Quantifying the impact of in-vehicle crowding on customer satisfaction in Public Transport

A Den Haag case study

M.P.J. Seerden

HTM

TU Delft

Quantifying the impact of in-vehicle crowding on customer satisfaction in Public Transport

A Den Haag case study

By

M.P.J. Seerden

in partial fulfilment of the requirements for the degree of

Master of Science
in Transport, Infrastructure and Logistics

At Delft University of Technology,
to be defended publicly on Wednesday August 7th, 2019 at 15:00.

Thesis Committee:	Prof. dr. ir. S.P. Hoogendoorn	TU Delft
	Dr. Ir. M. Kroesen	TU Delft
	Dr. Ir. N. van Oort	TU Delft
	S. Nijënstein MSc	HTM Personenvervoer N.V.

Final version
July 24, 2019

An electronic version of this thesis is available at http://repository.tudelft.nl

Preface

Almost seven years ago I started my studies in at TU Delft, and during these very enjoyable years I had the chance to experience many wonderful moments and learn a tremendous amount of things, not only academically but also personally. The project of which this thesis is a result thus not only marks the end of my studies but also the end of an era in my personal life. I feel happy about the results of this project: this thesis truly feels like an appropriate end to my period as a student. I did, however, not finish this project all by myself, and I owe a lot of people my thanks for their help and support.

Firstly, I would like to thank HTM for the opportunity to do my thesis' research with them. After many years of doing projects on hypothetical situation *X* using theoretical dataset *Y* it was a delight to work on a project which actually matters using input and data from the real world. Specifically I would like to thank Sandra for all her and her tendency to always look how things could be taken to the next level. I'd also like to thank Rien for his effort in helping me in collecting all relevant data from the mountain of data HTM has, Marije for her support in helping me understand HTMs Klantenpanel and Janiek for always knowing every answer to any question I had for him instantaneously. Without their support I would not have been able to finish this project. Furthermore, I'd like to thank the whole team for all laughs during work, which made me travel to HTM carrying a smile. As a result, I warmheartedly look back at my time with HTM.

My supervisors at TU Delft also played a key role in the creation of the thesis which you are reading at this moment. Niels helped me in finding this research position at HTM in the first place, and afterwards his enthusiasm and his experience on the process of thesis research helped me in my planning and in keeping my motivation at an acceptable level during difficult moments. Maarten was a beacon of knowledge with his expertise on data analysis and modelling, but also in reminding me at the correct moment that no research is perfect and at some point I just needed to accept its imperfections and carry on. Without this I think I'd still be working on improving each and every tiny detail. Lastly I'd like to thank Serge for ensuring committee meetings were productive and as pleasant as could be.

Besides my committee and HTM I'd also like to thank Menno Yap (TU Delft) and Mark van Hagen (NS) for their time, knowledge and enthusiasm, which encouraged my curiosity greatly.

Lastly, I'd like to thank Marit, my family and all my friends for helping me to forget all doubts, issues and stress about this project during weekends and holidays, as there is absolutely no necessity for thinking about your work 24 hours a day.

I wish the reader much enjoyment in reading this thesis and hopes it helps the reader in achieving the goal with which aim he/she is reading. Cheers!

Maarten Seerden
Den Haag, July 24th, 2019

Paper

Abstract

In an environment in which Public Transport plays an ever growing role in urban areas and passenger experience. In-depth knowledge on customer satisfaction is key in improving passenger experience. Crowding has shown to be an important aspect of route and mode choice, but its impact on customer satisfaction has not yet been quantitatively explored. This paper investigates this impact. Using a Structural Equation Model based on customer satisfaction survey data and corresponding occupancy and punctuality numbers a significant effect of occupancy levels on perceived crowding was found. In-vehicle crowding, both subjective and objective, seems to impact overall customer satisfaction indirect, with comfort being the main mediating variable. Model results were converted into a calculation tool which can be used to assess the impact of changes in passenger numbers, passenger distribution and frequencies on customer satisfaction. Further research can verify the linearity of the relation between in-vehicle crowding and customer satisfaction and deepen knowledge of external factors such as weather and disruptions have on experienced crowding.

Key words: Urban Public Transport, Structural Equation Modelling, Customer Satisfaction, In-vehicle Crowding

1. Introduction

In many large cities maintaining mobility is one of the main challenges which is faced today. The number of inhabitants continues to grow and congestion issues show that having everybody transport themselves by car is difficult. Due to its ability of carrying large numbers of passengers without using a lot of space Urban Public Transport offers a natural and logical alternative.

Passengers will only use Public Transport if they find this to be a comfortable way of travel. As a result, over the past years regulators have put an increasing focus on using passenger experience as a metric for operator performance. Dutch law allows regulators to reward or penalise operators in case of good or bad performance. As a result, lower bounds have been set for the minimum evaluation that passengers have to give a trip on average and benchmarks have been set to ensure the aim for continuous improvement. This can be seen, for example in the rail concession for the city of Den Haag, where operator HTM receives a fine if overall customer satisfaction is below 7.5 [1].

Crowding is one of the aspects known to impact how passengers experience a trip. It is clear that travelling in an overcrowded vehicle is much a much worse experience than travelling when enough seats are available. Research into the effect of crowding on customer satisfaction up to this moment is, however, limited. Having quantitative insight into the form and shape of this relation is useful as this helps operators in identifying potential overcrowding more accurately, which allows for quicker measures to solve this problems. The resulting main question which is to be answered in this paper is:

What is the relation between objective and subjective in-vehicle crowding in Public Transport and customer satisfaction?

2. Background

2.1. Literature review

Academic research into customer satisfaction in Public Transport remains a relatively young field: a large increase in the amount of research can be seen over the past 10 years. Customer satisfaction, in some studies also called service quality, is generally defined in literature as the gap between a customers' expectation of a service and his experience [2] [3]. The logical next question is what factors influence these experiences and expectations. Literature identifies two types of factors:

- Service-related factors, such as frequency and punctuality.
- Customer-related factors, such as age or gender.

In research into service-related factors, two approaches can be seen. Some research (e.g. [4]) identifies specific aspects (e.g. price, punctuality) which determine customer satisfaction and link these aspects directly to customer satisfaction. Other studies [5] choose a more layered approach. These studies categorise relevant aspects in a few factors (e.g. 'convenience', 'service planning and reliability') and state that these factors determine customer satisfaction. The second method has been slightly more often used. Table 1 provides an overview of relevant research into service-related factors affecting customer satisfaction in Public Transport. Regardless of the categorisation of attributes, recurring factors and aspects in literature include comfort, frequency, reliability, fare prices and travel speeds.

Table 1: literature review of service aspects affecting customer satisfaction

Author	Year	Modality	Dimensions
Fellesson & Friman	2008	Bus, Tram, Metro	System, comfort, staff, safety
Abenoza, Cats & Susilo	2017	Bus, tram, metro	Customer interface, operation, network, travel time
Redman, Friman, Gärling, Hartig	2013	None*	Frequency, fare prices, speed, reliability
Eboli & Mazzulla	2007	Bus	Service planning and reliability, comfort and other factors, network design
de Oña, Eboli & Mazzulla	2014	Bus	Fare, information, courtesy, safety, accessibility, cleanliness, space, temperature, proximity, speed, punctuality and frequency
Yaya, Fortià, Canals, Marimon	2015	Bus	Functional Quality, Physical Environment Quality, Convenience Quality
Morton, Caulfield & Anable	2016	Bus~	Convenience, Cabin Environment, Ease of Use
Olsson, Friman, Pareigis, Edvardsson	2012	Bus, Tram	Positive activation, positive deactivation, cognitive evaluation
Abenoza, Cats & Susilo	2018	Bus, tram, metro	Waiting times, Satisfaction with access and egress legs

**This is a literature review study, the dimensions found are an aggregate of other research.*
~ This study considered subjective attributes and thus did not find objective dimensions.

Table 2 shows the customer-related factors found to affect customer satisfaction. It can be seen most studies find factors such as gender, age, education level and income to play a role in some way.

Table 2: literature review of personal characteristics significantly affecting customer satisfaction

Author	Year	Modality	Factors
Morton, Caulfield & Anable	2016	Rural Bus	Economic status*, Age, Gender, Education level
Theler & Axhausen	2013	Urban Bus	Age, Frequency of PT Use
van Lierop & El-Geneidy	2016	Metro	Income, Car access
Van 't Hart	2012	Bus, Tram, Metro	Frequency of PT Use, Age, Gender, Location, Travel Purpose
Diana	2012	Bus (Urban and rural)	Frequency of PT Use, Location
Mouwen	2015	Bus, Tram, Metro, Train	Age, Past Experiences
Friman, Edvardsson & Görling	2001	Bus, Tram, Metro	Past Experiences
Abenoza, Cats & Susilo	2017	Bus, Tram, Metro	Frequency of PT Use, Age, Car Access
Yaya, Fortià, Canals, Marimon	2015	Bus, Tram, Metro	Age, Possession of Drivers' license, education
Koning, Haywood & Monchambert	2017	Metro	Income

** Defined as the main occupation of a respondent (employment/retired/student/etc.)*

Customer satisfaction and crowding have rarely be linked quantitatively in academic research up to now. Only Haywood et al. (2017), investigating the Paris metro, analysed the effect of perceived crowding on customer satisfaction [8]. They found this relationship to be linear. Nevertheless, from research into the effects of crowding it can be deduced that crowding has a definite effect on passenger experience. Experienced travel times and costs become much higher in case of crowding, which occurs when passenger numbers become high. The most often used metrics to measure crowding are Load Factor (the ratio between the number of passengers in a vehicle and the number of seats) and Standing Passenger Density (the ratio between the number of standing passengers in a vehicle and the space available for standing) [6], with the first in general being more usable on lower passenger numbers and the latter in crowded situations [7].

One possible reason for the lack of research on the effect of crowding on customer satisfaction might be the time investment coming with collecting enough data on occupancy. The introduction of Smart Card payment systems such as the OV-Chipkaart in the Netherlands opens up a lot of possibilities for gathering Smart Card data, which allows for much richer data set than used to be possible. As Yap et al. (2018), Hörcher et al. (2017), Hong et al. (2016), and Ticharini et al (2016) show, the increasing use of Smart Cards as payment measure in Public Transport provides a very rich data source for occupancy numbers in networks which was unavailable up to this point [9][10][11][12].

However, while some researchers (e.g. [13]) have explored some parts of the relationship between crowding and customer satisfaction no one has tried quantitatively to capture either:

- The exact relationship between objective and subjective crowding
- The effect of both objective and subjective crowding on customer satisfaction.

A framework has been developed which tries to capture both these relationships, as shown in figure 1. Rectangles show observed variables and ovals represent latent variables. Each colour also represents a category: green represents operational service performance, dark blue service characteristics, light blue characteristics which differ per passenger – the attributes mentioned in table 2 – and purple customer evaluations of (aspects of) the trip. Customer satisfaction is explained using a multi-layered structure, in line with, for example, Eboli and Mazzulla (2007) and Morton and al (2016) [5] [14]. Customer satisfaction is constructed as the sum of customer perception in three latent factors: service quality, comfort and safety. Table 3 provides an overview of the measurement variables per latent variable.

The framework is innovative in suggesting that the relation between occupancy and customer satisfaction is indirect. Previous studies such as those mentioned in table 1 do not consider crowding to be an important determinant of customer satisfaction. However, they do not search for indirect effects. It is logical that the effect of crowding on customer satisfaction is indirect: passengers do not dislike overcrowding because there are a lot of people in a vehicle, they dislike overcrowding because of the discomfort that comes with it. During analysis it has been tested whether this is indeed a correct method of modelling this relationship.

