allen et al., 2018c). In other words, the ΔCFI is the value at which the CFI deteriorates once we restrict for the SEM model parameters, except the regression estimates, to be the same across subpopulations (i.e. instead of all-loose). We determine which variables moderate the SEM measurement model and structural parameters. We use a threshold ΔCFI of >0.004, for the SEM-MGA ordinal Probit, in bold in Table 6-9. We found no specific recommendation in the literature for the ordinal Probit SEM-MG framework, and when we ran the MGA analyses, noticed that other fit indices started to deteriorate with this specific ΔCFI. Once we knew which categories affected the resulting SEM models, we ran the SEM-MG Case Analyses to assess for heterogeneity in the structural parameters of the regression equations.

Table 6-9: SEM-MGA ordinal Probit results for pooled data

Category	ΔCFI	MGA
YEAR	**0.006**	Yes
GENDER	0.003	No
AGE	**0.015**	Yes
BIRTHPLACE	**0.016**	Yes
MUNI	0.001	No
ZONE	0.002	No
HEADWAY	**0.004**	Yes
TIME OF DAY	0.002	No
NXB	**0.009**	Yes
DISCOUNT	**0.009**	Yes
Cut-offs	>0.004	----

We present the six specific Case Analyses in Tables 6-10 to 6-15, we focus on the standardised coefficients in our discussion. In Table 6-10, we analyse the year grouping variable. For the three years, the highest coefficient is for reliability, and it is very similar among the three groups (~0.41). We notice that for customer service and system, in the year 2015, the coefficients are higher than the other two years. The comfort coefficient is also lowest for 2015, this may indicate that comfort became less relevant in this year. This outcome may be explained by two causes, trips being faster, and as such the comfort construct became less relevant. Also, crowding reached its maximum in 2015, since bus service (i.e. supply) was reduced in 2014 while demand continuously grew during these years. This affected the comfort rating negatively, as comfort is valued more when it is high, hence its relevance was affected detrimentally in 2015.

Table 6-10: SEM-MG case analysis 1: YEAR

Regressions	Estimate	S.E.	Z-value	St.Coeff.	R^2
SB.LINE [2013]					0.62
RELI	0.56	0.03	21.62	0.41	
COMF	0.20	0.04	4.80	0.13	
SAFE	0.09	0.02	4.87	0.09	
CSER	0.06	0.02	2.43	0.06	
SYST	0.19	0.03	7.51	0.19	
SES	-0.16	0.02	-7.09	-0.12	
SB.LINE [2014]					0.56
RELI	0.52	0.02	24.39	0.41	
COMF	0.28	0.05	5.62	0.17	
SAFE	0.08	0.02	4.12	0.08	
CSER	0.05	0.02	2.20	0.05	
SYST	0.14	0.03	5.66	0.14	
SES	-0.09	0.02	-4.76	-0.07	
SB.LINE [2015]					0.61
RELI	0.54	0.02	25.53	0.42	
COMF	0.05	0.05	1.10	0.04	
SAFE	0.07	0.02	3.64	0.07	
CSER	0.10	0.02	4.82	0.10	
SYST	0.27	0.02	13.08	0.27	
SES	-0.09	0.02	-4.53	-0.07	

Table 6-11 focuses on the *age* grouping variable. The highest coefficient is for reliability in all cases. All groups, except those >56-year-old which had a nonsignificant coefficient value comfort highly. The system construct is valued higher by the 36-45-year-old group, but for this user group, customer service appears to be non-relevant. For the 46-55-year-old group, safety seems not relevant. Thus, we find some significant differences in satisfaction models across age groups, which is somehow expected since our necessities change when we age. Furthermore, the SES coefficient becomes more negative in older people, it has a higher magnitude, which is indicative of a multiplicative effect between SES and the >56-year-old users.

Table 6-11: SEM-MG case analysis 2: AGE

Regressions	**Estimate**	**S.E.**	**Z-value**	**St.Coeff.**	**R^2**
SB.LINE [AGE 15-35]					0.57
RELI	0.53	0.02	24.69	0.39	
COMF	0.22	0.05	4.74	0.13	
SAFE	0.08	0.02	4.81	0.08	
CSER	0.09	0.02	4.30	0.09	
SYST	0.16	0.02	7.94	0.16	
SES	-0.06	0.02	-2.92	-0.04	
SB.LINE [AGE 36-45]					0.64
RELI	0.53	0.03	20.66	0.41	
COMF	0.19	0.05	3.96	0.13	
SAFE	0.11	0.02	4.86	0.11	
SYST	0.28	0.03	9.69	0.28	
SES	-0.11	0.03	-4.00	-0.09	
SB.LINE [AGE 46-55]					0.62
RELI	0.55	0.03	17.61	0.43	
COMF	0.19	0.06	3.14	0.12	
SAFE	0.03	0.02	1.25	0.03	
CSER	0.11	0.03	3.54	0.11	
SYST	0.19	0.03	6.22	0.19	
SES	-0.11	0.03	-3.18	-0.07	
SB.LINE [AGE >56]					0.58
RELI	0.53	0.03	18.85	0.44	
SAFE	0.11	0.03	4.20	0.11	
CSER	0.12	0.03	4.63	0.12	
SYST	0.15	0.03	5.05	0.15	
SES	-0.15	0.03	-5.06	-0.10	

In Table 6-12, we analyse the *birthplace* grouping variable. There is a significant difference in the reliability coefficients. Spaniards value reliability more than foreigners; they also value comfort more, while foreigners value more system (with a larger coefficient) and safety and seem to assign null importance to customer service.

Table 6-12: SEM-MG case analysis 3: BIRTHPLACE

Regressions	Estimate	S.E.	Z-value	St.Coeff.	R^2
SB.LINE [Spain]					0.59
RELI	0.57	0.01	40.28	0.44	
COMF	0.18	0.03	5.89	0.11	
SAFE	0.06	0.01	5.35	0.06	
CSER	0.10	0.01	7.02	0.10	
SYST	0.16	0.01	11.18	0.16	
SES	-0.12	0.01	-9.51	-0.09	
SB.LINE [Foreign]					0.54
RELI	0.45	0.03	13.20	0.34	
COMF	0.10	0.06	1.66	0.07	
SAFE	0.13	0.03	4.97	0.13	
SYST	0.28	0.04	7.73	0.28	
SES	-0.20	0.05	-3.83	-0.11	

In Table 6-13, we analyse the *headway* grouping variable. The results show a significant difference in the reliability coefficient. Users who experienced long headways (>15 min) value reliability more (0.47), than with shorter ones (0.39). This result is intuitive, as we expect people that have less frequent service to suffer more if the service is unreliable. Interestingly, the >15 min group does not place a high value on comfort or customer service, which are hedonic attributes. We hypothesise that this is caused by low crowding inside the bus, since these are probably low demand services. In these low demand buses and bus-stops, safety becomes an issue. Indeed, these users value safety more (0.17), than the <15 min group (0.07).

In Table 6-14, we analyse the *bus-line type* grouping variable. Unexpectedly we did not find a big difference in the coefficients. However, we notice some subtleties in the regression coefficients. First, the regular bus users give a slightly higher value to reliability (0.41) compared to the *NXB* users (0.37). Additionally, the customer service is valued higher by *NXB* users (0.13) than by regular users (0.06); meaning that *NXB* users have a slightly higher appreciation for the customer service provided by their bus system.

Table 6-13: SEM-MG case analysis 4: HEADWAY

Regressions	Estimate	S.E.	Z-value	St.Coeff.	R^2
SB.LINE [0-14 min]					0.59
RELI	0.50	0.01	36.39	0.39	
COMF	0.21	0.03	7.22	0.13	
SAFE	0.07	0.01	6.02	0.07	
CSER	0.09	0.01	6.36	0.09	
SYST	0.20	0.01	13.57	0.20	
SES	-0.11	0.01	-8.23	-0.08	
SB.LINE [>15 min]					0.58
RELI	0.67	0.04	16.23	0.47	
SAFE	0.17	0.03	4.90	0.17	
CSER	0.05	0.04	1.38	0.05	
SYST	0.15	0.04	4.26	0.15	
SES	-0.17	0.03	-5.05	-0.12	

Table 6-14: SEM-MG case analysis 5: BUS-LINE TYPE

Regressions	Estimate	S.E.	Z-value	St.Coeff.	R^2
SB.LINE [REG.BUS]					0.59
RELI	0.54	0.02	33.69	0.41	
COMF	0.18	0.03	5.65	0.11	
SAFE	0.09	0.01	7.04	0.09	
CSER	0.06	0.02	4.21	0.06	
SYST	0.20	0.02	12.71	0.20	
SES	-0.13	0.01	-9.23	-0.10	
SB.LINE [NXB.BUS]					0.57
RELI	0.47	0.02	20.91	0.37	
COMF	0.17	0.05	3.30	0.10	
SAFE	0.07	0.02	3.51	0.07	
CSER	0.13	0.03	5.26	0.13	
SYST	0.17	0.03	6.55	0.17	
SES	-0.09	0.02	-3.86	-0.07	

Finally, in Table 6-15, we analyse the *ticket type* grouping variable. The reliability coefficient is similar for both cases, full ticket (0.41) and discount card (0.44). However, we notice that the full ticket users value comfort more than the discount card group. This effect may be due to the full-ticket group assigning more value to the bus service, as they pay more.

Table 6-15: SEM-MG case analysis 6: TICKET TYPE

Regressions	Estimate	S.E.	Z-value	St.Coeff.	R^2
SB.LINE [Full TICKET]					0.60
RELI	0.53	0.01	37.47	0.41	
COMF	0.20	0.03	6.98	0.13	
SAFE	0.08	0.01	6.53	0.08	
CSER	0.07	0.02	4.96	0.07	
SYST	0.19	0.02	12.87	0.19	
SES	-0.11	0.02	-7.16	-0.08	
SB.LINE [Discount CARD]					0.56
RELI	0.56	0.03	18.04	0.44	
SAFE	0.11	0.02	4.54	0.11	
CSER	0.09	0.03	3.24	0.09	
SYST	0.20	0.03	6.90	0.20	
SES	-0.21	0.03	-6.58	-0.13	

For all the MG Analyses, the SES construct obtained a significant and negative coefficient. We conclude that this construct allows correcting for the difference in *expectations* of the service by users of different socioeconomic status. We will further discuss the implications of the Case Analyses in the next section; specifically, in combination with the SEM-MIMIC, focusing on the *NXB* bus-lines and year categorical variables.

6.6. Discussion of Results

Our primary objective was to unveil *if*, *how*, and *why* performing a major network reform impacts PT user satisfaction by analysing one specific case study: Barcelona bus reform. Our objective is to determine also whether the impact is positive or negative, we base our study on specific SEM for user satisfaction models. We aim to discover differences in the satisfaction ratings between regular and NXB users. First, we discover that there is a clear satisfaction difference between regular and NXB users: 7.39 vs 7.85, respectively (see Table 6-1). Also, unequivocally, there is a more substantial difference between the average satisfaction for both types of services in the reliability related attributes: waiting time, frequency, and speed-reliability.

From the initial SEM models (see Table 6-7), we identify that the constructs that better explain the overall bus-line satisfaction are reliability, system, comfort, customer service, and safety. Also, we conclude from the standardised coefficients across all models tested, that the reliability satisfaction construct explains most of the variance for the bus-line satisfaction. We could argue now that we understand the *if* and *how*, but we are still missing the *why*.

Henceforth, we turn our attention to the SEM-MIMIC ordinal model (see Table 6-8) which provides a full SEM-MIMIC model that addresses heterogeneity in perceptions. All the original satisfaction constructs remain significant with slightly lower coefficients: reliability (0.38), system (0.18), comfort (0.10), customer service (0.08), and safety (0.08). Interestingly, the *Nova Xarxa de Bus* variable (NXB) stands out as a new decisive variable for explaining user satisfaction. It is ranked in the fourth position overall, after reliability, foreign-birthplace users, and system. Hence, after the two most essential satisfaction constructs, reliability and system, it is the single most crucial policy-related variable.

There is a fundamental factor in the NXB that causes its users to increase their satisfaction, which is not reflected in any of the SEM-MIMIC satisfaction constructs. This factor may be capturing additional mobility/accessibility attributes of the NXB network, that contributes to its users' perceived value. Notice that the customer is not directly asked whether he/she travels or not in an NXB line, since this is information obtained out of the interview. Also, the improvement within the reliability domain does not drive this increment since the effect is well-captured in our model. We hypothesise that this effect is related with the added mobility provided by more frequent services in a transfer-based network, especially for users travelling at least one of their trips on an NXB line (higher frequency). Note that the TMB bus system works relatively well, and people that live outside Barcelona, in other metropolitan cities, gives better scores to TMB than Barcelona citizens (see Table 6-8).

Additionally, based on the MGA, specifically the bus-line type (Table 6-14), we do not find evidence of differences in the satisfaction model. This fact indicates that the effect is mainly because of the NXB line. In other words, the satisfaction models are the same, but the NXB users perceive an additional value from the new NXB lines. Analysing the other groupings and significant variables we discover exciting results. First, robustly, all groups rank the reliability construct as the most important (see Tables 6-10 to 6-15); this is an essential policy-based conclusion since the NXB lines were designed to provide a more frequent/reliable transfer-based network. From the MG-year, Table 6-10, it is interesting to note that it is in the year 2015 that users start to value the system construct more. A possible interpretation is that once the NXB implementation was underway, it started improving the TMB system/image variable. This is a hedonic component that adds to the satisfaction perceived about the service, but we do not expect the effect to be immediate. Hence the lag seems reasonable.

From Table 6-12 we notice that foreigners place more value on the system than Spanish citizens. We also learned from the SEM-MIMIC that, in general, foreigners derive higher satisfaction from the bus system. Also, from Table 6-10, more satisfied users tend to value the image/system construct more. From Table 6-13 we notice two insightful results, the users of high frequency services (small headways) value comfort higher than users from low frequency lines, while the latter group value safety more. Unmistakeably, as users have to wait longer for the bus, they are more worried about safety (i.e. they feel unsafe). However, as users have to wait less for the bus, they feel more comfortable with their trip and rate it higher, as the comfort domain is a hedonic component. It appears that waiting less improves the comfort perception.

From all the SEM models including the SEM-MIMIC and the SEM-MGA models, the SES construct produces a significant and negative coefficient ~(-0.10). We believe this to be a key policy-related result, since it means that our framework allows us to disentangle this effect and correct satisfaction from the expectations component. High SES users have higher expectations, hence lesser satisfaction. We robustly proved this result with all SEM models presented in this study.

From the participation process regarding the implementation of NXB, the elderly perceive that the new network censors their travel patterns, as it forces them to make more transfers. Still, the SEM-MIMIC ordinal results show that users >56 years do not rate bus satisfaction negatively (see Table 6-8). Additionally, the satisfaction models do not identify a clear transfer penalty for bus-transfer. Furthermore, NXB lines increase users' satisfaction since the first year of implementation. As there were positive comments from the general public and only negative comments from a specific segment of users, the result stems out that participatory processes should be understood as an opportunity for certain groups to express their aspirations and requests, and to improve the system accordingly. In a nutshell, a randomly assigned customer satisfaction survey proves to be more efficient at determining the real causes of PT users' satisfaction, while a non-random self-selection participation process provides specific group's needs.

Most of the transfers to another transport mode do not affect bus-users satisfaction (see Table 6-8). We do not have information that clarifies whether the transfers happen between regular line, NXB line or a metropolitan one. Although we cannot disentangle this effect, transfers in the bus system are not penalised in the user satisfaction models. We believe this to be a noteworthy policy-related conclusion. For example, when the metropolitan bus-line runs along TMB routes in Barcelona, many bus-stops are shared with the TMB public company, increasing the easiness of transfers. TMB also owns Barcelona Metro system, providing incentives for the municipal company to facilitate transfers with bus-stops nearest to Metro-stops. The regional government of Catalonia operates FGC, a regional train service. However, users perceive FGC close to Metro lines due to a good connection with the Metro system, although it also connects metropolitan and non-metropolitan cities with Barcelona. An expected result corresponds to the negative satisfaction with Spanish central government rail service, operated by RENFE. The rail service has low levels of punctuality due to historical underinvestment compared with FGC.

Reducing bus supply can affect satisfaction. Due to budgetary constraints in the public sector, TMB was forced to reduce vehicle supply after 2013. TMB paid particular attention to NXB implementation and consolidation, and presumably, these cuts did not take place in this new network. On regular lines, satisfaction dropped especially between 2013 and 2014, when more substantial supply cuts took place, while NXB users' satisfaction remained stable. Also, the lines with headways longer than 15 min increased in 2014, compared with 2013, possibly damaging the satisfaction perception.

On the macroeconomic level, the Spanish economy grew after 2013. However, this effect is not significant on users satisfaction, or at least it was not possible to separate it from other effects, such as supply changes. Nevertheless, when analysing the vehicle availability statistics for Barcelona, we notice that 2014 marks the lowest vehicle ownership of the 2013-2015 period. This may have affected some of the public transport users and their travel satisfaction. However, we do not detect a significant empirical effect of the vehicle availability variable in the SEM-MIMIC model, so we discard it as an important role in determining the users' bus-line satisfaction.

We also observe that low-frequency affects user satisfaction negatively. This is supported by the survey statistics and the SEM models. Accordingly, if the regulator expects to increase satisfaction in low-frequency lines, the first recommendation is to increase the frequency. If this is not possible, the model suggests increasing bus safety and system satisfaction constructs. However, their effect will probably not be as powerful as increasing frequencies, since the most critical construct is the reliability-frequency construct. Regarding the type of users, there are no satisfaction differences between frequent bus users and occasional bus users. Thus, attracting more occasional users to use the system should not change overall satisfaction much. Also, frequent Metro users do not evaluate the bus service differently than the not frequent Metro users.

From Table 6-8, the SEM-MIMIC model suggests that not paying or having a discount on the bus ticket impacts satisfaction positively. We conclude that when public transport access does not come from out of pocket money, expectations drop, increasing the perceived satisfaction. In other words, students, the elderly, and low-income users that possess a discount car have lower expectations for the bus service. Thus, they have higher satisfaction, and this effect is incremental and different from the SES effect, as discount cards may be awarded to users that are not necessarily in a low SES (i.e. students).

6.7. Conclusions and Policy Recommendations

In this study, we characterise and model user satisfaction during a major PT network reform. Our approach is holistic and robust, allowing us to find answers to *if*, *how*, and *why* the user's perception of the system changes. Our results show that users perceive that NXB lines provide more satisfaction than regular lines. NXB increases user satisfaction *ceteris paribus* to all other attributes. We assert that the added value effect goes beyond the incremental frequency or improved reliability. The NXB lines provide an intrinsic value for the user since additional mobility/accessibility opportunities are provided through the network, which is perceived by the users.

Our results are also validated by the higher demand captured by the system, once it is implemented (see Table 6-1). Reducing bus supply usually has a negative impact on satisfaction. However, the effect can be compensated with a transfer based-network implementation. Coherent with previous literature (Badia et al., 2017), our results show that even though the NXB network increases transfers, these transfers do not penalize user satisfaction.

From our framework, we were able to disentangle the SES-effect. We show that users with higher SES levels perceive less satisfaction than low-SES users for the same service. Furthermore, high socioeconomic status has a more negative impact as people age, reaching its maximum adverse effect in the oldest stratum. Our results are generalizable in the sense that our framework allows disentangling the effect of transfer-based networks on users' perceptions. We argue that this type of network brought added-value mobility/accessibility from higher frequencies and increased connectivity. Again, for reproducing a similar study, operative variables (i.e. frequencies) in conjunction with a customer satisfaction survey should be analysed.

Regarding other PTAs, this work shows that transfer-based networks may be more efficient at producing satisfied customers by providing additional mobility, despite the increment in transfers per trip. We recommend to implement the network reforms in different stages, to avoid confusing the users of the system, such as the implementation of the Transantiago in its early years (see Muñoz et al., 2014), for example. For low frequency bus-lines, we consider essential to improve reliability by adhering their operation to fixed schedules. A limitation from our study is that we could not distinguish transfers between regular lines from those between NXB lines or between regular and NXB lines. If we could, we would have evaluated whether NXB transfers are valued differently than regular-line transfers.

This study contributed to the PT literature by analysing a major network reform from users' perspective, for the first time to the best of our knowledge. We believe more research is warranted along these lines, since these reforms are commonplace in growing Metropolitan areas, and it is of utmost importance to understand their effectiveness from users' point-of-view.

7. EFFECT OF OBJECTIVE ATTRIBUTES ON TRANSIT SATISFACTION

7.1. Introduction

Customer satisfaction surveys serve to measure users' satisfaction levels. Satisfaction is the actual performance of the service minus the expected one (Oliver, 2010, 1980). Such surveys allow PTA to obtain satisfaction scores for the overall service, service-specific (i.e. Metro-line), and attribute-specific items. Typically, the latter are contained in a specific domain (i.e. reliability, safety, comfort). There is an ample literature covering ways to reveal the specific attributes that users find more important when determining their satisfaction (Allen et al., 2018a, de Oña and de Oña, 2015, Eboli and Mazzulla, 2007). In such surveys, typically, the users' travel and socioeconomic characteristics are inferred or asked. Other questions can include specific items about their travel experience: such as the number of trains that pass before boarding, in a crowded system; and specific critical incidents (CI) that the user may have experienced while riding the PT system. In the literature regarding PT satisfaction, very few studies have attempted to determine how reported CI may affect transit satisfaction, see Allen et al., 2018b, Friman et al., 1998, 2001, for details. In fact, these three studies conclude that negative CI do affect transit satisfaction significantly and negatively.

Nevertheless, PTA know that specific life experience and events may influence users' perceptions and, also, that every individual's experience is completely different. Thus, the average satisfaction of a user of given socioeconomic characteristics using the system, may be refined if we knew the attributes of the specific services he/she takes. This may be further improved if we knew which specific vehicles this person took during which days, allowing us for instance to understand the impact of occasional operational failures experienced by the system (i.e. critical incidents). Henceforth, we consider that it is of utmost importance for

PTA to link objectively measurable attributes of the service provided to users with their overall PT satisfaction assessment. First, they should wish to determine *if* satisfaction changes according to the objective attributes, then to know *how* it changed (i.e. improves or deteriorates), and finally, *why* it changed. This *"why"* quantifies the direct impact of operative variables in users' transit satisfaction. It leads to direct policy-related variables that could improve PT policy knowledge and add insights and recommendations for other cities and PT systems to use.

In the PT satisfaction literature, there are few studies that consider objective attributes, alongside subjective ones. In fact, we found only one that successfully combined objective and subjective (i.e. perceived) attributes to determine transit performance (Eboli and Mazzulla, 2011). Another, by Tyrinopoulos and Aifadopoulou (2008), applied a combination of objective and subjective attributes to determine compound indicators. These authors suggest using econometric techniques, such as factor analysis and multinomial logistic regression, to test the influence of some performance indicators of the transit system on users' satisfaction.

Our study provides further insights, since we consider the complete Metro network in a comprehensive multiannual survey. We also have an extensive and highly disaggregated databank of operational variables of the PT system, which helps by providing variability in the levels of service. We disentangle the effects of the different travel and socioeconomic characteristics of users on transit satisfaction. For our purposes, we propose categorising different statistics for the objective attributes: means, standard deviations, coefficients of variation and percentiles, and assign them to users depending on the period when the survey was conducted. We categorise by Metro line, day of the week, season, year and period of the day, allowing us to capture the variability of the objective attributes. We consider this an exciting contribution to PT policy development.

The rest of the chapter is as follows. Section 7.2 presents a literature review relevant to our case study. First, we offer details on the few multiannual customer satisfaction surveys performed in specific PT contexts. Next, we detail specific studies that have attempted to capture objective and subjective attributes to explain transit satisfaction. Section 7.3 introduces the Santiago Metro case study. Section 7.4 presents the sample, survey and objective attributes. Section 7.5 presents the model results, and Section 7.6 discusses our primary results and their interpretation in detail. Finally, in Section 7.7 we provide the most important conclusions and offer policy recommendations for PTA from our findings.

7.2. Literature Review

Few researchers have studied the evolution of users' transit satisfaction over time. In particular, Susilo and Cats (2014) examined how the causes of overall satisfaction with PT evolve. They analysed data from Sweden's major PT systems, from 2001 to 2013. They show that users' satisfaction has diminished recently, mainly because of the deterioration in satisfaction with customer interface, the length of trip duration, and with the operation. Although the study is insightful, it does not incorporate additional information on objective attributes of the PT system.

De Oña et al. (2018) analysed the effect of an economic downturn on transit users' satisfaction. They used data for the bus transit service of Granada (Spain), and estimated SEM models for three years: 2008, 2011, 2014. They show that the price of the ticket lost relevance as the financial crisis gained strength, a counterintuitive result. Additionally, perceived service factors had lesser influence on users' satisfaction in 2011, when the economic downturn was more intense. Still, the authors do not provide information about the objective attributes during the economic downturn, so it is not possible to assess whether the changes in these attributes affected transit satisfaction, or just the economic downturn *per se*. Efthymiou and Antoniou (2017) similarly examined the effect an economic downturn, in satisfaction and demand, with data from two years: 2008 and 2013, the latter being the *aftershock* of the economic downturn. They concluded that improved satisfaction with PT quality of service, together with increasing car operating and maintenance costs led to an increase in PT demand. However, as no information about the objective variables of the system was provided, it is not possible to disentangle the economic-downturn effect from the level-of-service effect.

As there is almost no literature on multiannual customer satisfaction surveys that take into consideration objective attributes, we consider this to be a critical gap in the PT satisfaction literature. Using a multiannual survey, our study adds to the literature by analysing how the variability of the objective attributes affects the users' transit satisfaction. Next, we examine studies that combine both objective and subjective attributes in transit satisfaction research.

Eboli and Mazzulla (2011) proposed a methodology based on the calculation of an indicator for each service aspect of a bus system, in the southern Italian city of Cosenza. They proposed combining both subjective and objective indicators, to compute a final indicator. The indicator is determined by minimising the distance from the subjective and objective indicators. They used data collected in 2008 from 123 users of the bus system. This data was used to compute a subjective indicator for different service quality attributes. Additionally, objective indicators were computed from available data about PT service attributes. Their study is insightful. However, the study considers just one time period, so it is not possible to consider seasonal variability in level of service (i.e. the objective attributes).

Tyrinopoulos and Aifadopoulou (2008) presented a comprehensive methodology developed by the Hellenic Institute of Transport, and applied in Athens (Greece) considering both objective and subjective attributes of the PT system. Four hundred users were asked to rank satisfaction and importance of several subjective attributes. Moreover, specific operational attributes were collected to measure the system's performance: vehicle load, average passenger waiting time at terminals and stops, on-time performance, average line speed, and others. With all this data, compound indicators were derived; the methodology assumes that these indicators may serve as performance-based indicators, to set up goals for the PTA.

Although the abovementioned studies combine objective and subjective attributes to derive performance indicators, we consider that a critical gap in the literature remains, which is to assess whether including objective attributes improves the prediction of transit satisfaction. In our case study, we build up specific operational variables' statistics for all possible combinations of the time periods when users travel. Additionally, we cover 42 months, from January 2013 to June 2016. Hence, we consider our study innovative, as it uses one model to assess fairly different conditions in the objective attributes, allowing to capture *how* the variability of those attributes influences transit satisfaction.

7.3. Santiago Metro Case Study

The Santiago Metro is one of the most modern-day subway networks in Latin America. It is the third largest after Mexico City Metro and the Metropolitan System of Sao Paulo. Regarding frequency, it is the seventh worldwide (Metro de Santiago, 2018). Currently, six lines are operational (L1, L2, L4, L4A, L5 and L6), with 118 stations and 118 kilometres of rail network. Line 3 is currently under construction, and Line 6 opened in late 2017. The Santiago Metro carries ~2.5 million passengers every work-day. This figure represents over a million more than in 2007 when the Transantiago project was launched. Transantiago is the integrated bus-Metro PT system currently operating in Santiago. As Transantiago was launched in a *"Big Bang"* fashion (see Allen et al., 2018a, Muñoz et al., 2008), Metro had to accommodate this unexpected surge in demand, causing delays, problems when boarding, and more crowded trains. The highest passenger peak stands at 2.78 million passengers in one day.

Line 3 will connect Quilicura and La Reina with 21.7 km and 19 more stations. Expected demand is 120 million annual trips for this line. Operations are expected to begin the second semester of 2018. In 2014, additional extensions were announced: Line 3 would be expanded northwest, and Line 2 would be expanded further south, adding seven new stations and an additional 8.8 kilometres of tracks by 2020. In 2017, plans were announced for the construction of Line 7, which will connect Renca in the northwest of Santiago and Vitacura in the northeast. This line will add 24.8 kilometres and 21 new stations to the Metro network (Metro de Santiago, 2018).

The Santiago Metro operates seven models of rolling stock, including steel-wheeled (AS) and rubber-tyred (NS). Most train types used forced-air circulation as they are not fitted with air conditioning. In 2012, the NS2012 trains went into service in Line 1, the first to be built with air conditioning. The number indicates the year of design, in order of oldest to newest: NS74 for L2 and L5, NS88 for L2 (weekends), NS93 for L1 and L5, AS2002 for L4 and L4A, NS2004 for L2, NS2007 for L1, NS2012 for L1, and AS2014 for L3 and L6. L3 is not operational yet, so the only lines that provide air-conditioning are L6 and L1 (but not always). Santiago has relatively hot although dry summers, with temperatures rising in the vicinity of 35°C. Also, L6 has platform safety barriers and driverless operation, unlike the other lines.

Many and diverse services are provided within each Metro station: ticket offices, public telephones, Metro-network information, ATM, automatic recharge machines. Metro also offers a library: *Bibliometro,* with services in 21 stations. Further, customers can rent parking spaces for their bicycles at ~US$0.50 a day. Most underground Metro stations contain at least one shop or convenience store. Various private security agencies have responsibility for maintaining order and reducing petty crime and fare evasion. In general, staff is present at stations at all times (Metro de Santiago, 2018).

The Santiago Metro is part of Transantiago, the integrated PT system that also has feeder and trunk bus routes. Transantiago works with an integrated fare system, which allows passengers to make bus-bus or bus-metro transfers on a two-hour limit. Bip! cards cost ~US$2.67 and can be recharged. Fares depend on the time of use of the system: rush hour (07-09, 18-20) costs ~US$1.18, off-peak (6.30-7, 9-18, and 20-20.44) and all-day weekends costs ~US$1.08, and the low-use hours (5.35-6.29, 20.45-0.08) costs ~US$1.00. Metro operates daily from 5.35 am to 12.08 am, on Saturdays from 6.30 am, and on Sundays and public holidays from 8 am (Line 1 from 9 am) until 11.48 pm. Figure 7-1 presents the Metro Network map, including Line 3, with a dotted line; Line 7, which is not included, will run northwest to northeast.

7.4. Sample, Survey and Operational Variables

7.4.1 The Metro tracking survey: a satisfaction questionnaire

Data supporting this research was collected by Metro S.A. through customer satisfaction surveys, carried out exclusively for the Metro system. The data was collected with a standardised questionnaire, including seven modules, from January 2013 until June 2016. The data is standard, except for some specific questions that were omitted during 2013 and 2014. These items concerned the number of trains that passed before boarding and whether the user transferred while using the Metro network, for 2013 only. Both are subjective attributes. The survey measures the overall and attribute-specific users' satisfaction perceptions. Metro designed the survey based on their experience and a review of current state-of-the-art practices. They have consistently performed satisfaction surveys since the 1990s, to gain knowledge about which attributes are more relevant to their user base. They always commission the survey to a marketing firm.

Figure 7-1: The Santiago Metro network map

Source: http://www.metro.cl/

The survey is conducted systematically across the whole network, covering all weekly time periods, 1000 surveys are collected every month and are equally distributed over five Metro lines (L1, L2, L4, L4A, L5). L6 was not operational at the time of the surveys. The distribution of the surveys by stations in each line is weighted by the number of boardings at each station. The users are surveyed once they pass the turnstiles and enter the boarding and alighting areas of the Metro stations. In a few cases, the interviewer accompanies the users if they boarded the trains, while being surveyed.

The seven modules of the *questionnaire* are:

a) User travel data. This includes the following information about the stations used: current origin and destination, habitual origin and habitual destination. Users are asked if they transfer between Metro lines, and at which stations. Next, information about the user and its travel patterns: number of weekly Metro trips, exact age, and ticket type. Students and elderly ticket holders are not furtherly surveyed. Users are asked about their specific days of Metro use and use schedule. The interviewer reports the time of the survey: Monday to Friday peak, Monday to Friday off-peak, or weekends, and also the exact time. After this, the following perception questions are asked:

   i. On average, how many trains pass before boarding?
   ii. Including all transport modes, how long is your habitual trip in minutes?
   iii. Of this overall time, how many minutes do you spend in the Metro?

b) Initial overall satisfaction. Users are asked: how satisfied they are with the service delivered by Metro (P1); this item is rated on a 1 to 5 Likert scale.

c) Satisfaction with Metro trips. Users are asked to rate 19 Metro-satisfaction items, classified in specific domains for their regular trips in the last 15 days. Safety items include security against theft, safety against accidents, and two other items. Two personnel items refer to the professionalism and kindness of Metro staff. One information item is the availability of Metro information. Facility items are the availability of commercial premises and of complementary services, and whether the Bibliometro service is useful for the community. All items are rated on a 1-5 Likert scale.

d) Satisfaction with Metro and station conditions. Users are asked to rate 18 Metro-satisfaction items, classified in specific domains for their regular trips and station conditions in the last 15 days. First, information availability items: trip-planning, station conditions, waiting conditions at the platform and transfer conditions with other modes. Second, ticketing conditions items: self-service ticketing, regular-vendor ticketing, vendor attentiveness, turnstile mobility and easiness of the validation process. Third, alighting mobility items: easiness to move within transfer stations, safety within transfer stations, order when waiting at platforms, safety when boarding and alighting, and safety inside the train. Finally, accessibility conditions items: cleanliness and safety of access to the stations. All items are rated on a Likert scale from 1 to 5.

e) Service interruptions and Metro response. Items in this module include service interruption questions and satisfaction items about Metro's response to such interruptions. First, users are asked if they have experienced a service interruption in the last 15 days. If the answer is Yes, they are asked about the types of service interruptions (P33): here, seven options are assessed including an "other" option for a total of eight. Accordingly, any user can have from 0 to 8 possible types of service interruptions (i.e. critical incidents). Next, three items referring to Metro response when interruptions occur: Metro staff attentiveness to passengers, clarity of information provided, and timeliness of the information provided.

f) Users' socioeconomic traits, includes users' socioeconomic data. Items include: being head of the household (HHH); if the user is currently working, work status and education level. These last two items combine with the location (current station) to make an additional socioeconomic status (SES). Other items are the willingness to participate in future studies, gender, the exact time when the survey ends, and day of the week.

g) Final overall satisfaction. The user is asked the following: considering the overall interview, how satisfied they are with the service provided by Metro (P37). This item was also rated on a Likert scale from 1 to 5.

