The underlying effect of public transport reliability on users' satisfaction

Soza-Parra, Jaime; Raveau, Sebastián; Muñoz, Juan Carlos; Cats, Oded

DOI
10.1016/j.tra.2019.06.004

Publication date
2019

Document Version
Final published version

Published in
Transportation Research Part A: Policy and Practice

Transportation Research Part A 126 (2019) 83–93

Contents lists available at ScienceDirect

Transportation Research Part A

journal homepage: www.elsevier.com/locate/tra

The underlying effect of public transport reliability on users' satisfaction

Jaime Soza-Parra[a,*], Sebastián Raveau[a], Juan Carlos Muñoz[a], Oded Cats[b]

[a] Department of Transport Engineering and Logistic, Pontificia Universidad Católica de Chile, Chile
[b] Transportation & Planning Department, Delft University of Technology, the Netherlands

ARTICLE INFO

Keywords:
Public transport satisfaction
Headway irregularity
Crowding level

ABSTRACT

Service reliability has an important impact on the satisfaction stated by public transport users with the service they receive. The main source of unreliability is found in headway variance, which also affects waiting times and distributes passengers unevenly across vehicles. However, it is still unclear how headway irregularity, with its impact in waiting, crowdedness and reliability, affect travellers' service satisfaction. Different stated preference studies have identified non-linear impacts produced by overcrowding. However, none of these studies is directly related to users' satisfaction evaluation. In this study, we investigate the existence of this non-linearity in users' satisfaction caused by both the crowding level and the number of denied boardings through a post-service satisfaction survey of bus and metro users. An Ordered Logit Model was estimated, accounting for sample heteroscedasticity and preference heterogeneity. Overall, there is a significant and negative perception of the bus mode, keeping all other attributes equal. For users under 35 years old, comfort experienced almost always plays an important role in service satisfaction, while for those over 35 years old women are significantly more sensitive to this attribute. Most important, crowding has a negative and non-linear impact on how passengers evaluate their travel satisfaction. Using a Likert-type scale, this curve is convex. This relationship between crowding and satisfaction might bias service planning and delivery if performance indicators associated to service are not properly weighted by the number of passengers served. Improving level of service indicators in this direction might provide public transport agencies a clearer and more accurate perception of the actual users' experience.

1. Introduction

To achieve sustainable development, cities need its citizens to use public transport. This is easier when citizens have a positive feeling about their public transport system, which is understood as satisfaction. Within high-frequency public transport, travellers seek and highly value a trip with four fundamental operational attributes: speed, short waits, high transport capacity and reliability (Delgado et al., 2016; Redman et al., 2013). This reliability is related to the variability of the level of service experienced by a user making the same trip in different days. The relation between satisfaction and the first three trip attributes has been widely studied, but the relation with reliability has not been empirically underpinned. One possible explanation of this is that reliability is not perceived directly in any travel experience; a trip needs to be completed repeatedly for passengers to perceive an alternative's reliability. However, the lack of reliability affects other attributes which are perceived directly. Thus, the objective of this article is to

* Corresponding author.
E-mail address: jaime.soza@uc.cl (J. Soza-Parra).

https://doi.org/10.1016/j.tra.2019.06.004
Received 28 January 2019; Received in revised form 2 May 2019; Accepted 4 June 2019
0965-8564/

estimate the effect of metro and bus service lack of reliability effects on passengers' evaluation of the quality of service experienced irrespective of mode.

An element that strongly influences the reliability of a public transport service is its headway variance. This variability has a strong impact on users' satisfaction. For example, some studies have shown in Granada (Spain; de Oña et al., 2016), Calgary (Canada; Habib et al., 2011) and Santiago (Chile; DTPM, 2016) that headway regularity along with sufficient frequency was part of the core of public transport quality attributes. Unfortunately, the inherent variability in demand patterns and travel times causes headway instability leading to the well-known phenomenon of vehicle bunching. Headway variability has several harmful effects on travellers when compared with the same frequency being offered under regular headways. Among the most direct effects are an increase in average waiting time and in its variability, and comfort deterioration, since the demand is not homogeneously distributed among vehicles, causing more travellers to experience crowded vehicles than empty ones (Delgado et al., 2016).

To understand how these effects alter travellers behaviour, several stated and revealed preference studies reported in the literature have provided a direct monetary value for travel and waiting time induced by service unreliability (de Ortúzar and Willumsen, 2011). However, it is unclear what is the best methodology for valuing experienced comfort in public transport. Recently, different studies have been conducted in order to understand how overcrowding levels affect travellers' behaviour (Batarce et al., 2015; Cats et al., 2016; Kim et al., 2015; Li & Hensher, 2011; Tirachini et al., 2013, 2016). For instance, Batarce et al. (2016) found that the value of time of a user experiencing an overcrowded situation (i.e. six standing passengers per square metre) is 2.5 times larger than the value of time of empty seats available. The authors identify a non-linear relation between the value of travel time and the level of crowdedness the travellers suffered.

Still, it is unclear how different crowding levels, caused by headway irregularity in a high frequency context, and the uncertainty due to unknown waiting times affect travellers' service satisfaction. In this study, we analyse the relationship between users' satisfaction and the underlying effects of an irregular operation: both the crowding level experienced and the number of denied boardings, exploring whether these relations exhibit non-linear patterns.

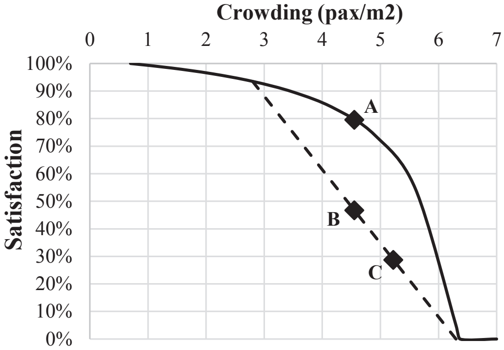
Public transport satisfaction has been studied extensively in the literature, focusing in its definition, its evolution over time, and its explanatory variables (Abenoza et al., 2017, 2018; Allen et al., 2018; Cats et al., 2015; De Oña and De Oña, 2014; Hensher et al., 2003; Tyrinopoulos & Antoniou, 2008). There is evidence to suggest that users value public transport service reliability the most over any other variable (Allen et al., 2018). Thus, it is especially important to unravel how service attributes caused by poor reliability (e.g. variations in on-board crowding) impact the overall satisfaction.