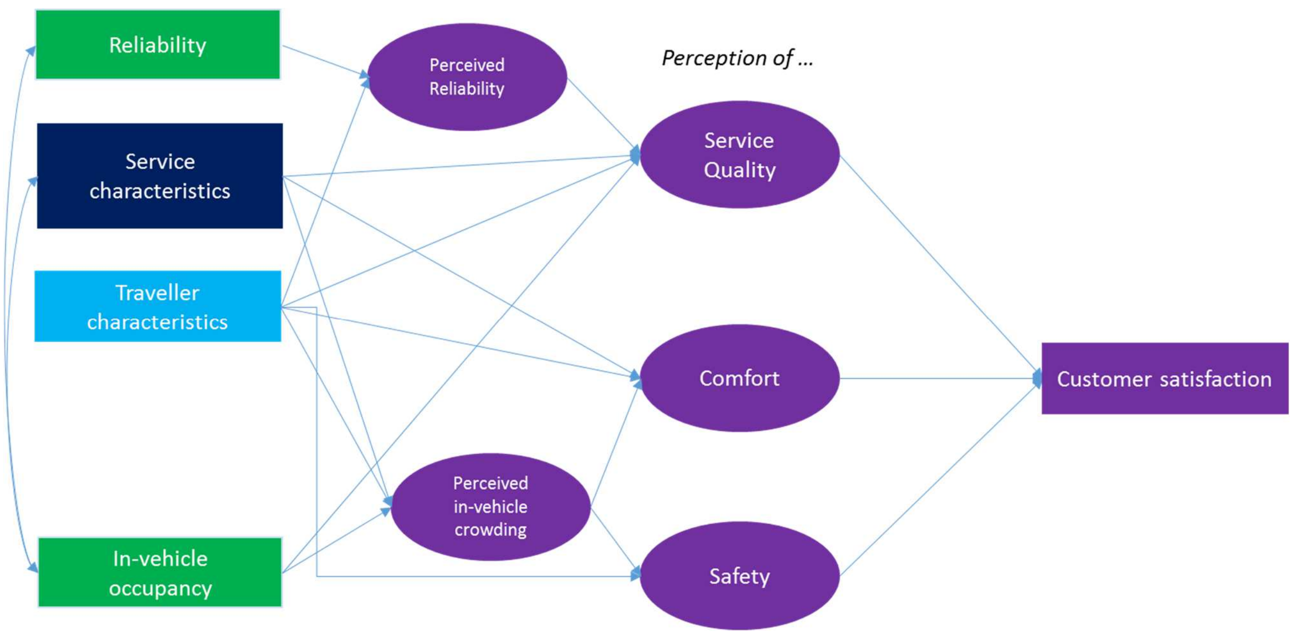


Figure 1: Framework used for analysis

Table 3: attributes per dimension

Factor	Attribute
Comfort	Comfort
	Cleanliness of vehicle
	Cleanliness of stop
	Friendliness of staff
	Ease of buying a ticket
	Driving style of driver
Customer Satisfaction	Overall customer satisfaction
Perceived Crowding	Probability of finding a seat
Perceived Reliability	Punctuality
Safety	Feeling of safety during this trip
	General feeling of safety in PT
Service Quality	Information supply on stop
	Information supply during delays or disruptions
	Frequency

2.2. Case Study Background

This research has been conducted at HTM Personenvervoer NV. HTM operates all urban rail lines in the Den Haag region and all urban bus lines within Den Haag, under concessions granted by the Metropoolregio Rotterdam-Den Haag (MRDH). The network as of 2018 consists of 12 tram lines and 14 bus lines. In 2017, HTM transported a total number of 98 million passengers in the Den Haag area: 82 million in trams, 16 million in urban buses [15]. HTM operates all urban rail lines in the Den Haag region and all urban bus lines within Den Haag. For its tram lines HTM uses three types of trams: from old to new these are GTL-8, Regio Citadis and Avenio. During the time period analysed (2018) HTM used one bus type for all bus lines. Starting in December 2018 a second, electric bus type was introduced.

HTM aims at offering its passengers a travel experience as pleasant as possible. As a result, it is useful for HTM to have knowledge of how passengers experience in-vehicle crowding. This insight can help HTM to further improve its services.

3. Methodology

A quantitative analysis of the framework presented in figure 1 requires data on both objective occupancy rates as well as subjective perceived in-vehicle crowding and customer satisfaction. For this study, data on vehicle occupancy is collected at HTM measuring transactions with the OV-chipkaart, a smart card. The usage of the OV-chipkaart provides enough data on occupancy rates for a large scale analysis of the impact of in-vehicle crowding on customer satisfaction. For customer satisfaction data HTMs own customer evaluation panel (HTM Klantenpanel) was chosen as a data source, using all data for the year 2018 as a basis. When evaluating a trip via the HTM Klantenpanel, a respondent is asked to give 13 aspects a mark from 1 (low) to 10 (high), with the possibility of answering 'I do not know' or 'did not apply to this trip' as well.

In order to analyse the effect of objective and subjective in-vehicle crowding on customer satisfaction objective data needs to be coupled to the customer satisfaction responses. Therefore each response was coupled to the service which it evaluated and occupancy and punctuality data was retrieved for this service. As a measurement of crowding the Load Factor was chosen as the average occupancy in the dataset was quite low, the modelling method used requires the use of just one metric per variable. If no corresponding occupancy data could be found or if there existed any ambiguity regarding what service was evaluated the data was deleted. As a result, evaluation of 2858 trips over the year 2018 was used for analysis.

The conceptual framework presented in figure 1 is complex and multi-layer. Because of this, it was chosen to use Structural Equation Modelling (SEM) to analyse the relation between crowding and customer satisfaction quantitatively. SEM is suited for this type of models and moreover is able to estimate relationships between unobserved constructions based on measured variables [16].

4. Results

4.1. Measurement Model

Before estimating the Structural Equation Model, a Confirmatory Factor Analysis (CFA) was carried out to test whether the proposed structure fits the data sufficiently enough. Table 4 shows the resulting weight at which each indicator loads on its respective factor and the average variance extracted on each factor.

Table 4: Standardised coefficients of measurement model

Factor	Indicator	*Weight*	*Average Variance Extracted (AVE)*
Perceived Safety	Feeling of safety during this trip	0.739	0.716
	General feeling of safety in PT	0.942	
Perceived Service Quality	Frequency	0.708	0.496
	Information supply during delays or disruptions	0.710	
	Information supply on stop	0.696	
Perceived Comfort	Cleanliness of vehicle	0.751	0.477
	Comfort	0.780	
	Driving style of driver	0.728	
	Cleanliness of stop	0.634	
	Friendliness of staff	0.721	
	Ease of buying a ticket	0.492	
Perceived Reliability	Punctuality	1*	1*
Perceived Occupancy	Probability of finding a seat	1*	1*

* Set to 1 per definition, as for these factors only one indicator is available

An indicator is said to load well to its corresponding factor if this weight is above 0.7 (marked with green) and satisfactory if this weight is above 0.5 (marked in yellow), the Average Variance Extracted (AVE) per factor ideally is at least 0.5. One indicator (marked in red) just fails to meet this threshold and the AVE is a tad low for the factors Perceived Service Quality and Perceived Comfort. Further analysis showed, however, that deleting the indicators with poorer fit to get the AVE above 0.5 significantly decreased overall model fit. Hence the decision was made not to leave out any indicators of the model.

Besides factor loading general model fit of the measurement model is also important to analyse. The Comparative Fit Index (CFI) of the measurement model is 0.916, above the threshold of 0.9 indicating good model fit. The Root Mean Square Error of Approximation (RMSEA) is 0.087, just above the threshold of 0.08 which indicates the upper bound for good model fit. In conclusion, the fit of the measurement model was considered to be good enough for the structural model to be estimated.

4.2. Structural Model

As the measurement model was found to be usable estimating the structural model as shown in figure 1 is possible. The results can be seen in figure 2, which gives a visual overview of the strength of the relations in the framework. All effects shown are standardised, which means the relative strength of relationships is shown well. The model fit of the structural model is good: the CFI is 0.910 and the RMSEA is 0.059.

Results show an evident effect of crowding on customer satisfaction: the standardised effect of the Load Factor on overall satisfaction was found to be -0.111, implying customer satisfaction drops by 0.1 standard deviation for each standard deviation the Load Factor increases. The effect of perceived crowding was even a bit stronger at 0.215. The effect of occupancy on subjective crowding was estimated to be -0.469 – one of the strongest weights in the model but still far from a one-on-one relation. Perceived crowding can thus not be fully explained using just occupancy. The relation between objective and perceived crowding has a negative sign due to the method of measurement: the occupancy is measured using the Load Factor, in which a higher number means more crowding. On the other hand, perceived crowding is measured using the mark given for the probability of finding a seat on boarding, in which a higher value means less crowding.

The hypothesis that the effect of occupancy on customer satisfaction is indirect seems to be verified by the results. Both perceived comfort and perceived service quality are found to be significant mediating factors, although when measured in coefficients the relation via perceived comfort is much stronger.

After testing, a linear relation between occupancy and overall satisfaction was found to be the best way of modelling. This was done by fitting a linear, quadratic and cubic polynomial on the data for these variables. While the explained variance found was quite low, all polynomials were found to be significant ($p = 0.000$) and using a non-linear polynomial resulted in little extra explained variance.

Looking at factors which further affect customer satisfaction indirectly, some interesting conclusions can be found. Vehicle type seems to have an impact on perceptions: the newer Avenio trams are evaluated as significantly more comfortable than other tram types and buses. Delay perception has a large impact on customer satisfaction as well: passengers who mention having experienced a delay or disruption during their trip evaluate their overall satisfaction with their trip 1.2 points lower than passengers who did not experience a delay or disruption.

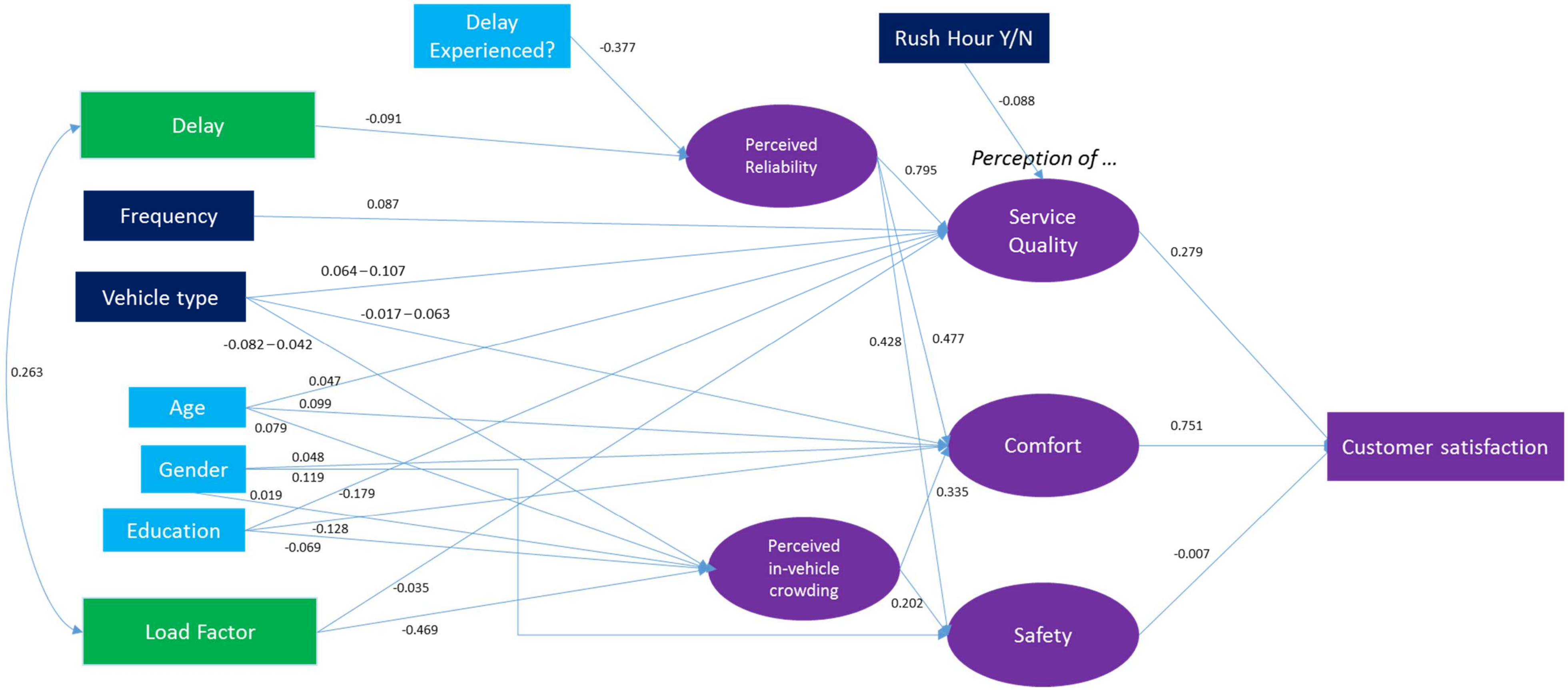


Figure 2: Estimated structural model with standardised effects

Looking at the effect of personal characteristics, elderly are more satisfied than younger travellers and females are slightly more satisfied than males. Interestingly the effect of education level on customer satisfaction is negative and strong: the model predicts a difference of a whole point on a scale of 1 to 10 between low and high education.