7.4.2 Sample and survey results

The sample consists of continuous applications of the survey every month, from January 2013 until June 2016. Approximately 1000 questionnaires were collected each month, for a total of 41,993 completed questionnaires, 12,000 per year. In 2016, we only had available surveys until June, so there are just 6,000. Users aged 18-60, that take four or more trips per week, and that do not possess a *student* or *elderly* discount pass are considered. In Chile, summer extends from January until March, and winter from July until September.

Table 7-1 presents the users' travel characteristics both pooling the data and by year. In the pooled sample, most users do not transfer (58%). However, this varies by year, and for 2013, this question was not asked. Moreover, many users transfer in 2014 (88%), much less in 2015 (34%), and about half (52%) in 2016. In summary, the samples are not equally distributed for this attribute throughout the various years. About half of the users (47%) travel ten times per week, less travel 4-9 times (21%) or 11-12 times (19%). The smallest group travel 13 or more times per week (13%). The distributions are similar across years. Half of the users (50%) travel during off-peak hours, and 35% do so during peak hours.

For the perceived number of trains that pass before boarding, again there is no data for 2013 and 2014. Comparing 2015 and 2016, we notice that for 2016 more users report trains passing before boarding (43%) than in 2015 (31%). For the total travel time, the most substantial stratum is for 40-60 min (37%), next, the 20-40 min (28%) stratum. For Metro travel time, the largest corresponds to the 0-20 min (37%) and the 20-40 min (38%) strata. Most users (75%) have a less-than-forty-minute regular Metro trip.

Table 7-1: Sample travel characteristics: Pooled data and by year

CATEGORY	UNIT	All	2013	2014	2015	2016
SAMPLE	n	41993	11999	11994	12000	6000
	%	100	29	28	29	14
SEASON	Summer	29	25	25	25	50
	Autumn	29	25	25	25	50
	Winter	21	25	25	25	0
	Spring	21	25	25	25	0
METRO	L1	20	20	20	20	20
LINE	L2	20	20	20	20	20
	L4	20	20	20	20	20
	L4A	20	20	20	20	20
	L5	20	20	20	20	20
TRANSFER	Yes	42	*0*	88	34	52
(METRO)	No	58	*100*	12	66	48
WEEKLY	4-9	21	20	25	18	21
USE	10	47	43	45	53	49
(TRIPS/WK)	11-12	19	18	19	19	18
	>13	13	19	11	10	12
HABITUAL	MF.Peak	35	36	35	34	38
TRIP	MF.Valley	50	50	50	50	47
	Weekends	15	14	15	16	15
No. TRAINS	0	85	*100*	*100*	69	57
PASSING	1	6	*0*	*0*	11	17
BEFORE	2	5	*0*	*0*	11	13
BOARDING	>3	4	*0*	*0*	9	13
TOTAL	0-20	16	18	23	11	12
TRAVEL	20-40	28	30	24	30	30
TIME(MIN)	40-60	37	36	32	41	39
	>61	19	16	21	18	19
METRO	0-20	37	40	30	41	33
TRAVEL	20-40	38	39	33	40	43
TIME(MIN)	>41	25	21	37	19	24
TIME	07-09	20	19	20	21	22
OF DAY	09-13	21	21	21	22	22
	13-18	29	29	29	28	28
	18-20	17	18	17	17	17
	20-00	13	13	13	12	11
CRITICAL	0	77	65	73	90	80
INCIDENT	1	12	20	14	5	8
(N TYPES)	2	7	11	7	3	6
	>3	4	4	6	2	6

Regarding time of day of the trip, the most prominent strata corresponds to the 13-18 interval (29%). The other intervals are similarly distributed, except the 20-00 stratum which is smaller (13%). For the perceived critical incidents (CI), most users (77%) do not perceive any, 12% perceive one CI, and 11% perceive at least two types of CI. However, the distributions are not the same for all years; 2013 had the most reported CI (35%), and 2015, the least (10%).

Users' socioeconomic characteristics are presented both for the pooled data and by year in Table 7-2. Although most users are Female (54%), the age groups are equally distributed for the pooled data and across years. Most users are currently employed (58%). However, notice that the distributions are very different across years. For 2013, half (50%) are employed, while for 2014 only 9% were employed, and in 2015-2016 a high majority (96%) were employed. One could hypothesise that the differences in employment may be due to an economic downturn being at its highest peak in 2014, in Chile. However, we suspect that some error may have been made when coding this variable. Thus, we will be cautious regarding this variable.

Most users (58%) are head of the household (HHH), and the distributions are similar in every year. For the work status categories, the most prominent strata are Unemployed-WS1 (27%) and Employee-WS5 (39%). However, notice again that 2014 is very different from other years. For 2014, 94% of HHH reported being unemployed. We conclude that for this year, a problem indeed occurred with the work status question. *We will probably discard this variable in the SEM models*.

Next, we discuss the educational levels. In general, most users have Secondary or University studies (79%), but again there are some variations across years, although not as disproportionate as the work status variable. We notice a high proportion of users with incomplete secondary studies in 2013 (35%).

Table 7-2: Sample socioeconomic characteristics: Pooled data and by year

CATEGORY	UNIT	ALL	2013	2014	2015	2016
GENDER	Male	46	48	47	44	46
	Female	54	52	53	56	54
AGE (YEARS)	18-24	20	22	22	17	18
	25-31	21	19	21	23	24
	32-39	20	21	19	20	20
	40-48	21	22	21	21	20
	49-60	18	16	17	19	18
CURRENTLY WORKS	Yes	58	50	*9*	97	95
	No	42	50	*81*	3	5
HEAD OF HOUSEHOLD (HHH)	Yes	58	61	57	57	57
	No	42	39	43	43	43
WORK STATUS OF HHH	No work-WS1	27	2	*94*	1	0
	Minor-WS23	8	9	*3*	10	9
	Skilled-WS4	23	26	*2*	38	28
	Employee-WS5	39	58	*1*	47	58
	Exec-WS67	3	5	*0*	4	5
EDUCATION OF HHH	Primary-ED.LV12	7	12	2	7	6
	Sec.Incom-ED.LV3	14	*35*	1	9	11
	Sec.Compl- ED.LV4	26	4	17	50	40
	Uni.Incom- ED.LV5	28	24	41	19	27
	Uni.Comp- ED.LV67	25	25	39	15	16
SOCIOECONOMIC STATUS (SES) (HOUSEHOLD)	E (low)	**2**	6	2	2	**1**
	D (mid-low)	**19**	25	17	16	**14**
	C3 (mid-mid)	**42**	34	44	48	**42**
	C2 (mid-high)	**32**	31	31	30	**39**
	ABC1 (high)	**5**	4	6	4	**4**

Next, we present the SES variable *GSE* (*Grupo socioeconómico*, in Spanish), which is used as a standard in Chile to categorise SES. Although many different factors are considered, in the survey this variable was recorded by the interviewer according to three factors: location (station), educational level, and work status of the user. Marketing companies use this variable for stratification purposes as a standard, and it is also used in academic circles. We believe that the GSE probably has a better measurement precision than the other two traits: educational level and work status. Confidence in this assertion is provided by the distribution of the SES;

for all, the biggest stratum is the C3 group (42%), second in size is the C2 group (32%), and third, the D group (19%). The distribution holds for all years, save for a slight variation in 2013. According to Metro's official data (Metro de Santiago, 2016), the real distribution for all groups for 2016 was: C3 with 33%, DE with 29%, C2 with 23%, and ABC1 (the highest income group) with 15%. The distributions in the survey differ due to the non-random sampling methodology, that some user groups were discarded, and because some SES (i.e. C2 and C3) users travel more frequently in the Metro system.

The average survey results, for the pooled data and each year, are presented in Table 7-3. We comment on the results for the pooled data by grouping them into specific domains: safety, personnel, information, additional facilities, ticketing, alighting mobility, accessibility conditions, and response to critical incidents. Finally, the overall satisfaction items are shown at the bottom of the table. All items are scored using a 1-5 Likert scale, except P37 that used a 1-7 range.

From the safety domain, the best-rated items (3.74 both) are *safety against accidents* (P3) and *travel time is easy to predict* (P4) and the worst, *security against theft* (P2, 3.63). Both items in the personnel domain are highly scored (3.80 and 3.77). In the information domain, the best-rated is *information about trip planning* (P17, 3.81), but the rest are highly scored also. The additional facilities items have the highest scores, and *Bibliometro is useful for the community* has the best score (P13, 4.15). In the ticketing domain, *turnstile mobility* (P22, 3.99) is the best rated. Again, the rest is highly rated too. In the alighting mobility domain, the best rated is *safety in transfer stations* (P25, 3.62) and the worst is *safety when boarding* (P28A, 3.42). Finally, in the accessibility conditions, *cleanliness in the access areas* (P31A, 3.86) is better assessed than the *safety of access areas* (P31B, 3.72). Notice that, so far, all average values are over 3.0, which could suggest a passing (acceptable) grade.

Table 7-3: Survey results (mean). All Likert type (1-5) except P37 (1-7)

DOMAIN	Service quality attributes	All	2013	2014	2015	2016
Safety	P2: Security against theft	3.63	3.58	3.56	3.76	3.59
	P3: Safety against accidents	3.74	3.76	3.73	3.77	3.67
	P4: Travel time is easy to predict	3.74	3.77	3.75	3.77	3.60
	P5: Metro cares about users' travel conditions	3.64	3.61	3.61	3.73	3.59
Personnel	P8A: Professionalism of Metro staff	3.80	3.73	3.83	3.80	3.90
	P8B: Kindness of Metro staff	3.77	3.70	3.78	3.76	3.92
Information	P9A: Metro information availability	3.80	3.78	3.81	3.84	3.75
	P17: Trip-planning information	3.81	3.73	3.73	3.84	3.73
	P18: Station-conditions information	3.80	3.79	3.83	3.83	3.67
	P27: Platform waiting-conditions information	3.75	3.65	3.78	3.82	3.74
	P30: Transfer with other modes information	3.75	3.68	3.80	3.83	3.63
Additional Facilities	P11: Commercial premises availability	3.91	3.96	3.91	3.87	3.86
	P12: Complementary services availability	4.02	4.09	4.05	3.96	3.94
	P13: *Bibliometro* is useful for the community	4.15	4.01	4.24	4.12	4.28
Ticketing	P19: Self-service ticketing	3.92	3.94	3.96	3.91	3.79
	P20: Regular-vendor ticketing	3.83	3.77	3.88	3.90	3.73
	P21: Regular-vendor attentiveness	3.93	3.93	3.95	3.94	3.83
	P22: Turnstile mobility	3.99	4.00	4.06	3.95	3.88
	P23: Validation process easiness	3.88	3.88	3.96	3.95	3.55
Alighting Mobility	P24: Easiness to move within transfer stations	3.56	3.41	3.47	3.76	3.63
	P25: Safety within transfer stations	3.62	3.46	3.63	3.78	3.60
	P26B: Order when waiting at platforms	3.61	3.45	3.53	3.80	3.69
	P28A: Safety when boarding and alighting	3.42	3.23	3.26	3.70	3.55
	P29: Safety inside the train	3.47	3.32	3.33	3.70	3.59
Accessibility Conditions	P31A: Cleanliness in access to Metro areas	3.86	3.84	3.88	3.88	3.82
	P31B: Safety in access to Metro areas	3.72	3.70	3.64	3.82	3.71
Response to Critical Incident	P34A: Metro staff attentiveness to passengers	3.02	3.06	3.00	2.99	3.02
	P35: Clarity of information provided	2.99	3.03	2.96	2.98	2.98
	P36: Timeliness of information provided	2.98	3.01	2.93	2.99	3.00
Overall	P1: Initial overall satisfaction with Metro (1-5)	3.69	3.76	3.72	3.74	3.42
	P37: Final overall satisfaction with Metro (1-7)	4.91	4.84	4.94	4.97	4.86

The items related with the response to critical incidents domain were only asked when users reported being involved in a service interruption during the last 15 days, so there are many missing values. We imputed a three (3) value for all these cases, a neutral score in the 1-5 Likert scale, to avoid introducing bias, as these users did not answer this questions. Unexpectedly, we notice that even after imputing all these missing values, the three items are scored the lowest in the questionnaire (3.02, 2.99, 2.98); only the first item, *Metro staff attentiveness to passengers* is over par. Thus, we can immediately conclude that these policy items could be improved, as they reveal that most users who endured a CI, scored it subpar in this domain.

The initial Metro overall satisfaction mean value is 3.69 (P1), and the final overall satisfaction is 4.91 (P37), although scored in a 1-7 Likert scale. The initial satisfaction was the first question (P1), and the final satisfaction the last question of the survey (P37). By standardising both values to a 0-10 ranking system the final satisfaction (P37, 6.52/10) is inferior to the initial satisfaction (P1, 6.73/10). On this issue, Dell'Olio et al. (2010) obtained an opposite result as, on average, the final assessment was better than the initial one in their study. As in our case, the items just before the final assessment were the critical incident questions (i.e. adverse events), it could be that users tended to lower their grades because of this. On another hand, we also believe that P37 may have a better measurement since the school grading system in Chile uses a 1-7 scale, with four (4) as the minimum passing grade. So, probably users felt more comfortable with the 1-7 scale, and thus gave a more precise grade.

7.4.3 Operational variables

In this subsection we explain the operational variables provided by Metro S.A. and our approach to linking them with the survey respondents. There are three datasets regarding crowding levels, speed, and critical incidents over five minutes (i.e. trains operations were stopped for more than five minutes). After cleaning the datasets from abnormal data and outliers, we used predictive mean-matching to impute some missing values. In all three datasets, the missing values were less than 9% for each variable that had any missing values. Next, we comment on the structure of the datasets.

The first dataset, includes 11 variables after cleaning: line, direction, station, day of the week, day of the month, month, year, hour, load (weight), crowding (passenger density), and frequency (trains/15 minutes). It includes 1,044,156 observations at 15-min intervals. The two relevant objective variables are crowding and frequency. The crowding variable (passengers/m^2) is estimated based on the load (weight) of the train, measured at certain stations that have a balance for weighing. Based on this information and user validations at each station, Metro estimates the load of each train at every other station. We obtained the data already processed.

As stated before, we eliminated abnormal values, and limited the maximum passenger/m^2 to 7, a high standard typically used in the public transport literature.

Given the distribution of the frequency variable, we decided to set a maximum of 10 trains per 15 min. The crowding statistics correspond to: a mean of 1.41 passengers/m^2, a median of 1.00 passengers/m^2, and the third quartile of 2.00 passengers/m^2. The 15-min frequency has a mean of 4.10, a median of 4.00, and the third quartile of 5.00 (all in trains/15 min). The data covers all time periods when the trains are operating, for the same 42 months (i.e. from January 2013 until June 2016).

The second dataset, the commercial speed dataset, also includes 11 variables after cleaning: line, one-hour frequency, year, month, day of the month, day of the week, hour, travel time (in seconds), distance, commercial speed (km/h), and standard deviation of the travel time (in seconds). It includes 436,735 observations at 1-hour intervals. The most relevant objective variables are commercial speed and one-hour frequency. The commercial speed (km/h) was cleaned from abnormal values, by limiting the minimum commercial speed to 10 km/h and the maximum to 50 km/h. The 1-hour frequency variable was left as given since it did not present abnormal values. The commercial speed has a mean of 35.9, a median of 35.8, the third quartile of 39.0, with a minimum of 10.1, and a maximum of 49.95, all in km/h. The 1-hour frequency has a mean of 13.3, a median of 13.0, the third quartile of 17.0, and a maximum of 33.0, all in trains/h.

The third and final dataset, the critical incidents (CI) dataset, includes 19 variables after cleaning: line, day of the week, direction, year, month, day of the month, hour, minutes of duration, and lost time, the latter two in minutes. In essence, each entry represents a CI of more than five-min duration in the Metro system. For the 42-month period, there were 1,012 CI. The other key variable is the duration, with a mean of 18.1, a median of 7.0, the third quartile of 11.0, and a maximum of 1105, all in minutes.

With the three clean datasets, we can obtain summary statistics for the specific users during their surveyed periods. We know that operational levels of service vary across seasons, time periods, years, lines, and type of day of the week. Also, it is highly likely that even though users were asked about their last 15 days of Metro use, they probably take into consideration at least one month when assigning their overall satisfaction. Also, we know that both the weather and the operational conditions change significantly across seasons. For these reasons and for easiness of interpretation, we grouped the data by season.

We obtain a summary of the distributions for the variables of interest. These are crowding, 15-minute frequency, commercial speed, and 1-hour frequency. We obtained the mean, standard deviation, coefficient of variation, 30% (70%) and 15% (85%) percentiles. We used either 15% and 30% or 70% and 85% depending on what the critical conditions for the users were (i.e. for crowding we used 70% and 85%).

For the CI duration, we obtained the length value (i.e. number of critical events), the mean duration, and the percentiles. Last, we standardised the number of CI by the number of hours in each period, to obtain a CI/h variable in the respective period. Approximately 700 sets of different categories were obtained for all variables according to the time periods and dates, allowing us to obtain variability in the summary statistics.

The final step was to link the users with these statistics by merging both datasets using the following variables: line, day of the week (Monday to Friday, or weekend), season, year, and period (7-9, 9-13, 13-18, 20-0, 0-7). The task was to assign a statistic for the operative variables matching the period, line and year when the survey was performed for each user.

7.5. Model Results

In this section, we present the model results; we worked with the pooled data in all cases. Included are: (i) PCA for the satisfaction items, (ii) SEM ordinal models with CI reported by the user, (iii) SEM-MIMIC ordinal model with travel characteristics and socioeconomic traits, and finally (iv) SEM-MIMIC ordinal model including all users' characteristics and their respective objective attributes. We used *R* (R Core Team, 2013) and its associated library packages; specifically, the *Lavaan* package for *R* (Rosseel, 2012) in all SEM models. In most tables the values are rounded to two decimal places for readability.

7.5.1. Principal component analysis: satisfaction domains

Firstly, we want to find out how users perceive the different satisfaction constructs. For this, best practice is to run an exploratory PCA (Hoyle, 2012, Jolliffe, 2014) on the satisfaction items. We can determine which items represent specific latent constructs, that is, which items appear to be grouped together by users.

Thus, we ran a PCA on the set of satisfaction items, obtaining eight components according to the Kaiser (1960) rule. When we ran the initial PCA, some items were left out because they loaded into two components. The criterion used was that the items had a resulting absolute loading >0.3, exclusively on one component.

Next, with the eight components, we computed Cronbach's alpha to assess internal validity. When two or more items were almost identical when assessed using Cronbach's alpha, we eliminated one. The results are presented in Table 7-4.

Table 7-4: PCA satisfaction items

	SAFE	PERS	INFO	FACI	TICK	MOBI	ACCE	CRIT	OVS
P2: Security against theft	**-0.56**	-0.01	0.04	0.02	-0.01	-0.02	-0.04	-0.01	
P3: Safety against accidents	**-0.55**	0.01	0.01	-0.01	0.04	0.05	-0.00	-0.00	
P4: Travel time is easy to predict	**-0.41**	-0.01	-0.08	-0.00	0.08	-0.03	0.13	-0.00	
P5: Metro cares about users' travel conditions	**-0.41**	0.02	-0.05	-0.02	-0.05	-0.10	0.02	0.01	
P8A: Professionalism of Metro staff	-0.03	**0.67**	-0.02	0.01	0.00	0.01	0.00	-0.00	
P8B: Kindness of Metro staff	0.02	**0.71**	0.01	0.00	-0.01	-0.01	-0.01	0.00	
P9A: Metro information availability	-0.10	0.16	**-0.36**	-0.04	-0.04	0.04	0.04	0.01	
P17: Trip-planning information	-0.00	0.00	**-0.52**	0.01	0.01	0.04	0.00	0.00	
P18: Station-conditions information	-0.01	-0.03	**-0.51**	0.00	0.03	0.03	0.03	-0.00	
P27: Platform waiting-conditions information	0.08	0.01	**-0.33**	-0.01	-0.01	-0.19	-0.00	0.00	
P30: Transfer with other modes information	0.05	-0.05	**-0.41**	0.00	-0.06	-0.06	-0.18	-0.01	
P11: Commercial premises availability	-0.04	-0.04	-0.06	**-0.56**	-0.04	-0.01	0.01	0.00	
P12: Complementary services availability	-0.01	-0.03	-0.03	**-0.61**	-0.01	0.02	0.00	-0.00	
P13: *Bibliometro* is useful for the community	0.05	0.06	0.10	**-0.55**	0.06	-0.02	-0.01	-0.00	
P19: Self-service ticketing	-0.02	-0.03	-0.10	-0.04	**0.39**	0.09	-0.05	-0.00	
P20: Regular-vendor ticketing	-0.04	0.01	0.01	0.02	**0.39**	-0.03	-0.07	0.01	
P21: Regular-vendor attentiveness	-0.03	-0.01	0.06	-0.00	**0.48**	-0.01	-0.06	0.01	
P22: Turnstile mobility	-0.03	-0.01	0.00	0.00	**0.48**	0.07	-0.05	0.00	
P23: Validation process easiness	0.04	-0.01	-0.01	-0.01	**0.41**	-0.14	0.11	-0.01	
P24: Easiness to move within transfer stations	0.05	-0.03	-0.07	0.06	0.14	**-0.36**	0.08	0.01	
P25: Safety within transfer stations	0.09	0.06	-0.06	0.02	0.13	**-0.39**	0.14	-0.01	
P26B: Order when waiting at platforms	0.00	0.04	0.03	-0.03	0.01	**-0.43**	-0.01	0.00	
P28A: Safety when boarding and alighting	-0.04	-0.03	0.07	-0.02	-0.06	**-0.49**	-0.07	0.00	
P29: Safety inside the train	-0.08	-0.02	0.00	-0.00	-0.11	**-0.43**	-0.09	-0.00	
P31A: Cleanliness in access to Metro areas	0.05	0.03	-0.02	-0.00	0.05	0.03	**-0.69**	-0.01	
P31B: Safety in access to Metro areas	-0.05	-0.02	0.01	0.01	-0.01	-0.05	**-0.64**	0.01	
P34A: Metro staff attentiveness to passengers	0.02	0.02	0.00	-0.01	0.01	0.03	-0.00	**0.58**	
P35: Clarity of information provided	-0.00	-0.01	-0.01	-0.00	0.00	0.00	0.00	**0.60**	
P36: Timeliness of information provided	-0.02	-0.02	0.00	0.01	-0.02	-0.03	0.00	**0.55**	
Alpha Cronbach (α)	**0.80**	**0.83**	**0.83**	**0.70**	**0.76**	**0.81**	**0.75**	**0.89**	**0.77**

Notice that we also include Cronbach's alpha for the overall satisfaction domain. The resulting final eight components are safety (SAFE), personnel (PERS), information (INFO), additional facilities (FACI), ticketing (TICK), alighting mobility (MOBI), accessibility (ACCE), and response to critical incidents (CRIT). All items have high validity (>0.7) measured via Cronbach's alpha. The overall satisfaction also has high validity, providing confidence in the measurement of the dependent variable: overall satisfaction (OVS).

7.5.2. Structural equations models: SEM ordinal Probit model with CI reported by the user

The PCA results were used to build our SEM ordinal models. The SEM ordinal treats the satisfaction items with an ordinal range (for details, see Allen et al., 2018b), as all satisfaction items have a Likert-type range. For all SEM models, we assumed that all the satisfaction constructs and the CI.TOTN (number of different type of CI reported by the user) variable regress on the overall satisfaction (OVS). In Table 7-5, we present the results of the SEM ordinal Probit model with CI reported by the user. We show estimates, standard errors (S.E.), Z-values, and standardised coefficients (Std.Coeff.). The latter refers to how many standard deviations a dependent variable changes, per standard deviation increase in the predictor variable. From the measurement model, we conclude that all satisfaction constructs have high reliability (Std.Coeff. >0.6) in the measurement; for these constructs, all values lie within the 0.67-0.92 range.

Specific items that load more on the latent constructs are the items the users consider more important in each case. For SAFE, it is *Metro cares about users' travel conditions* (P5), for PERS, it is *the professionalism of Metro staff* (P8A), for INFO, it is *trip-planning* and *station-conditions information* (P17 and P18). For FACI, it is *commercial premises* and *complementary services availability* (P11 and P12), for TICK, it is *regular-vendor attentiveness* (P21), for MOBI, it is *order when waiting at platforms* (P26B), and for ACCE, it is *safety in access to Metro areas*. For CRIT, it is *clarity of information provided* (P35), and for OVS, it is *the final satisfaction assessment* (P37).

For the regression part of the SEM model, the explained variance, 0.60, (R^2) is adequate (>0.6); we consider an R^2 >0.5 reasonably acceptable and >0.6 adequate.

Table 7-5: SEM ordinal Probit: Pooled data with Critical Incidents reported by the user

Measurement	Estimate	S.E.	Z-value	St.Coeff.	R^2
SAFE					
P2	1.00			0.75	0.56
P3	1.04	0.01	217.35	0.77	0.60
P4	0.99	0.01	185.06	0.74	0.54
P5	1.05	0.01	199.30	0.79	0.62
PERS					
P8A	0.92	0.00	396.87	0.92	0.85
P8B	0.85	0.00	343.00	0.85	0.72
INFO					
P9A	1.00			0.75	0.56
P17	1.04	0.01	228.62	0.78	0.61
P18	1.04	0.01	224.18	0.78	0.60
P27	0.99	0.01	207.23	0.74	0.55
P30	0.98	0.01	205.21	0.74	0.54
FACI					
P11	1.00			0.78	0.61
P12	1.00	0.01	155.88	0.78	0.61
P13	0.87	0.01	128.49	0.67	0.46
TICK					
P19	1.00			0.67	0.45
P20	1.05	0.01	171.88	0.70	0.50
P21	1.07	0.01	167.81	0.72	0.52
P22	1.02	0.01	154.50	0.69	0.47
P23	1.02	0.01	158.58	0.69	0.47
MOBI					
P24	1.00			0.70	0.49
P25	1.02	0.01	179.79	0.71	0.50
P26B	1.09	0.01	184.93	0.76	0.58
P28A	1.04	0.01	176.53	0.72	0.52
P29	1.03	0.01	173.15	0.72	0.52
ACCE					
P31A	0.81	0.00	313.93	0.81	0.66
P31B	0.84	0.00	336.17	0.84	0.70
CRIT					
P34A	1.00			0.84	0.70
P35	1.13	0.00	325.79	0.94	0.89
P36	0.96	0.00	416.11	0.81	0.65
OVS					
P1	0.49	0.00	171.07	0.75	0.57
P37	0.59	0.00	144.07	0.90	0.81
Regressions	**Estimate**	**S.E.**	**Z-value**	**St.Coeff.**	R^2
OVS					0.60
SAFE	1.07	0.02	58.42	0.50	
INFO	0.19	0.02	9.93	0.09	
MOBI	0.36	0.02	18.42	0.16	
CRIT	0.13	0.01	12.13	0.07	
CI.TOTN	-0.54	0.01	-54.48	-0.30	

CFI	TLI	GFI	AGFI	RMSEA	SRMR	WRMR
0.988	0.987	0.992	0.998	0.054	0.027	9.281

For the structural coefficients, values below 0.1 are considered very low, between 0.1 and 0.3, low, between 0.3 and 0.5, moderate, and above 0.5 high (Allen et al., 2018a, Currie and Delbosc, 2017). The most relevant construct is safety (0.50), and the second most relevant is CI.TOTN (-0.30). This variable represents the total number of different types of CI, from zero to eight, that the user reported on the survey. Considering the remaining satisfaction constructs, the only ones that were significant at 85% were information, alighting mobility, and Metro's response to critical incidents. From the fit indices, the CFI, TLI, GFI, and AGFI all comply with the recommended cut-off (>0.95) values proposed by Hu and Bentler (1999). The RMSEA (0.054), the most critical index, also complies with the cut-off value (<0.08).

7.5.3. Structural equations models: SEM-MIMIC ordinal Probit model with travel and socioeconomic traits

In this subsection, we present the SEM-MIMIC (Joreskog and Goldberger, 1975) ordinal model. We base this model on the previous one, the SEM ordinal, and follow the methodology proposed by Allen et al. (2018b). We introduced dummy variables into the OVS regression in an attempt to capture the heterogeneity of perceptions depending on the users' travel and socioeconomic characteristics. For all categories, we added *n-1* dummy variables, where *n* is the number of groups in each category. We introduced (base in parentheses) three regressors for year (2013), three for season (summer), four for Metro-line (L1), one for transfer (No), three for frequency of travel (4-9 trips/week), two for time period of the survey (MF Peak), three for total travel time (TTT0.20), two for Metro travel time (MTT0.20), one for being the head of the household (No), and one if the person was currently employed (Unemployed).

Additionally, we introduced four dummies for work status (WRK.ST1), one for gender (Male), four for the current time (07-09), four for education level (ED.LV12), four for the SES level (SES.E), and four for the age group (AGE18.24). Lastly, we introduced one numeric variable, TR.PASSN, representing the numeric value of the trains that the user has to let pass before boarding; this was coded from zero to four (including more than four).

In total, we included 44 dummy variables and one numeric one, and kept those that resulted significant at the 85% level (25 dummy variables in the model, and the two numeric ones), see Table 7-6. The variables that did not prove significant were (number of categories for that variable): Y2016 (1), Winter (1), Line 2 (1), TR.WK14 (1), the total travel time variables (3). The work status variables were also not significant (4), supporting our initial assumption that they may have had problems during the data collection process. Finally, ED.LV3 (1), SES.D (1) and being the head of the household (1) were also not significant.

Considering the significant variables, those with the higher negative coefficients (value in parenthesis) were: *TR.PASSN* (-0.14) and *MTT41* (-0.11). The dummies associated with the year of the survey show positive estimates for 2014 and 2015, indicating that there was a systematic improvement, not explained by the structural model or by other variables, in the years 2014 and 2015. There is also a systematic negative value for the spring season.

Surprisingly transfers are positively valued. Notwithstanding, notice that Line 1 has an inferior level of satisfaction concerning other lines (except Line 2) and Line 1 is the one with most transfers, more connectivity with other lines. This result may be a positive effect of connecting to another Metro-line, and it is low (0.02).

People that are currently employed are under satisfied; this may be related to having to arrive at a particular time, meaning that they are users with a higher value of time. Females have less transit satisfaction, a result which is consistent with the PT literature (see Allen et al., 2018a). An unexpected result is that all the age variables produced positive and increasing coefficients, indicating more satisfaction as users grow in age. We hypothesise that Metro offers a superior service in some aspects that are valued more as people age (i.e. hedonic attributes such as customer service).

Table 7-6: SEM-MIMIC ordinal Probit with travel and SES traits

Regressions	Estimate	S.E.	Z-value	St.Coeff.	R^2
OVS					0.65
SAFE	1.10	0.02	58.60	0.48	
INFO	0.15	0.02	7.57	0.07	
MOBI	0.44	0.02	20.78	0.18	
CRIT	0.14	0.01	12.85	0.07	
CI.TOTN	-0.50	0.01	-47.57	-0.26	
Y2014	0.08	0.03	2.26	0.02	
Y2015	0.16	0.02	6.76	0.04	
AUTUMN	0.03	0.02	1.47	0.01	
SPRING	-0.12	0.02	-5.21	-0.03	
LINE4	0.21	0.03	8.11	0.05	
LINE4A	0.50	0.03	20.15	0.12	
LINE5	0.25	0.02	10.46	0.06	
TRANSFER	0.08	0.02	3.13	0.02	
TR.WK10	-0.07	0.02	-3.40	-0.02	
TR.WK12	0.09	0.03	-3.40	-0.02	
MF.VALLEY	0.42	0.02	21.70	0.12	
WEEKENDS	0.46	0.03	16.49	0.10	
TR.PASSN	-0.29	0.01	-25.66	-0.14	
MTT20.40	-0.26	0.02	-12.38	-0.07	
MTT41	-0.42	0.02	-17.68	-0.11	
WORKS	-0.11	0.03	-4.13	-0.03	
FEMALE	-0.21	0.02	-11.66	-0.06	
AGE25.31	0.11	0.03	3.90	0.03	
AGE32.39	0.23	0.03	8.21	0.06	
AGE40.48	0.28	0.03	10.30	0.07	
AGE49.60	0.47	0.03	16.30	0.11	
ED.LV4	-0.19	0.03	-6.45	-0.05	
ED.LV5	-0.16	0.03	-4.80	-0.04	
ED.LV67	-0.27	0.04	-7.31	-0.07	
SES.C3	-0.12	0.03	-4.75	-0.03	
SES.C2	-0.15	0.03	-4.54	-0.04	

CFI	TLI	GFI	AGFI	RMSEA	SRMR	WRMR
0.980	0.979	0.982	0.978	0.041	0.027	7.713

Higher education levels have less satisfaction than lower levels. Also, the C2 and C3 SES groups are under satisfied. We conclude that with higher education and higher SES there is a higher expectation, but the effect is not present for the ABC1 as these are probably choice travellers (only 5% of respondents). As C2 and C3 are probably captive users, they have less satisfaction. Similar results have been reported by Allen et al. (2018a) for the bus system in Santiago.

All the satisfaction constructs and the CI.TOTN variable produced similar coefficients to the original SEM model, although some of them decreased slightly. This result attests to the fact that the SEM-MIMIC model allows discerning for heterogeneity in the transit satisfaction perception. The model complies with the recommended cut-off values (Hu and Bentler, 1999). We obtained an excellent fit to the data as the CFI, TLI, GFI and AGFI are all above 0.97. Also, the RMSEA is not only less than 0.05, but the model has a smaller RMSEA than the original SEM model, indicating a better fit. The WRMR also improves. Thus, adding all the dummy variables improves the model fit, and allows to capture the heterogeneity in travel and socioeconomic traits. The explained variance improves to 0.65, compared to the original SEM model (0.60), a motivating result.