Instead of explaining the average satisfaction evaluation value by different attributes, we aim in this study to estimate the impact associated with each of the values within the range of satisfaction scores. To this end, we estimate an Ordinal Logit model (McCullagh, 1980). One important characteristic of this model is the possibility to estimate the threshold associated with moving between consecutive scale levels rather than implying that they are all equal.

This study is structured as follows. Section 2 explains the motivation behind the idea of non-linear interaction between travel attributes and passengers' satisfaction. Section 3 describes the survey carried out and the methodology used to process the data. Section 4 shows the main results for the Ordered Logit model while Section 5 shows the satisfaction evaluation analysis. Finally, Section 6 presents our main conclusions, their potential implications and provides some guidelines for further research on public transport satisfaction matters.

2. Motivation

Let us assume there is a non-linear relationship between the vehicle load during a trip and the satisfaction of a user experiencing it. It is reasonable to assume that it is expected that the impact of an extra passenger onboard on the rest of the passengers inside the vehicle is not constant as it should depend on the current load level. One well-founded hypothesis is that this curve is concave, as the marginal rate of substitution between crowding and in-vehicle travel time (i.e. crowding multiplier) obtained in different discrete choice experiments (Batarce et al., 2015; Liu & Wen, 2016; Tirachini et al., 2017; Wardman & Whelan, 2011; Yap et al., 2018) increases. It is important to emphasize that this concavity might not hold when analysing the effect of crowding and satisfaction, as



		Average crowding	Average atisfaction
A	Perfectly regular	4.5 pax/m^2	79.6%
B	Irregular	4.5 pax/m^2	47.0%
C	Perceived	5.2 pax/m^2	28.9%

Fig. 1. Satisfaction decline due to headway irregularity.

there is no evidence suggesting a direct relationship between the value of time and satisfaction. Fig. 1 illustrates this relation in which service satisfaction drops non-linearly with increasing vehicle occupancy.

The impact of this non-linear relation on the level of service perceived by users is not understood completely if the service satisfaction is not analysed. We use Fig. 1 to illustrate the underlying damage to public transport service quality perception caused by headway irregularity. Let us consider a bus service that is planned to operate with an average headway of 6.5 min and that this implies an average passenger density of 4.5 passengers per square metre over the course of the entire route. The curve of Fig. 1 tells us that the expected satisfaction of users of this service should be 79.6% as long as the buses keep regular headways, and therefore, identical loads (letter A in Fig. 1, perfectly regular scenario). However, lets assume that the headways between buses are 4 and 9 min alternately, keeping an average headway of 6.5 min. According to this sequence, the expected bus load, considering seated passengers, will be 2.8 and 6.3 passengers per square metre respectively. The satisfaction of users of both types of buses will be quite different; while users of the first type will present a 94% satisfaction, in the second type it will be 0%. By averaging both evaluations over vehicles, the average satisfaction evaluation between all buses drops to 47%, as illustrated by the letter B in Fig. 1 (irregular scenario).

However, this average between average satisfaction of both vehicle types ignores that there are fewer travellers inside the first type of buses than in the second type, and our interest is to obtain the average evaluation perceived across users, not buses. Considering the number of travellers that each type of bus carry, the average crowding perceived by them rises to 5.2 passengers per square metre and the average evaluation drops to 28.9% (letter C in Fig. 1, perceived scenario). Thus, the system was planned for an average evaluation of 79.6%, while it dropped to 28.9% due to the headways' irregularity. In reality, quite often buses actually bunch.

This very worrying impact is aggravated as people tend to assign disproportional weights to their bad experiences over their good ones. Thus, level of service variability affects their appreciation by unbalancing it towards those experiences with long delays and big discomfort. It would not be surprising then that, in the experiment proposed, bad experiences loom over respondents' recollection when they are evaluating the system. This fact will be important not only in the methodology design but also in the analysis of the results.

3. Methodology

3.1. Survey description

In order to develop a methodology able to identify and model this non-linear effect, a survey was conducted among public transport users in Santiago de Chile who travel with services that are characterized by high headway variability and/or passenger density within the vehicle. The survey collected the perception or satisfaction perceived by users about the waiting time and travel comfort of the trip they just finished. The fact that they are evaluating their just ended experience (i.e. revealed preferences) make this study different and novel in comparison to the literature regarding comfort valuing (mostly based in stated preferences).

This survey was conducted between the 17th and 20th of July 2018, during the extended morning peak hour, from 07:00 am until 12:00 pm, to obtain observations in periods when capacity binds and when it does not. Users were asked to report their experience regarding their last trip-leg by metro or bus only (i.e. their most recent experience).

The goal was to characterise the effect that comfort and waiting have on travellers' satisfaction. The survey was carried right outside of four selected metro stations (from west to east: *República, Universidad de Chile, Pedro de Valdivia,* and *Manquehue*) and at their surrounding bus stops, approaching alighting travellers to guarantee the randomness of the sample (Fig. 2). These stations were selected for two different reasons. Firstly, they concentrate a high level of alighting passenger for both metro and bus. Secondly, these passengers represent different origin-destination paths through the city, which means they experience different crowding levels along their trip. This is confirmed in Fig. 3, which shows the reported crowding distribution for both metro and bus at the four different study zones. Only *Manquehue* station had a significantly lower crowding for bus and overall reported bus crowding was lower than in metro.

Regarding the survey itself, five surveyors worked for five hours each day, obtaining a total of 1150 responses. The survey was applied for both metro and bus users and gathered information about five aspects, which are detailed below:

(1) **Satisfaction**
Respondents provided a global satisfaction level, using a 1–7 scale (traditionally used for grading in the Chilean education system), to evaluate their perceived experience in the travel-leg they have just completed. In Chile 4 is the minimum passing grade.

(2) **Number of denied boardings**
To have a more precise estimation of waiting time, respondents were asked about how many vehicles they could not board due to insufficient capacity before boarding the vehicle they alighted from.