5. Implications

The main aim of this research was to gain insight in the relation between objective and subjective crowding and overall customer satisfaction. To use these insights in the context of growing passenger numbers which are expected in Den Haag. The model poses a linear relationship between objective and subjective crowding with a standardised coefficient of -0.469. Following all relations, to have overall satisfaction be lowered by 0.1 point an increase of the Load Factor with almost 20 percentage point is necessary.

Nevertheless, there are certainly instances in which significant gains can be made. Several lines on the HTM network, for example lines 3, 4 and 9 suffer from skewed occupancy between services during rush hour. It was calculated how much customer satisfaction could be gained by distributing passengers evenly. This way, overall customer satisfaction can be improved by up to 0.05 point and perceived crowding can be improved by up 0.3 point on a scale of one to ten. While not enormous this could certainly help in improving passenger experience and customer satisfaction, albeit slightly. It is thus advisable to put effort in trying to distribute passengers more evenly among trams during rush hour.

The effects of passenger growth on customer satisfaction are found to be highly dependent on the size of this growth. Current growth predictions at HTM estimate the yearly passenger growth to be between 1% and 3% [17]. Table 5 shows how long it takes before overall satisfaction drops with 0.1 due to increased passenger numbers for both the lower and upper bound of this estimate, assuming that all other variables (frequencies, delays, etc.) do not change. If passenger growth proves to be near the upper bound of the current estimates effects on customer satisfaction can be seen within five years. The crowded services of the network are much more vulnerable for these effects than the quiet services. In planning, HTM uses a nominal capacity during normal operations ('inzetnorm') which corresponds with a Load Factor of 175% to 200%, the crush capacity of HTMs vehicles corresponds with a Load Factor of approximately 250%.

Table 5: expected time (in years) before overall satisfaction drops 0.1 on average due to growth of passenger numbers

Current Load Factor (%)	Yearly growth	
	1%	3%
50	31	11
75	22	7
100	17	6
125	13	5
150	12	4
175	10	3.5
200	9	3
250	7	2.5

The effect of a change of frequency can also be estimated. Table 6 gives an overview of what happens with customer satisfaction in case of a change in frequency for a variety of current Load Factors. Lowering frequencies on relatively quiet lines in order to increase them on busy lines does not seem to affect customer satisfaction heavily. Increasing frequencies on busy lines has a higher effect on customer satisfaction, an effect which is strengthened by more people profiting in the busy services than suffer in the quiet services.

Table 6: Effect of frequency changes on overall satisfaction

Frequency change [veh/h/dir]	6 -> 5	6 -> 5	6 -> 7	6 -> 7
Load Factor per vehicle before change (%)	*25*	*50*	*175*	*250*
Change in overall satisfaction [1-10]	-0.045	-0.054	+0.153	+0.211

6. Conclusions and recommendations

This research quantified the relation between vehicle occupancy, in-vehicle crowding and overall customer satisfaction. The relation between these aspects was found to be significant though indirect: figure 3 shows an overview of the relation found. The strongest effect found is that if a traveller experiences crowding, this will lead to more discomfort which leads to lower overall satisfaction. Due to the variety of factors which affect customer satisfaction a modelling method which is able to model the complex nature of passenger experience. Structural Equation Modelling has proven to be an adequate method for doing so.

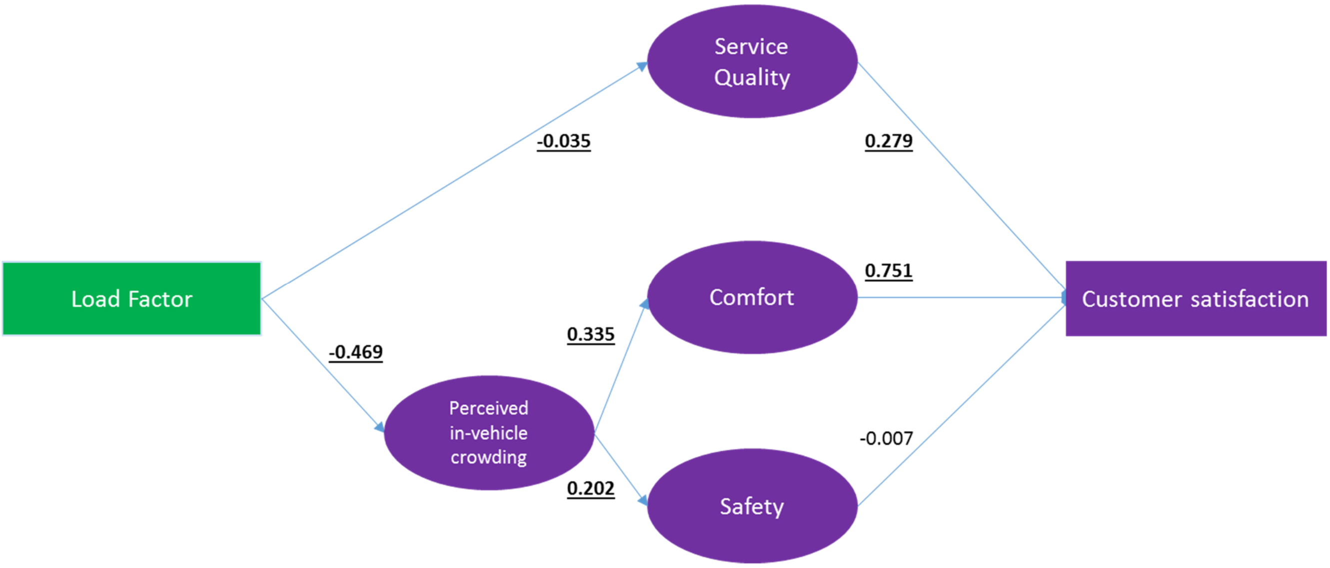


Figure 3: Link between occupancy and customer satisfaction. ***Bold, underlined relations*** *are significant at the 0.05 level.*

Based on the analysis of model results some recommendations can be given on how customer satisfaction can be optimised using small tweaks. These recommendations are as follows:

- Given service and passenger numbers, optimal customer satisfaction is reached when passengers are distributed evenly among services.
- The analysis of the impact of a (perceived) disruption illustrates that the effect of perceiving a disruption or delay by a passenger is much larger than the effect of an actual delay or disruption.
- The effect of passenger growth on customer satisfaction is highly dependent on the size of that growth and current passenger numbers. For services which already have high occupancy rates, small changes can make a large impact.
- The effect of lowering frequencies on quiet lines is smaller than increasing frequencies on busy lines on the level of an individual passenger. As much more passengers travel on busy lines compared to quiet lines, this effect becomes much larger when all passengers are considered instead of one.

7. Discussion and outlook

This research has quantified the relationship between objective crowding, subjective crowding and customer satisfaction using customer satisfaction survey data and corresponding smart card data. This has been a novelty with possibilities for further research.

Firstly, literature has used a variety of different categorisations to explain customer satisfaction in Public Transport. This study chose an often used approach in which attributes were first linked to latent constructs which in turn affected customer satisfaction. Results suggest that this structure might not be the best way to measure customer satisfaction: it might be better to just directly link relevant aspects to customer satisfaction instead. Experimenting with different set-ups of how customer satisfaction can be defined model wise is useful to get to know what the best construction is.

The model could also be made more accurate by expanding the factors which are known to have an impact on customer satisfaction but were not included. This includes, amongst others, the effect which the weather has on passenger numbers (on a rainy day people will tend to use PT more, on a sunny day they might prefer to bike – but little is known on the quantitative form of this relationship) as well as passenger experience. This was not included because there is insufficient knowledge on how weather affects perceptions exactly. In a sense, the same lack of knowledge applies here as was the case for the impact of in-vehicle occupancy levels: it might be an interesting topic for research.

This research modelled the relation between occupancy and perceived crowding using a linear relationship, in line with [8]. The evidence for a linear relationship was found not to be overly strong. It would be interesting to use more advanced methods to test whether more complex mathematical functions can quantify the relation between objective crowding, subjective crowding and customer satisfaction even better. A second issue in this line is the sole usage of Load Factor as measurement of occupancy, which literature has shown to be an imperfect metric.

Lastly, regarding data collection this research was conducted based on existing data provided by members of HTM Klantenpanel. This data proved to be imperfect. Ideally a customer satisfaction survey is set out based on the conceptual framework which is developed.

All in all, this research has shown that it is possible to properly quantify the relation between occupancy rates, perceived crowding and customer satisfaction in Public Transport. When calibrated to a specific case model results can help an operator in tweaking and thereby optimizing passenger flows in their network. The suggestions in this section could help to further improve the accuracy of an estimate.

References

[1] MRDH. (2017). *Concessiemonitor MRDH 2017.* Metropoolregio Rotterdam-Den Haag.
[2] Morfoulaki, M., Tyrinopoulos, Y., & Aifadopoulou, G. (2010). Estimation of Satisfied Customers in Public Transport Systems: A New Methodological Approach. *Journal of the Transportation Research Forum*.
[3] Yaya, L., Fortià, M., Canals, C., & Marimon, F. (2015). Service quality assessment of public transport and the implication role of demographic characteristics. *Public Transport, 7*, 409-428.
[4] 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*(3), 605-622.
[5] 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.
[6] Wardman, M., & Whelan, G. (2011). Twenty years of rail crowding valuation studies: Evidence and lessons from British experience. *Transport Reviews*.

[7] Tirachini, A., Hensher, D., & Rose, J. (2013). Crowding in public transport systems: Effects on users, operation and implications for the estimation of demand. *Transportation Research Part A: Policy and Practice, 53*, 36-52.
[8] Haywood, L., Koning, M., & Monchambert, G. (2017). Crowding in public transport: Who cares and why? *Transportation Research Part A: Policy and Practice, 100*, 215-227.
[9] Yap, M., Cats, O., & van Arem, B. (2018). Crowding valuation in urban tram and bus transportation based on smart card data. *Transportmetrica A: Transport Science*.
[10] Hörcher, D., Graham, D., & Anderson, R. (2017). Crowding cost estimation with large scale smart card and vehicle location data. *Transportation Research Part B: Methodological, 95*, 105-125.
[11] Hong, S., Min, Y., Park, M., Kim, K., & Oh, S. (2016). Precise estimation of connections of metro passengers from Smart Card data. *Transportation, 43*, 749-769.
[12] Tirachini, A., Sun, L., Erath, A., & Chakirov, A. (2016). Valuation of sitting and standing in metro trains using revealed preferences. *Transport Policy, 47*, 94-104.
[13 Mohd Mahudin, N., Cox, T., & Griffiths, A. (2012). Measuring rail passenger crowding: Scale development and psychometric properties. *Transportation Research Part F: Traffic Psychology and Behaviour, 15*, 38-51.
[14] Eboli, L., & Mazzulla, G. (2007). Service Quality Attributes Affecting Customer Satisfaction for Bus Transit. *Journal of Public Transportation, 10*(3), 21-34.
[15] HTM Personenvervoer NV. (2017). *Jaarverslag 2017.* HTM Personenvervoer NV, Den Haag.
[16] Byrne, B. (2013). *Structural Equation Modelling with AMOS: Basic concepts, applications and programming.* Routledge.
[17] Kruijff, J.S de (2019, July 23). Personal communication. (M. Seerden, Interviewer) Den Haag.