7.5.4. Structural equation models: SEM-MIMIC ordinal Probit model with subjective and objective attributes

First, we present the tentative objective attributes considered in the next SEM-MIMIC ordinal model. As mentioned in Section 7.3, we considered five primary variables: crowding, 15-min frequency, commercial speed, one-hour frequency, critical incidents (CI) per hour, and duration of CI in minutes. After merging the datasets for all respondents, we obtained the results shown in Table 7-7.

Table 7-7: Objective variables statistics for surveyed users

Objective attribute	MIN	1st Q	Median	Mean	3rd Q	MAX
CROWD.mean	0.40	1.21	**1.72**	**1.85**	2.47	3.83
CROWD.cv	0.33	0.45	**0.63**	**0.66**	0.79	1.43
CROWD.p85	0.90	2.00	**2.60**	**3.17**	4.40	6.00
FR15minute.mean	2.22	3.76	**4.37**	**4.42**	4.99	7.11
FR15minute.cv	0.16	0.20	**0.23**	**0.25**	0.28	0.46
FR15minute.p15	1.00	3.00	**4.00**	**3.62**	4.00	6.00
CSPEED.mean	24.32	34.28	**35.53**	**35.20**	37.55	43.58
CSPEED.cv	0.01	0.04	**0.05**	**0.06**	0.07	0.24
CSPEED.p15	21.89	32.66	**34.21**	**33.55**	35.81	42.44
FR1HR.mean	3.14	12.20	**15.85**	**16.16**	19.97	27.54
FR1HR.cv	0.02	0.13	**0.17**	**0.20**	0.25	1.48
FR1HR.p15	0.00	10.00	**13.00**	**13.96**	18.00	26.00
CIminute.mean	0.00	0.00	**6.73**	**10.39**	10.55	92.00
CIminute.p85	0.00	0.00	**7.00**	**14.12**	12.60	188.00
CI.NperHR	0.00	0.00	**0.50**	**0.58**	1.00	4.50

The mean crowding levels are not considered high; the mean is only 1.85 passengers/m^2, and the 85th percentile is 3.17 passengers/m^2. However, the mean coefficient of variation for crowding is high (0.66), implying differences within the time periods. The 15-min frequency has a mean of 4.42, and the one-hour frequency a mean of 16.16. Their coefficients of variation (CV) are of similar magnitude.

The commercial speeds vary from 24.3 to 43.6 km/h, and this fact is essential as L1 has a lower commercial speed than all the others, due to having less distance between stations and the highest demand. However, notice that the CV is low for this variable, implying regularity across most time periods. Still, we notice a rather high maximum CV (0.24) compared to the mean, medians, and quantiles. For critical incidents duration, the means are higher than the medians, indicating some exceptionally large values; we can also notice this from the maximum values. For CI.N/h (number of critical incidents/hour) the median is 0.50 and the mean 0.58, the third quantile is 1.00; this represents the number of CI per hour, in each specific period and season.

For the final model, the SEM-MIMIC ordinal Probit with objective and subjective variables, we included the original 44 variables, plus the new 15 objective attributes (Table 7-8). We expect these variables to provide significance in the SEM-MIMIC model; however, we do not expect them all to be significant. For example, the percentiles variables may not be significant, as we also include the means and CV.

We built the model by eliminating all variables that were not significant at the 85% level. First, we notice that the year and season significant dummies changed, in comparison with the first SEM-MIMIC model. Also, two of the three Line dummies decreased their coefficient. These facts indicate that by adding operational variables, we capture more of the satisfaction variability. Another variable that entered the model is TIME18.20, with a negative sign; this fact indicates that there is a systematic decrease in satisfaction in this period, compared to the 07-09 base period, over the operational variables. This effect was not captured in our original SEM-MIMIC model, but other variables have similar coefficients and signs.

We will now comment on the objective attributes' effects. First, for all variables the most significant statistics were the mean and the CV. This result is motivating, as it means that users perceive not only the mean levels of service but also their variability. In particular, the mean and CV of crowding have a negative sign. This result is expected; it means that users are less satisfied with more crowded trains, and even more under satisfied if there is considerable variability in this attribute.

Table 7-8: SEM-MIMIC ordinal Probit results: Subjective and objective variables

Regressions	Estimate	S.E.	Z-value	St.Coeff.	R^2	
OVS					0.65	
SAFE	1.09	0.02	58.41	0.48		
INFO	0.15	0.02	7.25	0.07		
MOBI	0.46	0.02	21.45	0.18		
CRIT	0.15	0.01	13.10	0.07		
CI.TOTN	-0.49	0.01	-47.56	-0.25		
Y2015	0.10	0.03	3.08	0.03		
Y2016	-0.07	0.04	-1.98	-0.02		
WINTER	-0.08	0.02	-3.38	-0.02		
SPRING	-0.18	0.02	-7.25	-0.04		
LINE4	0.14	0.04	3.95	0.03		
LINE4A	0.28	0.05	5.45	0.07		
LINE5	0.32	0.03	9.84	0.08		
TRANSFER	0.08	0.02	3.72	0.02		
TR.WK10	-0.07	0.02	-3.67	-0.02		
TR.WK12	-0.09	0.03	-3.57	-0.02		
MF.VALLEY	0.38	0.03	13.37	0.11		
WEEKENDS	0.47	0.05	9.85	0.10		
TR.PASSN	-0.29	0.01	-24.25	-0.14		
MTT20.40	-0.25	0.25	-12.27	-0.07		
MTT41	-0.42	0.02	-17.63	-0.11		
TIME18.20	-0.08	0.03	-3.02	-0.02		
WORKS	-0.13	0.03	-4.71	-0.04		
FEMALE	-0.21	0.02	-11.71	-0.06		
AGE25.31	0.11	0.03	4.03	0.03		
AGE32.39	0.24	0.03	8.46	0.06		
AGE40.48	0.29	0.03	10.42	0.07		
AGE49.60	0.47	0.03	16.50	0.11		
ED.LV4	-0.20	0.03	-6.56	-0.05		
ED.LV5	-0.16	0.03	-4.92	-0.04		
ED.LV67	-0.27	0.04	-7.55	-0.07		
SES.C3	-0.12	0.03	-4.67	-0.03		
SES.C2	-0.15	0.03	-4.80	-0.04		
CROWD.mean	-0.09	0.02	-5.20	-0.04		
CROWD.cv	-0.22	0.07	-3.31	-0.03		
CSPEED.mean	0.01	0.01	2.26	0.03		
CSPEED.cv	3.00	0.55	5.45	0.05		
FR1HR.mean	0.01	0.01	2.41	0.03		
FR1HR.cv	-0.60	0.13	-4.60	-0.04		
CI.NperHR	-0.14	0.02	-7.98	-0.05		
CFI	**TLI**	**GFI**	**AGFI**	**RMSEA**	**SRMR**	**WRMR**
0.981	0.980	0.981	0.978	0.036	0.027	6.933

For the commercial speed variable, we obtain a natural result for the mean, a significant and positive value, meaning that users are more satisfied with faster services over the reported travel times, which have negative signs. However, the CV has a positive sign; this is a counterintuitive result since we would expect users to be more satisfied with stable speeds. We checked this variable and found that most services with a high CV in commercial speed operate during weekends. We expect different surges in demand during weekends, as more variability is expected. We hypothesise this to be a weekend effect, as speeds during the week are significantly more stable.

As we had access to two different frequency datasets (i.e. 15 min and one hour), we entered the two frequency variables in the models and kept the most significant ones. The result is that the most significant one was the 1-hour frequency, both mean and CV. We hypothesise that this result stems from the fact that the one hour provides more variability than a 15-min interval. We notice this fact from the CV values for both types of frequencies, see Table 7-7. The variable FR.1HR.mean (i.e. mean-frequency) has a positive sign, an intuitive result, meaning that with more frequent service users are more satisfied than with infrequent ones. The CV of the frequency is negative, again as expected, indicating that with more irregular intervals during the time periods users are less satisfied. Finally, for the critical incidents' statistics, the duration of the CI was not significant. However, the number of CI per hour per period was significant and negative. This result means that indirectly the CI affect users' satisfaction, over the number of CI that users report (CI.TOTN), which is the subjective attribute. Notice that the coefficient for CI.TOTN remained almost the same.

The model, once more, complies with the recommended cut-off values (Hu and Bentler, 1999). In fact, we obtained an excellent fit to the data as the CFI, TLI, GFI and AGFI are all close to 0.98. Also, the RMSEA, is now less than 0.04, a value smaller than both for the original SEM model and the SEM-MIMIC model with travel and SES traits only. The WRMR also improves as compared to the previous models. Therefore, adding the objective indicators improved the model fit, considerably. Notwithstanding, the explained variance did not improve, but as stated before it is motivating that some of the seasonal, year and period variables changed significance, and that the Line dummies decreased their coefficients. This result indicates that by adding the operational variables, more variability was captured; the model improved as we added the objective attributes.

7.6. Discussion of Results

In this section, we follow the same order as with the model results above. First, from the PCA, it is interesting to see that the safety satisfaction construct has other components that would not be typically associated with safety: *travel time is easy to predict*, and *Metro cares about the passengers' travel conditions*. Users assess the Metro's responsiveness to travel conditions, and factor in the predictability of travel time when constructing their safety assessment. This result is beneficial for the Metro Administration when considering policy changes, since it means that in order to improve the safety perception, the reliability and the attentiveness to users' also need to be taken care of. They are related; hence they should be treated simultaneously. The rest of the satisfaction constructs provide natural and intuitive results in their composition.

From the SEM model with the CI.TOTN variable (i.e. number of types of critical incidents that the passenger endured in the last month) we were able to determine which satisfaction constructs had high relevance. The latent safety construct was the most important one and by far. The CI.TOTN variable was highly significant and negative, indicating that it is also a policy variable that should be addressed. The PTA should determine the most recurrent critical incidents, and set up a taskforce to assess why they happen and how to avoid them. Next, implement a working plan to minimise them. Not surprisingly, by minimising CI, the PTA would be able to increase satisfaction.

It is also interesting that the variable, *response to critical incidents* (CRIT) is significant and positive, as expected. This suggests that the PTA also would need to improve in this respect, to increase satisfaction; as shown in Table 7-3, these variables have unusually low scores.

It is important to mention that for Overall Satisfaction (OVS) item composition, the final satisfaction loading (0.81) is higher than the initial satisfaction one (0.57). This suggests that users modify their assessment slightly after answering the questionnaire. This is consistent with the findings of dell'Olio et al. (2010), however contrariwise, in our case the final satisfaction decreases. We hypothesise this is an effect of the critical incident items being ordered right before the final satisfaction item. Nevertheless, the coefficient for initial satisfaction loading is still reasonably high (>0.5), as such, it is appropriate to measure the OVS taking both items.

The SEM-MIMIC model with travel and socioeconomic users' traits allowed us to assess what types of conditions affect their transit satisfaction. Key policy variables that resulted in significant and negative signs were the perception variables: *number of trains that pass before boarding* and *Metro travel time*. Both variables represent some element of travel time, and it is intuitive that they are negative. Also, it is interesting to notice that the TR.PASSN (i.e. trains that pass before boarding) coefficient is larger than the coefficient of MTT41 (i.e. travel time of more than 40 min), indicating that users penalise satisfaction more when they have to wait longer than when their travel time is increased from 0-20 min to more than 40 min. This result is consistent with the findings of Allen et al. (2018a) for the Santiago bus system. In their case, the perceived waiting time variable was more significant than the perceived travel time variable.

All the age variables came out positive and increasing, meaning that as users age they like Metro more. One possible explanation for this result is that as users age they value the service reliability of Metro more, in comparison, say, with the bus system. We may also hypothesise that as users age their value of time increases and, indirectly, they value the reliability of travel time brought by the Metro system more. However, we commented above that the higher the education level and SES, the less satisfied users were. This result indicates that users increase their expectation of the system with more educational studies and higher SES levels. Interestingly, employment status did not improve the model, as it was not significant.

Finally, from the SEM-MIMIC model with objective and subjective attributes, we were able to demonstrate that by inserting these variables into the model, the goodness-of-fit increased. In particular, the most critical index, RMSEA, improved significantly. This result provides confidence in stating that by including objective attributes a more refined model can be obtained, than when using just reported (i.e. subjective) values. We had hypothesised that the commercial speed, frequency and number of actual critical incidents, and their variability, would have a repercussion in users' transit satisfaction; we obtained results to demonstrate this. In our final SEM-MIMIC model, most variables provided expected results and had consistent signs.

When adding the operational variables, the Line dummy coefficients decreased in magnitude, implying that the objective variables were able to capture more of the variability of the transit satisfaction. The only counterintuitive result was that the variability in commercial speeds had a positive effect. We hypothesise that this result is related to having more variability in commercial speed during weekends; this would need to be assessed by the Metro operator. For key policy variables, such as frequency and crowding levels, the CV is significant and negative, meaning that users perceive negatively the unreliability of level of service. Again, we consider this to be a critical result for policy design, as improving the stability of the operative level of services (frequency an crowding levels) would improve satisfaction.

7.7. Conclusions

In this section, we discuss some limitations from our study, policy recommendations, and future research questions that evolve from our findings. A specific limitation is that periods were grouped by season and Metro line, when one could argue that a more refined procedure could have been employed. Although this is feasible, the work involved in segmenting all variables by all stations is more complicated, since some of the statistics are presented for the complete lines. In essence, we believe that in a future study a more refined disaggregation process could be undertaken; one possibility would be to assess a specific subsample (i.e. L1 and L2, for six months).

We believe that we had some incorrect data for employment status in 2014. However, we overcame this problem by using only the education level variables and SES variables in our SEM-MIMIC models. In fact, we first tried the employment status variables, but since they were not significant we removed them. Nevertheless, we assert that best would had been to model the three sets of variables, when correctly measured. Both education level and SES variables were significant and with the expected signs. It is interesting to point out that the high-class SES group (ABC1) did not penalise the transit satisfaction. This result is consistent with the study by Allen et al. (2018a) for the Santiago bus system, where the same SES was also measured.

In both SEM-MIMIC models, we obtained that Metro lines L4, L4A and L5 produced more satisfaction than L1 and L2. We know that the latter are the lines with higher demand (and the oldest in the system), so we believe that a variable indicating the number of validations in a specific time period could be excellent at capturing this effect. In a sense, we would capture the effect of the number of users that demand the line at that time, at every specific station, this could be performed in a future study. On the other hand, one could argue that if the operational variables capture part of the variability, the Line dummies should lose significance.

As it happens, we were able to reduce most of their coefficients when adding the operational variables, except for that of L5 coefficient, which increased slightly.
To the PTA, we recommend focusing their efforts on the four most important variables: the safety construct, the alighting mobility construct, the CI perceived by the user, and the number of trains that pass before boarding. For the safety construct, we believe that the PTA needs to address the issue of safety when waiting at the platform. In Santiago, during peak-periods, there is a certain level of disorganisation when users wait for trains, and it can become even dangerous if people push each other; we deem that a passenger management system should be enforced. A similar recommendation is made by Suazo-Vecino et al. (2017), specifically from a Santiago Metro study about the dwell time at platforms. A maximum number of users allowed on the platform should be enforced. Note that the alighting mobility is also related with this issue.
The CI numbers need to be addressed; for this, we recommend contracting a study to establish which CI are most recurrent, and possible solutions to avoid them. If the PTA sets the objective of systematically minimising the number of CI per period, and implements a plan to accomplish this, it would substantially increase satisfaction. As currently all actions associated with *critical incidents response* are perceived as underperforming, the PTA could also offer users more accurate information about when CI occur. Finally, the *number of trains that pass* variable indicates that during certain periods, Metro demand surpasses the supply. We reiterate that the PTA should study mechanisms to prevent users reaching the platforms before they can board the train. The problem is the waiting time effect and the disorder on the platforms, which is highly and negatively valued by the users. Economic incentives could be offered to users that travel during off-peak periods. As lower off-peak fares already exist, the PTA could consider more significant discounts for off-peak users or, instead, charging even higher fares for people riding during peak hours.

As stated before, we are not certain if all the operative variables influencing transit satisfaction were considered. Thus, a clear research line is to address them. Specifically, the number of validations at each station during specific time periods is a candidate variable; another one is the number of people that transfer at transfer stations, but this is more difficult to measure. Another variable to consider is the temperature inside the trains and at the stations. As there is no air-conditioning in most lines and at all stations, the temperature attribute could prove important during the warmer months. Cold temperatures could also affect. Subsequently, a comprehensive temperature study should be commissioned: to answer whether the variability in temperature affects transit satisfaction. Furthermore, as the age of the trains was not considered, if it were possible to disaggregate this variable, an objective variable for *train condition* could be obtained. A final possible variable to consider is associated with the data collected by the PTA regarding claims. This information could be aggregated by Metro line, season, year and possibly time of day, and included afterwards when merging the operational variables.

We believe our framework could be employed in other cities and PT systems, independently of the type of operational attributes available. We showed that with our methodology we were able to improve the transit satisfaction models, over just including users' subjective perceptions. Combining both objective and subjective attributes allowed for the generation of a more powerful model. Our framework could be applied to bus and regular train systems as well.

8. EVASION BEHAVIOUR IN PUBLIC TRANSPORT: A SATISFACTION FRAMEWORK

8.1. Introduction

Governments and public transport administrators (PTA) strive to provide services that satisfy customers. Since public transport (PT) shares physical space with private transport modes, such as cars and motorbikes, governments and PTA should attempt to persuade their users to transfer to PT, to decrease externalities. Customer satisfaction surveys measure users' satisfaction levels. Satisfied users may reuse the services and recommend them to their friends and families. Satisfaction is defined as the real performance of the service minus the expected one (Oliver, 2010, 1980). Existing and extensive literature covers the topic of identifying which specific attributes users find more relevant when determining their overall satisfaction (Allen et al., 2018a, de Oña and de Oña, 2015, Eboli and Mazzulla, 2007). In such surveys, typically, the users' travel and socioeconomic characteristics are inferred or asked. Other specific items about their travel experience are asked: such as their perceived travel times, number of transfers, critical incidents (CI) they have experienced (Friman et al., 2001, 1998, Allen et al., 2018b), or even their mood (Gao et al., 2017).

Moreover, fare evasion in PT is a major problem which hampers the PTA's resources directly, especially in developing countries. Fare evasion is estimated to cost close to one billion euros per year to PTA worldwide (Bonfanti and Wagenknecht, 2010). Fare evasion refers to non-payment and underpayment of fares, and misuse of discount tickets/cards. We hypothesise that a direct link exists between users' transit satisfaction and their fare evading behaviour. We believe that the perception of *other users*' fare evading behaviour may directly impact transit satisfaction. In other words, a potentially vicious snowball effect may occur if the fare evading behaviour is allowed to increase in any major PT system. Thus, identifying the possible causes of this behaviour is of relevance, since allowing high rates of fare evasion implies smaller revenues and consequently, fewer funds for improving the quality of the system.

PTA should be interested in understanding how the user's overall and attribute-specific satisfaction influence their fare evading behaviour. Doing so, an indirect link between satisfaction and evasion could be established. In this sense it would be interesting to determine: (i) *if* satisfaction changes according to perceived *other users*' evading behaviour (i.e. their attitudes towards this behaviour); (ii) *how* it changed (we expect a deterioration); (iii) *if* overall or attribute-specific satisfaction influences the user's fare evading behaviour, and (iv) *how* - we expect a decrease in fare evasion with increasing satisfaction. These results would contribute with direct policy-related knowledge for PTA that could aid in decision-making and add insights and recommendations for other cities and PT systems.

We have found no studies that unveil a possible relationship between the user's perception of *other users*' fare evading behaviour and their PT satisfaction in the literature. Thus, we consider this a critical gap. On the other hand, a literature review about fare evading behaviour by Delbosc and Currie (2018), suggests that users' motivations and their attitudes, social norms and conditions motivating them to fare evade, are the key to understanding and managing this behaviour. In three studies, these same authors (Currie and Delbosc, 2017, Delbosc and Currie, 2016a, 2016b) attempted to segment users according to their fare evading behaviour and formulated a psychological fare evasion model for a Melbourne case study. We also found a few other studies linking the user's perceptions of the system and their fare evading behaviour. Two case studies from Italy are presented by Barabino et al. (2015) and Bucciol et al. (2013). Both use a single item for satisfaction. Empirically the former evidenced that unsatisfied users were more prone to evade, while the latter quite the opposite: dissatisfied customer were less prone to fare evade. One study in Santiago de Chile, by Buneder and Galilea (2017), reported that higher aggregate satisfaction ratings produced lower evasion rates. Further research is warranted to gain evidence on the relationship between satisfaction and fare evading intent.

In this study, we propose and develop a framework for a transit *satisfaction-evasion* behavioural model. The proposed models account for satisfaction with *other users'* fare evading behaviour, overall satisfaction, reuse intention and current fare evading behaviour. All are modelled through a single structural equation model (SEM). We compare two modes, Metro and bus, in Santiago. As the survey items are the same and the evasion items could be under-responded, we designed the data acquisition process in four steps. First, we conducted two focus groups with current Metro and bus users in order to determine all the relevant attributes regarding their needs in relation with their current trips. As the bus system in Santiago has a high rate of fare evasion (~28%) (Tirachini and Quiroz, 2016, Buneder and Galilea, 2017), in this mode we included specific evasion cues, aiming to trigger possible responses about their perception towards this behaviour as well as their potential *fare-evading* behaviour. We obtained positive responses, even admitting evasion behaviour. Second, we designed a customer satisfaction questionnaire including the principal items discussed in the focus groups and devised two items of satisfaction with *other users*' fare evading behaviour. At the end of the questionnaire, we included two items related with their current validating behaviour (i.e. the opposite of evasion).

Third, with the draft questionnaire, we conducted a pilot survey with 197 respondents intercepted at Metro stations and bus stops. The specific evading behaviour items were well-responded by users. Some modifications to the final survey were made for a better comprehension all-around according to users' responses and the interviewers' judgment. Finally, we implemented the final survey instrument and surveyed 2,002 users of the PT system. With the results, we estimated SEM-MIMIC models for both the Metro and bus systems. We disentangled the effects of the satisfaction items and the heterogeneous travel and socioeconomic traits of the users on transit satisfaction, reuse behaviour, and fare evading behaviour. The heterogeneity analysis allowed us to obtain a profile of the *fare-evader* for both modes, with significant differences in their rates of evasion, in the same city. Our results confirm that the satisfaction of (i.e. attitude towards) *other users*' fare evading behaviour negatively affects satisfaction. Specific satisfaction constructs affect users' fare evading behaviour, including the satisfaction with *other users*' fare evading behaviour and *reliability*. In a nutshell, the more satisfied a user is with other users' *evading* behaviour and the more dissatisfied with *reliability*, the more he/she evades.

We introduced three additional concepts: satisfaction with travel alternatives, *captivity*, which is the perception of not having alternatives, and *mood*. These three concepts proved significant in the models. Additionally, focusing on the overall satisfaction regression, we tested several types of models including mixture SEM models (i.e. latent class) and non-linearities (SEMM-NL). With the resulting best models, two-latent class models for both Metro and bus, we confirmed that *reliability* was, by far, the most crucial satisfaction construct regarding overall satisfaction. Furthermore, the non-linearities proved to be non-significant, indicating that *reliability* is a performance attribute for both Metro and bus users, and has a linear relationship with overall transit satisfaction. All of the above results contribute to PT policy development.

The rest of the chapter is organised as follows. Section 8.2 presents a literature review relevant to our case study. We offer details on the fare evasion in *Transantiago* and on the few studies that researched PT fare-evading behaviour from the users' perspective. Section 8.3 presents the Santiago bus and Metro case study, and Section 8.4, the sample and the survey. Section 8.5 presents the model results, and Section 8.6 discusses our principal results and interpretations in detail. Lastly, in Section 8.7, we provide the most important conclusions and offer policy recommendations for PTA from our findings.

8.2. Literature Review

The fare evasion rate in the bus component of Transantiago, Santiago's integrated bus-Metro PT system, is one of the highest in the world at 28%. In a fare integrated system, identifying the revenue losses due to fare evasion requires estimating the proportion of this trip legs that belong to trips in which other legs are not evaded (i.e. the fare is paid). Private estimates bring the losses in revenues (i.e. leakage) close to US$ 415 million per year, in 2015 (La Tercera, 2015, cited by Buneder and Galilea, 2017), which represents a figure close to 50% of the subsidy given to the system. These figures indicate a lack of efficiency in the PT system, and plenty to gain from understanding *why* and *how* users evade the fare, as this revenue could be used to improve the system.

Torres-Montoya (2014) offers a first academic glance at the fare evasion problem in Santiago's bus system. In a synthesis of the issue, the author highlights that a key concern ingrained in the collective perception of Santiaguinos, is that they do not value the service enough to pay the fare. This fact can be traced back to the implementation of *Transantiago*, as the model used to design and evaluate the system did neither incorporate an explicit penalty for transfers nor a penalty for the discomfort associated with travelling in crowded conditions. Longer travel times and crowded journeys during the first year generated a lot of bitterness and disapproval for the new system. Although the system has improved beyond recognition from its initial state, this reputational damage remains to date.

One of the first steps towards tackling fare evasion was accomplished by the design of new contracts that were implemented in June 2013, as these transferred a bulk of the demand risk to the private operators (Torres-Montoya, 2014). However, evasion remains a critical structural problem to date. Some of the recommendations given by Torres-Montoya (2014) are: *Transantiago* needs economic incentives for users, quality of service improvements, and better communications than what is currently offered. In order to deter fare evasion, the PTA needs to increase frequencies and improve bus quality, as these are persistent issues brought up by users. Implementation of loyalty programs could also bring evasion down. Plus, students should be targeted with deterrence campaigns as they are chronic evaders. On the other hand, a qualitative study commissioned by the Ministry of Transport (Factor Estratégico, 2010) identified four types of evaders:

i. Involuntary: evade because the smart card cannot be recharged.
ii. Uninhibited chronic: refuse to pay arguing that the system provides a poor service.
iii. Circumstantial: takes advantage of situations to evade but does not create them.
iv. Disguised chronic: saves money by inappropriate use of student or old age pensioner's passes or by abusing of transfers.

Both these studies represent the first efforts to understand the fare-evader in Santiago.

Tirachini and Quiroz (2016) analysed the international literature and compared Santiago with other cities, based on fare evasion rates, fines for fare-evading, actual fares, and inspection rates. They determined that Santiago had an unusually high evasion rate (28%), and also a low inspection rate (0.11%) compared to other cities. They give recommendations to tackle fare evasion in Santiago: (i) increase and speed up anti-evasion enforcement, (ii) develop a robust public awareness campaign against fare evading, and (iii) implement preferential fares for low-income users. Both Tirachini and Quiroz (2016) and Torres-Machado (2014), emphasise that the way to tackle fare evasion is by deploying an integrated package of measures; the former recommend effective strategies deployed in Melbourne, Australia (PTV, 2015).

Guarda et al. (2016a, 2016b) provide the first econometric research attempts to determine which variables are significantly related to fare-evasion in *Transantiago*'s bus component. Using a cross-sectional approach, with data from October 2012, they find that evasion rates in buses increase as: (i) more people board/alight at a given bus door, (ii) more passengers board by a rear door, (iii) buses have higher occupancy levels, (iv) buses have more doors, and (v) passengers experience longer headways (i.e. waiting times). Additionally, bus stops located in lower income municipalities produce higher evasion rates.

Two different studies utilised time series approaches to empirically determine relevant variables to help explain fare evasion, longitudinally. The first, by Troncoso and de Grange (2017), determined that: (i) a 10% increase in the fare raises evasion by two percentage points and (ii) a 10% increase in inspections lowers evasion by 0.8 percentage points. An increase in unemployment, the third explanatory variable in their model, tends to induce a decrease in evasion, a counterintuitive result. Buneder and Galilea (2017), determined that fare evasion was not driven up over time by increases in the bus fare or decreases in the public approval of the system, from an aggregate point-of-view. Their study found a spatial correlation between the index of social priority (IPS) and fare evasion; IPS is a municipality index that depends on the population's level of education, income, and health indicators. The higher the IPS (i.e. lesser levels of education, income, and health), the higher the fare evasion rates. This result aligns with the findings of Guarda et al. (2016b), as lower income municipalities produced higher rates of evasion. It is relevant to note that in both time series approaches, the overall perception (i.e. approval rate) of the system turned out to be non-significant; however, both studies only used overall-aggregate satisfaction.

These four previous studies used the evasion rates recorded by the authorities, which are calculated from field measurements using inspectors dressed as civilians. There are limitations to this methodology due to the probable human error associated with the measure of fare evasion (Buneder and Galilea, 2017). None of these studies procured an individual user self-report approach, where users would be surveyed regarding their regular PT use, perception of the PT system, and possible fare evading behaviour. We believe such an approach is essential to attain an understanding of users' perception of the system and its relationship with their fare evading behaviour, and it is missing to this date in Santiago.

The latest effort to explore the fare-evading behaviour in Santiago's bus system is a multi-disciplinary study of a single bus-line, BL-502 (LIP, 2018). The exploratory study used quantitative and qualitative methodologies to understand the phenomenon of fare evasion. In a first stage, a quantitative component was based on measurements of evasion on board buses in different days. Next, qualitative information was gathered using four different techniques: focus groups, interviews with principal actors, ethnographic observations on board of buses and at stops, and in-depth interviews with BL-502 service users. The study identified a new category of evader, namely an opportunistic evader, who sometimes pays and sometimes evades the fare. The opportunistic evader updates the previously identified evaders' categories: (i) opportunistic, (ii) circumstantial, (iii) disguised chronic, and (iv) uninhibited chronic.

The study (LIP, 2018) analysed the effect of different sociodemographic and operational variables on the evasion rates observed, by estimating logistic regression models. Bus occupancy rate, IPS, male gender, and being aged under 25 years are all associated with increased levels of evasion. Active paid zones and turnstiles inside the bus decrease the rates of evasion. All the coefficients' signs are intuitive and align with Guarda et al. (2016a) for the operative variables, and with Buneder and Galilea (2017) for IPS.

Next, we highlight the international PT literature on fare evading behaviour from the users' perspectives. A review by Delbosc and Currie (2018) defined three types of studies: (i) from the conventional transit system perspective, (ii) from the customer profile perspective, and (iii) from the customer motivation perspective. Studies in the first group refer to infrastructure and operational management; the four econometric approaches in Santiago fall into this category, and an interesting approach by Reddy et al. (2011) analysed and quantified the effectiveness of countermeasures and tactics applied to reduce fare evasion in the NYC Metro system. The second group of studies use customers' demographic characteristics to identify who is more likely to evade fares. This data is drawn from passenger surveys (onboard or at stations), population surveys (mostly online), and in-depth profiling (see Delbosc and Currie, 2018). Two exciting studies, both applied in Italy, fall into this category; we give further details on them, next.

Bucciol et al. (2013) randomly interviewed 541 passengers who used the bus in Reggio Emilia. In their sample, 43% of respondents travelled without a valid ticket. Specifically, young individuals (<25 years old), males and non-European immigrants were more likely to travel without a ticket. Interestingly, travelling with other people correlates with the probability of holding a valid ticket. It is plausible that this effect is related to *contagion* or *restitution*, a sociological aspect presented by Gino et al. (2009). Fare-evading behaviour may be most common when fellow *evaders* travel together, instead of travelling with others (i.e. non-evaders).

Barabino et al. (2015) analysed 2,177 onboard personal interviews collected at the Cagliari bus system. They performed logistic regression models to determine a *fare-evader* profile. Results indicate that males, younger than 26-year-olds, with a low education level, unemployed, or students, and without an alternative mode of transport besides the bus, were the most likely fare evaders. Moreover, people who make shorter than 15 min trips, who are systematic users and are not satisfied with the service are possible fare evaders. Finally, they found that a low level of inspection, knowledge of fines, and previous ticket violations are key elements to make people more prone to fare evade.

Finally, in the third group, we find an emerging area of research into the attitudes, motivations, and justifications for fare evasion. Although this perspective is relatively new in the PT fare evasion arena, it has been explored for several decades within criminology literature. Delbosc and Currie (2018) conclude that several studies have found that attitudes, social norms, customer satisfaction, and elements of the customer experience can all influence the intention of consumer misbehaviour (Alm, 2012, Reynolds and Harris, 2009, Tonglet, 2002).

In the PT literature, the common questions regarding consumers' perceptions towards fare evading behaviour have been: (i) what is fare evasion to customers? (ii) how extended is fare evasion? (iii) how common is fare evasion from users' reports? and (iv) why do people evade fares? (Currie and Delbosc, 2017, Factor Estratégico, 2010).

Quite a few studies state that increasing users' satisfaction is related to decreasing fare evasion (Bucciol et al., 2013, Barabino et al., 2015, Buneder and Galilea, 2017). Other studies, such as Tirachini and Quiroz (2016), openly state that the perceived service quality of the PT system is one of the causes of high evasion rates. Moreover, consumer misbehaviour literature states that customer satisfaction and elements of the customer experience can affect the rate of misbehaviour. However, we found no study that directly analysed the *transit satisfaction-evasion* relationship. We believe this to be a critical gap in PT literature. In this study we directly pursue the following research questions: (i) How does *other users*' fare evading behaviour affect transit satisfaction? (ii) How does overall transit and attribute-specific PT satisfaction relate to current fare evading behaviour?

8.3. Transantiago: Metro and Bus Case Study

Transantiago is the PT system serving the capital of Chile. It is considered the most ambitious transport reform undertaken by any developing country. It consisted of a major reform of the bus-component of the PT system and its complete fare integration with Metro. However, *Transantiago* was launched in a "*Big Bang*" fashion (Allen et al., 2018a, Muñoz et al., 2008), with a problematic initial implementation, as a decreased bus fleet and newer routes proved insufficient to serve the demand, which was also inadequately informed. Most user complaints related to the absence of buses and their irregular frequencies, absent or modest infrastructure (i.e. segregated corridors, prepaid areas, and bus stops), worsened network coverage, and that more transfers were needed for longer trips.