(3) **Location during the most heavily loaded section**
Given the differences in passenger density within the same vehicle, respondents were asked to indicate where (within the vehicle) they were located during the most heavily loaded moment of their travel-leg (diagram for metro in Fig. 4a).

(4) **Characterisation of the most heavily loaded section**
Finally, respondents characterised the passenger density experienced at the most heavily loaded moment of their travel-leg by choosing one of six images showing different crowding levels (diagram for metro in Fig. 4b).

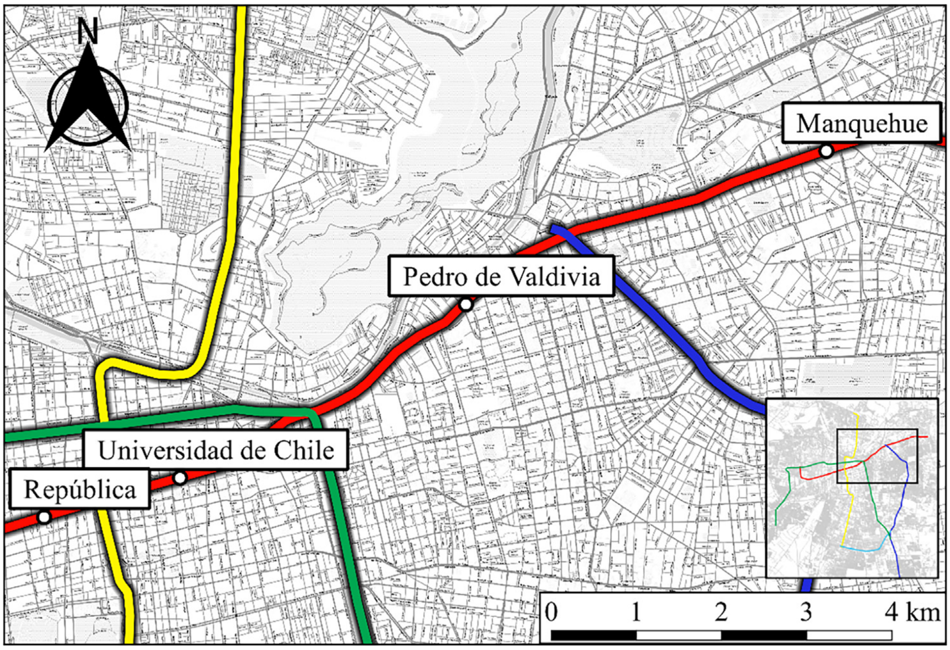


Fig. 2. Survey area of analysis.

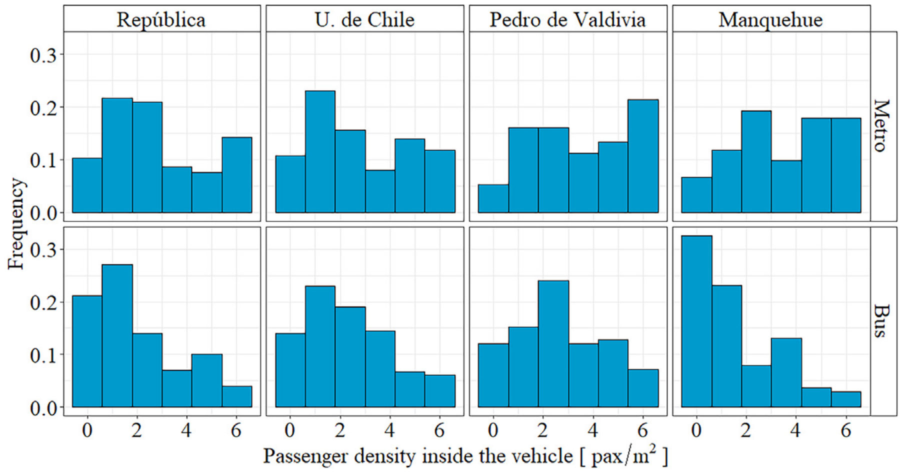


Fig. 3. Reported crowding distribution for different modes and surveyed areas.

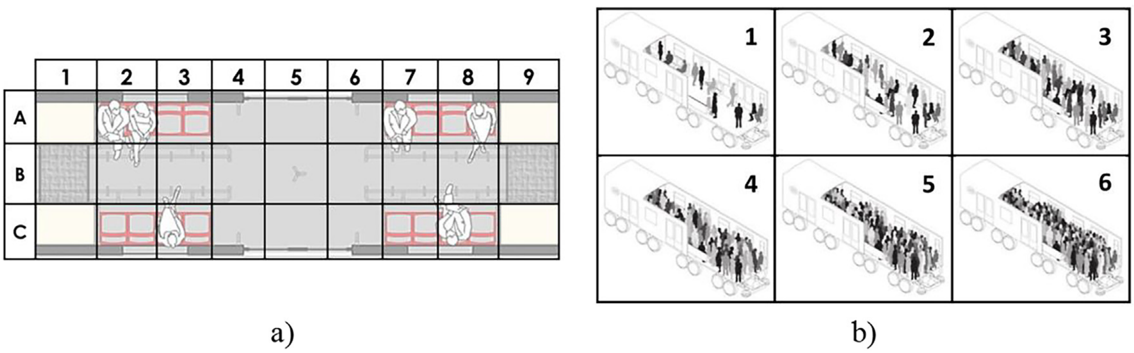


Fig. 4. Location and Crowding inside the vehicle diagrams for metro. *Similar diagrams were used for bus respondents.*

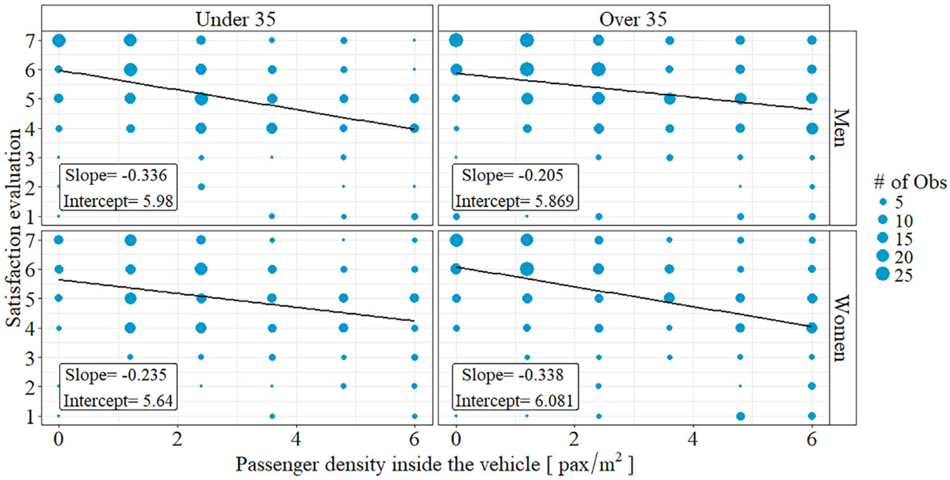


Fig. 5. Trend lines relationship between satisfaction evaluation and reported passenger density.