Table of Contents

1. INTRODUCTION 3

1.2. Research Question 4

1.3. Research methods 5

1.4. Report Structure 5

2. LITERATURE REVIEW 6

2.1. Customer satisfaction 6

2.2. Crowding 10

2.3. The link between customer satisfaction and crowding 14

2.4. Conclusion 18

3. METHODOLOGY 19

3.1. Case study introduction: Den Haag 19

3.2. Data collection: objective and subjective crowding 22

3.3. Statistical model building 26

3.4. Practical framework 28

3.5. Panel Effects 31

3.6. Conclusion 32

4. EXPLORATORY DATA ANALYSIS 33

4.1. Composition and representativeness 33

4.2. Survey Data HTM Klantenpanel 37

4.3. Crowding 39

4.4. Delays 43

4.5. Conclusion 44

5. RESULTS 45

5.1. Measurement model 45

5.2. Structural model 45

5.3. Panel Effects **53**

5.4. Conclusion **53**

6. IMPLICATIONS 55

6.1. Calculation model **55**

6.2. Impact of changes on customer satisfaction **57**

6.3. Growing passenger numbers and customer satisfaction **64**

6.4. Conclusion **66**

7. CONCLUSION AND DISCUSSION 67

7.1. Conclusion **67**

7.2. Policy Recommendations **69**

7.3. Academic Discussion **69**

BIBLIOGRAPHY 72

APPENDICES I

Appendix A: Questionnaires **i**

Appendix B: Search terms literature Review. **x**

Appendix C: Traveller characteristics and composition of HTM Klantenpanel **xi**

Appendix D: Data preparation **xv**

Appendix E: Structural Equation Model **xix**

Appendix F: Manuals of Excel tools **xxvii**

Appendix G: Translation of trip evaluation questions in HTM Klantenpanel **xxx**

1. Introduction

Cities continue to grow and the demand for urban mobility is ever increasing. Resulting from these trends, congestion in urban transport systems is posing a problem, growing in size. The effects of congestion differ per mode: for car traffic congestion leads to traffic jams, while in Public Transport (PT) other effects can be seen, of which crowding is an important one. Passenger frustration caused by crowding, for example, has led to PT operators being sued (van de Wiel, 2017). Crowding can have several negative impacts on the functioning of a PT system, including longer in-vehicle times and delays because of increased dwell time at stops, as well as increased passenger travel time due to denied boarding. Research has shown that the social cost of crowding can become equivalent to a significant portion of the uncrowded travel time (Hörcher et al., 2018). Of course, traffic jams can have a profound impact on PT as well in case of shared infrastructure (Tirachini et al., 2014). Crowding in a PT system can be a difficult problem to solve when the system operates near its capacity. In the Netherlands, several rail lines have this problem, such as the train line between Den Haag and Rotterdam and the Den Haag tram tunnel lines.

Currently, Public Transport operations in Europe are handled via concessions, which are given to an operator for a pre-determined amount of time, often a decade. During these years, services are continuously evaluated to see if the operator is indeed providing the service level which was promised during the tender. Several KPI's are used to determine this; these KPIs include a variety of factors. Objective indicators such as on-time percentages and cancellations are, of course, important, though most regulators prefer to also include subjective KPIs, which measure how people experience their journeys. An example is the concession for the Dutch main rail network (Ministerie van I&M, 2014).

The context described above introduces two critical concepts: the ever-growing demand for mobility in urban areas, leading to congestion as well as crowding in Public Transport, as well as the importance of customer satisfaction for PT operators working in an institutional alignment as seen in Europe. It is logical that congestion and crowding lead to lower customer satisfaction. As Freeman and Felleson (2009) concluded these subjective experiences do not match perfectly with objective measurements of performance. As a result, the effects of an increase in demand or capacity are difficult to translate one-on-one to their impact on customer satisfaction. This topic has yet to be thoroughly researched. The relation between these two factors will be the main topic of this thesis. The results of this analysis can be used to gain insight in at what level of crowding passenger satisfaction is at the minimum acceptable level set in a concession. Afterwards, these insights can be used to improve planning from an operator perspective. Moreover, information supply on crowding to passengers could be improved as a result of having more insight in their perception of crowding.

A valid question to ask in this context is why detailed insights in how travellers value crowding in Public Transport are necessary if capacity issues can also be solved by simply increasing capacity. The answer to this question is fairly simple: increasing capacity is often easier said than done. Problems in this case can be both of a technical as well as an economic nature. From a technical point of view capacity increases might be impossible due to the system operating at capacity, as described earlier. Other restrictions might be posed by the size of an operator's fleet or by stop layouts, which might for example prevent driving with longer vehicles. Changes to each of these fields take a long time to implement and often come with high investment costs.

However, even if technical capacity is present simply increasing the frequency can be unviable from an economic point of view. Often PT operators receive some sort of subsidy as part of their concession. In the Netherlands this is often done based on a fixed number of vehicle kilometres. Increasing frequencies without consultation of the governing authority therefore comes at high costs for an operator, which often results in not increasing frequencies being the financially more interesting option. As a result of both these arguments it is necessary to first optimise the current situation. In order to achieve this it is necessary to have detailed insights in how customer satisfaction is affected by crowding.

This research will be done based on a case study at HTM. HTM is responsible for all urban Public Transport in the city of Den Haag, under concessions granted by the Metropoolregio Rotterdam-Den Haag (MRDH). The most recent concessions given by MRDH to HTM started in December 2012 (urban bus) and December 2016 (tram). This concession included several requirements which HTM has to meet in order to have its operations being judged positively by MRDH. One of these requirements is that customers have to score their travel experience with HTM with at least 7.8 on a scale of ten on average. Besides, additional specific requirements are to be met on how customers score their probability of having a seat during their ride. It is known that these scores correlate. High customer satisfaction is therefore critical for HTM to be able to favourably keep its concession, both after the mid-term review in 2021 as well as after 2026. Over 2018, this requirement was met with an average satisfaction of 7.9 for HTMs tram and bus network and an average satisfaction of 7.8 for the RandstadRail network, the difference being explained by crowding on RandstadRail lines according to HTM (HTM Personenvervoer N.V., 2018). However, MRDH assumes continuous improvement in HTMs service level and as a result, the lower threshold of what is an acceptable customer satisfaction level increases from 7.8 now to 8.1 in 2026. However, in its vision HTM (2018) also notes that it, among others, strives to be 'an obvious choice towards their destination for even more people'. Passenger numbers over the past years also have been increasing and projections show that this will continue to be the case. As noted previously, transporting more passengers in the same system leads to higher load factors and thus lower customer satisfaction. These effects at some point will slow down growth. Simply solving this by increasing the frequency of services is difficult as HTM is paid by MRDH to drive a fixed number of vehicle kilometres, a construction which is not unique. Moreover, building new rail infrastructure to enable higher frequencies is expensive. If HTM provides more services they do not receive subsidies for those extra services which makes doing so often unprofitable and thus unattractive. Besides, HTM has limited fleet and thus the number of services provided is limited as well. It is therefore interesting to have more insight into the dynamics of crowding.

1.2. Research Question

Based on the previous section, a research question has been formulated.

What is the relation between objective and subjective in-vehicle crowding in Public Transport and overall customer satisfaction?

In order to be able to answer the research question a number of sub-questions has to be answered.

1. What affects customer satisfaction in Public Transport?
2. What is the definition of 'crowding' in an urban public transport network?
3. What role does crowding and seat availability play in overall customer satisfaction?
4. What factors affect customer satisfaction regarding seating opportunities?
5. How can crowding perception of a PT-traveller be predicted?
6. How does the predicted growth of travellers relate to the perception of crowding?
7. How does the prediction of crowding affect HTMs predicted performance?
8. How can HTM use this knowledge to improve its services?

The answers to these subquestions will be used to develop a conceptual model and empirically test this conceptual model afterwards. The results of this test can be used to answer the main research question. The next section will describe what methods will be used to answer these questions.

1.3. Research methods

All subquestions presented in the previous section will have to be answered in order to provide a full answer to the main research question. This section will deal with the methodology which will be used to answer these questions.

Questions 1, 2 and 3 provide a background into the topics of Public Transport Crowding and Customer Satisfaction in Public Transport. They aim to define key concepts. Moreover, these topics have been thoroughly investigated over the past decade or so, and it is crucial to gain an overview of what research has been done and to what extent valuable conclusions have been drawn. A critical literature review will be used to answer these subquestions.

Subquestions 4 to 7 are the questions on which this research will focus mainly. These subquestions require a less straightforward approach to be answered. These also are the questions which enable answering the main question. These sub-questions can be divided into two groups:

- Subquestions 4 and 5 aim at defining and understanding the relationship between crowding levels in an objective way and the customer perception of crowding (the first part of the main question). A statistical model will be built to achieve this. Structural Equation Modelling (SEM) is used as a modelling technique, as this technique is able to accurately model the complex relations between the variables involved and take into account latent variables.
- Subquestions 6 and 7 and 8 aim at how this knowledge can be used by HTM to improve their service levels in the future (the second part of the main question).

Chapter three will provide more information on the exact methodology which will be used.

1.4. Report Structure

This chapter has provided an overview of the context and goals of this thesis: it has provided the reader with the context in which this research is necessary, has elaborated on the main research questions that this thesis aims to answer and has provided a short overview of the methodology that will be used to answer these questions. Lastly, an overview of the report structure as well as the contents of each chapter will be provided.

The first part of this report will provide the reader with theoretical background and an overview of existing analysis. To start, chapter two provides an overview of existing research on the topics of crowding and customer satisfaction and provides an in-depth analysis of knowledge gaps existing in current research. This will result in a theoretical framework which will serve as a basis for the rest of the analysis. Chapter three will provide an overview of the methodology used in this research. It will provide an overview of which method will be used to answer which sub-question, along with some remarks on the methods used.

Chapter four to six will form the quantitative analysis. Chapter four will contain an exploratory data analysis on objective a subjective crowding for the Den Haag case, exploring a relationship between objective and subjective crowding. Chapter five will try to explain the relationship found by presenting and testing a statistical model. Chapter six will convert the insights which can be gained from this model to existing HTM growth models. Lastly, chapter seven will contain both practical and academic conclusions and recommendations for future work and research.

2. Literature Review

Research into the topics of customer satisfaction and crowding in Public Transport is not new. In the past a lot of different studies have been conducted from varied angles using various cases. The aim of this chapter is to develop a conceptual model based on and supported by existing research and literature on these two topics. Lastly, existing research on the relationship between these two will be discussed. It will be seen that the main research question as presented in chapter one has not yet been answered. Using the knowledge gained from all literature, a framework will be presented that does capture this relationship.

The literature used in this review was found based on several methods. Firstly, online databases were searched for relevant literature based on key words. If a paper was found, its abstract was read to determine its possible relevance. Besides searching for key words a search based on author was also conducted to find more work by key authors in the field. Lastly, literature recommended by experts was used if relevant, based on their suggestions.

2.1. Customer satisfaction

Customer satisfaction has become an important factor in evaluating the performance of a Public Transport system. This can be seen in practice, for example in the rail concession for the city of Den Haag (MRDH, 2017) and the Dutch main rail network (Ministerie van I&M, 2014), in which fines or bonuses are given if customer satisfaction falls below or rises above a defined threshold. A high level of customer satisfaction is thus crucial for a PT operator to maintain its position. Because of this increased institutional importance academic research into customer satisfaction in PT has also become much more relevant. This section gives an overview of this research.

To start with, the term 'customer satisfaction' has to be defined. Customer satisfaction is generally defined in literature as the gap between a customers' expectation of a service and his experience (Morfoulaki et al, 2010). Different research uses the term service quality for the same definition (Yaya et al, 2015), though service quality is often also used to describe the measurable performance of a Public Transport system (e.g. frequencies, travel speeds, reliability). This thesis will use the definition used by Morfoulaki et al. as a definition of customer satisfaction.

The factors affecting customer expectation and experience can be divided into two categories:

- Factors related to personal experiences and characteristics
- Factors related to PT system planning, operations and performance

This section will first discuss previous research into these two categories. A list of factors impacting customer satisfaction in PT results. For this list, it will then be investigated whether a hierarchy exists in these factors or if all factors can be considered to be equally important.

2.1.1. Traveller specific characteristics

Measuring customer satisfaction properly, however, is more easily said than done. To start with, the two components (experience and expectation) affect each other. How a traveller experiences his/her travels is modified by his or her expectations, while experiences also affect expectations for future travels (European Union RTD Programme - Project QUATTRO, 1998).

Moreover, large differences can exist between individuals in how they expect and experience an equal journey. These differences between travellers can be found both on a macro as well as a micro-level. On a micro-level, a variety of studies have shown demographic and socio-economic factors to be of influence on PT customer satisfaction, as well as general attitude and loyalty towards Public Transport. Table 2.1 shows an overview of these studies.

Table 2.1: Personal factors affecting PT Attitudes

Author	Year	Modality	Factors
Morton, Caulfield & Anable	2016	Rural Bus	Economic status[1], Age, Gender, Education level
Theler & Axhausen	2013	Urban Bus	Age, Frequency of PT Use
van Lierop & El-Geneidy	2016	Metro	Income, Car access
Van 't Hart	2012	Bus, Tram, Metro	Frequency of PT Use, Age, Gender, Location, Travel Purpose
Diana	2012	Bus (Urban and rural)	Frequency of PT Use, Location
Mouwen	2015	Bus, Tram, Metro, Train	Age, Past Experiences
Friman, Edvardsson & Görling	2001	Bus, Tram, Metro	Past Experiences
Abenoza, Cats & Susilo	2017	Bus, Tram, Metro	Frequency of PT Use, Age, Car Access
Yaya, Fortià, Canals, Marimon	2015	Bus, Tram, Metro	Age, Possession of Drivers' license, education
Koning, Haywood & Monchambert	2018	Metro	Income

As can be shown, a variety of different factors has been identified. Traveller characteristics which return the most are age, past experiences with Public Transport and frequency of PT used. Regarding past PT experiences it has been shown that negative experiences have a much stronger effect on attitudes towards PT than positive effects (Friman et al, 2001; Mouwen, 2015). Lastly, it has been found that mode choice availabilities explain why different people find different factors to be important, e.g. choice travellers valuing 'luxurious' aspects (e.g. comfort) more when compared to captive PT users (van Lierop & El-Geneidy, 2016).