Torres-Machado (2014) gave a historical review recounting the many issues regarding this *sub-par* implementation. On the other hand, Metro had to accommodate an unexpected surge in demand, causing delays, problems when boarding, and more crowded trains. Today, Santiago hosts close to 7 million people and the Metro carries ~2.7 million passengers every day. This figure represents over 1.2 million passengers more than in 2007 when Transantiago was launched. Currently buses carry ~3.3 million passengers daily.

Although Metro is the backbone of *Transantiago*, a set of trunk bus-lines complements the Metro trunk network skeleton, and a set of feeder bus-lines handles shorter trips to feed Metro and the trunk bus-lines. The operational speed of Metro ranges between 25 and 40 km/h depending on the line, while the operational speed of the bus reaches 25 km/h in segregated corridors and between 8 and 15 km/h for the regular bus service. Some bus-lines operate 24 hours.

In 2010, a governmental body called Metropolitan Transit Directory (DTPM) started managing *Transantiago*'s bus operations. DTPM aims to continually improve the bus-component, both at service levels and regarding overall system quality. Since 2012, modifications to the contracts with the private bus operators were implemented to allow the creation of incentives for better performance of the scheme; this included penalties for non-compliance of various service parameters and rewards for controlling fare evasion, which had rocketed (Torres-Machado, 2014). In June 2012, DTPM discarded the notion that different companies should operate the feeder and trunk lines. Connections between lines serving different zones were encouraged to eliminate unnecessary transfers and thus provide better services. DTPM formed seven Operating Units (U1-U7), each assigned to a concessionaire that included trunk and feeder lines. This configuration remains to date.

Currently, *Transantiago* has less than 80 km of designated bus-lines out of the 300 km planned initially. The system has approximately 11,000 bus stops and 6,500 buses. The segregated corridors lack off-level payment stations. Thus, they cannot be considered BRT corridors. Studies by Batarce et al. (2016) and Allen et al. (2018a) focusing on the *Transantiago* bus component, concluded that reliability and waiting time are a significant concern for users, and are the most critical attributes regarding users' satisfaction. This is consistent with Torres-Machado (2014), who stated that the main users' concerns were the low frequencies and poor bus quality. The bus component of *Transantiago* is clearly the least favourite mode for most PT users in Santiago. We hypothesise that users' (dis)satisfaction plays an essential role in triggering fare-evading behaviour.

8.4. Focus Groups, Sample and Survey Results

8.4.1. Focus groups

The first part of our data collection effort consisted of two separate focus groups with current Metro and bus users. Their objectives were to identify relevant dimensions and attributes that affect *Transantiago* users' satisfaction with Metro and bus transport modes. A script was designed to encourage respondents to mention those aspects that were significant and relevant to them, including a set of flexible, concrete questions, phrased in familiar language for the respondents. The focus groups were designed and implemented in September 2017 (Alcaíno and Torres, 2017).

Firstly, general aspects were consulted, associated with learning about the context of the usual trips of the participants and opinions about their travel experiences. Secondly, we inquired about user satisfaction, probing into four attributes: reliability, security, comfort, and customer service. The same approach was taken with other variables that emerged during the focus groups sessions. Bus users were also consulted regarding their opinion on evasion and its incidence on satisfaction with the service, since this is the mode with the highest evasion rate. Finally, participants were asked about the perception of their travel alternatives and the advantages and disadvantages of the chosen modes, over the alternatives. Each focus group was implemented with the participation of nine users and their sociodemographic characteristics are shown in Table 8-1.

Table 8-1: Focus groups sociodemographic characteristics

Socioeconomic traits	**Metro focus group (9)**	**Bus focus group (9)**
Gender	Male (5), Female (4)	Male (3), Female (6)
Employment status	Professional (3), technician, administrative (3), service workers and sellers (2)	Professional (2), technician (2), administrative, service workers and sellers (4)
Age	25-30 years (2), 30-39 years (4), 40-49 years (2), 50-55 years	25-30 years (3), 30-39 years (3), 40-49 years, 50-55 years (2)
***Comuna* / Municipality**	Santiago (2), Providencia (2), La Florida, Maipú, Huechuraba, San Miguel, Puente Alto	Maipú (2), Santiago, Lo Espejo, Quilicura, Peñalolén, La Florida, Quinta Normal, Conchalí
SES	C2 mid-high (5), C3 mid-mid (4)	C2 mid-high (2), C3 mid-mid (4), D mid-low (3)

The selection criteria for integrating the groups were: (i) to be use public transport (bus or Metro) users, (ii) to travel to work at least three times a week, and (iii) to be frequent users of these modes of transport during peak and non-peak hours (Alcaíno and Torres, 2017). For the socioeconomic (SES) variables, we used the Chilean standard to categorise SES (*Grupo socioeconómico*, in Spanish), into five strata: ABC1, C2, C3, D and E. The *thematic analysis* technique was used to analyse the focus groups' data. It involves identifying, quantifying, and reporting recurrent themes within the data (Braun and Clarke, 2006). First, common criteria were distinguished, defining relevant variables. In this way, a primary level (very relevant), a secondary level (relevant), and a tertiary level (less relevant) of attributes were determined, according to (i) number of appearances, (ii) spontaneous emergence, (iii) and emphasis assigned by the interviewees.

For the Metro system, the *primary* level attributes included reliability, travel time, comfort, security (absence of crimes), frequency and regularity of the services, and the interactions with other passengers. The *secondary* level attributes included access and egress, price, safety (absence of accidents), and environmental conditions. Finally, the *tertiary* level attributes included customer service and information availability about the service. One of the comments about reliability was:

"*It is reliable in the sense that I can organise my time according to when it [the train] is going to arrive, with the regularity that it has.*" (Female, C2 mid-high).

For the bus system, the *primary* level attributes included travel time, frequency and regularity, price, comfort, and security (absence of crimes). The *secondary* level attributes included access and egress, security (absence of accidents), reliability, and the interactions with other passengers. Finally, the *tertiary* level attributes included customer service, information availability about the service, environmental conditions, and flexibility. For both modes, the order of priorities was very similar. One minor difference is that for the bus mode the reliability component was on the *secondary* level. However, frequency and regularity were on the *primary* level for both modes. One of the comments about travel time for the bus was:

"*I think [the most important thing is] to get there faster, you have to get to your work on time.*" (Female, D mid-low)

Finally, for the bus system, specific cues were designed for respondents to speak about evasion behaviour. Evading behaviour was discussed by asking participants, about their perception of it and whether they consider that it had any influence on their assessment of service quality. This aspect was previously alluded to by one of the participants, when talking about the price of the ticket and considering that the high fare made evasion understandable. All respondents indicated having witnessed evasion in the buses and that it happens frequently, practically every day. One of the comments about evasion on the bus system was:

"*I'm not going to pay this money, because with that I can buy a kilo of bread, I can prepare* ***"el once"*** *[a coffee/tea and a snack] with the money that I'm saving with the roundtrip fare. Then, one begins to prioritize other types of things, it stops being relevant to pay the fare.*" (Male, C2)

Two respondents (out of nine) acknowledged having evaded the fare sometimes, arguing, as detailed above, that the value of the ticket was high. On the one hand, they argued that the service provided was not worth it, considering the low quality received in exchange, and on the other hand, that the money spent in the ticket could be used for other things of greater urgency or importance. On the contrary, one participant agreed with the deficiencies in service quality but believed these could not be resolved if people did not pay. Hence, evasion would affect future improvement in quality. In this regard, he also pointed out that, at times, the evasion problem has affected the service, causing the driver not to stop at a certain bus stops if he/she perceives that the passengers there will evade the fare (Alcaíno and Torres, 2017).

8.4.2. The satisfaction-evasion questionnaire

Bearing in mind the focus group results, we designed a customer satisfaction questionnaire according to their results. We attempted to unify the questionnaire to gain comparability between modes. We explicitly designed two items of satisfaction with *other users*' fare evading behaviour. Also, at the end of the questionnaire, we included two items about their current validating behaviour (i.e. the contrary of evasion). With the draft questionnaire, we conducted a pilot survey, where we intercepted 197 users at Metro stations and bus stops. All evading behaviour items were well-responded and only minor adjustments were made to improve comprehension. We carried out the surveys only during the Off-Peak, and PM-Peak periods, as the AM-Peak had a lower rate of response. We intercepted and surveyed 2,002 users of the PT systems.

The survey was carried out by DICTUC S.A., a specialist firm; the sample design and survey results are reported in DICTUC (2018). In summary, the sample size design resulted in 326 responses per period for Metro and 354 for bus. Sampling was divided equally among the five lines (L1, L2, L4/L4A, L5, L6) for Metro, and among the seven bus operators (U1-U7). The distribution of the surveys by stations and bus stops was approximately weighted by the number of boardings for each Metro-line or operator (see DICTUC, 2018, for details). Users were surveyed once they passed the turnstiles and entered the boarding/alighting areas of the Metro stations, and at bus-stops when they arrived. Data was collected during weekdays in March 2018 for the bus, and in in April 2018 for the Metro. The final *questionnaire* was as follows:

A: Initial module. Included a filter question for the frequency of use in days/week (>1); two mood items, measured with a Likert (1-7) scale, and travel characteristics including travel, waiting and door-to-door times, and occupancy levels (explained with specific images, see Appendix G).

B: Satisfaction module. Included initial satisfaction and attribute-specific (seven domains) satisfaction items, measured with a Likert (1-7) scale. Two items were related to satisfaction with *other users*' evading behaviour. Other items included: final overall satisfaction (1-7 satisfaction scale), an intent to recommend item (1-7 agreement scale), and an intent to reuse the service (1-much less, 7-much more) item.

C: Socioeconomic module. Included ownership of a discount card, age, gender, number of cars in the household, having a driver's license, education, employment status, *Comuna*/Municipality of residence, and income.

D: Evasion behaviour module. Two items: (i) How often does your family/friends validate the fare? (ii) How often have you validated the ticket this week? Notice that both items represent *anti-evasion* behaviour.

8.4.3. Sample and survey results

Next, in Table 8-2, we present the sample travel characteristics for both modes. We will comment on the pooled sample and discuss details for the two different modes when needed. The total sample consists of 2,002 surveyed users, approximately equally distributed among the Metro Off-Peak (26%), Metro PM-Peak (23%), bus Off-Peak (27%), and bus PM-Peak (24%).

Table 8-2: Sample travel characteristics: pooled data, by mode and period (%)

Category	Unit	ALL	Metro.Off	Metro.PM	Bus.Off	Bus.PM
SAMPLE	n	2002	522	465	544	471
	%	100	26	23	27	24
FREQ.USE	1-2	16	21	14	19	9
(days/week)	3-4	17	19	16	18	13
(Metro-bus)	5	51	43	57	45	61
	6-7	16	17	13	18	17
No.VEH.PASS	0	70	77	62	76	62
(Metro-bus)	1	14	13	16	13	15
	>2	16	10	22	11	23
WAITING	0-2	20	23	17	---	---
TIME-MEAN	3	21	21	20	---	---
(Metro)	4-5	43	42	44	---	---
	>6	16	14	19	---	---
WAITING	0-5	14	---	---	15	14
TIME-MEAN	6-10	38	---	---	41	35
(Bus)	11-15	22	---	---	20	23
	>16	26	---	---	24	28
WAITING	0-5	37	40	34	---	---
TIME-MAX	6-9	21	20	22	---	---
(Metro)	10	20	20	21	---	---
	>11	22	20	23	---	---
WAITING	0-15	23	---	---	26	20
TIME-MAX	16-20	23	---	---	24	22
(Bus)	21-30	32	---	---	31	33
	>31	22	---	---	19	25
OCCUPANCY	1-2	32	52	18	35	18.3
VISUAL	3	22	24	22	24	15.9
(Metro-bus)	4	19	11	23	21	23
	5-6	27	13	37	20	43
TRAVEL	0-10	23	23	22	22	25
TIME-MEAN	11-20	33	32	28	38	32
(Metro-bus)	21-30	21	22	22	21	22
	>31	23	23	28	19	21
TRAVEL	0-20	12	17	11	11	8
TIME-MEAN	21-40	29	34	31	26	28
DOOR-2-DOOR	41-60	32	31	34	33	26
(Metro-bus)	>61	27	18	24	30	38

For the pooled sample, most users (51%) travel five days per week. The distributions for 5-day-week travel are a little higher for the PM-Peak (57% for Metro, 61% for bus) than the Off-Peak (43% for Metro, 45% for bus). Regarding the number of trains/buses that have to let pass before boarding, 14% perceive one vehicle passing, while 16% perceive two or more. The distributions are higher in the PM-Peak, for two or more trains/buses, 22% and 23%, respectively for Metro and bus.

In Metro, 44% perceive 0-3 min of mean waiting time and 43% perceive 4-5 min. For bus, 38% perceive a 6-10 min wait and 47% perceive 11 min or more. For maximum perceived waiting time, in Metro, the biggest stratum was for 0-5 min (37%); for bus, it was 21-30 min (32%). Thus, there are clear differences in waiting times between both modes.

For the occupancy levels, there are differences across periods. In Metro, the biggest group is the lower occupancy levels (L1-L2, 52%) for the Off-Peak, and it is the highest occupancy levels (L5-L6, 36%) for the PM-Peak. For bus, the biggest group is the lower occupancy levels (L1-L2, 35%) for the Off-Peak, and the highest occupancy levels (L5-L6, 43%) for the PM-Peak. Overall, 33% of users have a 11-20 min travel time and 44% endure 21 or more minutes. For travel time door-to-door, the biggest group is the 41-60 min door-to-door (31%). However, differences across modes and periods appear, for example, in the bus PM-Peak the biggest group is >61 min (37%).

In Table 8-3, we present the sample sociodemographic characteristics for the pooled sample. A discount card is available for students (School and University) and the elderly (>60 years-old). Overall, 33% of users have it. Slight differences appear across modes and time periods: of Metro Off-Peak users, 40% possess a discount card and 25% have it in the PM-Peak for bus users. Overall, 53% of users are female, 57% for the bus Off-Peak and 48% for the bus PM-Peak.

Table 8-3: Sample sociodemographic characteristics: pooled data, by mode and period (%)

Category	Unit	ALL	Metro.Off	Metro.PM	Bus.Off	Bus.PM
FARE	Discount	33	39	34	33	25
	Regular	67	61	66	67	75
GENDER	Male	47	45	50	43	52
	Female	53	55	50	57	48
AGE	18-24	36	41	34	37	31
	25-31	20	22	25	16	19
	32-39	14	13	15	13	15
	40-48	13	10	15	13	13
	>49	17	14	11	21	22
EDUCATION	Primary.HS	36	27	27	44	47
LEVEL	Technical	21	23	22	20	19
	Univ.Grad	43	50	51	36	34
EDUCATION	Yes	61	58	63	58	65
COMPLETE	No	39	42	37	42	35
EMPLOYMENT	Non-other	6	8	3	9	5
(WORK)	Student	31	37	29	33	24
STATUS	Employed	52	40	59	46	64
	Entrepreneur	11	15	9	12	7
LICENSE	Yes	26	31	32	19	23
Current	No	74	69	68	81	77
No.CARS	0	54	49	52	57	57
HOUSEHOLD	1	36	39	38	34	34
	>2	10	12	10	9	9
INCOME	I1-I2 (<850)	39	36	37	44	40
HOUSEHOLD	Middle I3	37	36	36	37	37
(US$)	I4-I5 (>1500)	24	28	27	19	23
INCOME	1st Quartile	7.3	7.3	7.3	7.3	7.3
TOTAL-MEAN	Median	8.1	9.8	8.3	8.1	8.1
(COMUNA)	Mean	10.0	10.6	10.3	9.3	9.7
(xUS$150)	3rd Quartile	10.1	11.4	10.9	10.1	10.1

Overall, the biggest age stratum is 18-24 years old (36%), next, 25-31 years old (22%), and then >49 years old (17%). The bus mode has a bigger >49 years old group (22%) than Metro (13%). For education levels, the biggest group overall is for University studies (43%). Overall, 61% completed their last degree. Regarding employment status, the biggest stratum is Employed (52%), the next in size is Students (31%), then Entrepreneur (11%), and last, Other (6%). For the Metro Off-Peak, 40% are Employed, while this rises to 64% for the bus PM-Peak.

For the Metro PM-Peak, 32% currently have a driver's license, while this drops to 19% for bus Off-Peak. Regarding number of cars in the household, 54% do not own a car, 36% own one, and 10% own two or more.

For the household income categories overall, the two lowest categories (I1-I2) make up 39% of the sample, while the two highest (I4-I5), 24%. The I2-I3 threshold is US$850, while the I3-I4 threshold US$1,500 monthly, approximately. Metro Off-Peak users have the highest incomes, and bus Off-Peak users the lowest. For mean total income per HH, in the *Municipality/Comuna of residence*, the mean total incomes are slightly higher for Metro (~US$1550) than for bus (~US$1425). An analysis of all socioeconomic characteristics allows to conclude that Metro users have a higher socioeconomic status (SES) than bus users, and the difference is higher in the Off-Peak period.

In Table 8-4 we present the average survey results for the Metro sample, the pooled data and per period. We comment on the results for the pooled data and group them into the specific satisfaction domains for the first nine categories. The last five categories include: mood, overall satisfaction, reuse behaviour, validate behaviour, and *captivity* items. All items are scored using a 1-7 Likert scale, except validating behaviour (1-5) and *captivity* items (0-1).

For reliability, the best-rated item is speed (B05, 5.70), and the worst, frequency (B03, 5.36). For comfort, the best-rated item is comfort on stairs and tunnels (B07, 5.39), and the worst, comfort inside the trains (B10, 5.02). Regarding environment, the best-rated is air pollution (B12, 4.53), and the worst, temperature inside the trains (B13, 3.87). For information, the best-rated is respect and cordiality from the personnel (B15, 5.45), and the worst, information about station status (B16, 5.29).

Table 8-4: Metro survey results. Mean values. Likert (1-7)

DOMAIN	Service quality attributes	Metro	Metro.Off	Metro.PM
Reliability	B02: Easiness to access stations	5.51	5.57	5.45
	B03: Frequency	5.36	5.33	5.38
	B04: Regularity	5.39	5.42	5.34
	B05: Speed	5.70	5.72	5.67
	B06: Reliability	5.42	5.45	5.39
Comfort	B07: Stairs and tunnels, in general	5.39	5.37	5.41
	B08: Stairs and tunnels, at transfers	5.26	5.25	5.28
	B09: Platforms	5.37	5.36	5.37
	B10: Inside trains	5.02	5.02	5.01
Environment	B11: Noise	4.04	4.09	3.99
	B12: Air pollution	4.53	4.54	4.52
	B13: Temperature	3.87	3.98	3.75
Information	B14: Info for planning the trip	5.36	5.38	5.34
	B15: Respect and cordiality, personnel	5.45	5.46	5.45
	B16: Service status at the station	5.29	5.34	5.24
Ticketing	B17: Location of recharge points	5.27	5.31	5.21
	B18: Ease and speed at recharge points	5.13	5.17	5.08
Safety	B19: Security-theft on the way Metro	4.46	4.36	4.56
	B20: Security-theft inside station/train	5.02	4.96	5.09
	B21: Safety bumps/falls stations	4.93	4.91	4.95
	B22: Safety bumps/falls trains	4.84	4.85	4.83
User	B23: Respect cordiality, in general	4.03	4.05	4.00
Behaviour	B24: Respect cordiality, to women	4.03	4.04	4.02
	B25: To reduced mobility users	4.11	4.16	4.07
User	B26: Fare evasion at turnstiles	2.10	2.13	2.07
Evasion	B27: Fare evasion through exits	2.02	2.06	1.98
Transport	B28: Satisfaction with buses	3.62	3.66	3.58
Alternatives	B29: Satisfaction with taxi/auto	4.32	4.28	4.36
	B30: Satisfaction with shared taxi	4.48	4.46	4.51
Mood	MD1: Being in an excellent mood	5.78	5.78	5.78
	MD2: Feeling happy	5.92	5.91	5.93
Overall	OS1: Initial overall satisfaction	4.93	4.96	4.91
Satisfaction	OS2: Final overall satisfaction	5.18	5.18	5.17
	OS3: Meets personal needs	5.37	5.38	5.37
	REC: Recommend to friends and family	5.41	5.42	5.41
Reuse	LOY: Reuse more or less, next six months	4.70	4.71	4.70
Validate	VB1: Friends and family validate (1-5)	4.79	4.77	4.81
Behaviour	VB2: You validated, this past week (1-5)	4.95	4.94	4.95
Captivity	N28: Buses not available (0-1)	0.10	0.10	0.10
	N29: Taxi or auto not available (0-1)	0.16	0.17	0.14
	N30: Shared taxi not available (0-1)	0.20	0.20	0.20

For ticketing, location of recharge points (B17, 5.27) is better rated than ease and speed of using recharge points (B18, 5.13). In the safety domain, the best-rated item is security against theft inside the stations and trains (B20, 5.02), and the worst, security against theft on the way to the stations (B19, 4.46).

Regarding users' behaviour, respect and cordiality towards passengers with reduced mobility (B25, 4.11) is slightly better-rated than the other two items (4.03). For user evasion, both items are very low-rated, 2.10 and 2.02, indicating that users are highly dissatisfied with this behaviour. From the transport alternatives, the best-rated is shared taxi (B30, 4.48), and the worst, the buses (B28, 3.62).

We comment next on the last five categories. For mood, feeling happy (MD2, 5.92) was better-rated than being in an excellent mood (MD1, 5.78). For overall satisfaction, the best-rated item was recommending the Metro to friends and family (REC, 5.41), and the worst, initial overall satisfaction (OS1, 4.93). Reuse behaviour (LOY) was rated at a mean value of 4.70. The user's own validating behaviour was better-rated (VL2, 4.95, 96% scored a 5) than the friends and family's (VL1, 4.79, 84% scored a 5). The *captivity* items ranged between 0.10-0.20.

In Table 8-5 we present the average survey results for the bus sample, pooling the data and per period (i.e. off-peak, PM-peak). Again, we comment on the results from the pooled data. For the reliability domain, the best-rated item is the ease of access to bus stops (B02, 4.93), and the worst, frequency (B03, 3.93), which is below par (<4). For comfort, the best-rated item is comfort in bus stops with seats and roof (B08, 4.82), and the worst, comfort with bus stops without infrastructure (B07, 2.87). In the environment domain, the best-rated is air pollution (B11, 3.59), and the worst, temperature inside the buses (B12, 3.30), all below par. For information, the best-rated is information and signage at the bus stops (B15, 5.13), and the worst, respect and cordiality of the personnel (B14, 4.77).

.

Table 8-5: Bus survey results. Mean values. Likert (1-7)

DOMAIN	Service quality attributes	Bus	Bus.Off	Bus.PM
Reliability	B02: Easiness to access bus stops	4.93	4.95	4.90
	B03: Frequency	3.93	4.01	3.84
	B04: Regularity	4.07	4.12	4.00
	B05: Speed	4.59	4.63	4.55
	B06: Reliability	4.25	4.29	4.21
Comfort	B07: Stops without infrastructure	2.87	2.86	2.89
	B08: Stops with seats and roof	4.82	4.90	4.74
	B09: Inside the buses	4.07	4.18	3.93
Environment	B10: Noise	3.57	3.64	3.49
	B11: Air pollution	3.59	3.56	3.63
	B12: Temperature	3.30	3.34	3.25
Information	B13: Info for planning the trip	4.84	4.75	4.93
	B14: Respect and cordiality, personnel	4.77	4.72	4.83
	B15: Service status at the station	5.13	5.10	5.16
Ticketing	B16: Location of recharge points	4.28	4.35	4.20
	B17: Ease and speed at recharge points	4.60	4.65	4.54
Safety	B18: Security-theft on the way stops	3.61	3.63	3.58
	B19: Security-theft stops/inside buses	3.67	3.69	3.65
	B20: Safety bumps/falls stops/boarding	3.92	3.87	3.99
	B21: Safety bumps/falls trains	3.75	3.77	3.73
User	B22: Respect cordiality users, in general	4.03	4.08	3.97
Behaviour	B23: Respect cordiality users, to women	4.25	4.34	4.16
	B24: To reduced mobility users	4.27	4.37	4.16
User	B25: Through front door and fare evade	2.11	2.21	2.00
Evasion	B26: Through back door to fare evade	1.97	2.06	1.87
Transport	B27: Satisfaction with Metro	4.80	4.98	4.60
Alternatives	B28: Satisfaction with taxi/auto	4.77	4.84	4.69
	B20: Satisfaction with shared taxi	4.82	4.91	4.71
Mood	MD1: Being in an excellent mood	5.60	5.63	5.56
	MD2: Feeling happy	5.71	5.73	5.67
Overall	OS1: Initial overall satisfaction	4.08	4.21	3.93
Satisfaction	OS2: Final overall satisfaction	4.22	4.31	4.12
	OS3: Meets personal needs	4.42	4.51	4.31
	REC: Recommend to friends and family	4.13	4.28	3.96
Reuse	LOY: Reuse more or less, nextsix6 months	4.34	4.40	4.27
Validate	VB1: Friends and family validate (1-5)	4.40	4.41	4.40
Behaviour	VB2: You validated, this past week (1-5)	4.71	4.69	4.73
Captivity	N28: Buses not available (0-1)	0.19	0.18	0.20
	N29: Taxi or auto not available (0-1)	0.22	0.23	0.22
	N30: Shared taxi not available (0-1)	0.25	0.24	0.25

In ticketing, ease and speed of using recharge points (B17, 4.60) is better rated than location of recharge points (B16, 4.28). For safety, the best-rated item is safety against bumps and falls at bus stops and when boarding the bus (B20, 3.92), and the worst, security against theft on the way to bus stops (B18, 3.61). Regarding user behaviour, respect and cordiality towards passengers with reduced mobility (B24, 4.27) is better-rated than the rest. For the user evasion domain, both items are very low-rated, 2.11 and 1.97. For the alternatives, the best-rated is shared taxi (B29, 4.82), and the worst, taxi/auto (B28, 4.77); Metro lies in between (B27, 4.80). Finally, we comment on the last five categories. From the mood domain, feeling happy (MD2, 5.71) was better-rated than being in an excellent mood (MD1, 5.60). From the overall satisfaction domain, the best-rated item was meeting your personal needs (OS3, 4.42), and the worst, initial overall satisfaction (OS1, 4.08). Reuse behaviour (LOY) was rated at a mean value of 4.34. For the validating behaviour, the user's own validating behaviour was better-rated (VL2, 4.71, 81% scored a 5) than the friends and family's (VL1, 4.40, 58% scored a 5). The *captivity* items ranged between 0.19-0.25.

8.5. Model Results

We now present the model results for Metro and bus. Included are (i) PCA, (ii) SEM ordinal Probit models, (iii) SEM-MIMIC ordinal Probit models with travel and socioeconomic characteristics, and finally (iv) SEMM-NL analysis and models. We used the *R* (R Core Team, 2013) and its associated library packages. In particular, we used the *Lavaan* package for *R* (Rosseel, 2012) for SEM and SEM-MIMIC models, and the *nlsem* package for *R* (Umbach et al., 2017) for the SEMM-NL models.

8.5.1. Principal component analysis

Firstly, we want to ascertain if users perceive the different satisfaction, overall satisfaction, and behavioural constructs as we designed them in the questionnaire. We wish to determine if the items represent the specific latent constructs as we expect. Thus, we performed four separate PCA on the items. In each mode, a separate PCA was performed on the satisfaction items and on the behavioural items. We used the Kaiser (1960) rule for all the PCA. Additionally, we assessed for Cronbach's (1951) alpha. The behavioural items behaved as expected and nine separate components were formed, not shown due to space availability.

Regarding the PCA for overall satisfaction and behavioural constructs, we present the results and Cronbach's alpha in Tables 8-6 and 8-7. The resulting final five components are mood (MOOD), overall satisfaction (OVSAT), reuse (REUSE), validate behaviour (VALID), and *captivity* (CAPTI). All items have an adequate validity (>0.5) measured via Cronbach's alpha, and have loadings higher that |0.4| for all items and components (see Tables 8-6 and 8-7).

Table 8-6: PCA overall satisfaction and behavioural constructs: Metro

	MOOD	OVSAT	REUSE	VALID	CAPTI
MD1: Being in an excellent mood	**0.70**	-0.01	0.03	-0.00	
MD2: Feeling happy	**0.70**	0.01	-0.01	0.00	
OS1: Initial overall satisfaction	0.01	**-0.47**	-0.04	0.02	
OS2: Final overall satisfaction	-0.00	**-0.52**	-0.02	0.03	
OS3: Meets personal needs	0.01	**-0.52**	-0.03	-0.04	
REC: Recommend to friends and family	-0.01	**-0.49**	0.09	-0.01	
LOY: Reuse more or less, next six months	0.01	0.00	**0.99**	0.00	
VB1: Friends and family validate	0.08	0.00	-0.08	**0.69**	
VB2: You validated, this past week	-0.07	-0.00	0.07	**0.72**	
Cronbach's alpha (α)	**0.90**	**0.89**	**------**	**0.52**	**0.86**

Table 8-7: PCA overall satisfaction and behavioural constructs: bus

	MOOD	OVSAT	REUSE	VALID	CAPTI
MD1: Being in an excellent mood	**-0.70**	-0.02	-0.01	-0.00	
MD2: Feeling happy	**-0.71**	0.02	0.01	0.00	
OS1: Initial overall satisfaction	0.00	**-0.48**	-0.01	0.00	
OS2: Final overall satisfaction	-0.00	**-0.53**	-0.05	-0.02	
OS3: Meets personal needs	-0.00	**-0.49**	0.04	0.04	
REC: Recommend to friends and family	-0.00	**-0.50**	0.04	-0.02	
LOY: Reuse more or less, next six months	0.00	-0.00	**1.00**	-0.00	
VB1: Friends and family validate	0.01	-0.01	-0.01	**0.70**	
VB2: You validated, this past week	-0.01	0.01	0.01	**0.71**	
Cronbach's alpha (α)	**0.89**	**0.88**	**------**	**0.63**	**0.84**

The overall satisfaction has high validity, assuring the measurement is correct. It is interesting to highlight that the recommend to friends and family correlated highly with the overall satisfaction items, as such we kept it in the OVSAT construct.

8.5.2. Structural equation models: SEM ordinal Probit models

From the results of the four PCA, we built our SEM ordinal Probit models, one for Metro and one for bus. The SEM ordinal Probit model treats all items with an ordinal range (for details, see Allen et al., 2018b), as all the items have a Likert-type range. First the measurement models were proposed and tested for the nine satisfaction constructs, overall satisfaction, and the attitudinal and behavioural constructs. The Measurement system for Metro is presented in Appendix H (Figure 8-3). For the bus system there is one minor modification as one less item was used for comfort. The results of the SEM Measurement models are presented in Appendix I (Tables 8-14 and 8-15). We present estimates, standard errors (S.E.), Z-values, and standardised coefficients (Std.Coeff.). The standardised coefficients refer to how many standard deviations in a dependent variable will change per standard deviation increase in the predictor variable. Most of the standardised coefficients present high validity (Std.Coeff. >0.6). However, for both models there are four items with low validity (Std.Coeff. <0.5); as they are only a few, we maintain our working models. Notice that these measurement models are part of the SEM to be presented next.

Next, in Figure 8-1, we present the full SEM system.

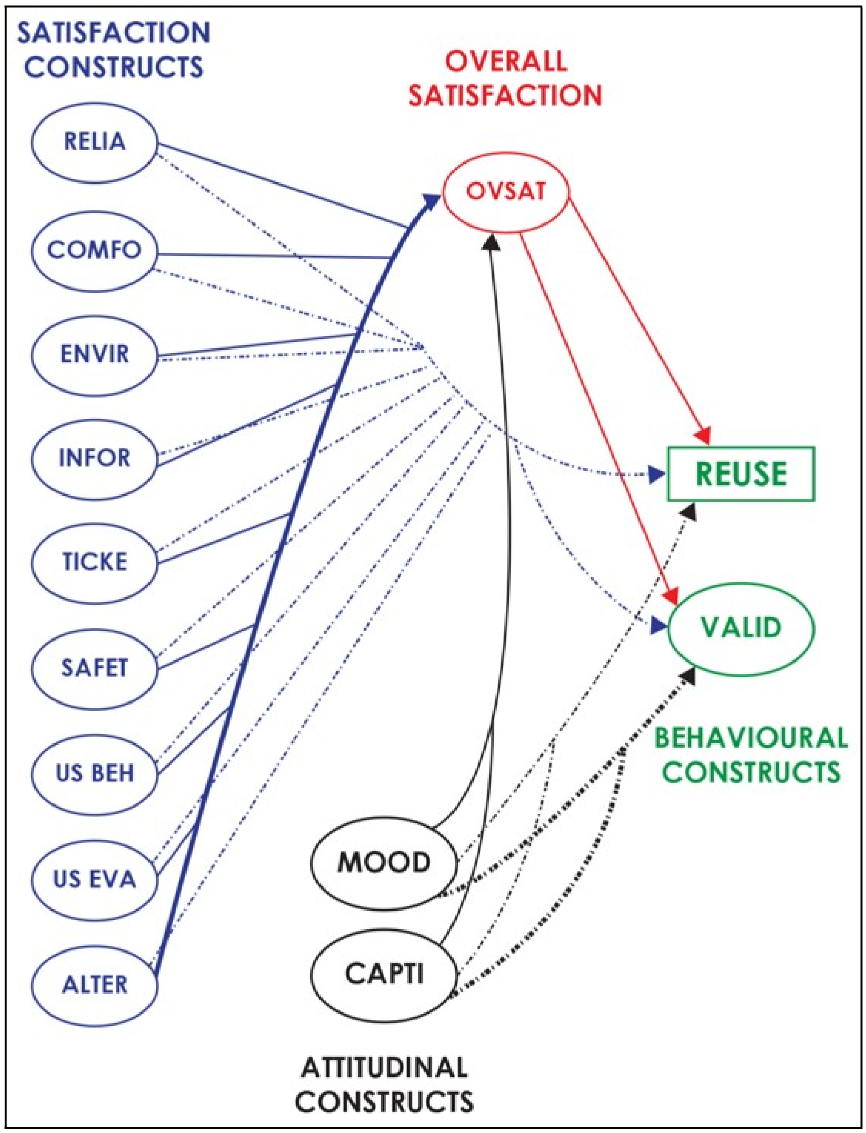

Figure 8-1: SEM for Metro and bus

We hypothesise that all the satisfaction constructs and the attitudinal constructs (MOOD, CAPTI) regress on overall satisfaction (OVSAT), reuse (REUSE), and validate behaviour (VALID). Next, OVSAT additionally regresses on REUSE and VALID. With this structure we can capture the effects of all satisfaction and attitudinal constructs on *evasion behaviour* ("-VALID"), and we concurrently estimate the effects they have on *satisfaction* and *reuse* behaviour. We can answer our research questions with this structure, providing critical results for policy-design.