Income was not directly asked, but instead, their commune of residence was. In Santiago de Chile, average income is very heterogeneous between the different communes and people living in one of them tend to have a similar income. However, this categorization was not found to be significative in the models.

It is important to highlight that respondents were asked only about their global satisfaction rather than adding questions about the level of satisfaction with specific service characteristics. This was done for two reasons. Firstly, we wanted to avoid inducing any possible bias for those subsequent perception questions. This is also the reason why the question concerning global satisfaction appears early on the survey. Secondly, we wanted to estimate a model which explains satisfaction based on operational characteristics instead of those characteristics' satisfaction level. This way, it is possible to estimate the change in global satisfaction as a direct consequence of operation modifications.

3.2. Exploring user categories

To get an idea on how crowding affects different socioeconomical groups' travel satisfaction, scatter plots with a linear trend lines were created (Fig. 5). Two sets of characteristics were selected, which are sex (men and women) and age (under 35 and over 35) to distinguish between four different groups.

Overall, men over 35 years old show a lower slope in this linear relationship, which might be interpreted as a smaller sensitivity to crowding. Besides, women over 35 and men under 35 both have a greater slope and intercept, meaning their satisfaction is more influenced by passenger density. Finally, even though the slope is not greater for women under 35, their intercept is lower, implying that people in this category are less satisfied with crowding, everything else being the same.

However, this analysis is not conclusive as these relations are not sufficiently strong and rigid. Instead, latent classes will be analysed.

3.3. Latent class Ordered Logit

Ordered Logit Models are estimated to explain the satisfaction grade given to the just finished trip based on the conditions of the trip and the respondent's socio-economical information. Several models were separately calibrated for metro and bus users, as an initial exploratory approach, in order to analyse potential mode-specific effects. However, to compare the impacts of different variables obtained for bus and metro users, a model considering both modes simultaneously is also calibrated. Finally, to address preference heterogeneity, a Latent Class model is estimated (Train, 2009).

Since both databases may have different variances (i.e. heteroscedasticity across samples), a first model with common parameters and a scale factor λ_{Bus} for bus users is estimated. Also, a shift parameter Δ_{Bus} is considered, acting as an alternative specific constant. This parameter will test if, ceteris paribus, there is a difference in the evaluation given by users to the level of service experience inside a bus compared to metro. This parameter is expected to be negative, since bus services are found in the literature to be perceived more negatively than rail-bound services (this is commonly called "rail factor"; Scherer, 2010). This difference may not only stem from psychological factors, but also all those differences related with the operation (i.e. stopping at traffic lights, not constant speed) and the experience (i.e. noise, vibration, cleanliness). Moreover, it has been found that in some scenarios, this strong preference for rail actually hides significant level of service differences (Ben-Akiva & Morikawa, 2002). In the case of Santiago de Chile, both kind of differences are specially noticeably between busses and metro and thus, all the conclusions are specific for this context and cannot be generalized to every bus system.

To test the potential existence of a non-linear relation between crowding and satisfaction, two alternative approaches are tested. The first one consists of incorporating an exponent parameter. If this parameter turns out to be significantly different than 1, this means that the relation exercises non-linearity. The second one consists of estimating five different parameters, one for each crowding level, i.e. image (setting the first image as 0). Thus, if the hypothesis that the difference between consecutive parameters is not constant can be statistically rejected, then non-linearity exists. For the sake of readability, all the following models correspond to the first alternative, with an exponent, as it is easier to understand the existence of non-linearity this way and the fit is found not to be significantly better when considering specific crowding parameters for each level.

4. Modelling and results

4.1. Variables

Based on the answers gathered during the survey, the following are the variables considered for modelling:

woman	Equals to 1 if woman and 0 otherwise.
under35	Equals to 1 if under 35 years old and 0 otherwise.
peakhour	Equals to 1 if the respondent was travelling during the morning peak hour and 0 otherwise.
Dens_k	Reported density in mode k.
Door_k	Equals to 1 if the respondent was located in front of the door in mode k and 0 otherwise.
Seat_k	Equals to 1 if the respondent travelled seated in mode k and 0 otherwise.
Veh_k	Equals to the number of denied boardings the respondent experienced in mode k

4.2. Model specification

Regarding the class membership model, after trying different alternative specifications, the best resulting model consists of two classes: (i) considering all the attributes (Class 1) and (ii) without considering crowding nor being seated in their satisfaction evaluation (Class 2). Class 1 can be interpreted as being more sensitive to crowding when users evaluate their travel satisfaction because Class 2 lacks any comfort related attribute. This is known in the literature as attribute non-attendance or attribute ignoring (Nguyen et al., 2015). This way, the differences between classes might be interpreted in terms of their comfort sensitivity.

The class membership function is a Multinomial Logit Model, where the Class1 membership systematic utility is:

$$V_{Class1} = \text{ASC}_{Class1} + \beta_{woman}\cdot\text{woman} + \beta_{age}\cdot\text{under35} + \beta_{hour}\cdot\text{peakhour}$$

Specific parameters for each mode are tested for passenger density, the location inside the vehicle, the number of denied boardings and the chance to get a seat. Regarding passenger density and the number of denied boardings, no significant difference between bus and metro parameters is found. This suggests that those attributes are perceived equally negatively regardless of the mode when passengers are asked to evaluate their satisfaction. The location inside the vehicle is found only significant for metro.