It is interesting that the studies mentioned in table 2.1 find quite different characteristics. Three arguments can explain these differences. Firstly, often only a limited number of demographic or socio-economic characteristics was considered when trying to explain customer satisfaction. For example, past PT experiences as an explaining factor was only considered by Mouwen (2015) and Friman et al. (2001). Travel purpose was only mentioned by van 't Hart (2012), while Morton et al. (2016) were the only one to mention economic status. Gender is the only factor taken into account on which existing research has been unclear whether it is significant or not. Authors do not describe why they do or do not include certain characteristics.

A second explanation is that some studies aim to answer very specific questions, for example explaining differences in customer satisfaction between travellers (Yaya et al, 2015), while others focus on customer satisfaction in general. For example, Theler & Axhausen (2013) ask PT travellers to judge whether they find a bus full or not. They do not consider other aspects of the PT service. Their main conclusion, however, is that terms as 'full' or 'empty' do not mean much without context, which is a useful lesson to take into account.

Lastly, the system which was investigated can be a cause for different results. In rural networks network characteristics are different and this results in a different type of PT user as well as a different evaluation when compared to urban PT networks (van 't Hart, 2012; Diana, 2012). However, between urban networks differences can be found as well. For example, Haywood et al. (2017) found that having car access made no difference on customer satisfaction in Paris while van Lierop & El-Geneidy (2016) and Abenoza et al. (2017) concluded otherwise for Vancouver and Sweden, respectively. Even in comparable systems differences can exist. Based on a comparison between 9 European cities[2] it has

[1] Defined as the main occupation of time by a respondent (employment/retired/student/etc.)

[2] The cities included in this study were Barcelona, Copenhagen, Geneva, Helsinki, Vienna, Berlin, Manchester and Oslo.

been found that for different cities different factors were found to be having a significant impact on customer satisfaction, (Fellesson & Friman, 2008). This suggests that geographical and cultural characteristics can also have an impact on how PT is looked at. Factors such as city lay-out might be a reason for differences (e.g. in heavily congested cities with good PT systems such as Paris travelling by car might simply not be attractive enough for car accessibility to be important) as well as cultural differences (e.g. in North America people's attitude towards PT might simply be more negative than in Europe for cultural reasons) are good examples of geographical and cultural aspects. These need to be kept into mind when comparing different networks.

2.1.2. Service-related characteristics

As for traveller specific characteristics, literature has found a variety of service-related characteristics affecting customer satisfaction in Public Transport. This can be seen in literature on what affects customer satisfaction: a variety of factors has been identified as significantly having impact on customer satisfaction. Often these factors have been grouped into dimensions. Table 2.2 provides an overview of studies on PT customer satisfaction.

Table 2.2: Overview of service quality dimensions affecting customer satisfaction

Author	Year	Modality	Dimensions
Fellesson & Friman	2008	Bus, Tram, Metro	System, comfort, staff, safety
Abenoza, Cats & Susilo	2017	Bus, tram, metro	Customer interface, operation, network, travel time
Redman, Friman, Gärling, Hartig	2013	None[3]	Frequency, fare prices, speed, reliability
Eboli & Mazzulla	2007	Bus	Service planning and reliability, comfort and other factors, network design
de Oña, Eboli & Mazzulla	2014	Bus	Fare, information, courtesy, safety, accessibility, cleanliness, space, temperature, proximity, speed, punctuality and frequency
Yaya, Fortià, Canals, Marimon	2015	Bus	Functional Quality, Physical Environment Quality, Convenience Quality
Morton, Caulfield & Anable[4]	2016	Bus	Convenience, Cabin Environment, Ease of Use
Olsson, Friman, Pareigis, Edvardsson[3]	2012	Bus, Tram	Positive activation, positive deactivation, cognitive evaluation
Abenoza, Cats & Susilo	2018	Bus, tram, metro	Waiting times, Satisfaction with access and egress legs

The factors shown can be divided into two types of attributes: objective attributes, which can be measured (e.g. price, frequency, speed) and subjective attributes, which are purely the interpretation of the customer (e.g. comfort, safety). Some of the studies mentioned in table 2 tried to generalise the identified relevant attributes into a few latent constructs, some did not try to generalise, some considered all possible aspects of customer satisfaction, and some only considered subjective

[3] This is a literature review study, the dimensions found are an aggregate of other research.
[4] These study considered subjective attributes and thus did not find objective dimensions.

attributes. It can be seen that the number of dimensions obtained varied heavily between 3 (Eboli & Mazzulla, 2007; Yaya et al., 2015), for studies which used latent constructs, and 12 (de Oña et al., 2014) for studies which did not. What most studies did not consider is that perceptions are influenced by the journey as a whole, not just by the in-vehicle time (Abenoza et al., 2018).

For a long time it was assumed that the objective factors were the dominant attribute in determining customer satisfaction. However, after testing this assumption Fellesson and Friman (2009) conclude that this relationship is weaker than expected, finding amongst others a counterintuitive negative relationship between average PT speed and travel time satisfaction. They thus conclude that customer perception must play an important role in explaining customer satisfaction besides objective system performance. For example, an analysis of PT customer satisfaction in Sweden over the years 2001-2013 concluded that, besides travel time, the attitude of an operator towards its customers regarding information and disruption management was a driving factor behind an overall decrease of customer satisfaction (Cats et al., 2015). However, others have found more objective attributes such as price (Eboli & Mazzulla, 2007) or (perceived) speed and frequency (Morfoulaki et al., 2010) to be the most important for achieving higher PT service quality rather than subjective attributes.

It is interesting to note that almost all studies used survey data as a basis for analysis of customer satisfaction. For factors such as punctuality or occupancy, this means that not the objective attribute is measured but rather the perception of the attribute. The use of disaggregate objective data (e.g. actual punctuality, reliability and occupancy numbers) has not yet been used for PT customer satisfaction research. Aggregate data (i.e. average punctuality and occupancy) has been used but failed to produce results. It has been suggested that this results from detail which is lost because of aggregating data (Redman et al., 2013): if 95% of services is on time and 5% is severely disrupted the effects of this 5% on customer satisfaction will most likely be neglected. Whether disaggregate objective data (i.e. punctuality and occupancy per service) is sufficient is also debatable: De Oña and de Oña (2015) conclude that even if the data is not aggregated its heterogeneity and subjectivity results in a broad range of possible factors, while other studies found disaggregate data to be sufficiently explanatory (Şimşekoğlu et al., 2015). The best conclusion to draw based on all these studies seems to be that a combination of subjective and objective factors explain customer satisfaction. In a literature review on the topic Redman et al. (2013) also come to this conclusion.

2.1.3. Hierarchy of aspects affecting customer satisfaction

The identification of all these different factors affecting PT customer satisfaction does not mean, of course, that all factors are equally important. To some extent a hierarchy can be made. For example, travellers will not use a PT service which they feel is unsafe or unreliable regardless of its performance on all other factors.

Figure 2.3 gives a visual overview of such a hierarchy as developed by van Hagen (2011). Van Hagen states that five levels of factors which determine PT experience exist, which can be ranked qualitatively in a hierarchal structure (van Hagen & Bron, 2014; van Hagen & Sauren, 2014). This resembles the ranking of human life needs as developed by Abraham Maslow resulting in his well-known pyramid (Maslow, 1943). In Van Hagens (2011) theory, a higher level of comfort and passenger experience only becomes relevant after lower, basic needs such as safety have been met. The lower three categories can be described as 'dissatisfiers', which need to be at an adequate level to prevent discontent, the upper two levels as 'satisfiers' which positively affect a journey. It should be noted that this theory has yet to be empirically verified.

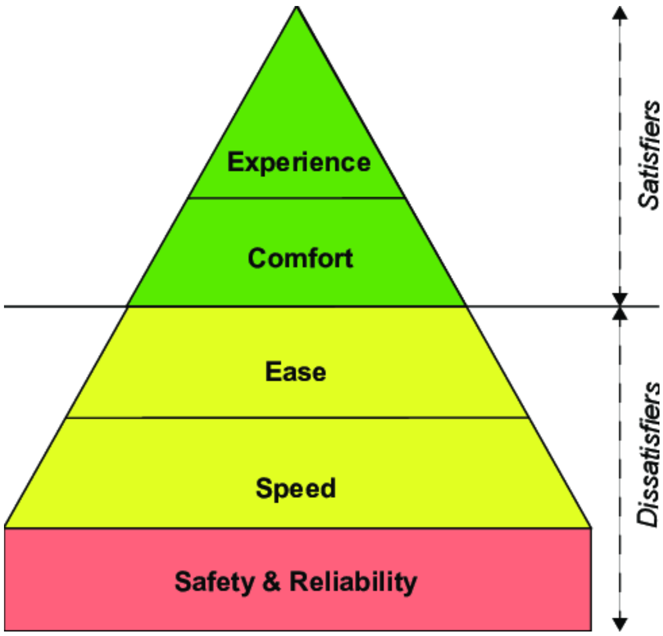


Figure 2.3: Hierarchy of PT Service attributes according to van Hagen (2011)

The same topic has been investigated in a quantitative way as well. Del Castillo and Benitez (2012) note that not every dimension has the same weight in determining 'overall' satisfaction and search for a model that can be used to determine this. While most of recent research tries to determine the importance of each factor statistically, Guirao et al. (2016) show that simply asking travellers to rank attributes in order of what they find important results in an almost identical order compared to calculating an hierarchy using much more advanced mathematical methods.

Lastly, the weather seems to have an effect on Public Transport experience and passenger numbers. Research by the Dutch Kennisinstituut voor Mobiliteit has shown that people in the Netherlands use Public Transport more on rainy days (Jonkeren et al, 2018). Different research by the same institute has shown that perception of waiting and travel times changes with the weather (Bakker et al, 2015). In conclusion weather seems to have an effect on customer perception, and thus on customer satisfaction. It should, however, be noted that little is known on the shape and size of this relationship.

To sum up, customer satisfaction in Public Transport can be defined as the difference between a customers' expectations and the service they experience. These expectations and experiences are affected by past PT experiences as well as socio-demographic factors such as age, as well as external factors such as weather. Network lay-out and cultural factors seem to play a role as well. In research, almost all studies only use survey data to find the different factors explaining customer satisfaction rather than objective data on crowding.

2.2. Crowding

As has been described in chapter 1, passenger numbers in Public Transport continue to increase in The Netherlands. Of course, each vehicle has limited capacity and at some point it is full. In academic research the term 'crowding' is often used to show how full a vehicle is compared to its maximal capacity. This section will provide an overview of relevant existing research on crowding.

In a broad sense crowding occurs when more people are located within a given space than is considered tolerable. As this is quite a general definition, it is important to be able to measure adequately what 'tolerable' exactly means. How crowded a specific number of passengers in a specific vehicle is can be measured using a variety of metrics. The two most commonly used ones are the Load Factor and standing passenger density (Tirachini et al., 2013). These are defined as follows:

- $Load\ Factor\ (\%) = \frac{number\ of\ passengers/veh}{number\ of\ seats/veh}$ (Whelan and Crockett, 2009).
- $Standing\ passenger\ density\ \left(\frac{pax}{m^2}\right) = \frac{number\ of\ standing\ passengers/veh}{space\ available\ for\ standing\ passengers/veh}$ (Wardman & Whelan, 2011).

Literature emphasises the necessity of using both Load Factor and Standing Passenger density as metrics for crowding, as both complement each other's strengths and weaknesses well. The density of standees is an unusable definition in the case of free seats resulting in a standing density of zero, while crowding effects might arise even before all seats are occupied (Wardman & Whelan, 2011). On the other hand, the Load Factor does not adequately accommodate heavily crowded situations in which a lot of passengers have to stand (Tirachini et al., 2013). This is the case because the number of seats does not say anything about the standing capacity of a vehicle. To illustrate this example two extremes are shown below in figure 2.4: on the left an image of the London 1996 tube stock, with a low number of seats and much standing space; on the right a NS VIRM, with many seats and little standing space. A load factor of 200% (i.e. twice as much passengers as seats) for the tube stock is still acceptable while the VIRM will be extremely overcrowded in such a situation.