In Tables 8-8 and 8-9, we present the results of the regression component of the SEM ordinal Probit model for Metro. We briefly comment on them; we kept variables with at least an 80% significance ($|z| > 1.28$, two-tailed). In the OVSAT regression both explained variances (R^2) are high, 0.72 for Metro and 0.70 for bus. We deem an R^2 >0.3 moderate, >0.5 acceptable, >0.6 adequate, and >0.7 high. For the structural coefficients (St.Coeff.), values within the range 0.1-0.3 are low, within the range 0.3-0.5, moderate, and >0.5 high (Allen et al., 2018a, Currie and Delbosc, 2017). The most relevant construct is RELIA (0.50 and 0.42). The rest of the constructs present values around 0.1, and USEVA present a negative sign (-0.07 for both), as expected. This fact indicates that as users are more satisfied with *other users*' evading behaviour, they are more dissatisfied.

For the REUSE regression both explained variances (R^2) are low, 0.20 for Metro and 0.16 for bus. The most relevant construct is OVSAT (0.42 and 0.29), as expected, and is coherent with satisfaction theory (Oliver, 2010). For Metro, TICKE and CAPTI are also relevant. For bus, COMFO, ALTER, and CAPTI are relevant. The USEVA construct did not prove significant for either mode. This fact indicates the perception of this behaviour does not deter users from reusing the system.

Table 8-8: SEM ordinal Probit model: Metro

Regressions	Estimate	S.E.	Z-value	St.Coeff.	R^2	
OVSAT					**0.72**	
RELIA	0.52	0.04	14.37	0.50		
COMFO	0.07	0.03	2.72	0.08		
ENVIR	0.13	0.03	4.01	0.12		
TICKE	0.09	0.02	4.36	0.11		
SAFET	0.12	0.03	3.47	0.11		
USBEH	0.10	0.02	4.45	0.11		
USEVA	-0.06	0.02	-3.12	-0.07		
REUSE					**0.20**	
OVSAT	0.53	0.10	5.60	0.42		
TICKE	0.11	0.05	2.12	0.11		
CAPTI	0.32	0.04	8.91	0.32		
VALID					**0.31**	
RELIA	0.30	0.14	2.07	0.19		
SAFET	0.42	0.18	2.35	0.27		
USEVA	-0.54	0.08	-6.84	-0.45		
ALTER	-0.39	0.25	-1.55	-0.10		
CFI	**TLI**	**GFI**	**AGFI**	**RMSEA**	**SRMR**	**WRMR**
0.997	0.996	0.996	0.994	0.038	0.040	1.286

For the VALID regression both explained variances (R^2) are low to moderate, 0.31 for Metro and 0.26 for bus. The most relevant construct for both cases is USEVA with high coefficients (-0.45 and -0.50), as expected, since the more users are satisfied with fare evading behaviour, the more inclined to evade the fare (i.e. *not validate*) they are. Excitingly, we also discover that for both modes satisfaction with *reliability* is relevant and with a positive sign for validate behaviour (0.19 and 0.12). Interestingly satisfaction with safety (SAFET) in Metro also increases VALID.

We also discover that the OVSAT does not affect validate behaviour in both modes. This represents a valuable policy-related result since it means that attribute-specific satisfaction constructs (i.e., *reliability*, *safety*) may be the key to deterring evasion behaviour instead of overall transit satisfaction, as one could have suspected initially.

Table 8-9: SEM ordinal Probit model: bus

Regressions	Estimate	S.E.	Z-value	St.Coeff.	R^2	
OVSAT					**0.70**	
RELIA	0.50	0.04	13.02	0.42		
COMFO	0.33	0.10	3.44	0.19		
ENVIR	0.09	0.04	2.48	0.09		
SAFET	0.16	0.03	5.97	0.17		
USBEH	0.11	0.03	4.20	0.11		
USEVA	-0.05	0.02	-2.63	-0.07		
CAPTI	-0.03	0.02	-1.62	-0.04		
MOOD	0.11	0.02	5.77	0.14		
REUSE					**0.16**	
OVSAT	0.35	0.08	4.42	0.29		
COMFO	0.45	0.23	1.96	0.21		
USBEH	0.11	0.06	1.80	0.09		
ALTER	0.27	0.14	1.96	0.14		
CAPTI	0.26	0.07	3.59	0.26		
MOOD	0.08	0.04	1.93	0.08		
VALID					**0.26**	
RELIA	0.21	0.07	2.97	0.12		
USEVA	-0.58	0.06	-10.00	-0.50		
CFI	**TLI**	**GFI**	**AGFI**	**RMSEA**	**SRMR**	**WRMR**
0.994	0.993	0.993	0.990	0.044	0.043	1.400

From the fit indices for both modes, the CFI, TLI, GFI, and AGFI all comply with recommended cut-off (>0.95) values by Hu and Bentler (1999). In fact, they are all above 0.99, indicating excellent fits to the data. The RMSEA, the most critical index, complies with the strictest cut-off value (<0.05) at 0.038 and 0.044, for Metro and bus respectively. The WRMR is a fit index for SEM ordinal models; in general, the lower it is, the better when comparing between different models estimated from the same data. We use all indices to make comparisons with the SEM-MIMIC models, which we present next.

8.5.3. Structural equations models: SEM-MIMIC ordinal Probit models with travel and socioeconomic traits

In this subsection, we present the SEM-MIMIC (Joreskog and Goldberger, 1975) ordinal Probit models, for Metro and bus. We base this model on the previous one, the SEM ordinal. We follow the methodology presented in Allen et al. (2018b). We introduce dummy variables into the OVSAT, REUSE, and VALID regressions, which endeavour to capture the heterogeneity of the perceptions depending on users' travel and socioeconomic traits. For categorical variables, we add *n-1* dummy variables, where *n* is the number of groups in each category. For Metro, we introduced (base in parentheses): four regressors for Metro-line (Line 1), two for day of the week (Monday): TUE.WED and THU.FRI; three for frequency of travel (5 days/week), and one for period (Off-Peak). We introduced the following travel characteristics as numeric variables: No.PASS, WT.MEAN, WT.MAX, OCCUP.N, TT.MEAN, and TT.D2D. Additionally, we introduced one dummy for travel card (Regular), four for age group (AGE<24), and one for gender (Male). We introduced the numeric variable CARS.N for number of cars.

We included dummies: one for driver's license (No license), two for education level (Primary or High school), three for employment status (Non-other), and three for transfer (No): Metro transfer, bus transfer, and other transport transfer. Last we included the INCOME.N numeric variable using the given groups, and the mean household Income by *Comuna* of residence was inserted as it stood. We hypothesise the income effects to be different. For bus, the same regressors were used, except that instead of Metro-line variables, we used six dummies for operator (U1).

In total, we included 34 variables for Metro and 36 for bus. For the MIMIC variables, we kept those that proved significant at a 75% level (|z| >1.15, two-tailed). We used a less strict threshold, as these variables usually have lower levels of significance than the main satisfaction constructs.

We present the results for Metro in Table 8-10. The results are presented for all three regressions: OVSAT, REUSE, and VALID. The estimates are grouped into three groups for each regression: satisfaction constructs, travel characteristics, and socioeconomic traits. For Metro, in the OVSAT regression, the same satisfaction constructs remained significant.

Again, USEVA has a negative coefficient (-0.07). For REUSE and VALID, again the same satisfaction constructs remained significant. However, the explained variance (R^2) improved significantly in all three regressions (0.73, 0.29, 0.50). Also, the standardised coefficients in the VALID regression are more prominent than for the other two, indicating marked differences in evasion behaviour according to sociodemographic traits. Age, gender, having a license, employment status, and household income variables all play a critical role in explaining fare evading behaviour. We discuss this further in Section 8.6.

Table 8-10: SEM-MIMIC ordinal Probit model: Metro

Regressions	**Estimate**	**S.E.**	**Z-value**	**St.Coeff.**	**R^2**
OVSAT					**0.73**
RELIA	0.51	0.04	14.70	0.46	
COMFO	0.05	0.03	1.97	0.05	
ENVIR	0.13	0.03	4.30	0.12	
TICKE	0.10	0.02	4.78	0.12	
SAFET	0.12	0.03	3.81	0.10	
USBEH	0.10	0.02	4.61	0.10	
USEVA	-0.06	0.02	-3.29	-0.07	
LINE6	0.39	0.09	4.50	0.17	
WT.MAX	-0.01	0.00	-1.56	-0.06	
OCCUP.N	-0.11	0.02	-5.92	-0.20	
DISCOUNT	0.14	0.11	1.35	0.09	
FEMALE	-0.10	0.05	-1.93	-0.06	
LICENSE	-0.15	0.06	-2.35	-0.08	
INCOME.N	-0.05	0.03	-1.66	-0.06	
INC.MUNI	0.00	0.00	1.23	0.04	

Table 8-10: SEM-MIMIC ordinal Probit model: Metro (continued)

Regressions	Estimate	S.E.	Z-value	St.Coeff.	R^2	
REUSE					**0.29**	
OVSAT	0.53	0.09	5.79	0.40		
TICKE	0.11	0.05	2.25	0.10		
CAPTI	0.23	0.04	5.63	0.21		
LINE6	0.38	0.13	2.91	0.13		
THU.FRI	0.23	0.10	2.21	0.10		
FR.WEEK34	-0.17	0.11	-1.47	-0.06		
FR.WEEK67	0.18	0.11	1.63	0.06		
WT.MAX	0.02	0.01	2.99	0.13		
TT.D2D	0.00	0.00	3.46	0.12		
AGE25.31	-0.33	0.12	-2.89	-0.13		
AGE32.39	-0.36	0.16	-2.34	-0.12		
AGE40.48	-0.56	0.17	-3.38	-0.17		
AGE49	-0.62	0.18	-3.51	-0.19		
LICENSE	-0.11	0.09	-1.22	-0.05		
TECH.PROF	0.11	0.09	1.19	0.04		
INCOME.N	0.06	0.04	1.42	0.05		
VALID					**0.50**	
RELIA	0.38	0.12	3.22	0.20		
USBEH	0.31	0.09	3.47	0.19		
USEVA	-0.47	0.07	-6.52	-0.33		
ALTER	-0.76	0.24	-3.23	-0.20		
CAPTI	-0.33	0.10	-3.31	-0.23		
LINE2	0.48	0.25	1.95	0.12		
FR.WEEK12	-0.29	0.24	-1.21	-0.08		
FR.WEEK34	-0.42	0.23	-1.85	-0.11		
FR.WEEK67	-0.33	0.25	-1.33	-0.08		
No.PASS	0.12	0.09	1.31	0.09		
WT.MEAN	-0.06	0.03	-2.10	-0.13		
AGE25.31	0.33	0.26	1.30	0.10		
AGE32.39	1.02	0.32	3.19	0.25		
AGE40.48	0.80	0.80	2.11	0.19		
AGE49	1.18	0.35	3.37	0.28		
FEMALE	0.38	0.17	2.20	0.13		
LICENSE	0.37	0.22	1.69	0.12		
STUDENT	0.66	0.37	1.82	0.22		
EMPLOYEE	0.90	0.34	2.64	0.32		
ENTREPRE	0.64	0.39	1.63	0.15		
INCOME.N	0.17	0.10	1.75	0.12		
CFI	**TLI**	**GFI**	**AGFI**	**RMSEA**	**SRMR**	**WRMR**
0.991	0.991	0.985	0.982	0.036	0.041	1.355

The results for bus, in Table 8-11, are shown for all three regressions: OVSAT, REUSE, and VALID.

Again, the estimates are grouped into three groups for each regression: satisfaction constructs, travel characteristics, and socioeconomic traits. For bus, in the OVSAT regression, the same satisfaction constructs remained significant and, once more, USEVA had a negative coefficient (-0.07). For REUSE and VALID, the significant constructs changed. For REUSE, only OVSAT, CAPTI, and MOOD remained, indicating that the MIMIC variables provide better explanatory power for this regression. For VALID, only RELIA, USEVA, ALTER, and CAPTI remained.

Table 8-11: SEM-MIMIC ordinal Probit model: bus

Regressions	**Estimate**	**S.E.**	**Z-value**	**St.Coeff.**	**R^2**
OVSAT					**0.73**
RELIA	0.47	0.04	12.03	0.37	
COMFO	0.35	0.09	3.72	0.20	
ENVIR	0.08	0.03	2.40	0.08	
TICKE	0.06	0.02	2.45	0.07	
SAFET	0.14	0.03	5.56	0.14	
USBEH	0.11	0.03	4.37	0.11	
USEVA	-0.06	0.02	-2.89	-0.07	
MOOD	0.12	0.02	6.27	0.14	
UNIT2	-0.22	0.10	-2.17	-0.09	
WT.MEAN	-0.01	0.00	-2.38	-0.09	
WT.MAX	-0.01	0.00	-3.92	-0.17	
OCCUP.N	-0.11	0.02	-5.48	-0.19	
TT.D2D	-0.00	0.00	-1.43	-0.05	
AGE25.31	-0.13	0.09	-1.35	-0.06	
AGE32.39	-0.17	0.11	-1.51	-0.07	
AGE49	-0.14	0.10	-1.34	-0.07	
FEMALE	-0.10	0.06	-1.72	-0.06	
TECH.PROF	-0.24	0.08	-3.12	-0.11	
STUDENT	-0.29	0.14	-2.12	-0.16	
EMPLOYEE	-0.17	0.08	-2.15	-0.10	
METRO.TR	-0.14	0.07	-2.12	-0.08	
BUS.TR	-0.10	0.06	-1.52	-0.05	

Table 8-11: SEM-MIMIC ordinal Probit model: bus (continued)

Regressions	Estimate	S.E.	Z-value	St.Coeff.	R^2	
REUSE					**0.28**	
OVSAT	0.42	0.06	7.59	0.32		
CAPTI	0.14	0.04	3.38	0.13		
MOOD	0.06	0.04	1.62	0.05		
UNIT5	-0.30	0.14	-2.06	-0.09		
UNIT6	-0.36	0.17	-2.18	-0.11		
THU.FRI	0.61	0.28	2.21	0.25		
FR.WEEK12	-0.22	0.14	-1.56	-0.07		
PM.PEAK	-0.12	0.10	-1.20	-0.05		
AGE40.48	-0.28	0.18	-1.53	-0.09		
AGE49	-0.25	0.17	-1.49	-0.09		
CARS.N	-0.08	0.07	-1.15	-0.05		
STUDENT	0.31	0.20	1.54	0.13		
INCOME.N	-0.06	0.04	-1.48	-0.06		
VALID					**0.50**	
RELIA	0.31	0.09	3.59	0.14		
USEVA	-0.52	0.06	-9.15	-0.37		
ALTER	-0.21	0.15	-1.36	-0.07		
CAPTI	-0.15	0.08	-1.72	-0.10		
UNIT2	0.45	0.19	2.30	0.11		
UNIT3	1.71	0.45	3.78	0.43		
UNIT4	0.32	0.20	1.56	0.08		
UNIT5	0.53	0.21	2.56	0.13		
UNIT6	0.50	0.22	2.31	0.12		
UNIT7	1.57	0.45	3.48	0.38		
FR.WEEK12	-0.25	0.17	-1.45	-0.06		
FR.WEEK67	-0.25	0.14	-1.77	-0.07		
WT.MEAN	-0.02	0.01	-1.82	-0.09		
TT.D2D	0.00	0.00	1.83	0.09		
DISCOUNT	0.37	0.21	1.75	0.12		
AGE25.31	0.36	0.19	1.91	0.10		
AGE32.39	0.49	0.22	2.29	0.12		
AGE40.48	0.92	0.23	4.05	0.22		
AGE49	1.21	0.22	5.57	0.35		
FEMALE	0.47	0.12	4.08	0.17		
TECH.PROF	0.21	0.15	1.39	0.06		
UNIVERSITY	0.42	0.16	2.59	0.14		
STUDENT	-0.33	0.26	-1.25	-0.10		
INCOME.N	0.19	0.06	3.42	0.14		
INC.MUNI	0.00	0.00	1.47	0.06		
CFI	**TLI**	**GFI**	**AGFI**	**RMSEA**	**SRMR**	**WRMR**
0.982	0.981	0.973	0.967	0.039	0.046	1.458

Again, the MIMIC variables provide more explanatory power. Additionally, the explained variance (R^2) improved significantly in the three regressions (0.73, 0.28, 0.50). The standardised coefficients in the VALID regression are greater than for the other two, indicating clear differences in evasion behaviour according to sociodemographic traits. Age, gender, having a license, employment status, and household income variables all play a critical role in explaining fare evading behaviour. Additionally, the INC.MUNI variable became significant. We discuss each variable further in Section 8.6.

Both model results attest to the fact that the SEM-MIMIC allows us to discern heterogeneity in transit satisfaction perception. Both models comply with the recommended cut-off values (Hu and Bentler, 1999). We obtain an excellent fit to the data as the CFI, TLI, GFI, and AGFI are all still above 0.98 for Metro and 0.96 for bus (>0.95). Also, the RMSEA remains less than 0.05. Both models improve on the RMSEA compared to the original SEM models, indicating a better fit. To be precise, 0.036 for the MIMIC against 0.038 for Metro, and 0.039 for the MIMIC against 0.044 for bus. The WRMR did not improve in either case; however, adding all the dummy variables improved the model fit according to the explained variance and the RMSEA. Explicitly, the explained variance improves to 0.50 for the VALID (*validating behaviour*) for both Metro and bus in comparison to the original SEM model (0.31, 0.26), this is an appealing result. The model fit is enhanced by adding the travel and socioeconomic traits, and the results also proves that evasion behaviour is highly heterogeneous, and demographic-dependent.

8.5.4. Structural equations models: SEMM-NL latent class analysis

We provide additional insight by introducing the finite mixture with non-linearities SEMM-NL models. The objective of running these models was to assess whether there are two or more classes with different satisfaction models in the Metro and bus samples. We want to test for non-linearities in the most relevant attributes. In essence, we should be able to separate the latent classes in regard to the relative weight they assign to the satisfaction domains. This would provide insight on whether some attributes are more important for different groups. We use the *nlsem* (Umbach et al., 2017) library package from the *R* (R Core Team, 2013) statistical programming language. We run the analysis for Metro and bus.

First, we tested using one-class models with seven and five variables. The five variable-model provided a more parsimonious model according to the AIC and BIC indices: the smaller, the better (Vrieze, 2012). Next, we tested for non-linearities for *reliability*, the most relevant variable, to assess whether this helped improve the fit indices.

Then we tested a two-class model, which was superior. Subsequently, we tested a two-class model with non-linearities, resulting in a model with higher AIC and BIC indices than the former. As such, we retain that the best SEMM-NL model is the two-class model with non-linearities, shown in bold in Table 8-12. A summary of the fit indices are presented in Table 8-12; the same type of models proved to be the best for both Metro and bus. To estimate these models, the satisfaction ratings must be entered as numeric, so the ordinal nature of the ratings is lost; however as seven categories are used, this may not be problematic (Hoyle, 2012).

Table 8-12: SEMM-NL analysis: AIC and BIC for Metro and bus latent class models

Mode	No.Class	No.LV	No.NL	AIC	BIC
METRO	1	7	0	80651	81188
METRO	1	5	0	63808	64190
METRO	1	5	1	49999	50385
METRO	**2**	**5**	**0**	**24852**	**25924**
METRO	2	5	2	26370	27455
BUS	1	7	0	84635	85157
BUS	1	5	0	71807	72205
BUS	1	5	1	65180	65583
BUS	**2**	**5**	**0**	**29738**	**30855**
BUS	2	5	2	30121	31252

Subsequently, in Table 8-13, we present the final model results for the SEMM, for Metro and bus. For Metro, two separate classes were found. For both classes, RELIA is the most significant variable and by far (0.50, 0.54); the next most relevant variable is ENVIR (0.14) for class1 and SAFET (0.15) for class2. Both of these variables are *safety/protection* variables. The results are intuitive as reliability is the most fundamental attribute and after this both these *safety/protection* variables take relevance. The USEVA variable was lost in all models; it is not that relevant for overall satisfaction, result that is coherent with the previous SEM models where the USEVA was significant but with a small coefficient than other variables.

Table 8-13: SEMM Overall Satisfaction 2-Latent Class models: Metro and bus

Metro	**St.Coeff.**	**S.E.**	**Z-value**	**Prob.Class**
class1.RELIA	0.50	0.06	7.82	P=0.55
class1.ENVIR	0.14	0.05	3.11	
class1.TICKE	0.07	0.04	2.09	
class1.SAFET	0.08	0.04	1.77	
class1.USBEH	0.08	0.03	2.28	
class2.RELIA	0.54	0.05	10.01	P=0.45
class2.ENVIR	0.06	0.04	1.57	
class2.TICKE	0.09	0.04	2.46	
class2.SAFET	0.15	0.04	3.50	
class2.USBEH	0.03	0.02	1.25	
Log L.: -12424.03				
Bus	**St.Coeff.**	**S.E.**	**Z-value**	**Prob.Class**
class1.RELIA	0.40	0.06	7.15	P=0.55
class1.COMFO	0.24	0.13	1.83	
class1.ENVIR	0.07	0.06	1.17	
class1.SAFET	0.10	0.04	2.76	
class1.USBEH	0.11	0.04	2.58	
class2.RELIA	0.48	0.06	8.48	P=0.45
class2.COMFO	0.19	0.14	1.38	
class2.ENVIR	0.05	0.04	1.16	
class2.SAFET	0.19	0.04	4.26	
class2.USBEH	0.04	0.04	1.19	
Log L.: -14866.98				

For bus, again, two separate classes were for found both classes. RELIA is the most significant variable and by far (0.40, 0.48). However, the next variable in relevance is COMFO (0.24) for class1 and COMFO and SAFET (0.19) for class2.
We interpret that RELIA, as the most fundamental attribute, is the most relevant. For class1, *safety* is not that relevant, hence the COMFO variable takes relevance. Conversely, for class2, *safety* is relevant. In short, for both modes, *reliability* is the most relevant policy-related variable, across the user base, enhancing previous results, from the OVSAT regressions.

8.6. Discussion of Results

In this section, we follow the same order as the model results (Section 8.5) while discussing the most important results and their interpretation. First, from the measurement models (see Annex I, Tables 8-14 and 8-15), we obtain which are the most relevant items in each domain according to the users' perceptions. We comment on the most relevant first for the satisfaction constructs in Metro. For RELIA, it is frequency and reliability (B03, B06); for ENVIR, it is temperature (B13); for TICKE, it is ease and speed to use recharge points (B18); and for USEVA it is fare evasion through exits (B27). Next, we comment for the satisfaction constructs in bus. For RELIA, it is reliability (B06), for COMFO, it is comfort inside the buses (B09), for SAFET, it is security against theft at bus stops and inside buses (B19), and for USEVA, it is fare evasion through the front door (B25). These results provide policy-making information for the PTA, as they indicate which specific items are the most important for users.
For the behavioural and attitudinal constructs, results are similar across modes. For OVSAT, it is final overall satisfaction that proved to be the most relevant (OS2), indicating that users can assess their satisfaction better once they have answered most of the questionnaire. This result is coherent with dell'Olio's (2010), signifying that users modify their assessment slightly after performing the

questionnaire. For MOOD, it is being in an excellent mood (MD1), for VALID, it is the users' validate behaviour (VB2), and for CAPTI, it is not having the shared taxi alternative. The PCA (Tables 8-6 and 8-7) allowed us to discern the composition of the OVSAT variable, as the recommend to family and friends item (REC) proved to be associated with OVSAT. The measurement model confirmed this result.

From the SEM ordinal Probit models (Tables 8-8 and 8-9) we gain insight into which satisfaction and attitudinal constructs help explain OVSAT as well as the behavioural constructs REUSE and VALID. For OVSAT, across both modes, USEVA proved to be significant with a negative coefficient (-0.07). The correct interpretation is that users who are satisfied with *other users evading behaviour* will have less overall satisfaction. Even though the coefficient is not that high (<|0.10|), it is significant and as such, the result suggests that less evasion will imply more satisfaction, as most users ranked USEVA low. For OVSAT, the most relevant was RELIA with a high coefficient for both modes; this result is coherent with Allen et al. (2018a). Again, this is a critical policy-related result. For the bus regression, COMFO and SAFET turned out to be moderately significant, implying that these would also be relevant for the Administration.

For REUSE, TICKE, and CAPTI were important for Metro users, while COMFO and CAPTI were relevant for bus users. OVSAT was relevant for REUSE, as expected from satisfaction theory (Oliver, 2010). It is important to note that the explained variance indices were low for REUSE, which could indicate that other variables are missing in order to explain this behaviour.

For VALID (i.e. validating behaviour) in both modes, USEVA proved to be highly negative and the most significant variable. This result indicates that as expected, as users are more satisfied with fare evading behaviour, they will more likely *evade fare*s. Additionally, for both modes, reliability is highly relevant for validating behaviour. This result is directly related to policy-making since it implies that to increase users' likelihood to validate, USEVA must be deterred and satisfaction with reliability must increase. We believe that this result is aligned with past studies which were analysed in the literature review, such as Guarda et al. (2016b) which determined that increasing frequencies decreases fare evasion. The link between regularity-frequency and decreasing evasion was mentioned by Torres-Montoya (2014) as well as by Tirachini and Quiroz (2016). Our study provides empirical proof to this connection directly from the users' perspective.

Additionally, we learned that in the VALID regression, OVSAT did not prove relevant; this is a crucial result. It suggests that overall satisfaction is not a direct predictor of fare evasion. Nevertheless, from our previous result, we can infer that it is the fundamental notion of *reliability* in the system that the user assesses in order to decide if they evade the fare or not. Moreover, studies by Buneder and Galilea (2017) and Troncoso and de Grange (2017) empirically determined that the overall aggregate satisfaction did not have a causal effect on fare evasion; our study is coherent with both. From our results, we conclude that the most relevant policy-related actions needed in order to cut back evasion will be to deter the USEVA and increase RELIA. In other words, increase satisfaction with reliability and decrease satisfaction with *other users*' fare evading behaviour.

From the SEM-MIMIC models that included the travel and socioeconomic traits of the users (Tables 8-10 and 8-11), we were able to assess what kind of different conditions affect the users' overall satisfaction, their reuse behaviour, and their validate behaviour. We will detail the results and their interpretation.

For the Metro model, we start with the OVSAT regression. The following variables turned out to have positive coefficients: Line 6, having a discount, and Municipality Income. All of them make intuitive sense; L6 is the newest Metro-line which has air conditioning and also has the most extended spacings between stations, providing the fastest travel times in the current network. Having a discount affects the expectation, as users pay less for the service they have lower expectations, hence higher satisfaction. Satisfaction is the real service minus the expected one. We hypothesise that Income by Municipality being positive indicates better services offered, as it increases. With negative coefficients, we find the following variables: maximum waiting time, vehicle occupation, female, license, and income. It is expected that as the maximum waiting time increases the user perceives less satisfaction; the same is expected when the vehicle occupation increases. Notice that this coefficient has a high magnitude (-0.20). Being female is adverse for satisfaction, as women tend to have higher expectations in some of the transit attributes; this result is coherent with Allen et al.'s (2018a, 2018b). Finally, having a driver's license and higher income indicate that users have a higher expectation from the service and hence, a lower satisfaction, an intuitive result.

Next, we detail on the REUSE regression for Metro. The following variables have positive signs: Line 6, day of the week: Thursday or Friday, frequency of use: 6-7 days a week, waiting time max, travel time door-to-door, having a technical degree, and income. L6 is the newest line, hence has a higher quality service as explained before. The day of the week may be a *close to the weekend* effect and no further interpretation can be gained. The frequency of use is intuitive, users that travel very frequently now will most probably travel frequently in the future. The waiting time max and travel time door-to-door being positive are counter-intuitive; these variables may be showing a *captivity* effect. Having a technical degree and higher income relates to having a higher SES, an indicator that higher SES users have a preference to reuse Metro more often. With negative signs, we find the following variables: frequency of use 3-4 days a week, all age groups (>24), and having a license. Again the frequency of use is intuitive; as users are less frequent, they will tend to use the Metro less. The age groups indicate that as users age, they tend to use the Metro less; we believe this can be associated with general mobility decreasing as we age. Finally, having a license would indicate higher probability of an alternate transport mode.

Next, we detail on the VALID regression for Metro. The following variables have positive signs, an indication that users are less prone to evade fares: Line 2, number of trains that pass before boarding, all the age groups (>24), female, license, student or employee, entrepreneur, and income. We have no interpretation for L2 having more validations than other lines. For number of trains that pass, we hypothesise that these users have a fixed schedule, hence employment, and as such are less prone to evade fares. All age groups indicate that young users (18-24) are more prone to evade fares, *ceteris paribus*. It is interesting to point out that the 25-31 and the 40-48 age groups are slightly more inclined to evade-fare than the other two groups. In general, females are less prone to evade fares. Both these results resonate with Bucciol et al.'s (2013) and Barabino et al.'s (2015) where males and under 26 were found to be more likely to evade-fare. Being occupied (student, employed, entrepreneur) all indicate a sense of employment, which makes these users less likely to evade fares. Having a license and higher incomes indicate higher SES, again making users less prone to evade fares. The following variables have negative signs, an indication that users are more prone to evade fares: frequency of use 1-4, 6-7 days/week, and mean waiting time. It may be that the 1-4 days/week denote some underemployment. For heavy users, it may be a medium to save on fare. Finally, waiting more would cause dissatisfaction with reliability, hence users are more prone to evade-fare.

For the bus model, we start with the OVSAT regression. The MOOD variable proved positive and significant. No variables from the MIMIC ones turned out to have positive coefficients. The following have negative coefficients: Unit 2, waiting time mean and max, vehicle occupation, travel time door-to-door, age 25-39 and >49, female, having a technical degree, being a student or an employee, and having to transfer to Metro or bus. The U2 negative sign may indicate that this operator has a lower quality of service than the rest. The waiting times, occupation, and door-to-door travel time all provide an intuitive sense; the higher they are, the less satisfaction. The age groups indicate that as we age, we are less satisfied with the bus service. Yet, the 40-48 did not turn out significant. Having a technical degree, and being occupied indicate higher SES, hence less satisfaction due to higher expectations. Plus, having to transfer in the case of the bus system indicates less satisfaction.

Next, we detail on the REUSE regression for bus. The following variables have positive signs: day of the week: Thursday and Friday, and being a student. Again, the day may be a *close to the weekend* effect. Being a student implies having to travel every day of the week, hence the increase in *reuse* behaviour. The following variables have negative signs: Unit 5, Unit 6, frequency of use 1-2 days/week, PM-Peak, age >40, number of cars, and income. The U5 and U6 are related to the differences in operators' service quality. These two may have lower levels. The frequency of use is intuitive as less frequent users are more prone to use the service less in the future. The PM-Peak is more crowded in general than the Off-Peak in Santiago and also, buses suffer congestion during the PM-Peak. The over 40 group may find travelling by bus harder than the other groups due to restrictions on mobility. Finally, having more cars and higher income indicate higher SES. The higher SES users tend to prefer other modes of transportation. This result is reflected in these variables.

Next, we detail on the VALID regression for bus. The following variables have positive signs, indicating that users are less prone to evade fares: all Units (2-7), travel time door-to-door, discount, all age groups, female, technical degree, University, income, and mean Municipality Income. The Units variables indicate that all operators face significant differences in users' willingness to evade fares; Unit 3 and Unit 7 possess the highest validate behaviour, while Unit 1 the least. This information is key for the PTA. The travel time door-to-door may be related to the users valuing the trip as it is long, hence more willingness to pay. Also, a longer trip may involve a connection with Metro, which would make it more difficult to evade the fare. Having a discount makes it more probable to validate, as the fare is lower. Being a female and ageing makes it more probable to validate; again the result is consistent with past literature (Delbosc and Currie, 2018), where males and under 26 years of age are reported more prone to evade fares. Having a technical or University education and higher income implies higher SES, hence, less fare evasion behaviour. Now, the higher Municipality Income also indicates a higher SES, as reported by Guarda et al. (2016b) and Buneder and Galilea (2017); lower income Municipalities report higher fare evasion rates in the bus system. Our result confirms this statement not only by household income but also by mean Income by Municipality (*Comuna*) of residence.

The following variables have negative signs, an indication that users are more prone to evade fares: frequency of use days/week 1-2/6-7, waiting time mean, being a student. In short, very infrequent users and heavy users are more prone to evade fares. The former may be dissatisfied by the service, and we hypothesise that the latter are very avid users that take advantage of the opportunity to evade fares as they use the bus system more; they believe that they are more *entitled* to evade fares. Having to wait more increases the chances of fare evading, again due to dissatisfaction. Also, being a student makes users more prone to evade fares; this result is coherent with Bucciol et al. (2013) and Barabino et al. (2015). Indeed, a *contagion* effect is plausible amongst students that make them more prone to evade fares when travelling together.

Both SEM-MIMIC models allowed us to unveil the heterogeneities in overall satisfaction, reuse behaviour, and validating behaviour, all in one model. The results are all important for policy decision-making. Finally, from the SEMM two-class models, we learned that *reliability* is the most important attribute across modes. No non-linearities were found in *reliability* towards OVSAT, indicative of a performance attribute; in other words, *reliability* is relevant across all the range of its scale of satisfaction. This result is crucial for the PTA, as it indicates that in general, most policies should be related to improving *reliability*.

8.7. Conclusions

In this section, we present some limitations of our study, conclusions, and policy recommendations from our findings. One of the limitations of our study is the sample size, even though we were able to obtain adequately fitting SEM-MIMIC models. Nevertheless, we utilised a 75% significance level for these last models. We believe that with a minimum sample size of 2,000 per mode we could have achieved higher significance for some variables. Moreover, some may have been introduced categorically, such as income and vehicle occupation levels. Nevertheless, our sample size allowed us to obtain minimum significance in most of the sociodemographic variables, giving us confidence in the results obtained. Additionally, most results make intuitive sense and are comparable with the existing literature.