Passenger's density inside the vehicle, the number of denied boarding and the possibility of travelling seated are expected to have a direct impact on satisfaction. In addition, the location inside the vehicle is expected to affect the perception of passenger density. Thus, the systematic utilities for each class and alternative are specified as follows:

Class 1

$$V_{m,C1} = \theta_{com}\cdot(1 + \theta_{com}\cdot\text{Door}_m)\cdot\text{Dens}_m^{\gamma_{com,m}} + \theta_{veh,C1}\cdot\text{Veh}_m$$

$$V_{b,C1} = \lambda_{bus}\cdot(\Delta_{bus,C1} + \theta_{com}\cdot\text{Dens}_b^{\gamma_{com,b}} + \theta_{seatb}\cdot\text{Seat}_b + \theta_{veh,C1}\cdot\text{Veh}_b)$$

Class 2

$$V_{m,C2} = \theta_{veh,C2}\cdot\text{Veh}_m + \theta_{age}\cdot\text{under35}$$

$$V_{b,C2} = \lambda_{bus}\cdot(\Delta_{bus,C2} + \theta_{veh,C2}\cdot\text{Veh}_b + \theta_{age}\cdot\text{under35})$$

4.3. Results

Regarding the class membership model, we observe that, as expected, women are more likely to belong to Class 1, which is in line with previously reported results by research on gender mobility (Allen et al., 2017). This is arguably explained due to other factors related with overcrowding has a larger relative importance for women than for men, such as security and safety.

The same is observed for people under 35 years old as well as with people travelling during the morning peak hour. It is important to emphasise the absence of endogeneity in the latter classification, as there are no significant differences in the distribution of reported densities over time. To complete this analysis, the probabilities of belonging to Class 1 are calculated (Table 1).

Overall, people under 35 years old mostly belong to Class 1, regardless of sex and the time of travel. However, when it comes to people over 35 years old, women are significantly more likely to belong to this class, which reinforces that they are more sensitive to crowding. Besides, this was not found in a previous model without classes (tested as taste variations), which confirms the presence of preference heterogeneity in the sample. Finally, based on the socioeconomical distribution in the sample (Table 2), we compute the average Class 1 membership probability, which is 85.0%.

Table 1
Class 1 membership probabilities.

		Peak hour	Non-peak hour
Women	Under 35	99.6%	98.4%
Men	Under 35	98.3%	93.6%
Women	Over 35	93.6%	78.6%
Men	Over 35	77.6%	46.6%

Table 2
Socioeconomical distribution in the sample.

		Peak hour	Non-peak hour
Women	Under 35	10.9%	10.0%
Men	Under 35	12.4%	10.6%
Women	Over 35	15.0%	9.8%
Men	Over 35	19.1%	12.4%
Average Class 1 membership probability in the sample: 85.0%			

This Latent Class Ordinal Logit Model is estimated using PythonBiogeme (Bierlaire, 2016). The results of this process are summarized in Table 3.

The estimated thresholds are not reported as they do not provide further insights in terms of analysis. However, it is important to highlight that the set of parameters presented are consistent across all of them (which is known as the proportional odds assumption (McCullagh, 1980). Furthermore, the dependent variable is ordinal, the independent variables are continuous and categorical and there is no multicollinearity, fulfilling thus the main assumptions behind Ordinal Logit Models.

Overall, there is a significant and negative shift for bus evaluation ($\Delta_{Bus,Ci}$), which means that users have a more negative perception of the level of service experienced inside a bus than inside metro everything else being the same. Also, the impact of travelling seated is larger for bus users. Class 2 has a more negative perception of buses, as its shift parameter $\Delta_{Bus,C2}$ is approximately 1.70 times larger than Class 1 shift parameter. When considering people under 35, this perception is even worse as Class 2 members have a significant and negative taste variation θ_{age}.

Regarding crowding perception in Class 1, the most interesting result is that this perception is found to be equal for bus and metro. This was tested against three additional models (equal θ_{Com} and different γ_{com}, different θ_{Com}- and equal γ_{com}, different θ_{Com} and different γ_{com}) and none of them turned out to be significantly different to the model presented here. The only exception are the passengers located in front of the door in metro (locations C4, C5 and C6 in Fig. 4). These users perceive 25% more negatively this attribute in comparison. Besides, the reported density parameter, θ_{Com}, in Class 1 was found to be larger than the parameters found in previous models without classes. As Class 2 is not sensitive to comfort, these previous models ended up averaging the sensitivities of both classes, which resulted in a lesser estimated parameter. This confirms again the heterogeneity in user preferences.

Respecting the hypothetical non-linearity, γ_{com} is significantly different from 1, which confirms the proposed hypothesis. However,

Table 3
Latent class calibrated parameters.

Attribute		Parameter	Estimate	*t*-test
Class Membership Model	Alternative specific constant Class 1	ASC_{Class1}	−0.137	−0.32
	Woman	β_{woman}	1.440	1.87
	Under 35 years old	β_{age}	2.820	2.23
	Morning peak hour	β_{hour}	1.380	2.44
Class 1	Reported crowding	θ_{Com}	−1.000	−4.97
	Door impact - metro	θ_{Door}	0.250	2.63
	Reported crowding	γ_{Com}	0.576	6.40
	Travelling seated - bus	$\theta_{seat,b}$	1.050	2.93
	Number of denied boardings	$\theta_{veh,C1}$	−0.316	−4.97
	Shift parameter – bus	$\Delta_{Bus,C1}$	−1.450	−4.69
Class 2	Number of denied boardings	$\theta_{veh,C2}$	−1.270	−4.91
	Under 35 taste variation	θ_{age}	−2.460	−1.83
	Shift parameter – bus	$\Delta_{Bus,C2}$	−2.470	−4.63
Scale factor – bus		λ_{Bus}	0.758	6.62
Number of observations			1150	
Log-likelihood			−1786.8	

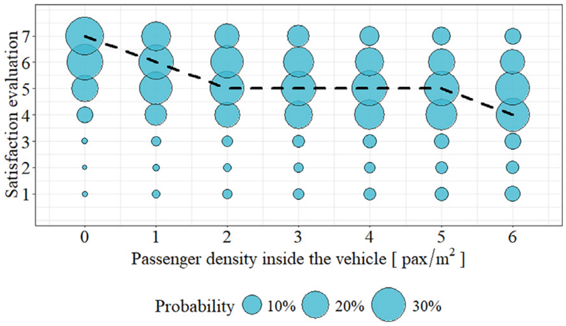


Fig. 6. Satisfaction evaluation probabilities for bus services.

the parameter value is lower than 1, which means that the relationship has the opposite curvature to the one proposed in Section 2. An explanation for this finding and its implications will be further discussed in the following Sections.