Figure 2.4: Interior of refurbished London 1996 tube stock (left, source: [Wikimedia](), author: Aroura465) and NS VIRM4 (right, source: [Wikimedia](), author: Maurits90)

PT crowding has only become an increasingly explored topic over the past decade or so. Mostly, this research has been focused on the topic of 'crowding costs'. Crowding is then defined in an econometric way as a multiplier which is applied to the perceived travel time or costs in crowded circumstances. This enables the incorporation of crowding effects in transport models, which are used to estimate passenger numbers. Accurately estimating the crowding multiplier is difficult, as a large part of crowding costs consist of comfort losses, which are difficult to measure compared to time losses (Prud'homme et al., 2012). In transport models these comfort losses also form an iterative process: more passengers on a route means more crowding, which makes the route less attractive, which results in less passengers, which means less crowding, which makes the route more attractive, et cetera. For a long time, computers were unable to properly handle this iterative process. As a result, PT crowding was left out of transport models and consequently its effect often missed in transport research.

However, current computers are able to incorporate these crowding effects and as a result a variety of research into crowding multipliers has been conducted over the past few years. Research on crowding multipliers can also indirectly give insight in the effect of crowding on customer satisfaction. The main conclusion to be drawn is that a large difference in experience seems to exist between sitting and standing. A large difference in disutility (i.e. the experienced extra travel time or costs) was found to exist between standing and sitting, having to stand leading to an extra multiplier of around 1.2 in Hong Kong (Hörcher et al., 2017), 1.375 in Santiago de Chile (Tirachini et al., 2017) and 1.55 in Singapore (Tirachini et al., 2016). Moreover, respondents were found to be willing to pay more to sit instead of stand than they were to stand comfortably instead of cramped (Björklund & Swärdh, 2017).

The studies mentioned use crowding over a vehicle as a whole as a basis for estimating crowding effects. This method has a tendency to underestimate the negative impact of in-vehicle crowding. This can be explained by the fact that it neglects the effect of variations in passenger loads across vehicles (Cats et al., 2016). Figure 2.5 provides an example of this, using a vehicle which consists of three separate carriages with equal capacity. The load factor per carriage is shown in the figure, the load factor of the vehicle as a whole is:

$$LF = \frac{1.5 + 0.4 + 0.1}{1 + 1 + 1} \approx 66.7\%$$

Compared to the aggregate load factor which will be used in models a majority of passengers, who are travelling in the first carriage, experience a much higher load factor and thus more crowding effects.

Figure 2.5: Example of variations in passenger loads within a vehicle

Besides insight in the height of a crowding penalty, it is necessary to know at what load factor crowding starts to play a role. All studies show that when crowdedness increases the willingness-to-pay to mitigate crowding, either paid in time or money, increases as well (Li & Hensher, 2011). The graphical form of this increase differs per study. Wardman and Whelan (2011) found that crowding effects generally come into play at a load factor between 60% and 90%, depending on circumstances. For a higher load factor they assume that the size of these effects grows linear. This hypothesis was confirmed by an analysis of the Paris metro network by Haywood & Koning (2015), who found the relationship between passenger density and crowding costs to be linear. A different study on the Paris network by Kroes et al. (2013) finds similar conclusions. They compare their results with those by Wardman and Whelan (2011) and find a much lower multiplier for standing. Kroes et al. (2013) suggest the multiplier estimated by Wardman and Whelan simply is too high. A different explanation could be that Wardman and Whelan interviewed long distance rail travellers while Kroes et al. studied urban metro travellers. The former often travel much longer and as a result might experience much more discomfort from standing in the form of time loss.

While the previous authors often assume linear relationships, current computing capacities allow for non-linear approaches (Qin, 2014). De Palma et al. (2015) make use of this and disagree with the conclusion of a linear relationship, using:

- A constant penalty which is applied a passenger who has to stand.
- An additional exponentially growing penalty which increases when the passenger has to stand on a smaller surface.

A different research by Tirachini et al. (2016) investigates the trade-off between sitting and standing and travel time, in the process estimating how much higher the crowding multiplier becomes when travelling standing. It is found that given an equal number of passengers the crowding multiplier for standing passengers is much higher than for sitting passengers. This is in line with earlier studies, which often estimate different multipliers for sitting and standing (Wardman & Whelan, 2011; Kroes et al., 2013). However, this crowding multiplier is always estimated at a macro level (for a whole line or even system) and provides no information on individual passengers' experiences.

Research connecting crowding to socio-psychological factors such as customer experience (i.e. *why* does crowding have negative impacts?) is limited. This type of research can be able to tell much more in-depth how individual passengers experience crowding. Susilo et al. (2012) evaluate passengers' valuation of their travel time whilst travelling by train. They come to the conclusion that passengers appreciate their travel time significantly better if they are able to work or study whilst travelling, which becomes more difficult in crowded circumstances. Later research by Haywood et al. (2017) on why passengers experience discomfort because of crowding found having to stand, invasion of personal space and loss of time to be the main factors. Cantwell et al. (2009) found based on a Stated Preference survey that commuters travelling on crowded lines experienced significantly more stress than those who travelled on less crowded lines. Tirachini et al. (2013) and Mohd Mahudin et al. (2012) come to the same conclusion, adding that psychological factors (e.g. anxiety or discomfort due to limited space) play a key role in how a passenger experiences crowding besides objective measurements such as passenger density. Li and Hensher (2013) confirm these findings and add that the definition of crowding ideally is differed per mode: for a short bus ride it is acceptable to stand, while on intercity train services every passenger should be able to sit. This confirms discomfort is a result of a combination of high passenger densities and psychological factors. In line with earlier mentioned studies this suggests that crowding has consequences such as having to stand or having too many people around you. These result in passengers experiencing discomfort.

The conclusions as described above have generally been reached using data gathered from two different methods: stated preference (SP) experiments and smart card data. For a long time, estimations on (effects of) crowding have been done using stated preference experiments, as researchers were limited by a lack of information on actual occupancy. It is both very time-consuming and difficult to measure the actual number of passengers in a vehicle at large scale. In a SP survey which people were asked what they feel or do in a hypothetical situation, as is concluded by an overview on crowding research by Li and Hensher (2011). However, it is well-known that these stated preferences (intentions) do not always match with actual behaviour, as is explained by Ajzen (2005).

The growing use of smart cards in the PT industry, such as OV-chipkaart in the Netherlands, provides new opportunities for revealed choice analysis, in which actual behaviour and data is used. Smart card data can then be used to determine the actual number of passengers in a vehicle on a given moment. This type of data has been used over the past few years, mainly to be able to estimate crowding multipliers. Van Oort et al. (2016) have used smart card data to improve the estimation of the crowding multiplier in Public Transport, and afterwards Yap et al. (2018), Hörcher et al. (2017), Hong et al. (2016), and Ticharini et al (2016) have used smart card data. All authors try, for different situations, to quantify the effect of crowding this way. The resulting outcome is either a travel time multiplier for individual travellers or total travel time loss on the network. Their findings show that using smart card data can be a viable methodology to research crowding. However, their research primarily focuses on the economic factors of crowding.

To sum up, crowding in Public Transport can be defined as a high number of passengers compared to vehicle capacity. Most crowding studies estimate crowding effects from an econometric and macro point of view. Their aim is to find a crowding multiplier, which can be either a constant or a function of the number of passengers, which symbolises the extra travel time or costs a passenger 'experiences' whilst travelling a crowded service. It has been suggested that this penalty is not caused by crowding itself but by the discomfort caused by crowding. This discomfort is experienced as a result of loss of valuable time as well as invasion of personal space. This thesis will test whether this is indeed the case. This is useful in traffic models to represent how people choose modes and routes at a macro level. Research on the psychological aspect of crowding and its effects on individual passengers is limited. It is known that people experience negative emotions such as stress from crowding.

In all these crowding studies a major research gap can be seen. While a lot of studies have been conducted to convert objective data to crowding multipliers for traffic models and some studies have asked how customers experience crowding, objective data (such as smart card data) has never been connected to passenger experience.

2.3. The link between customer satisfaction and crowding

Sections 2.1 and 2.2 have elaborated on the topics of customer satisfaction and crowding. Intuitively, one can say that those two must in some way correlate, based on personal experiences as well as the number of complaints on busy trains which can be seen in the Netherlands (van de Wiel, 2017). This section explores current knowledge on the relationship between customer satisfaction and crowding. It will be demonstrated that while the link between these two variables is clear large knowledge gaps exist in this relationship. A framework for this relationship will then be presented that will serve as a basis for the analysis in this thesis which aims to narrow these gaps.

Firstly, section 2.2 has shown that a difference exists between the measurable (objective) crowding and the perceived (subjective) crowding. The role of objective in-vehicle crowding on PT services can vary. Firstly, high passenger numbers can have effects on travel speed and reliability: large numbers of passengers getting in and out of vehicles can lead to longer dwelling times which might affect punctuality (and thus reliability). Moreover, overcrowded vehicles might lead to slower travel speeds as drivers can decide to drive more careful for safety reasons. This affects system performance on a macro scale. Of course, this leads to passengers scoring reliability and speed lower when asked. This can be both due to the system actually performing worse but also because unoccupied, uncertain or uninformed extra waiting time (i.e. higher dwell times in PT) seem to take longer than they actually do (Maister, 1985).[5]

Regarding subjective crowding section 2.1 has shown that large differences can exist between people. An interesting follow-up question to ask is *why* this is the case. This 'why'-question is an important question to ask, as understanding why these difference exist can help in achieving better solutions for possible problems. Literature has found that differences in socio-economic, cultural and geographical background can provide an explanation, as well as past experiences with Public Transport.

[5] Maister gives little scientific evidence to support this claim, for which I paraphrase three of his eight 'waiting propositions' here. However, a variety of later research showed him to be right for at least seven of his eight propositions, including the three mentioned here. See Van Hagen (2011) for an overview of research on Maister's propositions.

A problem in asking this question is that academic debate does not yet agree on why people travel the way they do. In his lecture series on travel behaviour, Kroesen (2017) shows that a variety of underlying paradigms has been used to explain why people travel the way they do. For example, travel behaviour can be described as the outcome of a rational choice process or be driven by psychological factors. The question how travel satisfaction is determined is similar.

Often, it has been assumed that rational choice processes form the basis for choices people make. Most crowding studies which have been discussed in section 2.2 are based on this assumption. In the Rational Choice model used in all those studies, people weigh each aspect of each possible journey, based on travel time, comfort and the number of transfers, for example. Differences in customer satisfaction between passengers for the same trip are then explained by different valuations. For example, business travellers see travel time as more valuable than leisure travellers (Kennisinstituut voor Mobiliteitsbeleid, 2013).

This rational approach, however, is not the only one available. Psychological studies, such as Kahneman (2012), show that most of our choices are not based on rational thought. Van Hagen (2011) suggests psychology explains these differences better, mainly referring to Wundt (1910). Wundt developed the so-called inverted U-curve to note how people experience a service with arousal being the explanatory variable. Wundt (1910) notes that for a comfortable experience someone needs the right level of arousal: not too few and not too much. Figure 2.6 shows this relationship graphically.