Our study focused on obtaining a holistic *behavioural satisfaction-evasion* model. Two focus groups aided in building a standard transit satisfaction questionnaire. The questionnaire was answered by 2,002 users, for Metro and bus, and at the end they were asked about their fare evading behaviour. One may argue that a more focused approach could have been used. However, we believe that with our approach we were able to gain more accurate responses from users, specifically in the last two items (i.e. evasion behaviour). We argue that by asking them first about their peers' (i.e. friends and family) behaviours, we were able to obtain a more *realistic* response about their own behaviour. We learned that only 4% of Metro users admit to evading fares, while they admit that their peers evade 15% of the time. For the bus, 19% of bus users admitted to evading fares, while they state that their peers evade 42% of the time. Since the evasion rate for the bus system reported in the literature approximates was 28%, we are confident on the results obtained from our surveys.

Our main results indicate the following. First, for overall satisfaction, the most relevant attribute is *reliability*. This result stands for both Metro and bus. Additionally, we tested possible SEMM-NL, i.e. SEM with finite mixtures and non-linearities. The best fitting models indicated two-latent classes, for both modes, where the *reliability* is the most relevant attribute for both, and by far (see Tables 8-12 and 8-13). This result provides robustness to our initial results for overall satisfaction (OVSAT). From the users' overall satisfaction point-of-view, *reliability* is the domain were policy-related actions need to take place.

From the SEM and SEM-MIMIC models (Tables 8-8 and 8-11) we learn that for the REUSE regression, overall satisfaction is the most relevant variable, a result in accord with satisfaction theory (Oliver, 2010). However, we notice that some of the satisfaction constructs loose significance when the MIMIC variables are introduced, indicating that the heterogeneity (i.e. MIMIC) variables take on more explanatory power, and thus aid in providing a better representation of the phenomenon. We attest to this using the explained variances indices. All of the R^2 increase significantly for the SEM-MIMIC models. For the VALID (i.e. validating behaviour), we learn that the two most significant variables are USEVA (i.e. satisfaction with *other users*' fare evading behaviour), with a negative coefficient, and RELIA (i.e. reliability) with a positive coefficient. Both results are influential in relation to policy-making; the USEVA coefficient is high ~0.50 for both modes. This indicates that a policy is needed to educate the user base on the wrongdoing of evading, plus, reliability needs to be improved significantly to deter fare evasion. As reliability is the most relevant variable for overall satisfaction, its importance is enhanced since it is also useful to deter fare evasion. We believe that the USEVA effect may also be explained by a *contagion effect* (Gino et al., 2009). As fellow fare evaders tend to travel together, and as more users perceive fare evasion, they become more used to it and at some point, find it acceptable enough to commit it themselves.

Finally, we can obtain a more holistic profile of the fare evader. For Metro, the *fare evader* is captive or rates other alternatives high (see Table 8-10), is a low-frequent user (1-4 days/week) or conversely a heavy and avid user (6-7 days/week), and his mean waiting time is higher than the average. He is aged 18-24 years on average, male, does not possess a driver's license, unemployed and does not *study*, and has a low household income. Additionally, the *fare evader* gives reliability a low rating and is highly satisfied with *other* users' fare evading behaviour.

For the bus, the *fare evader* is captive or rates other alternatives high (see Table 8-11), is a very low-frequent user (1-2 days/week) or conversely a heavy and avid user (6-7 days/week), his mean waiting time is higher than the average, the door-to-door travel time is lower than average, and does not have a discount card. He is aged 18-24 years on average, male, has a low level of education, is a *student*, has a low household income, and lives in a lower income Municipality. Additionally, the *fare evader* rates reliability low and is highly satisfied with *other* users' fare evading behaviour.

From all of the above, we learn that the fare evasion problem has a hefty socioeconomic component, as low educated users who live in lower income households tend to evade fares. Accordingly, in order to deter fare evasion, we recommend:

a) Invest in marketing campaigns for educating the userbase on the wrongdoing of fare evasion.
b) Invest in marketing campaigns for educating the userbase on the benefits of paying the fare to improve the system.
c) Implement stratified levels of fares for low-income users with unique personalised cards.
d) Focus PTAs resources in reliability improving strategies, such as designated corridors, bus only.
e) Decrease waiting times by deploying schedule base services on low-frequency routes, bus only.
f) Implement incentives for bus operators to improve the regularity and frequencies of the bus service.

Finally, we consider that our approach allowed us to identify motivational components of *fare evaders*, even in Metro, a system with a low rate of fare evasion. As such, we believe that a similar approach could be implemented in other PT systems with low to middle fare evasion rates. Excitingly, we were able to combine satisfaction theory applied to PT with fare evading perceptions and self-reported evading behaviour. We deem this to be an important contribution to the PT literature.

9. CONCLUSION

In this section, we address the general and specific objectives of the dissertation and show how all of them were achieved. Also, specific recommendations are given regarding how to improve or complement the results obtained. For this, we synthesise the results obtained in each chapter and deliver, in contrast with the specific objectives, more generalisable conclusions. Subsequently, we state the thesis contributions and suggest future research avenues. The last subsection presents the papers produced during this investigation.

9.1. General Objective

To develop a general model of users' satisfaction with public transport, which explains the satisfaction of users in different contexts (both modes of transport and cities), and is consistent with Maslow's theory of human motivation.

9.2. Specific Objectives

a) To develop a satisfaction model which captures the heterogeneity associated with the sociodemographic characteristics of users and their types of travel.

b) To develop a satisfaction model which includes non-linear links between service attributes and overall satisfaction.

c) To formulate a model that allows capturing heterogeneous subpopulations through specific satisfaction models that differ transversally in the subpopulations.

d) To establish empirically whether Maslow's theory of human motivation offers a plausible theoretical foundation for models of users' satisfaction with public transport.

e) To determine empirically whether including critical incidents enhances public transport satisfaction models.

f) To determine if and how a major network reform affects the perceived satisfaction of public transport users.

g) To determine if (and how) including operational variables improves models of the perceived satisfaction of the public transport users.

h) To determine if a satisfaction-evasion relation exists and to develop a behavioural fare-evader profile, by modelling public transport satisfaction and evasion, accounting for heterogeneity.

9.3. Accomplishment of Objectives

1. In Chapters 2 and 3 we developed part of the methodological toolkit for analysing public transport satisfaction. We proposed a SEM-MIMIC framework that allowed us to incorporate travel and sociodemographic heterogeneity into satisfaction models. Specific policy-related recommendations were obtained for two different case studies. The approach provided excellent statistical flexibility, allowed us to obtain critical policy-related information, and is widely applicable and generalizable to other case studies.
2. In Chapters 5 and 8, we tested non-linear in the parameter functions for overall satisfaction models. In Chapter 5, they did not turn out to provide a better fit than regular SEM models. In Chapter 8, additionally, we tested non-linear and finite mixture (i.e. latent class) SEM models, finding that a non-linear model provided a better fit than the linear model. However, when testing 2-class models (i.e. with latent classes), the linear model provided a better fit for two databanks examined, Metro and bus. This 2-class linear model proved to be the best fitting model of all. Related with this result, it is important to clarify that as reliability is regarded as a performance (i.e. critical) attribute, it should have

a linear relationship with satisfaction according to our theory, and we empirically confirmed this. Having said that, we believe that in a PT system where reliability is high, other (hedonic) attributes should gain importance and we should observe non-linearities in their specific cases. More research is warranted to attempt unveiling non-linear relations between attributes and satisfaction. In conclusion, although non-linearities may arise, it is clear that this is not the case for the most critical attribute in all our case studies: *reliability*.

3. In Chapter 4 we developed the last part of the methodological toolkit of the dissertation, a SEM-Multigroup framework, that allowed us to specify different satisfaction models across subpopulations, in terms of travel and sociodemographic characteristics of users. With this model, we were able to determine differing priorities across subgroups, leading to essential policy-related knowledge. Further, the methodology is applicable and generalizable to other case studies.
4. We argued that previous results reported in the PT literature had not been theoretically justified. In Chapter 5 we proposed and tested Maslow's hierarchy of human motivation as a theoretical foundation for the resulting satisfaction models. We obtained robust results proving that the proposed *hierarchy of transit needs* holds for the case studies analysed. We examined different cities, and in only one of them found a slight divergence with the order of priority suggested by the theory. Still, we argue that this is an instrument-city specific deviation, and as such should not deter from the conclusions gained from the other cities. Additionally, we found evidence that the *hierarchy* applied to different groups regarding travel characteristics and socioeconomic characteristics, both for the pooled data and for two cities in particular. Finally, additional finite mixture (i.e. latent class) models estimated also support the existence of the *hierarchy*. We believe that these results combine to be a

critical first step, to determine whether the proposed *hierarchy* is a proper theoretical foundation for PT satisfaction models.

5. In Chapter 2, we obtained strong evidence suggesting that attribute specific critical incidents (CI) helped in explaining attribute-specific satisfaction levels. Our model determines that the CI affect overall satisfaction and loyalty indirectly. As no studies had considered the effect of CI on loyalty behaviour before, we consider this a significant contribution, since CI represent a direct policy variable that may be targeted by PT authorities, to increase ridership.
6. In Chapter 6 we addressed a major network reform in a PT system from the users' satisfaction point-of-view. We estimated a working model that allowed us to understand how users perceive a major reform from the satisfaction perspective. Our framework extracts the *network reform effect*, and as such gives policy-directed insight to the PT administrators. We confirmed a positive perception of the *network reform*, as an added-value effect which was perceived by the users.
7. In Chapter 7 we found that adding operational variables to the proposed satisfaction models produced a better model fit to the data. Nevertheless, the *perceived* (i.e. subjective) variables still play an essential role in the satisfaction models.
8. We obtained empirical evidence for the *behavioural satisfaction-evasion* relationship, in Chapter 8. Although we did not find a direct link between users' overall transit satisfaction and fare evading behaviour, we found a significant link between the users' satisfaction *with reliability*, their (dis)satisfaction *with* other users' fare evading behaviour, *and* their *own* fare evading behaviour. We also obtained a *behavioural fare-evader profile* from our framework. Our analysis produced results that are consistent with the PT evasion literature.

In synthesis, we accomplished all the specific objectives of this thesis. In particular, we produced evidence that *Maslow's hierarchy of transit needs* works as an appropriate theoretical foundation for public transport satisfaction models. Next, we synthesise the resulting contributions of the dissertation.

9.4. Contributions

In this subsection we take a bird's-eye-view of the results of the thesis, beyond the accomplishment of specific objectives. For this, we highlight the contributions of the dissertation to the prevalent public transport satisfaction literature. We focus on the specific contributions of each chapter first (Table 9-1), showing the scope, the summarised results, the contributions, and their relevance.

Following, we comment on how, taken together, the contributions can impact the established literature. Finally, we synthesise the expected overall influence of the thesis in the state-of-the-art for the coming years.

The contributions of the thesis can be separated in four: (i) methodological, (ii) theoretical, (iii) specific applications, and finally, (iv) a general satisfaction model.

First, the thesis proposed a methodological toolkit that contributes and extends the current state-of-the-art in PT satisfaction. We show that with flexible SEM models we may pose several policy-related hypotheses and test them. In particular, note that our models incorporate the heterogeneity of PT travel, which is one of its most important characteristics.

Second, and this we consider our most important contribution; we proposed, tested and found evidence for a theoretical justification in the order of priorities of attributes in PT satisfaction models.

Table 9-1: Summary of Contributions

Chapter	Scope	Results	Contribution	Relevance
Ch.2	Methodology	SEM-MIMIC PT satisfaction model, service-specific and global system satisfaction.	A PT satisfaction model including heterogeneity and a correction for it. **Novel**	This framework is flexible enough to account for different travel conditions and sociodemographic characteristics. Operational variables may be added too.
Ch.3	Methodology	SEM-MIMIC ordinal Probit PT satisfaction model. Critical Incidents (CI) for attribute specific satisfaction constructs.	Ordinal SEM-MIMIC. is shown to be superior to the numeric one. Attribute specific CI highly impact all attribute constructs. **Novel**	This methodology is the most flexible to treat heterogeneous conditions and provides the most accurate account of the phenomenon. CI is a highly relevant policy variable.
Ch.4	Methodology	SEM-MGA. Attribute specific constructs relevant for loyalty. Different populations provide differing satisfaction models.	The SEM-MGA allows unveiling differences among the population, and to find which variables produce it. **Novel** Specific attributes impact loyalty directly.	Attribute specific variables are relevant for loyalty. Also, the framework allows to estimate different satisfaction models in the population.
Ch.5	Theory	We proposed and tested a theoretical foundation for PT satisfaction models, based on Maslow's theory, and found evidence for it.	We provide evidence for Maslow's hierarchy of Transit Needs. This is the first such attempt in the literature. **Novel**	Provides a stablished working hierarchy of which attributes have to be taken care of first.Very relevant for PT administrators.
Ch.6	Application	We tested whether a major network reform improves satisfaction.	A major network reform was assessed from the users' perspectives, during implementation. **Novel**	Our proposed framework for analysing major network reform, from the users' perspective, provides evidence for other PTAs, of the benefits provided in a transfer-based bus network.
Ch.7	Application	We show that adding operational variables improves SEM-MIMIC ordinal models for PT satisfaction	We include both subjective and objective attributes in the models, along different time periods and service lines. **Novel**	Allows for testing the effect of objective policy variables in the transit satisfaction models. Gives a framework that could be applied in other PT systems.
Ch.8	Application	We derive a *behavioural evasion-satisfaction* relation, using our proposed framework. We produce a *behavioural* fare evader profile.	The *evasion-satisfaction* link is directly explored. **Novel**. The model is tested using data for two modes in the same city, providing consistent and robust results.	Evasion is a major problem in some major PT systems. We provide insight of a *behavioural* relation between *evasion*and *satisfaction*. This is a key result for PT administrators worldwide that have similar problems.

Any PT authority should now have a theoretical base to understand which attributes are more important to users. In a sense, without even gathering information, a hypothesis of the order of the attributes can be made, and customer satisfaction surveys can be targeted with more precision. The *hierarchy of transit needs* brings a plausible theoretical base to PT satisfaction models.

Third, we demonstrated with three additional applications how our framework can be used to obtain policy-related conclusions and knowledge regarding particular PT scenarios. We showed: (i) how a major network reform affects transit satisfaction, (ii) how adding objective operational variables improves the satisfaction models, and finally, (iii) we unveiled an *evasion-satisfaction* relation and obtained a sensible fare-evader profile.

Fourth, we proposed a general satisfaction model, based on the results obtained in the thesis, which could serve as a guideline for future studies. The model is shown in Figure 9-1. The critical incidents (CI) act as antecedents of the satisfaction constructs and are attribute specific. The CI and the satisfaction constructs are divided in four main domains: reliability, safety, customer services, and comfort. In turn, the satisfaction constructs act as antecedents to the Overall Satisfaction, and to the behavioural intent constructs: evasion behaviour and loyalty. While Overall Satisfaction also acts as antecedent of the behavioural intent constructs. Finally, heterogeneity in both sociodemographic and travel characteristics can be considered using the SEM-MIMIC or the SEM-MGA paradigms. In particular we advocate the SEM ordinal Probit framework as it provides the most accurate results.

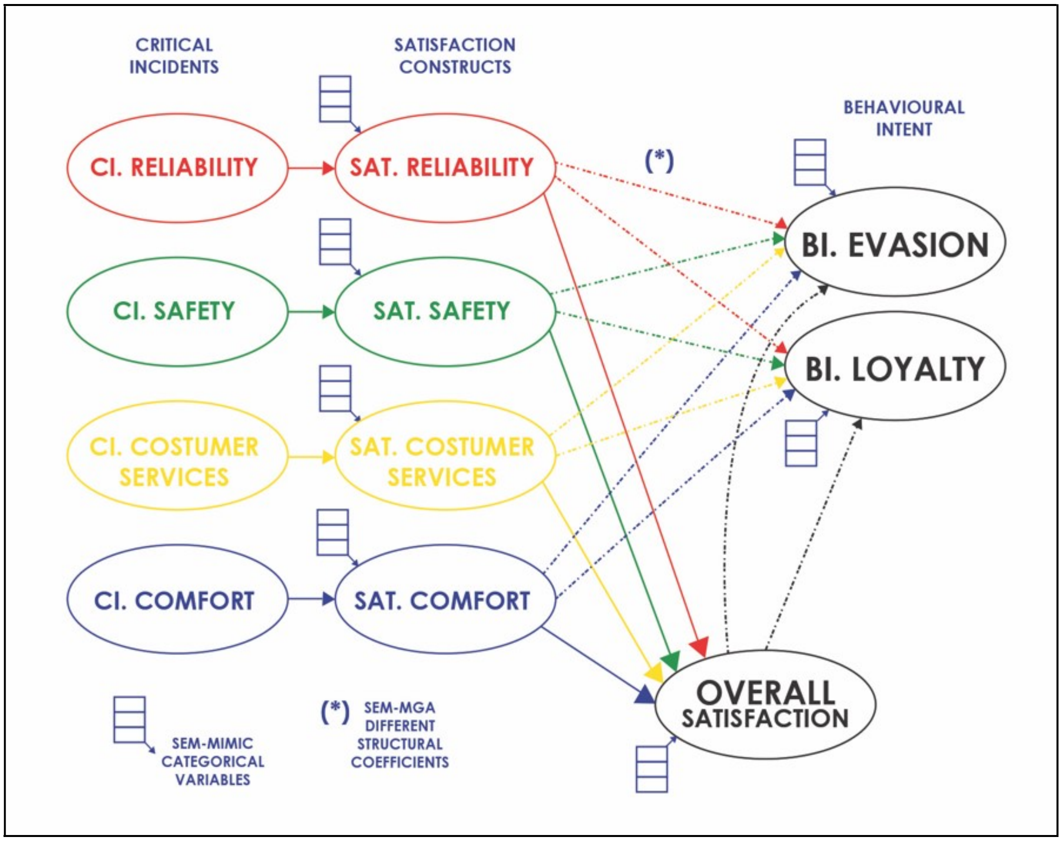

Figure 9-1: General SEM ordinal Probit Satisfaction Model

The contributions of this thesis can impact the established literature in two key aspects: the methodological framework using the proposed SEM (MIMIC or MGA) models, and the theoretical underpinning developed. We expect that further developments and refinements of these models will be achieved in the future. We expect the results presented herein to impact the state-of-the-art public transport satisfaction literature, by encouraging new investigations along the research lines suggested below.

9.5. Future Research

The primary objective of the dissertation was to deliver a theoretical foundation for public transport satisfaction models. We believe that our results provide evidence in this avenue. In future research, the *hierarchy of transit needs* should be tested in different settings. Ideally, a common questionnaire should be designed and tested as a pilot in multiple cities, to ensure standardisation of the satisfaction constructs. Once the pilot is complete, the survey may be implemented in the different PT contexts.

In this thesis we proved that SEM ordinal Probit models provided a better fit across the board over their numeric counterparts. This comparison could be made in other settings, to obtain a generalisation of the results. We recommend using CI in all future PT satisfaction studies, when possible. We also showed that CI have a significant adverse, and very relevant, effect on satisfaction. As such, it is convenient to measure this by including an additional CI for the attribute-specific items evaluated in the survey, see Table 9-1.

We also gained knowledge about the fact that not including specific items in the questionnaire could bias the results; as we compared different satisfaction models, with completely different items, in a Metro satisfaction survey. We recommend further exploring this avenue of research, in other settings.

When testing non-linear finite mixture models, our results demonstrated that reliability was the most relevant attribute across the board, and that it holds a linear relationship with satisfaction. Further research along the lines of testing non-linearities in PT systems that have an excellent level of reliability is warranted. Our hypothesis is that once the reliability expectation is fulfilled (i.e. highly graded), users will give more relevance to hedonic attributes, and some of these attributes will have a non-linear relation with satisfaction. The SEM with finite mixture and the SEM with non-linearities may help in this endeavour.

We discovered that the bus network reform provided higher satisfaction to users of the *new* lines. An intuitive research avenue would be to explore this in other network reform cases, in different cities. We also discovered that objective operative variables could be added to improve the models (i.e. the number of boardings at each station). From our results, regarding the analysis of the *evasion-satisfaction link*, we recommend replicating a similar study in another PT setting, perhaps with a lower fare-evasion rate and in a different country. The aim would be to generalise the results.

For future studies regarding modelling PT satisfaction, we recommend employing the methodological toolkit developed here: the SEM-MIMIC models and the SEM-Multigroup analysis using the ordinal Probit framework.

A possible research avenue, not explored in this research, would be to test the main hypotheses presented here, regarding satisfaction constructs, with discrete choice models (DCM). One of the disadvantages of the SEM models used in this dissertation is that we only assess current users. The DCM framework could allow to integrate non-users, by presenting them hypothetical scenarios from the current PT settings.

Additionally, the theoretical hypothesis underpinning the priorities in the satisfaction models could be applied in other industries and settings (i.e. marketing science). As such, we believe that this dissertation could open a new line of research in testing *Maslow's hierarchy of needs.* We believe that *Maslow's theory* is **testable**, and we formulated an adequate framework for testing it. Further examination of this theory, through our proposed operationalisation, could provide a deeper generalisation of the psychological foundation of human motivations. In the future, this framework may prove, to be the most significant contribution of this dissertation to the overall scientific literature. Recall that need is an expectation and satisfaction is the accomplishments of needs. Hence, *Maslow's theory* is an intuitive fit of a theoretical foundation for general satisfaction models.

9.6. Scientific Papers

The following is the list of papers produced in this dissertation:

Allen, J., Eboli, L., Forciniti, C., Mazzulla, G., Ortúzar, J. de D., 2018c. The role of critical incidents and involvement in transit satisfaction and loyalty. *Transport Policy* (*under review*).

Allen, J., Eboli, L., Mazzulla, G., Ortúzar, J. de D., 2018b. Effect of critical incidents on public transport satisfaction and loyalty: an ordinal Probit SEM-MIMIC approach. *Transportation*. https://doi.org/10.1007/s11116-018-9921-4

Allen, J., Muñoz, J.C., Ortúzar, J. de D., 2018a. Modelling service-specific and global transit satisfaction under travel and user heterogeneity. *Transportation Research Part A: Policy and Practice* **113**, 509–528. https://doi.org/10.1016/j.tra.2018.05.009

Allen, J., Muñoz, J.C., Ortúzar, J. de D., 2019a. Understanding public transport satisfaction: using Maslow's hierarchy of (transit) needs. *Transport Policy* (*under review*).

Allen, J., Muñoz, J.C., Ortúzar, J. de D., 2019c. Effect of objective attributes on transit satisfaction (*to be submitted*).

Allen, J., Muñoz, J.C., Ortúzar, J. de D., 2019d. Evasion behaviour in public transport: A satisfaction framework (*to be submitted*).

Allen, J., Muñoz, J.C., Rosell, J., 2019b. Effect of a major network reform on bus transit satisfaction. *Transportation Research Part A: Policy and Practice* (*under review*).

BIBLIOGRAPHY

Alcaíno, P., Torres, M.P., 2017. Estudio: Satisfacción usuaria en sistemas de transporte público. Informe resultados grupos focales. Santiago, Chile.

Allen, J., Eboli, L., Forciniti, C., Mazzulla, G., Ortúzar, J. de D., 2018c. The role of critical incidents and involvement in transit satisfaction and loyalty. Transport Policy (*under review*).

Allen, J., Eboli, L., Mazzulla, G., Ortúzar, J. de D., 2018b. Effect of critical incidents on public transport satisfaction and loyalty: An ordinal Probit SEM-MIMIC approach. Transportation. https://doi.org/10.1007/s11116-018-9921-4

Allen, J., Muñoz, J.C., Ortúzar, J. de D., 2018a. Modelling service-specific and global transit satisfaction under travel and user heterogeneity. Transportation Research Part A: Policy and Practice 113, 509–528. https://doi.org/10.1016/j.tra.2018.05.009

Alm, J., 2012. Measuring, explaining, and controlling tax evasion: Lessons from theory, experiments, and field studies. International Tax and Public Finance 19, 54–77. https://doi.org/10.1007/s10797-011-9171-2

Anable, J., 2005. 'Complacent car addicts' or 'Aspiring environmentalists'? Identifying travel behaviour segments using attitude theory. Transport Policy 12, 65–78. https://doi.org/10.1016/j.tranpol.2004.11.004

Anderson, J.C., Gerbing, D.W., 1988. Structural equation modeling in practice: A review and recommended two-step approach. Psychology Bulletin, 103, 411–423. https://doi.org/ 10.1037/0033-2909.103.3.411

Angrist, J.D., Krueger, A.B., 1999. Chapter 23 - Empirical Strategies in Labor Economics, in: Ashenfelter, O.C., Card, D. (Eds.), Handbook of Labor Economics. Elsevier, 1277–1366. https://doi.org/10.1016/S1573-4463(99)03004-7

Antonucci, L., Crocetta, C., d'Ovidio, F.D., Toma, E., 2014. Passenger satisfaction: A multi-group analysis. Quality and Quantity 48, 337–345. https://doi.org/10.1007/s11135-012-9771-7

Ashok, K., Dillon, W.R., Yuan, S., 2002. Extending discrete choice models to incorporate attitudinal and other latent variables. Journal of Marketing Research 39, 31–46. https://doi.org/10.1509/jmkr.39.1.31.18937

Backhaus, K., Bauer, M., 2001. The impact of critical incidents on customer satisfaction in business-to-business relationships. Journal of Business-to-Business Marketing 8, 25-54 https://doi.org/10.1300/J033v08n01_02

Badia, H., Argote-Cabanero, J., Daganzo, C.F., 2017. How network structure can boost and shape the demand for bus transit. Transportation Research Part A: Policy and Practice 103, 83–94. https://doi.org/10.1016/j.tra.2017.05.030

Barabino, B., Salis, S., Useli, B., 2015. What are the determinants in making people free riders in proof-of-payment transit systems? Evidence from Italy. Transportation Research Part A: Policy and Practice 80, 184–196. https://doi.org/10.1016/j.tra.2015.07.017

Batarce, M., Muñoz, J.C., Ortúzar, J. de D., 2016. Valuing crowding in public transport: Implications for cost-benefit analysis. Transportation Research Part A: Policy and Practice 91, 358–378. https://doi.org/10.1016/j.tra.2016.06.025

Beaujean, A.A., 2014. Latent Variable Modeling Using R: A Step-by-Step Guide, First Edition. Routledge, New York.

Bei, L.-T., Chiao., Y.-C., 2001. An integrated model for the effects of perceived product, perceived service quality, and perceived price fairness on customer satisfaction and loyalty. Journal of Consumer Satisfaction, Dissatisfaction, and Complaining Behaviour 14, 125-140.

Beirão, G., Sarsfield Cabral, J.A., 2007. Understanding attitudes towards public transport and private car: A qualitative study. Transport Policy 14, 478–489. https://doi.org/10.1016/j.tranpol.2007.04.009

Bel, G., Rosell, J., 2016. Public and private production in a mixed delivery system: Regulation, competition and costs. Journal of Policy Analysis and Management 35, 533–558. https://doi.org/10.1002/pam.21906

Beltrán, P., Gschwender, A., Palma, C., 2013. The impact of compliance measures on the operation of a bus system: The case of Transantiago. Research in Transportation Economics 39, 79–89. https://doi.org/10.1016/j.retrec.2012.05.026

Berry, L.L., Zeithaml, V.A., Parasuraman, A., 1990. Five imperatives for improving service quality. Sloan Management Review 31, 9-38.

Bitner, M.J., Booms, B.H., Tetreault, M.S., 1990. The service encounter: Diagnosing favourable and unfavorable incidents. Journal of Marketing 54, 71-84. https://doi.org/10.2307/1252174

Bollen, K.A., 1989. Structural Equations with Latent Variables. John Wiley and Sons, Chichester.

Bonfanti, G., Wagenknecht, T., 2010. Human factors reduce aggression and fare evasion. Public Transport International 59, 28-32.

Bordagaray, M., dell'Olio, L., Ibeas, A., Cecín, P., 2014. Modelling user perception of bus transit quality considering user and service heterogeneity. Transportmetrica A: Transport Science 10, 705–721. https://doi.org/10.1080/23249935.2013.823579

Bowen, N.K., Guo, S., 2012. Structural Equation Modeling. Oxford University Press, New York. http://trove.nla.gov.au/version/190046392

Braun, V., Clarke, V., 2006. Using thematic analysis in psychology. Qualitative Research in Psychology 3, 77-101. http://dx.doi.org/10.1191/1478088706qp063oa

BRT Centre of Excellence, 2018. https://BRTdata.org

Bucciol, A., Landini, F., Piovesan, M., 2013. Unethical behavior in the field: Demographic characteristics and beliefs of the cheater. Journal of Economic Behavior and Organization 93, 248–257. https://doi.org/10.1016/j.jebo.2013.03.018

Buneder, C., Galilea, P., 2017. Temporal and spatial analysis of fare evasion in Transantiago. Presented at the Transportation Research Board 96th Annual Meeting. Transportation Research Board.

Byrne, B.M., 2009. Structural Equation Modeling with AMOS: Basic Concepts, Applications, and Programming, Second Edition, 2nd. Edition. Routledge, New York.

Cadotte, E.R., Turgeon, N., 1988. Dissatisfiers and satisfiers: Suggestions for consumer complaints and compliments. Journal of Consumer Satisfaction, Dissatisfaction, and Complaining Behaviour 1, 74-79.

Cats, O., Abenoza, R.F., Liu, C., Susilo, Y.O., 2015. Evolution of satisfaction with public transport and its determinants in Sweden. Transportation Research Record: Journal of the Transportation Research Board 2538, 86–95. https://doi.org/10.3141/2538-10

Chen, H.K., 2016. Structural interrelationships of group service quality, customer satisfaction, and behavioral intention for bus passengers. International Journal of Sustainable Transportation 10, 418-429. https://doi.org/10.1080/15568318.2014.965372

Cheung, G.W., Rensvold, R.B., 2002. Evaluating goodness-of-fit indexes for testing measurement invariance. Structural Equation Modeling: A Multidisciplinary Journal 9, 233-255. https://doi.org/10.1207/S15328007SEM0902_5

Chou, J.S., Kim C., 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. https://doi.org/10.1016/j.eswa.2008.08.056

Cronbach, L.J., 1951. Coefficient alpha and the internal structure of tests. Psychometrika 16, 297–334. https://doi.org/10.1007/BF02310555

Cronin, J., Brady, M., Hult, G., Tomas, M., 2000. Assessing the effects of quality, value, and customer satisfaction on consumer behavioral intentions in service environments. Journal of Retailing 76, 193–218. https://doi.org/10.1016/s0022-4359(00)00028-2

Currie, G., Delbosc, A., 2017. An empirical model for the psychology of deliberate and unintentional fare evasion. Transport Policy 54, 21–29. https://doi.org/10.1016/j.tranpol.2016.11.002

Currie, G., Wallis, I., 2008. Effective ways to grow urban bus markets – A synthesis of evidence. Journal of Transport Geography 16, 419–429. https://doi.org/10.1016/j.jtrangeo.2008.04.007

Daganzo, C.F., 2010. Structure of competitive transit networks. Transportation Research Part B: Methodological 44, 434–446. https://doi.org/10.1016/j.trb.2009.11.001

de Oña, J., de Oña, R., 2015. Quality of service in public transport based on customer satisfaction surveys: A review and assessment of methodological approaches. Transportation Science 49, 605–622. https://doi.org/10.1287/trsc.2014.0544

de Oña, J., de Oña, R., Eboli L., Forciniti, C., Mazzulla, G., 2016b. Transit passengers' behavioural intentions: The influence of service quality and customer satisfaction. Transportmetrica A: Transport Science 12, 385-412. https://doi.org/10.1080/23249935.2016.1146365

de Oña, J., de Oña, R., Eboli, L., Mazzulla, G., 2013. Perceived service quality in bus transit service: A structural equation approach. Transport Policy 29, 219–226. https://doi.org/10.1016/j.tranpol.2013.07.001

de Oña, J., de Oña, R., Eboli, L., Mazzulla, G., 2015. Heterogeneity in perceptions of service quality among groups of railway passengers. International Journal of Sustainable Transportation 9, 612–626. https://doi.org/10.1080/15568318.2013.849318

de Oña, J., de Oña, R., Eboli, L., Mazzulla, G., 2016a. Index numbers for monitoring transit service quality. Transportation Research Part A: Policy and Practice 84, 18-30.

de Oña, R., de Abreu e Silva, J., Muñoz-Monge, C., de Oña, J., 2018. Users' satisfaction evolution of a metropolitan transit system in a context of economic downturn. International Journal of Sustainable Transportation 12, 66–74. https://doi.org/10.1080/15568318.2017.1328546

de Oña, R., de Oña, J. 2015. Analysis of transit quality of service through segmentation and classification tree techniques. Transportmetrica A: Transport Science 11, 367-387. https://doi.org/10.1080/23249935.2014.1003111

de Oña, R., Eboli, L., Mazzulla, G., 2014. Key factors affecting rail service quality in the Northern Italy: A decision tree approach. Transport 29, 75-83. http://dx.doi.org/10.3846/16484142.2014.898216

del Castillo, J.M., Benitez, F.G., 2012. Determining a public transport satisfaction index from user surveys. Transportmetrica A: Transport Science 9, 713-741. https://doi.org/10.1080/18128602.2011.654139

Delbosc, A., Currie, G., 2012. Modelling the causes and impacts of personal safety perceptions on public transport ridership. Transport Policy 24, 302–309. https://doi.org/10.1016/j.tranpol.2012.09.009

Delbosc, A., Currie, G., 2016a. Cluster analysis of fare evasion behaviours in Melbourne, Australia. Transport Policy 50, 29–36. https://doi.org/10.1016/j.tranpol.2016.05.015

Delbosc, A., Currie, G., 2016b. Four types of fare evasion: A qualitative study from Melbourne, Australia. Transportation Research Part F: Traffic Psychology and Behaviour 43, 254–264. https://doi.org/10.1016/j.trf.2016.09.022

Delbosc, A., Currie, G., 2018. Why do people fare evade? A global shift in fare evasion research. Transport Reviews. https://doi.org/10.1080/01441647.2018.1482382

dell'Olio, L., Ibeas, A., Cecín, P., 2010. Modelling user perception of bus transit quality. Transport Policy 17, 388–397. https://doi.org/10.1016/j.tranpol.2010.04.006

dell'Olio, L., Ibeas, A., Cecín, P., 2011. The quality of service desired by public transport users. Transport Policy 18, 217–227. https://doi.org/10.1016/j.tranpol.2010.08.005

Diamantopoulos, A., Siguaw, J.A., 2000. Introducing LISREL. London: Sage Publications.