In terms of the possibility of travelling seated, it is found not to be significant for metro users whereas it is positive and significant for bus users. This means that metro users do not consider travelling seated when the evaluate their global satisfaction. As it is similar with opposite sign to the shift parameter in Class 1, this means traveling seated helps to reduce the breach between buses and metro, showing a more similar level of satisfaction for the same travel conditions.

Finally, the impact of not being able to board a vehicle was found to be linear, same for both modes and significantly different between classes. In terms of denied boardings, Class 2 values approximately four times more the impact of denied boardings. As Class 2 only accounts for this variable in evaluating their satisfaction, it could be expected that they would be more sensitive to this.

5. Satisfaction evaluation analysis

With the Latent Class Model calibrated, it is possible to construct the relationship between the crowding level and satisfaction evaluation. As described in the previous Section, Class 2 is not sensitive to comfort, and most of the passengers belong to Class 1. Besides, as shown in Table 2, the average membership probability in the sample is 85%. Because of this, the following analysis will focus on Class 1.

First, the probability to evaluate travel satisfaction with a grade from 1 to 7 for densities between 0 and 6 passengers/m^2 for non-seated passengers in metro and bus services is calculated based on model estimation results. The results are displayed in Figs. 6 and 7 respectively. The dashed line in those figures represent the statistical mode for each passenger density level.

From the comparison of the two figures, it is noticeable that the distribution of probabilities for metro services is concentrated in the higher part of the satisfaction scale. This means that, on average, people are highly satisfied with this mode even when travelling in overcrowded situations and are more satisfied than when travelling with the bus under the same circumstances. This is explained by the negative shift parameter $\Delta_{Bus,C1}$ described in the previous Section.

The share of users indicating low satisfaction rates (i.e. satisfaction levels 1–4) increases with passenger density, whereas the share indicating very high satisfaction (i.e. satisfaction level 7) decreases. However, the situation is different for satisfaction levels 5 and 6. Bus service's satisfaction level 5 increases and then remains almost constant, and level 6 is always decreasing, while metro service's satisfaction level 5 is always increasing and level 6 increases and then decreases.

The scale limits may influence the results when people try to give a higher or lesser evaluation. This limitation is referred in the literature as the ceiling effect (Castle & Engberg, 2004). This is not observed in the satisfaction evaluation distributions, even though metro's are mostly in the higher scale limit for densities lesser than two passengers per square metre.

As it might be difficult to analyse this probability distributions for every different number of denied boardings, we computed the average value for each reported density value and mode, and plotted one curve for zero, one and two denied boardings. The obtained

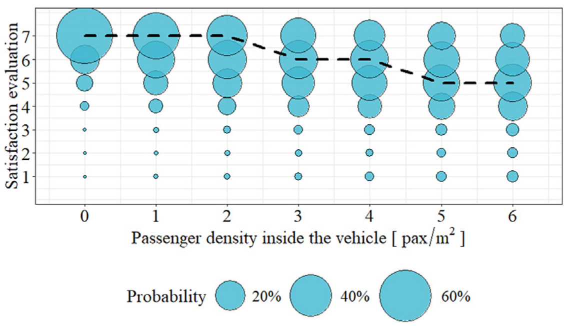


Fig. 7. Satisfaction evaluation probabilities for metro services.

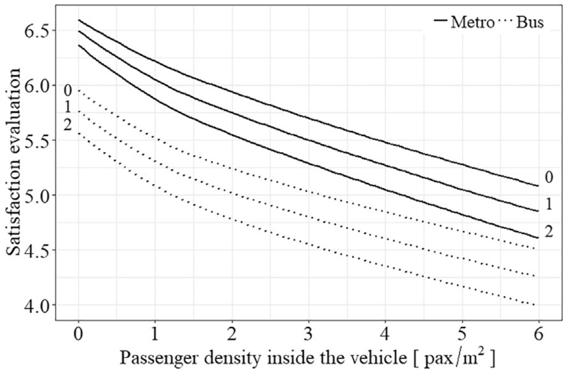


Fig. 8. Satisfaction rating curves per number of denied boardings and mode.

satisfaction curves are presented in Fig. 8.

Metro and bus's curves exhibit some similarities. For example, both modes' curves are decreasing and convex, which is explained by the power parameter in density γ_{Com} and the threshold parameters in the Ordinal Logit Model. Moreover, the difference between denied boardings' curves for each mode has a greater spread with increasing passenger density.

However, the differences between bus and metro curves is more substantial than among curves stemming from the same mode with a different number of denied boarding experiences. There is no single satisfaction evaluation point in bus services in the curves which reflects a higher satisfaction value than the value obtained by the worse situation in metro for identical crowding. Thus, more than two denied boardings in metro are needed to obtain the same predicted evaluation value for bus and metro for a given density level. Furthermore, satisfaction rises 0.25 points on average when passenger density decreases 1 passenger per square metre.

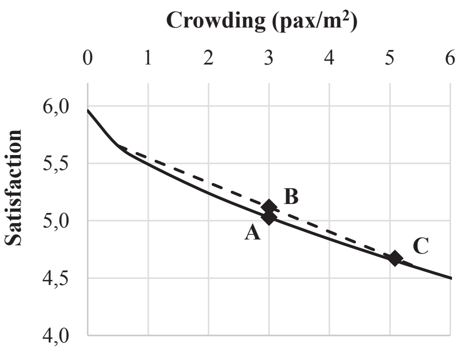
Finally, we perform an analysis similar to the one exposed in Section 2. Since each user has a different satisfaction curve, the analysis will focus on non-seated passengers which board the first vehicle for bus services, as can be seen in Fig. 9. This way, the analysis is centred round the unreliability effects and without compounding it with denied boardings.