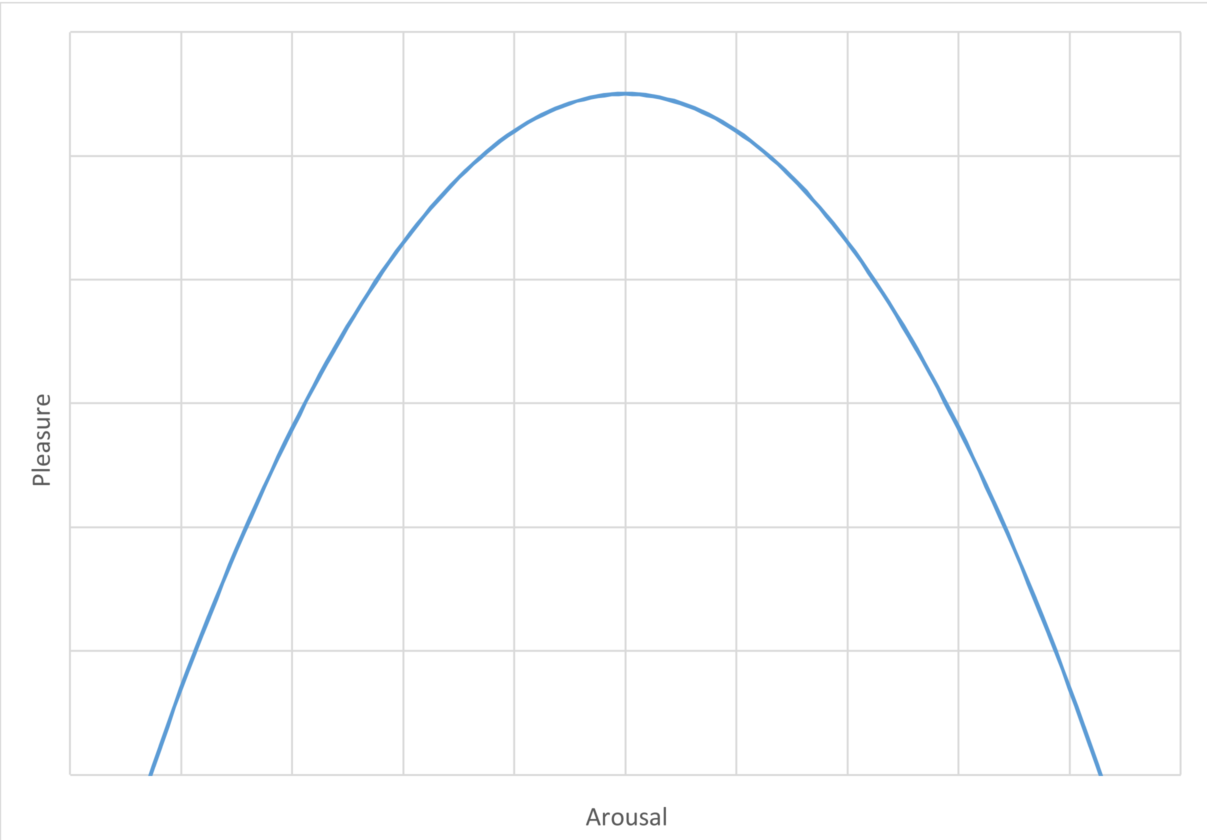


Figure 2.6: Wundt curve on the relation between pleasure and arousal

In the context of a passenger travelling on Public Transport, this means the following:

- When a passenger receives too few arousals, he or she gets bored or even anxious, which results in an unpleasant trip.
- If a passenger receives too much arousals, due to crowdedness, big loads of information, unclear information, etc., he or she gets stressed, also resulting in an unpleasant trip.

This theory suggests discomfort in crowded situations is due to an overload of arousal. This opens up possibilities for some out-of-the-box solutions to crowding discomfort, suggesting that lowering the total arousal helps in lowering the effects of crowding. Moreover, it offers a psychological framework to explain why different people (gender, ages, and groups) can experience the same trip in a different way, as the optimal arousal level can differ between people. In line with these conclusions, Mohd Mahudin et al. (2012) conclude that a passenger's experience of crowding consists of passengers feelings' beside passenger density and thus that subjective crowding differs from objective crowding. The additional factors found, mainly regarding comfort and safety, are in line with van Hagen (2011).[6] They conclude that their results 'challenge the norm (the generally used definition for crowding) and make the case for the role and significance of the psychological components in assessing crowding experience among passengers' and advise to change the conceptualisation of crowding accordingly. This is in line with the conclusions from section 2.2 which states psychological effects affect crowding.

While it is clear that a link might exist between both objective and subjective crowding and customer satisfaction, research on this link is limited. Haywood et al. (2017) are the only ones to directly ask how crowding affects customer satisfaction. Their conclusions, however, are of limited use. Based on surveys carried out during rush hour in the Paris metro a linear relationship between customer satisfaction and crowding was found. Income was found to be the only economic or demographic factor significantly having an effect on this relation, people with higher income were more dissatisfied due to crowding effects. However, due to the nature of the Paris transport network passenger densities in their data are often near the upper boundary of the spectrum and being able to sit is very unlikely. As a result, their findings are difficult to generalise to less crowded networks in which having a seat is a genuine possibility, or to off-peak situations in which crowding is not expected.

To conclude, research both shows that a link between customer satisfaction and crowding exists, most likely an indirect one. However, while some researchers (e.g. Mohd Mahudin et al. (2012)) have explored some parts of the relationship between crowding and customer satisfaction no one has tried quantitatively to capture either:

- The exact relationship between objective and subjective crowding
- The effect of both objective and subjective crowding on customer satisfaction.

This thesis will try to find and quantify this relationship. For this, a framework has been developed which explains the relationship between occupancy levels and customer satisfaction. Figure 2.7 shows this framework. All rectangle variables shown can be measured or calculated whereas all oval variables are latent variables. The relation which forms the main research goal of this thesis, the relation between occupancy rates, perceived in-vehicle crowding and overall customer satisfaction, is marked in red.

[6] The terms Mohd Mahudin et al. used to express the feelings affecting subjective crowding include cluttered, chaotic and disorderly – terms which match with the arousal overload van Hagen suggests in his work.

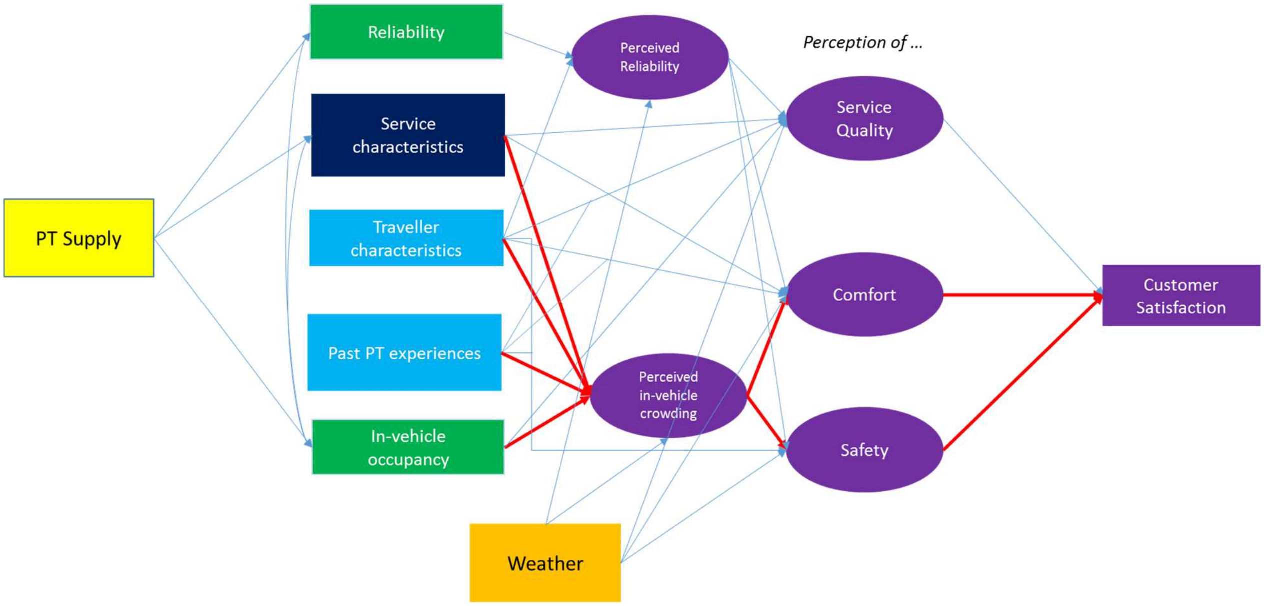


Figure 2.7: Conceptual model for explaining customer satisfaction including crowding

The framework is based on a wide variety of literature, which inevitably results in having to make choices and trade-offs. These will be discussed from right to left, starting with overall customer satisfaction. The first question to be asked is how to categorise the attributes which have a relevant influence on customer satisfaction. In literature, as shown in table 2.2, two approaches can be seen: either the relevant attributes are categorised into a few factors or all relevant attributes are linked directly to overall customer satisfaction. A majority of studies chose to categorise relevant attributes are into a few factors. Here, a construction with three explanatory dimensions is chosen:

- Service quality, which includes customer perception of the general service delivered. This includes speed, reliability and frequency, for example, but also information supply. This corresponds with levels 2 and 3 of PT service as defined by van Hagen (2011).
- Comfort, which includes all factors regarding travel comfort such as having a seat and enough space to relax or work. This corresponds with the fourth level of PT service as defined by van Hagen (2011).
- Safety, which states how safe a passenger feels during his/her journey) (corresponding with the lowest level of van Hagen's (2011) pyramid).

These factors all consider a variety of customer perceptions. In section 2.1, customer satisfaction has been described as the difference between expectation and experience. In order to properly model these perceptions, they have to be defined by both customer experience and customer expectation. As a result, these three factor are influenced by factors which determine customer experience and expectation at the level of an individual person:

- The performance of the service evaluated. This includes variables at both the tactical level (e.g. line frequency, vehicle type) as well as the operational level (e.g. actual delays, occupancy). These influence customer experience.
- Past Experiences with Public Transport. Literature emphasises the importance of this factor, which strongly affects customer expectation (Friman et al, 2001).
- Personal characteristics (e.g. age, gender, education) which give a personal framework in which the person experiences and colours his experiences and expectation.
- External factors which affect experiences. This includes, for example, the weather, which seems to affect customer perception (e.g. Bakker et al, 2015).

Most of these factors can be measured objectively – the frequency on the line travelled and the age of a traveller can be perfectly measured, for example – but it can still be questioned whether an additional layer of customer perception is optimal. For two factors, occupancy and reliability, it was chosen to include additional factors for perceived crowding and perceived reliability, as it seems that the perception of crowding and reliability can differ per person and service, regardless of objective occupancy and punctuality.

Zooming out even more the performance of the system (timetable, vehicle type, punctuality, etc.) is determined by high level choices on network lay-out, line and stop densities and infrastructure. These variables are captured in one variable in the model 'PT supply' and this reflects the indirect impact strategic choices have on customer satisfaction. Including this factor results in a complete framework with regard to all PT aspects which define PT customer satisfaction.

In the framework, it has thus been decided to model the relation between objective and subjective crowding and overall customer satisfaction indirect. This is in line with the conclusions found in section 2.2: not the number of passengers itself but the (dis)comfort and possible feeling of unsafety experienced as a result from high (low) passenger numbers has an effect on customer satisfaction.

2.4. Conclusion

This chapter aimed to answer three sub-questions as have been presented in chapter one:

1. What affects customer satisfaction in Public Transport?
2. What is the definition of 'crowding' in an urban public transport network?
3. What role does crowding and seat availability play in overall customer satisfaction?

This chapter has tried to answer these questions as thoroughly as possible based on existing literature.

Firstly, it has been found that customer satisfaction in Public Transport can often be defined as the difference between experience and expectation. Both are influenced by two main factors:

- The performance of the system. This includes both the lay-out of the system, which includes the route and timetable of the line, vehicle type used, etc., as well as the performance of the specific service which is used by the respondent, which includes, for example, in-vehicle occupancy and punctuality of that service.
- A variety of characteristics, such as age or past PT experiences, which vary per customer which define how the customer perceives the experience of the system.

In an urban public transport network, crowding occurs when too many passengers travel with a certain vehicle compared to the capacity of this vehicle. The optimal way to measure crowding uses a combination of two metrics: the load factor and standing passenger density. Literature is divided on the exact moment when crowding effects start to come into play. It is certain, however, that travelling standing or sitting has an effect on how a customer experiences crowding.

Lastly, a framework has been presented that explains how customer satisfaction in Public Transport can be explained. A major difference with earlier models on explaining customer satisfaction in Public Transport is the explicit addition of crowding, both objective and perceived crowding. Existing research has shown crowding to have a clear negative impact on customer satisfaction. In spite of this knowledge, the quantitative effect of crowding on customer satisfaction remains unclear: the size and shape of this relation is yet to be determined.

3. Methodology

Chapter one has introduced the reader to the topic with which this thesis will deal. Section 1.3 has already provided a very short overview of research methods, showing that the sub-questions can be divided into three categories: overview of existing research, exploring and quantifying the relation between objective and subjective crowding and customer satisfaction, and policy implications of the resulting knowledge. Chapter two has provided an overview of existing research. This chapter will elaborate on the methodology which has been used to answer the research questions as stated in section 1.2. This will be done by first introducing the case study, which is the city of Den Haag and it urban Public Transport network. Afterwards, the different data sources used will be discussed.