DICTUC, 2018. Diseño y aplicación de encuestas de satisfacción a usuarios de Transantiago. Informe Final para BRT Centre of Excellence. Santiago, Chile.

DiStefano, C., Morgan, G.B., 2014. A comparison of diagonal weighted least squares robust estimation techniques for ordinal data. Structural Equation Modeling: A Multidisciplinary Journal 21, 425–438. https://doi.org/10.1080/10705511.2014.915373

Donoso, P., Munizaga, M., Rivera, J., 2013. Measuring user satisfaction in transport services: Methodology and application. In J. Zmud, M. Lee-Gosselin, M. Munizaga, J.A. Carrasco (Eds.). Transport Survey Methods: Best Practice for Decision Making, 603-624. Emerald, Bingley.

Eboli, L., Forciniti, C., Mazzulla, G. 2018. Formative and reflective measurement models for analysing transit service quality. Public Transport 10, 107-127. https://doi.org/10.1007/s12469-017-0168-9

Eboli, L., Mazzulla G., 2015. Relationships between rail passengers' satisfaction and service quality: A framework for identifying the key service factors. Public Transport, Planning and Operations 7, 185-201. http://dx.doi.org/10.5038/2375-0901.10.3.2

Eboli, L., Mazzulla, G., 2007. Service quality attributes affecting customer satisfaction for bus transit. Journal of Public Transportation 10, 21-34. https://doi.org/10.5038/2375-0901.10.3.2

Eboli, L., Mazzulla, G., 2011. A methodology for evaluating transit service quality based on subjective and objective measures from the passenger's point of view. Transport Policy 18, 172–181. https://doi.org/10.1016/j.tranpol.2010.07.007

Edvardsson, B., 1992. Service breakdowns: A study of critical incidents in an airline. International Journal of Service Industry Management 3, 17-29. https://doi.org/10.1108/09564239210019450

Edvardsson, B., 1998. Causes of customer dissatisfaction-studies of public transport by the critical-incident method, Managing Service Quality 8, 189-197. https://doi.org/10.1108/09604529810215675

Edvardsson, B., Strandvik, T., 2000. Is a critical incident critical for a customer relationship? Managing Service Quality 10, 82-91. https://doi.org/10.1108/09604520010318272

Edwards, J.R., 2011. The fallacy of formative measurement. Organizational Research Methods 14, 370–388. https://doi.org/10.1177/1094428110378369

Efthymiou, D., Antoniou, C., 2017. Understanding the effects of economic crisis on public transport users' satisfaction and demand. Transport Policy 53, 89–97. https://doi.org/10.1016/j.tranpol.2016.09.007

EMTA, 2017. EMTA Barometer of Public Transport in the European Metropolitan Areas 2015. https://www.emta.com/spip.php?article267

Estrada, M., Roca-Riu, M., Badia, H., Robusté, F., Daganzo, C.F., 2011. Design and implementation of efficient transit networks: Procedure, case study and validity test. Transportation Research Part A: Policy and Practice, Select Papers from the 19th International Symposium on Transportation and Traffic Theory (ISTTT) 45, 935–950. https://doi.org/10.1016/j.tra.2011.04.006

Factor Estratégico. 2010. Estudio Cualitativo de Opinión – El Fenómeno de la Evasión. Santiago, Chile.

Falk, T., Hammerschmidt, M., Schepers, J.J.L., 2009. The service quality-satisfaction link revisited: Exploring asymmetries and dynamics. Journal of the Academy of Marketing Science 38, 288–302. https://doi.org/10.1007/s11747-009-0152-2

Fellesson, M., Friman, M., 2008. Perceived satisfaction with public transport service in nine European cities. Journal of the Transportation Research Forum 47, 93-103. http://dx.doi.org/10.5399/osu/jtrf.47.3.2126.

Figler, S., Iraj, P.S., Welch, E.W., Yavuz, N., 2011. Customer loyalty and Chicago, Illinois, Transit Authority buses: Results from 2008 customer satisfaction survey. Transportation Research Record: Journal of the Transportation Research Board 2216, 148-156. https://doi.org/10.3141/2216-16

Flanagan, J.C., 1954. The critical incident technique. Psychological Bulletin 51, 327-357. http://dx.doi.org/10.1037/h0061470

Friman, M., Edvardsson, B., Garling, T., 1998. Perceived quality of public transport service: Inferences from complaints and negative critical incidents. Journal of Public Transportation 2, 69-91.

Friman, M., Edvardsson, B., Garling, T., 2001. Frequency of negative critical incidents and satisfaction with public transport services. Journal of Retailing and Consumer Services 8, 95-104. https://doi.org/10.1016/S0969-6989(00)00003-5

Gao, Y., Rasouli, S., Timmermans, H., Wang, Y., 2017. Effects of traveller's mood and personality on ratings of satisfaction with daily trip stages. Travel Behaviour and Society 7, 1–11. https://doi.org/10.1016/j.tbs.2016.11.002

Gino, F., Gu, J., Zhong, C.-B., 2009. Contagion or restitution? When bad apples can motivate ethical behavior. Journal of Experimental Social Psychology 45, 1299–1302. https://doi.org/10.1016/j.jesp.2009.07.014

Golob, T.F., 2003. Structural equation modeling for travel behavior research. Transportation Research Part B: Methodological 37, 1–25. https://doi.org/10.1016/S0191-2615(01)00046-7

Gómez-Lobo, A., Briones, J., 2013. Incentive structure in transit concession contracts: The case of Santiago, Chile, and London, England. Clean Air Institute, Washington D. C. http://www.cleanairinstitute.org/cops

Greene, M., Ortúzar, J. de D., 2002. Willingness to pay for social housing attributes: A case study from Chile. International Planning Studies 7, 55–87. https://doi.org/10.1080/13563470220112607

Gremier, D.D., Gwinner, K.P., 2008. Rapport-building behaviours used by retail employees. Journal of Retailing 84, 308-324. https://doi.org/10.1016/j.jretai.2008.07.001

Gremler, D.D., 2004. The critical incident technique in service research, Journal of Service Research 7, 65-89. https://doi.org/10.1177/1094670504266138

Grison, E., Gyselinck, V., Burkhardt, J.M., 2013. Using the critical incidents technique to explore variables related to users' experience of public transport modes. Proceedings of the 31st European Conference on Cognitive Ergonomics, Toulouse, France, August 26-28, 2013. https://doi.org/10.1145/2501907.2501952

Guarda, P., Galilea, P., Handy, S., Muñoz, J.C., Ortúzar, J. de D., 2016a. Decreasing fare evasion without fines? A microeconomic analysis. Research in Transportation Economics, Competition and Ownership in Land Passenger Transport (selected papers from the Thredbo 14 conference) 59, 151–158. https://doi.org/10.1016/j.retrec.2016.06.001

Guarda, P., Galilea, P., Paget-Seekins, L., Ortúzar, J. de D., 2016b. What is behind fare evasion in urban bus systems? An econometric approach. Transportation Research Part A: Policy and Practice, Practical Applications of Novel Methodologies to Real Cases: Selected Papers from the XIII Pan-American Conference on Traffic and Transportation Engineering 84, 55–71. https://doi.org/10.1016/j.tra.2015.10.008

Guirao, B., García-Pastor, A., López-Lambas, M.E., 2016. The importance of service quality attributes in public transportation: Narrowing the gap between scientific research and practitioners' needs. Transport Policy 49, 68–77. https://doi.org/10.1016/j.tranpol.2016.04.003

Habib, K.M.N., Kattan, L., Islam, M. T., 2011. Model of personal attitudes towards transit service quality. Journal of Advanced Transportation 45, 271–285. https://doi.org/10.1002/atr.106

Hauser, R.M., Goldberger, A.S., 1971. The treatment of unobservable variables in path analysis. Sociological Methodology 3, 81–117. https://www.ssc.wisc.edu/wlsresearch/publications/files/public/Hauser-Goldberger_Treatment.Unobservable.V.P.A.pdf

Hensher, D.A., Stopher, P., Bullock, P., 2003. Service quality - Developing a service quality index in the provision of commercial bus contracts. Transportation Research Part A: Policy and Practice 37, 499–517. https://doi.org/10.1016/S0965-8564(02)00075-7

Henson, J.M., Reise, S.P., Kim, K.H., 2007. Detecting mixtures from structural model differences using latent variable mixture modeling: A comparison of relative model fit statistics. Structural Equation Modeling: A Multidisciplinary Journal 14, 202–226. https://doi.org/10.1080/10705510709336744.

Herrera, J.C., Hensher, D., Mulley, C., Li, Z., Lindau, L., 2016. Assessing corridor performance. In Muñoz J.C. and Paget-Seekins L. (Eds.), Restructuring Public Transport Through Bus Rapid Transit: An International and Interdisciplinary Perspective (pp. 299-316). Policy Press at the University of Bristol, Bristol. http://www.jstor.org/stable/j.ctt1t89fj8.21

Hooper, D., Coughlan, J., Mullen, M. 2008. Structural equation modeling: Guidelines for determining model fit. Electronic Journal of Business Research Methods 6, 53-60. http://arrow.dit.ie/libart/4

Hoyle, R.H., 2012. Handbook of Structural Equation Modeling. Guilford Publications, New York.

Hu, L., Bentler, P.M., 1999. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal 6, 1-55. https://doi.org/10.1080/10705519909540118

Huse, C., Evangelho, F., 2007. Investigating business traveller heterogeneity: Low-cost vs full-service airline users? Transportation Research Part E: Logistics and Transportation Review 43, 259–268. https://doi.org/10.1016/j.tre.2006.10.005.

IDESCAT, 2018. Motor vehicle fleet in Barcelona. Statistical Institute of Catalonia. https://www.idescat.cat/pub/

IERMB, 2015. Encuesta de Movilidad en Día Laborable (EMEF). Retrieved June 28, 2017. https://iermb.uab.cat/es/encuestas/encuestas-de-movilidad.

Imaz, A., K. Habib, A. Shalaby, Idris, A., 2015. Investigating the factors affecting transit user loyalty. Public Transport 7, 39-60. https://doi.org/10.1007/s12469-014-0088-x

Iseki, H., Taylor, B.D., 2010. Style versus service? An analysis of user perceptions of transit stops and stations. Journal of Public Transportation 13, 23–48.

Jedidi, K., Jagpal, H.S., DeSarbo, W.S., 1997a. Finite-mixture structural equation models for response-based segmentation and unobserved heterogeneity. Marketing Science 16, 39–59. https://doi.org/10.1287/mksc.16.1.39.

Jedidi, K., Jagpal, H.S., DeSarbo, W.S., 1997b. STEMM: A general finite mixture structural equation model. Journal of Classification 14, 23–50. https://doi.org/10.1007/s003579900002

Joewono, T.B., Kubota, H., 2007. The multigroup analysis regarding user perception of paratransit service. Proceedings of the Eastern Asia Society for Transportation Studies. https://doi.org/10.11175/easts.7.1651

Jolliffe, I., 2014. Principal Component Analysis. John Wiley and Sons, Chichester. http://onlinelibrary.wiley.com/doi/10.1002/9781118445112.stat06472/abstract

Jones, M.A., Suh, J., 2000. Transaction-specific satisfaction and overall satisfaction: An empirical analysis. Journal of Services Marketing 14, 147–159. https://doi.org/10.1108/08876040010371555

Joreskog, K.G., Goldberger, A.S., 1975. Estimation of a model with multiple indicators and multiple causes of a single latent variable. Journal of the American Statistical Association 70, 631–639. https://doi.org/10.2307/2285946

Kaiser, H.F., 1960. The application of electronic computers to factor analysis. Educational and Psychological Measurement 20, 141–151. http://dx.doi.org/10.1177/001316446002000116.

Kano, N., Seraku, N., Takahashi, F., Tsuji, S., 1984. Attractive quality and must-be quality. Journal of the Japanese Society for Quality Control 14, 147–156.

Kenny, D.A., McCoach, D.B. 2003. Effect of the number of variables on measures of fit in structural equation modeling. Structural Equation Modeling: A Multidisciplinary Journal 10, 333-351. https://doi.org/10.1207/S15328007SEM1003_1

Kenrick, D.T., Griskevicius, V., Neuberg, S.L., Schaller, M., 2010. Renovating the pyramid of needs: Contemporary extensions built upon ancient foundations. Perspectives on Psychological Science 5, 292–314. https://doi.org/10.1177/1745691610369469.

Kline, R.B., 2015. Principles and Practice of Structural Equation Modeling. Guilford Publications.

Kolenikov, S., Angeles, G., 2009. Socioeconomic status measurement with discrete proxy variables: Is principal component analysis a reliable answer? The Review of Income and Wealth 55, 128–165. https://doi.org/10.1111/j.1475-4991.2008.00309.x

La Tercera, 2015. Estudio: pérdidas por evasión representa cerca de la mitad del subsidio al Transantiago. http://www2.latercera.com/noticia/estudio-perdidas-por-evasion-representa-cerca-de-la-mitad-del-subsidio-al-transantiago/. Accessed October 08, 2018.

Laboratorio de Innovación Pública, 2018. Informe Final Laboratorio 502: Estudio exploratorio sobre las causas y posibles soluciones de la evasión en Transantiago. https://politicaspublicas.uc.cl/wp-content//uploads/2018/04/180307_LIP_Laboratorio_502_Informe_Final_vf.pdf

Lai W.T., Chen, C.F., 2011. Behavioral intentions of public transit passenger. The roles of service quality, perceived value, satisfaction and involvement. Transport Policy 18, 318-325. https://doi.org/10.1016/j.tranpol.2010.09.003

Laroche, M., Ueltschy, L.C., Abe, S., Cleveland, M., Yannopoulos, P.P., 2004. Service quality perceptions and customer satisfaction: Evaluating the role of culture. Journal of International Marketing 12, 58–85. https://doi.org/10.1509/jimk.12.3.58.38100

Likert, R., 1932. A technique for the measurement of attitudes. Archives of Psychology 22, 5-55. http://psycnet.apa.org/record/1933-01885-001.

Lindau, L.A., Barcelos, M.M., Costa, D., Berti, M.B., Caten, T., S, C., Silva, D., Moreira, C.A., Pereira, B.M., 2017. Benchmarking focused on the satisfaction of bus transit users. Presented at the Transportation Research Board 96th Annual Meeting. Washington, D.C.

Litman, T., 2007. Valuing transit service quality improvements: Considering comfort and convenience in transport project evaluation. Victoria Transport Policy Institute (www.vtpi.org).

Machado-Leon, J.L, de Oña, J., de Oña, R., 2016. The role of involvement in regards to public transit riders' perceptions of the service. Transport Policy 48, 34-44. https://doi.org/10.1016/j.tranpol.2016.02.014

MacKinnon, D.P., Fairchild, A.J., Fritz, M.S., 2007. Mediation analysis. Annual Review of Psychology 58, 593–614. https://doi.org/10.1146/annurev.psych.58.110405.085542

Mahmoud, M., Hine, J., 2016. Measuring the influence of bus service quality on the perception of users. Transportation Planning and Technology 39, 284–299. https://doi.org/10.1080/03081060.2016.1142224

Maslow, A.H., 1943. A theory of human motivation. Psychological Review 50, 370–396. https://doi.org/10.1037/h0054346.

Maslow, A.H., 1954. Motivation and Personality. Harper and Brothers, New York.

Matzler, K., Bailom, F., Hinterhuber, H.H., Renzl, B., Pichler, J., 2004. The asymmetric relationship between attribute-level performance and overall customer satisfaction: A reconsideration of the Importance–Performance analysis. Industrial Marketing Management 33, 271–277. https://doi.org/10.1016/S0019-8501(03)00055-5.

Metro de Madrid, 2016. Informe Anual de Gestión y RC 2016. https://www.metrodemadrid.es

Metro de Madrid, 2017a. Carta de Servicios. https://www.metrodemadrid.es

Metro de Madrid, 2017b. Evolución de la demanda. Cierre año 2017. Servicio Planificación y Estudios de Operación - Área Control y Planificación de la Operación. https://www.metrodemadrid.es

Metro de Santiago, 2016. Reporte de sostenibilidad: 2016. https://www.metro.cl/documentos/reporte_2016.pdf

Metro de Santiago, 2018. https://www.metro.cl/

Miller, J.A., 1977. Studying satisfaction, modifying models, eliciting expectations, posing problems, and making meaningful measurements. In: Hunt, H.K. (Ed.). Conceptualization and Measurement of Consumer Satisfaction and Dissatisfaction. Marketing Science Institute, Cambridge, MA, 72-91.

Minser, J., Webb, V. 2010. Quantifying the benefits. Application of customer loyalty modeling in public transportation context. Transportation Research Record: Journal of the Transportation Research Board 2144, 111-120.

Mittal, V. and Kamakura, W.A., 2001. Satisfaction, repurchase intent, and repurchase behavior: Investigating the moderating effect of customer characteristics. Journal of Marketing Research 38, 131–142. https://doi.org/10.1509/jmkr.38.1.131.18832

Mittal, V., Katrichis, J.M., Kumar, P., 2001. Attribute performance and customer satisfaction over time: evidence from two field studies. Journal of Services Marketing 15, 343–356. https://doi.org/10.1108/EUM0000000005655.

Moore, J.C., Stinson, L.L., Welniak, E.J., 2000. Income measurement error in surveys: A review. Journal of Official Statistics 16, 331-361.

Morton, C., Caulfield, B., Anable, J. 2016. Customer perceptions of quality of service in public transport: Evidence for bus transit in Scotland. Case Studies on Transport Policy 4, 199–207. https://doi.org/10.1016/j.cstp.2016.03.002

Mouwen, A., 2015. Drivers of customer satisfaction with public transport services. Transportation Research Part A: Policy and Practice 78, 1–20. https://doi.org/10.1016/j.tra.2015.05.005

Mulley, C., Ho, C., 2013. Evaluating the impact of bus network planning changes in Sydney, Australia. Transport Policy 30, 13–25. https://doi.org/10.1016/j.tranpol.2013.07.003

Muñoz, J.C., Ortúzar, J. de D., Gschwender, A., 2008. Transantiago: The fall and rise of a radical public transport intervention. Travel Demand Management and Road User Pricing: Success, Failure and Feasibility, 151–172, Ashgate, Farnham.

Muthén, B., 1989. Latent variable modeling in heterogeneous populations. Psychometrika 54, 557–585. https://link.springer.com/article/10.1007/BF02296397

Oakes, J.M., Rossi, R.H., 2003. The measurement of SES in health research: Current practices and steps toward a new approach. Social Science and Medicine 56, 769–784. https://doi.org/10.1016/S0277-9536(02)00073-4

Ofir, C., Simonson, I., 2001. In search of negative customer feedback: The effect of expecting to evaluate on satisfaction evaluations. Journal of Marketing Research 38, 170–182. https://doi.org/10.2307/1558622

Oliver, R.L., 1980. A cognitive model of the antecedents and consequences of satisfaction decisions. Journal of Marketing Research 17, 460–469. https://doi.org/10.2307/3150499

Oliver, R.L., 2010. Satisfaction: A Behavioral Perspective on the Consumer. M.E. Sharpe, Armonk, NY. https://doi.org/10.1108/09564231011066132

Olsen, M.J.S., 1996. The critical episode model as a tool for organizational learning in service organizations. In Edvardsson, B., Modell, S. (Eds.), Service Management, Stockholm: Nerenius and Santérus Förlag Ab.

Olsen, S.O., 2007. Repurchase loyalty: The role of involvement and satisfaction. Psychology and Marketing 24, 315–341. https://doi.org/10.1002/mar.20163

Ortúzar, J. de D., Willumsen, L.G., 2011. Modelling Transport. John Wiley and Sons, Chichester.

Parasuraman, A., Zeithaml, V.A., Berry, L.L., 1985. A conceptual model of service quality and its implications for future research. Journal of Marketing 49, 41–50. https://doi.org/10.2307/1251430

Parasuraman, A., Zeithaml, V.A., Berry, L.L., 1988. SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality. Journal of Retailing 64, 12–40.

Park, J. W., Robertson, R., Wu, C. L., 2004. The effect of airline service quality on passengers' behavioral intention: A Korean case study. Journal of Air Transport Management 10, 435-439. https://doi.org/10.1016/j.jairtraman.2004.06.001

Paulssen, M., Birk, M.M., 2007. Satisfaction and repurchase behavior in a business-to-business setting: Investigating the moderating effect of manufacturer, company and demographic characteristics. Industrial Marketing Management 36, 983–997. http://dx.doi.org/10.1016/j.indmarman.2007.05.011.

Peek, G.-J., van Hagen, M., 2002. Creating synergy in and around stations: Three strategies for adding value. Transportation Research Record: Journal of the Transportation Research Board 1793, 1–6. https://doi.org/10.3141/1793-01.

Perone, J., Winters, P., Read, M., Sankah, I., 2005. Assessing Hierarchy of Needs in Levels of Service. Final Report FDOT Project BD 549-1 NCTR Project 527-08, National Center for Transit Research, University of South Florida. https://www.nctr.usf.edu/pdf/527-08.pdf.

Petrick, J.F., 2004. The roles of quality, value and satisfaction in predicting cruise passengers' behavioral intentions. Journal of Travel Research 42, 397-407. http://doi.org/10.1177/0047287504263037.

Public Transport Victoria, 2015. Network Revenue Protection Plan, Public Transport Victoria, State Government Melbourne, Australia.

R Core Team, 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/

Raveau, S., Álvarez-Daziano, R., Yáñez, M.F., Bolduc, D., Ortúzar, J. de D., 2010. Sequential and simultaneous estimation of hybrid discrete choice models. Transportation Research Record: Journal of the Transportation Research Board 2156, 131–139. https://doi.org/10.3141/2156-15

Reddy, A., Kuhls, J., Lu, A., 2011. Measuring and controlling subway fare evasion. Transportation Research Record: Journal of the Transportation Research Board 2216, 85–99. https://doi.org/10.3141/2216-10

Redman, L., Friman, M., Gärling, T., Hartig, T., 2013. Quality attributes of public transport that attract car users: A research review. Transport Policy 25, 119–127. https://doi.org/10.1016/j.tranpol.2012.11.005

Reichheld, F., 2004. The one number you need to grow. Harvard Business Review 81, 46-55.

Reynolds, K. L., Harris, L. C., 2009. Dysfunctional customer behaviour severity: An empirical examination. Journal of Retailing 85, 321–335. https://doi.org/10.1016/j.jretai.2009.05.005

Rohde, M., 2017. "World Metro Database - metrobits.org". www.mic-ro.com

Roos, I., 1999. Switching processes in customer relationships. Journal of Service Research 2, 68-85. https://doi.org/10.1177/109467059921006

Roos, I., 2002. Methods of investigating critical incidents: A comparative review. Journal of Service Research 4, 193-204. https://doi.org/10.1177/1094670502004003003

Rosseel, Y., 2012. lavaan: An R package for structural equation modeling. Journal of Statistical Software 48, 1–36. https://www.jstatsoft.org/v48/i02/.

Stauss, B., Hentschel, B., 1992. Attribute-based versus incident-based measurement of service quality: Results of an empirical study in the German car service industry. In P. Kunst and J. Lemmink (eds.), Quality Management in Services. Assen/Maastricht.

Stauss, B., Weinlich, B., 1997. Process-oriented measurement of service quality: Applying the sequential incident method. European Journal of Marketing 31, 33-55. https://doi.org/10.1108/03090569710157025

Steg, L., 2005. Car use: Lust and must. Instrumental, symbolic and affective motives for car use. Transportation Research Part A: Policy and Practice, Special Issue: Positive Utility of Travel 39, 147–162. https://doi.org/10.1016/j.tra.2004.07.001

Suazo-Vecino, G., Dragicevic, M., Muñoz, J.C., 2017. Holding boarding passengers to improve train operation on basis of an economic dwell time model. Transportation Research Record: Journal of the Transportation Research Board 2648, 96–102. https://doi.org/10.3141/2648-11

Sumaedi, S., Bakti, I.G.M.Y., Medi Y., 2012. The empirical study of public transport passengers' behavioral intentions: The roles of service quality, perceived sacrifice, perceived value, and satisfaction (Case study: Paratransit passengers in Jakarta, Indonesia). International Journal for Traffic and Transport Engineering 2, 83-97. http://ijtte.com/uploads/2012-03-20/d4c8811d-88a5-92b1p8.pdf

Sundling, C., Nilsson M.E., Hellqvist, S., Pendrill, L.R. Emardson, R., Berglund, B., 2016. Travel behaviour change in old age: The role of critical incidents in public transport. European Journal of Ageing 13, 75-83. https://doi.org/10.1007/s10433-015-0358-8

Susilo, Y.O., Cats, O., 2014. Exploring key determinants of travel satisfaction for multi-modal trips by different traveler groups. Transportation Research Part A: Policy and Practice 67, 366–380. https://doi.org/10.1016/j.tra.2014.08.002

Susilo, Y.O., Joewono, T.B., Santosa, W., 2010. An exploration of public transport users' attitudes and preferences towards various policies in Indonesia: Some preliminary results. Paper presented at the 8th Eastern Asia Society of Transportation Studies (EASTS) Conference, Surabaya. https://www.jstage.jst.go.jp/article/easts/8/0/8_0_1230/_pdf

Tarigan, A.K.M., Susilo, Y.O., Joewono, T.B., 2014. Segmentation of paratransit users based on service quality and travel behaviour in Bandung, Indonesia. Transportation Planning and Technology 37, 200–218. https://doi.org/10.1080/03081060.2013.870792

Tirachini, A., Quiroz, M., 2016. Evasión del pago en transporte público: Evidencia internacional y lecciones para Santiago. Documento de Trabajo, Departamento de Ingeniería Civil, Universidad de Chile, Julio de 2016. https://doi.org/10.13140/RG.2.1.4212.8884.

TMB, 2018. Transports Metropolitans de Barcelona, 2018. Private communication.

Tonglet, M. 2002. Consumer misbehaviour: An exploratory study of shoplifting. Journal of Consumer Behaviour 1, 336–354. https://doi.org/10.1002/cb.79.

Torres-Montoya, M., 2014. Tackling fare evasion in Transantiago: An integrated approach. Presented at the Transportation Research Board 93rd Annual Meeting. Transportation Research Board.

Transport Metropolitans de Barcelona, 2018. – TMB. http://www.tmb.cat/.

Troncoso, R., de Grange, L., 2017. Fare evasion in public transport: A time series approach. Transportation Research Part A: Policy and Practice 100, 311–318. https://doi.org/10.1016/j.tra.2017.04.029.

Tyrinopoulos, Y., Aifadopoulou, G., 2008. A complete methodology for the quality control of passenger services in the public transport business. European Transport\Trasporti Europei 38, 1–16.

Tyrinopoulos, Y., Antoniou, C., 2008. Public transit user satisfaction: Variability and policy implications. Transport Policy 15, 260–272. https://doi.org/10.1016/j.tranpol.2008.06.002.

Umbach, N., Naumann, K., Brandt, H., Kelava, A., 2017. Fitting nonlinear structural equation models in R with package nlsem. Journal of Statistical Software 77, 1-20. https://doi.org/10.18637/jss.v077.i07.

UNI 11098, 2003. "Linee guida per la rilevazione della soddisfazione del cliente e per la misurazione degli indicatori del relativo processo". http://qualitiamo.com/ambiti/UNI%2011098.html

van der Waerden, P.J.H.J., Timmermans, H.J.P., 2003. Key events and critical incidents influencing transport mode switching behavior - An exploratory study, Proceedings of the 82nd Annual Meeting of the Transportation Research Board.

van Lierop, D., Badami, M.G., El-Geneidy, A.M., 2018. What influences satisfaction and loyalty in public transport? A review of the literature. Transport Reviews 38, 52–72. https://doi.org/10.1080/01441647.2017.1298683

van Lierop, D., El-Geneidy, A., 2016. Enjoying loyalty: The relationship between service quality, customer satisfaction, and behavioral intentions in public transit. Research in Transportation Economics 59, 50-59.

Vargo, S.L., Lusch, R.F., 2007a. Service-dominant logic: Continuing the evolution. Journal of the Academy of Marketing Science 36, 1–10. https://doi.org/10.1007/s11747-007-00696.

Vargo, S.L., Lusch, R.F., 2007b. Why "service"? Journal of the Academy of Marketing Science 36, 25–38. https://doi.org/10.1007/s11747-007-0068-7.

Vrieze, S. I., 2012. Model selection and psychological theory: A discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). Psychological Methods 17, 228-243. https://doi.org/10.1037/a0027127.

Vyas, S., Kumaranayake, L., 2006. Constructing socio-economic status indices: How to use principal components analysis. Health Policy and Planning 21, 459–468. https://doi.org/10.1093/heapol/czl029

Wachs, M., 1976. Consumer attitudes toward transit service: An interpretive review. Journal of the American Institute of Planners 42, 96–104. https://doi.org/10.1080/01944367608977708

Webb, V., 2010. Customer loyalty in the public transportation context. Massachusetts Institute of Technology. http://hdl.handle.net/1721.1/58083

Winters, P.L., Cleland, F., Mierzejewski, E., Tucker, L., 2001. Assessing level of service equally across modes. White Paper, Contract No. BC353 RPWO#15. http://www.fdot.gov/research/completed_proj/summary_pl/fdot_bc353_15_rpt.pdf.

Wong, A., Sohal, A., 2003. A critical incident approach to the examination of customer relationship management in a retail chain an exploratory study. Qualitative Market Research: An International Journal 6, 248-262

Yaya, L.H.P., Fortià, M.F., Canals, C. S., Marimon, F., 2014. Service quality assessment of public transport and the implication role of demographic characteristics. Public Transport 7, 409-428. https://doi.org/10.1007/s12469-014-0099-7

Zeithaml, V.A., Berry, L.L., Parasuraman, A.,1996. The behavioral consequences of service quality. Journal of Marketing 60, 31-46. https://doi.org/10.2307/1251929.

Zhao, J., Webb, V., Shah, P., 2014. Customer loyalty differences between captive and choice transit riders. Transportation Research Record: Journal of the Transportation Research Board 2415, 80-88. https://doi.org/10.3141/2415-09.

A N N E X E S

ANNEX A: Covariance Matrix (Measurement Model, CFA)

Table 2-13: Covariance matrix for the measurement model, CFA

	Estimate	S.E.	Z-value
BL1			
BL2	0.725	0.012	61.61
BL3	0.778	0.007	114.17
BL4	0.628	0.011	59.51
BL5	0.702	0.011	66.26
BL6	0.643	0.012	55.35
SC1	0.631	0.010	61.54
SC2	0.689	0.010	65.74
SC3	0.605	0.007	83.52
SC4	0.284	0.010	29.77
BL2			
BL3	0.861	0.012	74.66
BL4	0.728	0.015	48.59
BL5	0.795	0.015	51.56
BL6	0.742	0.016	46.66
SC1	0.642	0.014	45.14
SC2	0.667	0.015	45.18
SC3	0.552	0.011	49.47
SC4	0.306	0.011	27.42
BL3			
BL4	**0.936**	0.010	89.40
BL5	0.802	0.011	74.19
BL6	0.762	0.012	64.09
SC1	0.696	0.010	66.74
SC2	0.682	0.011	63.85
SC3	0.537	0.008	68.05
SC4	0.317	0.010	30.74

	Estimate	S.E.	Z-value
BL4			
BL5	0.635	0.014	46.32
BL6	0.637	0.015	43.76
SC1	0.660	0.014	48.40
SC2	0.544	0.013	41.81
SC3	0.423	0.010	43.04
SC4	0.270	0.010	26.95
BL5			
BL6	0.898	0.016	55.89
SC1	0.621	0.013	46.77
SC2	0.709	0.014	50.98
SC3	0.532	0.010	52.33
SC4	0.337	0.011	29.82
BL6			
SC1	0.572	0.014	41.58
SC2	0.613	0.014	43.21
SC3	0.459	0.011	41.97
SC4	0.322	0.011	28.84
SC1			
SC2	0.880	0.016	56.60
SC3	0.662	0.010	66.69
SC4	0.273	0.010	27.20
SC2			
SC3	0.754	0.010	74.16
SC4	0.310	0.011	28.69
SC3			
SC4	0.222	0.008	26.77

ANNEX B: Mediation Analysis Scheme For Full SEM-MIMIC MODEL

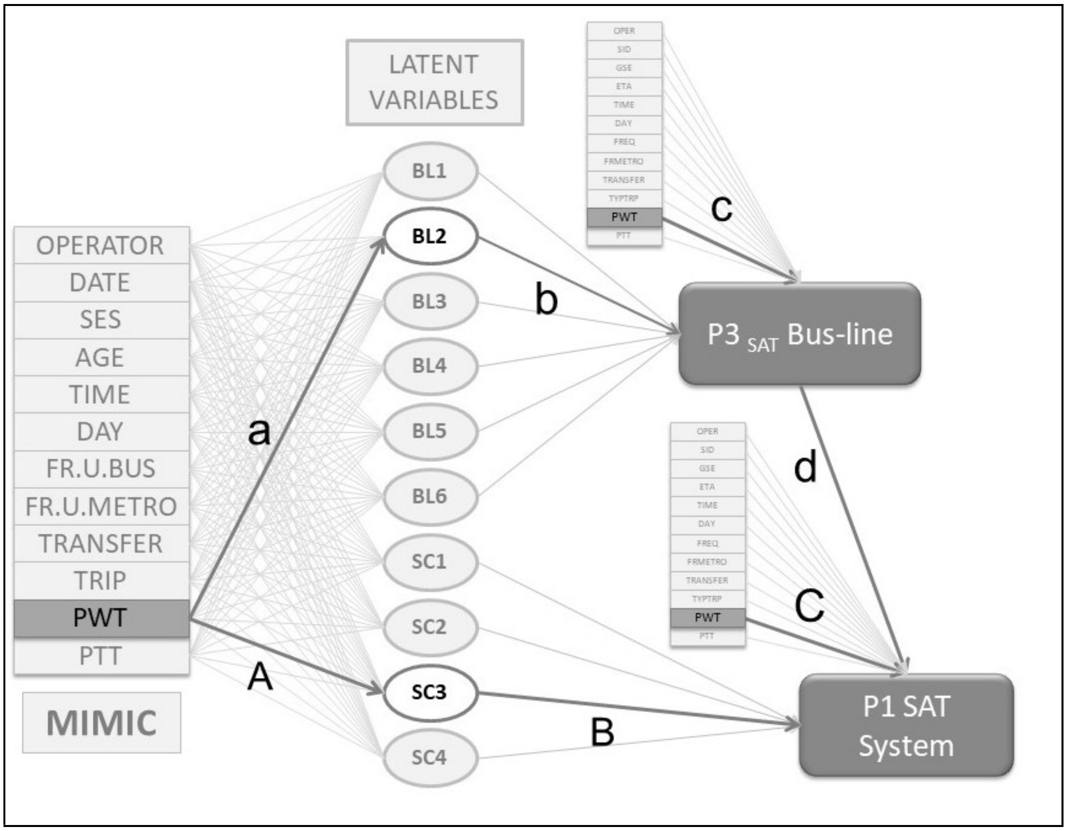

Figure 2-5: Mediation analysis scheme Full SEM-MIMIC Model. Effect of PWT through two latent constructs.