In this example, if the service maintains a perfectly regular headway, buses would have the same load, which in this case would be of 3.0 passengers per square metre. Under this situation, this model predicts an average satisfaction evaluation of 5.03 based on the estimated model (letter A in Fig. 9, perfectly regular scenario). However, as in the previous example, we will assume that this service operates irregularly with two alternating headways which causes loads of 0.5 and 5.5 passengers per square metre. If we consider the average evaluation per vehicle, we would observe the same average passenger density of 3 passengers per square metre but this time a satisfaction evaluation of 5.12, higher than the one under regular conditions (letter B in Fig. 9, irregular scenario). This occurs because of the convexity of the curve now presented. Nevertheless, when weighting the load in each vehicle, the situation moves in the opposite direction: the average passenger density rises to 5.1 passengers per square metre and the satisfaction evaluation drops to 4.67 (letter C in Fig. 9, perceived scenario).

This situation substantiates the claim that performance indicators must be weighted properly by demand. If it is not the case, public transport agencies might be perceiving they are offering a good service (even better than expected) while passengers are experiencing the opposite, i.e. a deterioration in service satisfaction.

6. Conclusions

Providing evidence confirming the relationship proposed between headway reliability and traveller's satisfaction could lead to a change in the perspective public transport systems are planned and operated. This research indicates that waiting time reliability and crowding levels have a very strong impact on users' satisfaction evaluation. Irregular headways generate heterogeneity in vehicles'



		Average crowding	Average satisfaction
A	Perfectly regular	3.0 pax/m^2	5.03
B	Irregular	3.0 pax/m^2	5.12
C	Perceived	5.1 pax/m^2	4.67

Fig. 9. Satisfaction fall due to headway irregularity.

level of service. An often-ignored problem here is that more travellers experience the more crowded vehicles, reducing the average satisfaction index further than if simply averaging over bus vehicles. A second issue is that crowdedness and waiting time are strongly correlated which should also be incorporated in the model.

The impact of unreliability and crowding on passenger experience is further exacerbated by the non-linear relation between satisfaction and crowding level revealed in this study. Using a Likert-like semantic grade scale, the curve obtained is convex. This curve shape might bias public transport agencies if they do not consider evaluation metrics weighted by the number of users, as gaps between the level of service believed to be offered and perceived by passengers will occur.

The results of this study also confirm that users evaluate, *ceteris paribus*, worse the level of service in bus than in metro. In addition, regarding socio-economical heterogeneity, people under 35 years old almost always evaluate the service taking into account comfort. However, women over 35 years old are significantly more sensitive to the comfort level, which is in line with our current knowledge of women mobility preferences.

Regarding the methodology employed in this study, we conclude that it is possible to obtain a crowding/satisfaction curve with a simple survey. It would be important to replicate this kind of experiment as the results are limited to the specific Santiago de Chile context. Presumably, since buses and metro offer a very different level of service in the case of the study area, it would be important to analyse how different satisfaction is perceived when buses perform significantly better.

Our results are consistent with other studies that have identified a preference by transit users for metro to the detriment of buses. Understanding how much of this satisfaction or preference is explained by the difference in level of service experienced by users may encourage more affordable and cost-effective alternatives with high level performance, as Bus Rapid Transit (Delgado et al., 2016; Hidalgo & Gutiérrez, 2013).

Future research should pursue at least two new directions of analysis. Firstly, it would be important to characterize respondents by income, to test preferences' differences in the context of satisfaction evaluation. In many cities low income households are located far away the city centre, with poor public transport service. Thus, a better understand of their preferences and needs would enhance public policy application. Secondly, provided with reliable load information for both metro and buses (i.e. weights or APC), the comparison between stated peak passenger density and actual density measures could be analysed. This may expand the current knowledge we have about crowding perception and its relationship with satisfaction.

To make large cities more sustainable public transport should be one of the preferred modes to use. Thus, transport planners should measure, monitor and respond to traveller's satisfaction with their service experience. As we have shown in this study, if transport agencies are evaluated based on performance indicators, these must be properly weighted by the actual number of passengers served to understand the real conditions they face and the perceived quality of the service they deliver.

Acknowledgements

This research was supported by the Centro de Desarrollo Urbano Sustantable, CEDEUS (Conicyt/Fondap 15110020), the Bus Rapid Transit Centre of Excellence funded by the Volvo Research and Educational Foundations (VREF), the FONDECYT project number 1150657 and the scholarship funded by CONICYT for Ph.D. studies (CONICYT-PCHA/Doctorado Nacional/2016). The authors acknowledge the support of the Lee Schipper Memorial Scholarship for Sustainable Transport and Energy Efficiency for the completion of this work. Besides, we would like to thank Sam Zimmerman for his selfless assistance during the production of the article and two "anonymous" reviewers for their insights.

References

Abenoza, R.F., Cats, O., Susilo, Y.O., 2017. Travel satisfaction with public transport: Determinants, user classes, regional disparities and their evolution. Transport. Res. Part A: Policy Pract. 95, 64–84. https://doi.org/10.1016/j.tra.2016.11.011.

Abenoza, R.F., Cats, O., Susilo, Y.O., 2018. How does travel satisfaction sum up? An exploratory analysis in decomposing the door-to-door experience for multimodal trips. Transportation 1–28. https://doi.org/10.1007/s11116-018-9860-0.

Allen, H., Pereya, L., Sagaris, L., Cadenas, G., 2017. Ella de mueve segura - She moves safely. FIA Foundation.

Allen, J., Muñoz, J.C., Ortúzar, J. de D., 2018. Modelling service-specific and global transit satisfaction under travel and user heterogeneity. Transport. Res. Part A: Policy Pract. 113 (May), 509–528. https://doi.org/10.1016/j.tra.2018.05.009.

Batarce, M., Muñoz, J.C., Ortúzar, J. de D., 2016. Valuing crowding in public transport: Implications for cost-benefit analysis. Transport. Res. Part A: Policy Pract. 91, 358–378. https://doi.org/10.1016/j.tra.2016.06.025.