ANNEX C: Principal Component Analysis

Table 3-6: PCA for Critical Incident items

		CI1	CI23	CI4	CI5A	CI5B	CI6	CI7	CI8
C01	Safety of the journey (accidents)	**0.49**	-0.04	0.01	0.02	-0.05	0.06	0.01	0.04
C02	Security (thefts, harassments) on board	**0.62**	0.00	0.01	-0.01	0.03	-0.01	0.00	-0.01
C03	Security (thefts, harassments) at the station	**0.61**	0.03	-0.02	0.00	0.00	-0.02	-0.01	-0.02
C04	Cleanliness of the carriage	0.00	**0.37**	0.20	-0.05	0.10	0.02	0.04	0.01
C05	Cleanliness and general conditions of the seats	0.00	**0.37**	0.20	-0.05	0.10	0.02	0.03	-0.01
C06	Cleanliness of the toilets	0.00	**0.43**	0.06	0.00	0.04	0.05	0.03	0.01
C07	Cleanliness at station	0.01	**0.53**	-0.08	0.01	-0.03	-0.02	-0.01	0.00
C08	Station maintenance and decorum	0.00	**0.51**	-0.11	0.02	-0.07	-0.02	-0.03	0.00
C09	Overcrowding	0.01	-0.07	**0.52**	-0.08	0.01	0.05	0.02	0.01
C10	Temperature on board	0.00	-0.01	**0.57**	0.01	-0.02	-0.01	-0.01	-0.01
C11	Windows and doors working on board	-0.01	0.03	**0.53**	0.10	-0.12	-0.05	-0.04	0.01
C13	Timetable (frequency and daily distribution)	0.00	0.03	-0.05	**-0.47**	-0.21	0.03	0.05	0.01
C14	Train punctuality	0.00	-0.01	0.09	**-0.56**	0.06	0.01	-0.02	0.00
C15	Train regularity (absence of cancellations)	0.00	-0.01	-0.08	**-0.58**	-0.10	-0.03	-0.05	0.00
C16	Price integration with other public transport	-0.02	0.01	0.04	-0.08	**-0.60**	-0.01	-0.01	-0.01
C17	Distribution of stations in the region	0.01	0.00	0.03	-0.04	**-0.63**	0.00	0.04	0.00
C18	Parking at the station of departure	-0.01	-0.02	0.04	-0.07	0.12	**0.65**	0.04	-0.01
C19	Bike on board	0.01	-0.01	-0.07	0.06	-0.14	**0.56**	-0.01	0.02
C20	Facilities for PRM	0.00	0.06	-0.02	0.07	-0.06	**0.50**	-0.08	-0.02
C21	Information at station	0.01	0.00	0.00	-0.17	0.17	-0.01	**-0.50**	0.00
C22	Information on board	0.00	0.00	0.00	-0.10	0.13	0.00	**-0.54**	0.00
C23	Communication with operator (complaints, info requests)	0.00	0.00	0.03	0.14	-0.14	0.01	**-0.48**	0.01
C24	Information on connections with other transports	0.00	0.00	0.02	0.15	-0.21	0.01	**-0.45**	0.00
C25	Courtesy and competence of the personnel on board	0.03	-0.02	0.02	-0.02	0.00	0.00	0.03	**0.59**
C26	Frequency and attentiveness of the ticket check	-0.01	0.01	0.02	0.02	0.03	-0.01	-0.03	**0.55**
C27	Courtesy and competence of the personnel at the station	-0.02	0.01	-0.03	0.00	-0.01	0.00	-0.01	**0.60**

Table 3-7: PCA for Satisfaction items

		CS1	CS2	CS3	CS4	CS5	CS6	CS7	CS8	CS9
S01	Safety of the journey (accidents)	**-0.51**	-0.05	-0.02	0.06	-0.08	0.01	-0.06	0.03	0.00
S02	Security (thefts, harassments) on board	**-0.62**	0.04	-0.05	-0.01	0.02	-0.01	0.01	0.00	-0.01
S03	Security (thefts, harassments) at the station	**-0.59**	0.02	0.03	-0.03	0.04	-0.01	0.04	-0.02	0.01
S04	Cleanliness of the carriage	-0.05	**0.50**	0.12	0.02	-0.01	0.02	-0.07	0.02	-0.04
S05	Cleanliness and general conditions of the seats	-0.02	**0.51**	0.11	0.03	-0.01	0.02	-0.07	0.02	-0.04
S06	Cleanliness of the toilets	0.02	**0.50**	0.12	0.00	0.01	0.04	-0.02	0.01	-0.05
S07	Cleanliness at station	0.01	0.12	**0.62**	-0.01	0.00	-0.03	0.02	-0.01	0.03
S08	Station maintenance and decorum	0.02	0.06	**0.64**	-0.01	0.00	-0.04	0.05	-0.02	0.04
S09	Overcrowding	0.00	-0.01	-0.03	**0.60**	-0.01	-0.01	-0.01	0.00	0.02
S10	Temperature on board	0.01	0.02	-0.01	**0.61**	0.03	0.01	0.04	-0.02	0.00
S11	Windows and doors working on board	-0.01	0.04	0.05	**0.48**	-0.02	0.02	0.01	0.02	-0.01
S13	Timetable (frequency and daily distribution)	0.00	0.01	-0.01	0.00	**-0.47**	-0.01	-0.04	-0.02	0.01
S14	Train punctuality	0.05	0.17	-0.17	0.05	**-0.43**	-0.08	0.04	0.01	-0.01
S15	Train regularity (absence of cancellations)	0.01	0.07	-0.08	0.02	**-0.46**	-0.07	0.04	0.00	-0.02
S16	Price integration with other public transport	0.00	-0.09	0.12	-0.08	**-0.44**	0.07	0.02	0.00	-0.01
S17	Distribution of stations in the region	-0.04	-0.19	0.17	-0.01	**-0.42**	0.11	-0.05	0.02	0.00
S18	Parking at the station of departure	0.00	-0.02	-0.03	0.07	0.00	**0.57**	0.00	-0.04	0.03
S19	Bike on board	0.00	0.00	-0.03	-0.02	-0.02	**0.63**	-0.02	0.03	0.01
S20	Facilities for PRM	0.01	0.18	-0.05	-0.04	0.05	**0.48**	0.11	-0.03	-0.02
S21	Information at station	-0.03	-0.08	0.05	0.05	-0.03	0.00	**0.46**	0.00	0.00
S22	Information on board	-0.02	-0.02	0.00	0.06	0.01	0.00	**0.49**	0.03	-0.04
S23	Communication with the operator	0.00	-0.04	0.06	0.00	0.00	0.00	**0.49**	0.02	-0.02
S24	Information on connections with other transports	0.01	-0.07	0.09	-0.02	0.00	0.05	**0.49**	0.02	-0.04
S25	Courtesy of the personnel on board	-0.03	-0.02	0.02	0.01	-0.05	0.01	0.00	**0.54**	0.02
S26	Frequency and attentiveness of the ticket check	0.02	0.07	-0.08	0.00	0.03	-0.01	0.00	**0.62**	-0.03
S27	Courtesy of the personnel at the station	0.01	-0.03	0.05	-0.02	0.01	0.00	0.04	**0.56**	0.06
E32	Information under abnormal conditions	-0.01	0.17	-0.19	-0.07	-0.02	-0.07	0.21	-0.06	**0.44**
E34	Purchasing ticket system	0.00	-0.15	0.13	0.04	0.02	0.03	-0.07	0.03	**0.72**
E35	Interventions for improving service quality	0.01	0.24	-0.14	-0.03	-0.03	0.00	0.01	0.00	**0.52**

ANNEX D: SEM Comparison

Table 3-8: SEM Explained variance indices (R^2)

	SEM-numeric	SEM-ordinal
CS1 Safety	0.113	0.662
CS2 Cleanliness on Board	0.197	0.816
CS3 Cleanliness at Station	0.123	0.625
CS4 Comfort	0.240	0.881
CS5 Reliability and Accessibility	0.185	0.641
CS6 Additional Services	0.095	0.648
CS7 Information	0.164	0.799
CS8 Personnel	0.155	0.730
CS9 Added-Value Services	0.148	
OVS Overall Satisfaction	**0.505**	**0.771**
E38 Loyalty	**0.158**	**0.379**

ANNEX E: Threshold Parameters

Table 3-9: SEM-MIMIC ordinal: Threshold Parameter estimates CI and Satisfaction (S01, S02, S30, E38)

	Estimate	S.E.	Z-value	Std.Coeff.
C01|t1	1.914	0.032	60.60	1.914
C02|t1	1.851	0.029	63.51	1.851
C03|t1	1.855	0.030	62.80	1.855
C04|t1	1.263	0.020	61.64	1.263
C05|t1	1.223	0.020	60.25	1.223
C06|t1	1.339	0.021	62.90	1.339
C07|t1	1.596	0.024	67.33	1.596
C08|t1	1.719	0.025	68.33	1.719
C09|t1	1.051	0.020	52.94	1.051
C10|t1	1.258	0.021	60.76	1.258
C11|t1	1.447	0.023	64.24	1.447
C13|t1	1.550	0.024	63.92	1.550
C14|t1	1.062	0.020	54.40	1.062
C15|t1	1.455	0.022	64.99	1.455
C18|t1	1.715	0.026	65.27	1.715
C19|t1	2.112	0.036	59.22	2.112
C20|t1	1.982	0.031	63.54	1.982
C21|t1	1.630	0.024	66.80	1.630
C22|t1	1.689	0.026	65.71	1.689
C23|t1	1.856	0.029	62.99	1.856
C24|t1	1.989	0.032	62.54	1.989
C25|t1	1.830	0.029	62.83	1.830
C26|t1	1.810	0.028	63.96	1.810
C27|t1	1.884	0.030	62.65	1.884
S01|t1	-2.106	0.016	-127.79	-2.083
S01|t2	-1.874	0.016	-120.26	-1.874
S01|t3	-1.606	0.015	-107.14	-1.606
S01|t4	-1.388	0.015	-95.01	-1.373
S01|t5	-0.957	0.014	-67.73	-0.946
S01|t6	-0.471	0.014	-33.66	-0.465
S01|t7	-0.053	0.014	-3.81	-0.052
S01|t8	0.552	0.014	39.42	0.546
S01|t9	1.087	0.014	75.86	1.075

Table 3-9: SEM-MIMIC ordinal: Threshold Parameter estimates CI and Satisfaction (S01, S02, S30, E38) (continued)

S02|t1	-1.991	0.016	-128.34	-1.962
S02|t2	-1.712	0.015	-114.51	-1.687
S02|t3	-1.446	0.015	-98.92	-1.426
S02|t4	-1.184	0.014	-82.33	-1.167
S02|t5	-0.743	0.014	-52.67	-0.732
S02|t6	-0.242	0.014	-17.30	-0.239
S02|t7	0.194	0.014	13.85	0.191
S02|t8	0.756	0.014	53.29	0.745
S02|t9	1.271	0.015	86.48	1.252
S30|t1	-0.287	0.016	-17.98	-0.287
S30|t2	1.405	0.017	83.42	1.405
E38|t1	-0.399	0.017	-23.41	**-0.385**

ANNEX F: SEM-MIMIC ORDINAL: Satisfaction latent constructs regressions

Table 3-10: SEM-MIMIC ordinal: Satisfaction latent constructs regressions (continued) (CS3-CS4)

CS3 CLEANLINESS-STATION	**(R^2 = 0.644)**	Estimate	S.E.	Z-value	Std.Coeff.
CRITICAL INCIDENT	CI23	-1.687	0.010	-169.87	**-0.782**
SURVEY DATE	SPRING.2011	-0.168	0.021	-8.15	-0.038
	AUTUMN.2011	-0.468	0.020	-23.89	-0.106
	SPRING.2012	-0.186	0.020	-9.37	-0.042
	SPRING.2013	-0.304	0.022	-13.81	-0.056
	AUTUMN.2013	-0.366	0.023	-15.61	-0.062
	SPRING.2014	-0.322	0.026	-12.63	-0.061
TIME OF DAY	PM.PEAK	-0.121	0.015	-8.01	-0.028
ACCESS MODE	ACCESS.CAR	-0.113	0.015	-7.63	-0.029
SERVICE	SUBURBAN	0.098	0.012	8.30	0.029
LINE	NORTHERN.REGION	-0.444	0.014	-31.01	-0.107
	WESTERN.REGION	-0.498	0.023	-21.23	-0.073
	SOUTHERN.REGION	-0.278	0.018	-15.77	-0.054
TICKET TYPE	TICKET.ONE-WAY	0.202	0.016	12.91	0.049
TYPE OF USER	COMMUTER.STUDENT	-0.079	0.017	-4.77	-0.021
GENDER	FEMALE	-0.214	0.011	-18.81	-0.063
AGE	AGE>65	0.100	0.030	3.32	0.011
EDUCATION LEVEL	ELEM.SCHOOL.DEGREE	0.060	0.017	3.52	0.012
	UNIVERSITY.DEGREE	-0.038	0.013	-2.83	-0.010

Table 3-10: SEM-MIMIC ordinal: Satisfaction latent constructs regressions (continued) (CS3-CS4) (continued)

CS4 COMFORT	**(R^2 = 0.905)**	Estimate	S.E.	Z-value	Std.Coeff.
CRITICAL INCIDENT	CI4	-1.282	0.009	-136.13	**-0.924**
SURVEY DATE	SPRING.2011	-0.074	0.010	-7.24	-0.037
	AUTUMN.2011	-0.175	0.009	-19.86	-0.088
	SPRING.2012	-0.103	0.010	-10.76	-0.052
	SPRING.2013	-0.140	0.010	-13.76	-0.058
	AUTUMN.2013	-0.087	0.010	-8.32	-0.033
	SPRING.2014	-0.132	0.012	-11.42	-0.056
TIME OF DAY	AM.PEAK	0.032	0.008	4.07	0.016
	PM.PEAK	-0.067	0.007	-9.71	-0.034
ACCESS MODE	ACCESS.CAR	-0.041	0.007	-6.02	-0.024
	ACCESS.CYCLE	-0.057	0.011	-5.38	-0.024
SERVICE	SUBURBAN	0.037	0.005	6.86	0.024
LINE	NORTHERN.REGION	-0.135	0.007	-20.82	-0.073
	WESTERN.REGION	-0.230	0.011	-21.35	-0.076
	SOUTHERN.REGION	-0.126	0.008	-15.83	-0.055
TICKET TYPE	TICKET.ONE-WAY	0.252	0.007	34.87	0.138
TYPE OF USER	COMMUTER.STUDENT	-0.107	0.008	-14.14	-0.063
GENDER	FEMALE	-0.116	0.005	-22.37	-0.077
AGE	AGE>65	0.059	0.015	4.04	0.014
MONTHLY NET INCOME	<1000	-0.028	0.009	-3.19	-0.013
	1001.1500	-0.061	0.008	-7.67	-0.033
	1501.3000	-0.033	0.009	-3.80	-0.016
	>3000	0.045	0.013	3.40	0.012
EDUCATION LEVEL	ELEM.SCHOOL.DEGREE	0.045	0.008	5.75	0.020
	UNIVERSITY.DEGREE	-0.053	0.006	-8.59	-0.032

Table 3-11: SEM-MIMIC ordinal: Satisfaction latent constructs regressions (continued) (CS5-CS6)

CS5 RELIAB./ACCESSIBILITY	**(R^2 = 0.694)**	Estimate	S.E.	Z-value	Std.Coeff.
CRITICAL INCIDENT	CI5A	-1.169	0.010	-118.27	**-0.809**
SURVEY DATE	AUTUMN.2011	-0.177	0.009	-20.55	-0.090
	SPRING.2013	-0.133	0.010	-13.21	-0.056
	AUTUMN.2013	-0.167	0.010	-16.46	-0.064
	SPRING.2014	-0.184	0.011	-16.40	-0.079
TIME OF DAY	PM.PEAK	-0.053	0.007	-7.98	-0.028
ACCESS MODE	ACCESS.CAR	-0.032	0.007	-4.82	-0.018
	ACCESS.CYCLE	-0.038	0.011	-3.65	-0.016
LINE	NORTHERN.REGION	-0.068	0.006	-10.75	-0.037
	WESTERN.REGION	-0.182	0.010	-17.48	-0.060
	SOUTHERN.REGION	-0.085	0.008	-11.15	-0.037
TICKET TYPE	TICKET.ONE-WAY	0.158	0.007	22.69	0.087
TYPE OF USER	COMMUTER.STUDENT	-0.053	0.007	-7.28	-0.031
GENDER	FEMALE	-0.116	0.005	-23.01	-0.077
AGE	AGE>65	0.072	0.014	5.14	0.018
MONTHLY NET INCOME	1501.3000	0.073	0.008	8.67	0.036
	>3000	0.121	0.013	9.51	0.033
EDUCATION LEVEL	ELEM.SCHOOL.DEGREE	0.026	0.008	3.32	0.012

CS6	**(R^2 = 0.726)**	Estimate	S.E.	Z-value	Std.Coeff.
CRITICAL INCIDENT	CI6	-1.086	0.013	-84.97	**-0.837**
SURVEY DATE	AUTUMN.2011	-0.173	0.008	-20.49	-0.097
	SPRING.2013	-0.116	0.009	-12.28	-0.054
	AUTUMN.2013	-0.133	0.010	-13.31	-0.057
	SPRING.2014	-0.139	0.011	-12.46	-0.066
TIME OF DAY	AM.PEAK	0.024	0.007	3.50	0.013
ACCESS MODE	ACCESS.CAR	-0.036	0.006	-5.64	-0.023
	ACCESS.CYCLE	-0.075	0.010	-7.66	-0.035
	ACCESS.PT	-0.031	0.006	-5.01	-0.021
LINE	NORTHERN.REGION	-0.062	0.006	-10.08	-0.037
	WESTERN.REGION	-0.086	0.010	-8.32	-0.032
	SOUTHERN.REGION	-0.041	0.008	-5.39	-0.020
TICKET TYPE	TICKET.ONE-WAY	0.066	0.007	9.67	0.040
GENDER	FEMALE	-0.103	0.005	-20.61	-0.076
MONTHLY NET INCOME	1001.1500	-0.032	0.008	-4.24	-0.019
EDUCATION LEVEL	ELEM.SCHOOL.DEGREE	0.061	0.008	7.98	0.031
	UNIVERSITY.DEGREE	-0.066	0.006	-11.30	-0.044

Table 3-12: SEM-MIMIC ordinal: Satisfaction latent constructs regressions (continued) (CS7-CS8)

CS7 INFORMATION	**(R^2 = 0.844)**	Estimate	S.E.	Z-value	Std.Coeff.
CRITICAL INCIDENT	CI7	-1.255	0.009	-146.11	**-0.906**
SURVEY DATE	SPRING.2011	-0.049	0.011	-4.51	-0.021
	AUTUMN.2011	-0.176	0.010	-17.32	-0.076
	SPRING.2012	-0.055	0.010	-5.37	-0.024
	SPRING.2013	-0.134	0.011	-12.14	-0.048
	AUTUMN.2013	-0.152	0.012	-12.54	-0.050
	SPRING.2014	-0.138	0.013	-10.39	-0.051
TIME OF DAY	PM.PEAK	-0.062	0.008	-7.91	-0.027
ACCESS MODE	ACCESS.CAR	-0.026	0.008	-3.41	-0.013
LINE	NORTHERN.REGION	-0.090	0.007	-12.07	-0.042
	WESTERN.REGION	-0.158	0.012	-12.81	-0.045
	SOUTHERN.REGION	-0.076	0.009	-8.43	-0.028
USE FREQUENCY	WEEKLY	0.033	0.008	4.01	0.014
TICKET TYPE	TICKET.ONE-WAY	0.186	0.008	22.95	0.087
TYPE OF USER	COMMUTER.STUDENT	-0.051	0.009	-5.76	-0.025
GENDER	FEMALE	-0.095	0.006	-15.95	-0.054
AGE	AGE>65	0.060	0.016	3.79	0.013
MONTHLY NET INCOME	1001.1500	-0.049	0.009	-5.41	-0.023
	1501.3000	-0.041	0.010	-4.17	-0.017
EDUCATION LEVEL	ELEM.SCHOOL.DEGREE	0.032	0.009	3.51	0.012
	UNIVERSITY.DEGREE	-0.079	0.007	-11.41	-0.041

Table 3-12: SEM-MIMIC ordinal: Satisfaction latent constructs regressions (continued) (CS7-CS8) (continued)

CS8 PERSONNEL	**(R^2 = 0.790)**	Estimate	S.E.	Z-value	Std.Coeff.
CRITICAL INCIDENT	CI8	-1.237	0.011	-112.23	**-0.874**
SURVEY DATE	AUTUMN.2011	-0.163	0.011	-15.10	-0.069
	SPRING.2013	-0.086	0.012	-7.12	-0.030
	AUTUMN.2013	-0.058	0.013	-4.61	-0.019
	SPRING.2014	-0.098	0.014	-7.03	-0.035
TIME OF DAY	PM.PEAK	-0.049	0.008	-5.95	-0.021
ACCESS MODE	ACCESS.CAR	-0.047	0.008	-5.83	-0.023
SERVICE	SUBURBAN	-0.048	0.006	-7.46	-0.027
LINE	NORTHERN.REGION	-0.059	0.008	-7.53	-0.027
	WESTERN.REGION	-0.199	0.013	-15.31	-0.055
	SOUTHERN.REGION	-0.155	0.009	-16.42	-0.058
USE FREQUENCY	WEEKLY	0.046	0.009	5.12	0.020
TICKET TYPE	TICKET.ONE-WAY	0.177	0.009	20.60	0.082
TYPE OF USER	COMMUTER.STUDENT	-0.083	0.009	-9.21	-0.041
GENDER	FEMALE	-0.050	0.006	-8.00	-0.028
AGE	AGE>65	0.159	0.017	9.42	0.033
MONTHLY NET INCOME	<1000	0.041	0.010	4.05	0.017
	1501.3000	0.041	0.010	4.01	0.017
	>3000	0.062	0.016	3.97	0.014

ANNEX G: Occupation Levels Shown During The Surveys.

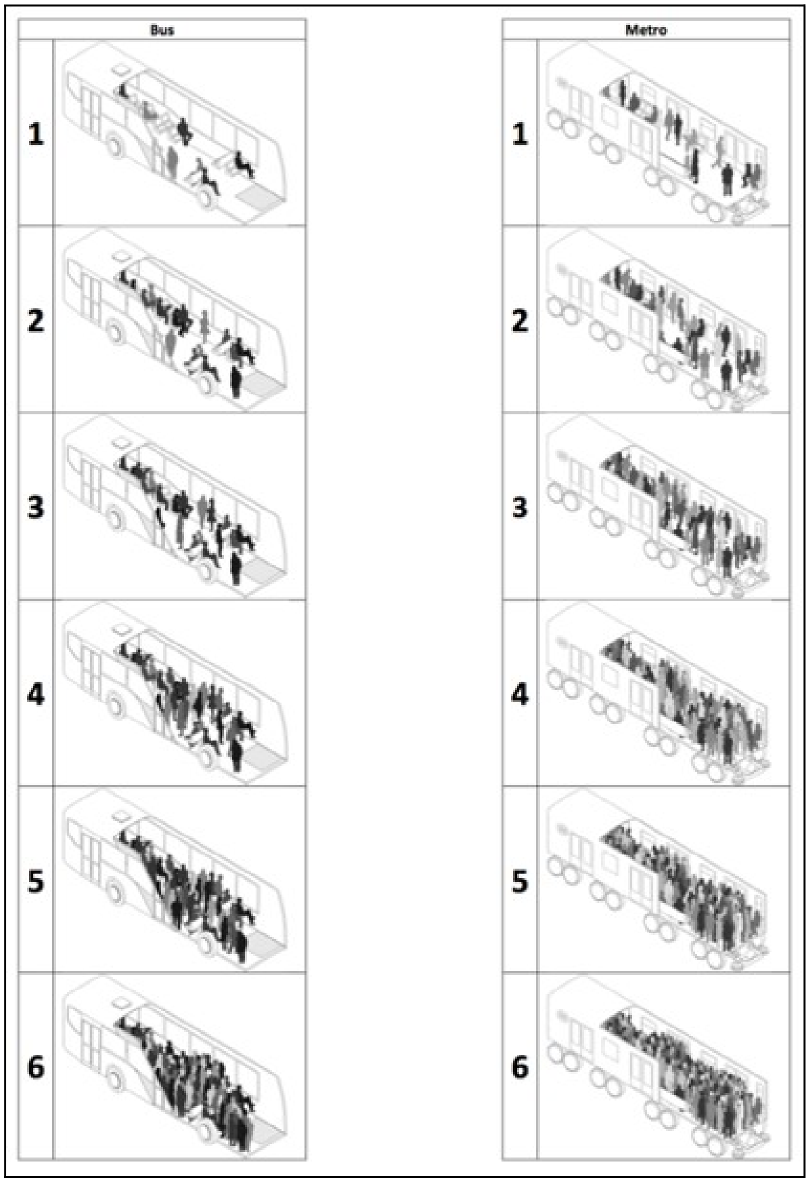

Figure 8-2: Visual occupation levels. Source: Batarce et al. (2016)

ANNEX H: Measurement System: SEM Metro

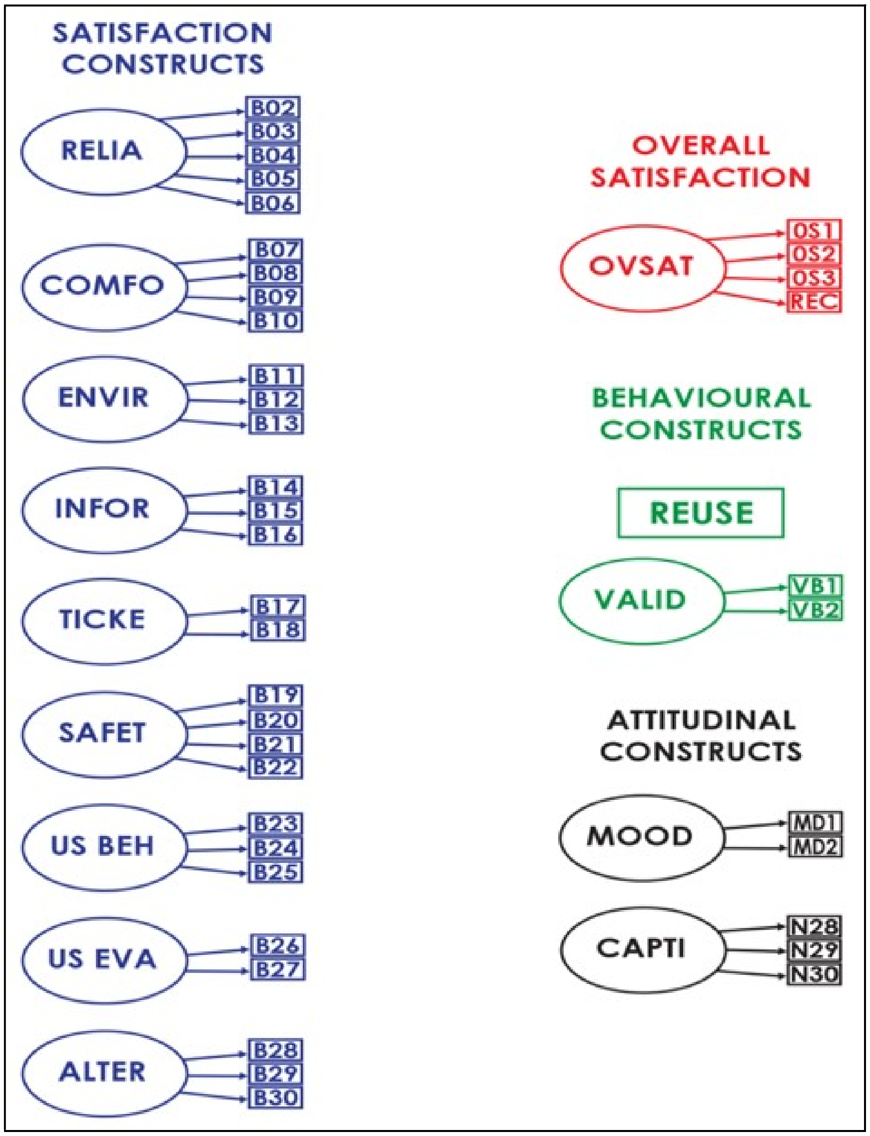

Figure 8-3: Measurement system for the SEM Metro

ANNEX I: SEM Measurement Models

Table 8-14: SEM Measurement model: Metro

Measurement	Estimate	S.E.	Z-value	St.Coef.	R^2
RELIA					
B02	1.00			0.75	0.56
B03	1.03	0.03	33.01	0.77	0.60
B04	0.97	0.03	30.69	0.73	0.53
B05	1.00	0.03	30.62	0.75	0.56
B06	1.03	0.03	30.84	0.77	0.60
COMFO					
B07	1.00			0.88	0.77
B08	0.99	0.02	61.84	0.87	0.75
B09	0.95	0.02	52.90	0.83	0.69
B10	0.96	0.02	54.56	0.84	0.70
ENVIR					
B11	1.00			0.75	0.56
B12	1.02	0.04	27.00	0.76	0.58
B13	1.06	0.04	26.79	0.79	0.62
INFOR					
B14	1.00			0.64	0.40
B15	1.17	0.07	17.92	0.75	0.56
B16	1.22	0.06	19.32	0.77	0.60
TICKE					
B17	0.84	0.02	54.52	0.84	0.71
B18	0.92	0.01	67.65	0.92	0.84
SAFET					
B19	1.00			0.76	0.58
B20	1.08	0.03	32.52	0.82	0.68
B21	1.11	0.04	29.40	0.85	0.72
USBEH					
B23	1.00			0.86	0.74
B24	1.08	0.02	53.34	0.93	0.87
B25	0.98	0.02	56.13	0.84	0.71
USEVA					
B27	1.00			1.00	1.00
B26	0.97	0.00	352.64	0.97	0.94
ALTER					
B28	1.00			0.31	0.10
B29	2.59	0.32	8.21	0.81	0.65
B30	2.56	0.32	7.94	0.80	0.64

Table 8-14: SEM Measurement model: Metro (continued)

OVSAT					
OS1	1.00			0.78	0.61
OS2	1.17	0.02	53.04	0.91	0.83
OS3	1.13	0.02	51.68	0.88	0.78
REC	1.10	0.02	49.83	0.86	0.73
MOOD					
MD1	0.98	0.04	26.87	0.98	0.95
MD2	0.89	0.03	26.77	0.89	0.80
VALID					
VB1	0.64	0.06	10.36	0.77	0.60
VB2	0.77	0.07	10.66	0.93	0.86
CAPTI					
N30	1.00			1.00	1.00
N29	0.98	0.01	179.99	0.98	0.97
N28	0.88	0.02	47.59	0.88	0.78

Table 8-15: SEM Measurement model: bus

Measurement	Estimate	S.E.	Z-value	St.Coef.	R^2
RELIA					
B02	1.00			0.69	0.48
B03	1.10	0.04	30.66	0.76	0.58
B04	1.05	0.04	28.66	0.73	0.53
B05	1.08	0.04	29.91	0.75	0.56
B06	1.19	0.04	30.84	0.82	0.67
COMFO					
B07	1.00			0.46	0.22
B08	1.23	0.10	12.90	0.57	0.33
B09	1.66	0.11	14.58	0.77	0.59
ENVIR					
B10	1.00			0.84	0.70
B11	0.94	0.03	35.13	0.78	0.61
B12	0.99	0.03	36.25	0.83	0.69
INFOR					
B13	1.00			0.52	0.27
B14	1.23	0.09	13.26	0.64	0.41
B15	1.53	0.11	14.03	0.80	0.64
TICKE					
B16	0.87	0.02	41.20	0.87	0.75
B17	0.88	0.02	41.53	0.88	0.78
SAFET					
B18	1.00			0.87	0.76
B19	1.03	0.02	65.28	0.90	0.82
B20	0.96	0.02	60.86	0.84	0.71
B21	0.99	0.02	62.03	0.86	0.74
USBEH					
B22	1.00			0.84	0.70
B23	1.09	0.02	46.51	0.92	0.84
B24	0.98	0.02	45.67	0.82	0.68
USEVA					
B25	1.00			1.00	1.00
B26	0.93	0.01	148.33	0.93	0.86
ALTER					
B27	1.00			0.51	0.26
B28	1.64	0.10	15.91	0.84	0.71
B29	1.58	0.10	16.61	0.81	0.65
OVSAT					
OS1	1.00			0.82	0.67
OS2	1.09	0.02	61.39	0.89	0.78
OS3	0.99	0.02	49.09	0.80	0.65
REC	1.01	0.02	50.43	0.83	0.68

Table 8-15: SEM Measurement model: bus (continued)

MOOD					
MD1	0.98	0.03	29.08	0.98	0.96
MD2	0.87	0.03	27.66	0.87	0.76
VALID					
VB1	0.59	0.03	17.52	0.69	0.47
VB2	0.80	0.05	15.08	0.93	0.87
CAPTI					
N29	1.00			1.00	1.00
N28	0.99	0.00	224.62	0.99	0.97
N27	0.78	0.03	26.40	0.78	0.61