Batarce, M., Muñoz, J.C., Ortúzar, J. de D., Raveau, S., Mojica, C., Ríos, R.A., 2015. Use of mixed stated and revealed preference data for crowding valuation on public transport in Santiago, Chile. Transport. Res. Record: J. Transport. Res. Board 2535, 73–78. https://doi.org/10.3141/2535-08.

Ben-Akiva, M., Morikawa, T., 2002. Comparing ridership attraction of rail and bus. Transp. Policy 9 (2), 107–116. https://doi.org/10.1016/S0967-070X(02)00009-4.

Bierlaire, M. (2016). PythonBiogeme : a short introduction, Technical report TRANSP-OR 160706. Transport and Mobility Laboratory, ENAC, EPFL.

Castle, N.G., Engberg, J., 2004. Response formats and satisfaction surveys for elders. Gerontologist 44 (3), 358–367. https://doi.org/10.1093/geront/44.3.358.

Cats, O., Abenoza, R.F., Liu, C., Susilo, Y.O., 2015. Evolution of Satisfaction with Public Transport and Its Determinants in Sweden. Transport. Res. Record: J. Transport. Res. Board 2538, 86–95. https://doi.org/10.3141/2538-10.

Cats, O., West, J., Eliasson, J., 2016. A dynamic stochastic model for evaluating congestion and crowding effects in transit systems. Transport. Res. Part B: Methodol. 89, 43–57. https://doi.org/10.1016/j.trb.2016.04.001.

De Oña, J., & De Oña, R. (2014). Quality of Service in Public Transport Based on Customer Satisfaction Surveys: A Review and Assessment of Methodological Approaches, January 2015.

de Oña, J., de Oña, R., Eboli, L., Mazzulla, G., 2016. Index numbers for monitoring transit service quality. Transport. Res. Part A: Policy Pract. 84, 18–30. https://doi.org/10.1016/j.tra.2015.05.018.

Delgado, F., Muñoz, J.C., Giesen, R., 2016. BRRT: adding an R for reliability. Restructuring Public Transport through Bus Rapid Transit.

DTPM, 2016. ¿Cuál es tu parada? Sé parte de la solución. Ministerio de Transportes y Telecomunicaciones, Gobrierno de Chile.

Habib, K.M.N., Kattan, L., Islam, T., 2011. Model of personal attitudes towards transit service quality. J. Adv. Transport. 45 271–185.

Hensher, D.A., Stopher, P., Bullock, P., 2003. Service quality - developing a service quality index in the provision of commercial bus contracts. Transport. Res. Part A: Policy Pract. 37 (6), 499–517. https://doi.org/10.1016/S0965-8564(02)00075-7.

Hidalgo, D., Gutiérrez, L., 2013. BRT and BHLS around the world: Explosive growth, large positive impacts and many issues outstanding. Res. Transport. Econ. 39 (1), 8–13. https://doi.org/10.1016/j.retrec.2012.05.018.

Kim, K.M., Hong, S., Ko, S., Kim, D., 2015. Does crowding affect the path choice of metro passengers? Transp. Res. Part A 77, 292–304. https://doi.org/10.1016/j.tra.2015.04.023.

Li, Z., Hensher, D.A., 2011. Crowding and public transport: A review of willingness to pay evidence and its relevance in project appraisal. Transp. Policy 18 (6), 880–887. https://doi.org/10.1016/j.tranpol.2011.06.003.

Liu, J., & Wen, H. (2016). Public transport crowding valuation: evidence from college students in Guangzhou. 19(3), pp. 78–97.

McCullagh, P., 1980. Regression models for ordinal data. J. Roy. Statist. Soc. Series B (Methodological) 42 (2), 109–142. https://doi.org/10.1079/IVPt200454IN.

Nguyen, T.C., Robinson, J., Whitty, J.A., Kaneko, S., Nguyen, T.C., 2015. Attribute non-attendance in discrete choice experiments: A case study in a developing country. Econ. Anal. Policy 47, 22–33. https://doi.org/10.1016/j.eap.2015.06.002.

Ortúzar, J. de D., Willumsen, L.G., 2011. Modelling Transport, fourth ed. John Wiley Sons Ltd.

Redman, L., Friman, M., Gärling, T., Hartig, T., 2013. Quality attributes of public transport that attract car users: A research review. Transp. Policy 25, 119–127. https://doi.org/10.1016/j.tranpol.2012.11.005.

Scherer, M., 2010. Is light rail more attractive to users than bus transit? Transport. Res. Record: J. Transport. Res. Board 2144 (1), 11–19. https://doi.org/10.3141/2144-02.

Tirachini, A., Hensher, D.A., Rose, J.M., 2013. Crowding in public transport systems: Effects on users, operation and implications for the estimation of demand. Transp. Res. Part A 53, 36–52. https://doi.org/10.1016/j.tra.2013.06.005.

Tirachini, A., Hurtubia, R., Dekker, T., Daziano, R.A., 2017. Estimation of crowding discomfort in public transport: Results from Santiago de Chile. Transp. Res. Part A 103, 311–326. https://doi.org/10.1016/j.tra.2017.06.008.

Tirachini, A., Sun, L., Erath, A., Chakirov, A., 2016. Valuation of sitting and standing in metro trains using revealed preferences. Transp. Policy 47, 94–104. https://doi.org/10.1016/j.tranpol.2015.12.004.

Train, K.E., 2009. Discrete Choice Methods with Simulation, second ed. Cambridge University Press, Cambridge.

Tyrinopoulos, Y., Antoniou, C., 2008. Public transit user satisfaction: Variability and policy implications. Transp. Policy 15 (4), 260–272. https://doi.org/10.1016/j.tranpol.2008.06.002.

Wardman, M., Whelan, G., 2011. Twenty years of rail crowding valuation studies: Evidence and lessons from British experience. Trans. Rev. 31 (3), 379–398. https://doi.org/10.1080/01441647.2010.519127.

Yap, M., Cats, O., van Arem, B., 2018. Crowding valuation in urban tram and bus transportation based on smart card data. Transport. A: Trans. Sci. 1–20. https://doi.org/10.1080/23249935.2018.1537319.