3.1. Case study introduction: Den Haag

As has been introduced in chapter 1, the research carried out in this thesis will be done based on a case study of the Dutch city of Den Haag. Den Haag is the third city of the Netherlands, the municipality having 540.297 inhabitants on April 1st, 2019 (Gemeente Den Haag, 2019). Its metropolitan area, however, is much larger. Currently Den Haag is considered to be part of the joint metropolitan area of Rotterdam-Den Haag, which had approximately 2.3 million inhabitants in 2017 (MRDH, 2019) in an area just over 1200 km^2.

HTM is responsible for all urban Public Transport in the city of Den Haag, under concessions granted by the Metropoolregio Rotterdam-Den Haag (MRDH). The network as of 2018 consists of 12 tram lines and 8 (up to December 2018) or 10 (starting December 2018) bus lines. Regarding passenger numbers, the latest year on which accurate numbers are available is 2017. In that year, HTM transported a total number of 100 million passengers in the Den Haag area: 84 million in trams, 16 million in urban buses (HTM Personenvervoer N.V., 2018). Figures 3.1 and 3.2 provides an overview of the HTM tram and bus network. As can be seen, the extent of HTMs network extends beyond the city boundaries of Den Haag to the neighbouring cities of Delft and Zoetermeer.

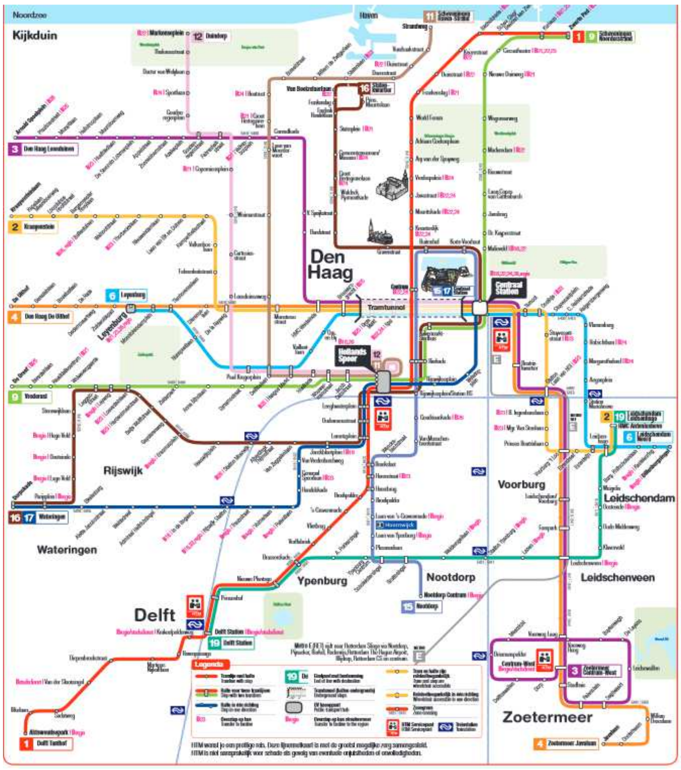


Figure 3.1: HTM Tram network (2018)

Figure 3.2: HTM Bus network (state up to December 2018)

This research will use the whole HTM network as a basis for all analysis considered. To be able to do this, detailed data is necessary, both on customer satisfaction and perception as well as on crowding levels on the HTM network. Section 3.2 will deal with data collection in achieving this.

HTM uses a variety of vehicles for its services. For its tram lines three tram types were used as of 2018:

- Figure 3.3 shows the GTL-8, the oldest type in use, introduced in the eighties. These trams were used on lines 1, 6, 12 and 16.
- Figure 3.4 shows the Alstom Regio Citadis, which were acquired by HTM mostly in 2006 when they took over the interurban lines to Zoetermeer from the Dutch national railways (NS). They are used on lines 3, 4, 19 and 2 (partly).
- Figure 3.5 shows the Siemens Avenio, which is the newest tram type owned by HTM. Its introduction started in 2015. They are used on lines 2 (mostly), 9, 11, 15 and 17.

In 2018, HTM used one bus type on all bus lines until the introduction of an electric bus in December, which will be left out of analyses. Figure 3.6 shows the MAN Lion City, the bus which operated all bus lines in all bus lines and trips included in the scope of the thesis. The rights for the photos used for figures 3.3 to 3.6 are all owned by HTM.

Figure 3.3: GTL-8 tram

Figure 3.4: Alstom Regio Citadis

Figure 3.5: Siemens Avenio

Figure 3.6: MAN Lion's City

As often is the case in large cities frequencies are high in Den Haag. Most tram lines have a frequency of 6 services per hour during weekdays and a minimum of 4 services per hour on late evenings and during weekends. Bus lines always have a minimum frequency of 4 services per hour during weekdays and 2 services per hour in the evening and during weekends.

This research will use the whole HTM network as a basis for all analysis considered. To be able to do this, detailed data is necessary, both on customer satisfaction and perception as well as on crowding levels on the HTM network. Section 3.2 will deal with data collection.

3.2. Data collection: objective and subjective crowding

To be able to quantitatively analyse the framework presented in section 2.4, data has to be used regarding crowding, both objective and subjective, and customer satisfaction. This section will deal with methodology on data collection and selection. It will both note which data is used as well as the advantages and disadvantages that come with it.

3.2.1. Objective crowding: Estimating crowding using Smart card data

To get insight into occupancy levels in the HTM network, OV chipkaart data will be used. The OV chipkaart is a smart card which is used by PT users in the Netherlands to pay for their travels. A user taps his smart card when boarding and leaving and the system uses this to calculate the costs of the trip. A main advantage of the system as applied in Den Haag is that the tap-in and tap-out points are located within vehicles. It is thus possible to determine at high levels of accuracy how many people have been travelling in a vehicle at a given moment. As a result, load factor or densities per vehicle as well as delays can be computed from raw data with no need for estimations. This does, however, also mean that less information is available on arrival patterns at stops and waiting times

Using smart card data as a basis for evaluating in-vehicle occupancy levels is not new. As section 2.2 has described, existing research on crowding has shown that smart card data can be a valuable source to get insight into occupancy levels in Public Transport. More specific, research by Yap et al. (2018) has shown that this also holds for OV-chipkaartdata in the Den Haag area. As a result it can be concluded that for this study occupancy data of the Den Haag network can be retrieved based on OV-chipkaart data.

3.2.2. Data on subjective crowding: Estimating customer satisfaction

Subjective crowding can be measured in customer satisfaction surveys. A variety of factors affect customer satisfaction. Section 2.1 has provided an overview of existing research on the factors that affect customer satisfaction. Here the methods used to measure customer satisfaction at HTM will be discussed, as these datasets will be combined with HTMs OV-chipkaart data in the analyses carried out later in this thesis.

HTM uses two survey methods to measure customer satisfaction, the nationwide 'OV-Klantbarometer' and its own customer panel ('HTM Klantenpanel'). These questionnaires can be found in appendix A. The OV-Klantbarometer consists of surveys carried out in-vehicle on a quarterly basis. This survey provides data on a variety of aspects regarding customer satisfaction, including several customer specific characteristics such as age, gender, travel purpose and travel frequency. These surveys are carried out in-vehicle and used by both HTM and the MRDH. They form the basis for the evaluation of customer satisfaction by all these institutions, and thus a basis for the evaluation of the Key Performance Indicators used to evaluate the subjective performance of an operator, being customer satisfaction. The exact line and trip number in which a participant fills in the survey are noted, which means the actual performance of the trip in which a survey was filled in can be traced afterwards. However, the exact location *within* the trip remains unknown. This poses problems when analysing lines with varying characteristics, such when comparing the urban and interurban segments of lines 3 and 4, as well as long lines with huge differences in load factor along long routes (in Den Haag line 1 might be an example). A more recent issue on using the OV-Klantbarometer data is that trip numbers were no longer recorded starting from 2018. This means that if this type of data is to be used the 2017 data is the latest usable data.

Using the data in this survey also has several advantages. Firstly, it offers the most representative dataset. Moreover, this survey uses the data which is also used by MRDH to measure customer satisfaction in evaluating HTM, which ensures more practical usability of the results for HTM. Secondly, as this survey has been carried out by HTM for quite some time, it enables the usage of a large data set. Thirdly, the way this survey is done enables analysis of customer satisfaction in-depth at a detailed level, even vehicle level for the in-vehicle surveys as these were up to 2017 linked to the exact vehicle and time used. However, when weighing pros and cons the problems described with adequately coupling responses to the actual occupancy levels at the moment of responding mean that in its current form OV-Klantbarometer survey data are not suitable for the research carried out here.

A second source of data comes from HTMs own customer panel (HTM Klantenpanel). Anyone can voluntarily subscribe for this panel. When subscribing a respondent provides his or her personal characteristics. Afterwards, HTM asks its members to evaluate a trip every month. These responses are then coupled to the provided personal characteristics for analysis. The basic questions on trip experience asked are the same as in the OV-Klantbarometer, but overall the trip evaluation data of HTM Klantenpanel is much richer: it provides more detailed information on the type and length of the trip, by providing the exact stop and time at which a customer boarded and alighted the vehicle. Moreover, respondents can comment on their results, which can provide more information on why specific answers are given in a certain situation. As subscribing to the panel is voluntary and members are asked to evaluate a trip once a month, using the HTM Klantenpanel data is prone to panel effects, due to respondents answering the survey more than once. Most statistical analysis methods by default assume independence between responses and this is not the case if one respondent is responsible for multiple responses.

Moreover it is reasonable to assume a bias, as people with a stronger opinion will be more likely to subscribe to a panel. Over the year 2018, the HTM Klantenpanel received approximately 3800 responses by 400 respondents evaluating trips. Section 4.1 will explore the representativeness of this group of respondents.

While HTM uses its Klantenpanel for more surveys than just trip evaluation, the trip evaluation survey will be used in this thesis. Table 3.7 gives an overview over which factors a customer is asked to rate for a trip when evaluating one, presenting both the Dutch label as well as the corresponding English translation. These aspects are sorted as asked when responding to the survey. A respondent is asked to rate each of these aspects with a mark ranging from 1 to 10. Besides, a respondent can answer 'I do not know' or 'this question does not apply for my trip'. These marks can be used in analysis as a measurement of a customers' satisfaction with a certain aspect. For easy translation, table 3.7 is also to be found in appendix G.

In order to be able to match these opinions to the service used the respondent is also asked to provide the following information on the service he or she is evaluating:

- The line travelled
- The direction travelled
- At which stop the respondent boarded the vehicle
- At which stop the responded left the vehicle
- The date and time, accurate to the minute, at which the respondent boarded the vehicle.

This provides enough information to couple the response to a specific trip number.

Table 3.7: Overview of satisfaction questions in the HTM Klantenpanel trip evaluation survey

Item	Translation
Totaaloordeel	Overall satisfaction
Zitplaatskans	Probability of finding a seat
Veiligheid OV	General feeling of safety in PT
Comfort	Comfort during trip
Netheid Voertuig	Cleanliness of vehicle
Netheid Halte	Cleanliness of stop
Info halte	Information supply on stop
Info Vertraging	Information supply during delays or disruptions
Klantvriendelijkheid	Friendliness of staff
Rijstijl	Driving style of driver
Gemak vervoersbewijs	Ease of buying a ticket
Stiptheid	Punctuality
Frequentie	Frequency
Veiligheid Rit	Feeling of safety during this trip

In conclusion, two main sources of data exist for the analysis of customer satisfaction on the HTM network. Each of these two sources has its advantages and disadvantages: the OV-Klantbarometer uses data from respondents which can be seen as more representative and objective, while the HTM Klantenpanel data is easier to use due to providing richer information. Given the current form of both surveys the OV-Klantbarometer is not suitable for the analysis as will be done in this thesis. As a result, the responses to HTM Klantenpanel will be used as customer satisfaction data in all analyses.

3.2.3. Combining objective and subjective crowding data

Having concluded that valid and rich data sources exist both for objective crowding, subjective crowding as well as customer satisfaction this data can be analysed. Several steps need to be taken, however, before this can be done:

- The survey responses require data preparation before data analysis is possible.
- In order to analyse the effect of objective crowding on subjective customer experience, survey responses need to be coupled to corresponding data on occupancy and reliability.

This section will describe how this has been done. Appendix D further elaborates on the changes and assumptions which were made on the data to ensure its usability in analysis.

To start with, the survey data needs to be made uniform in a way that allows for meaningful and correct analysis. The following list provides an overview of the changes and assumptions made:

- Respondents fill in their year of birth. Their age as of January 1st, 2019 can be calculated based on this and was used during analysis.
- The moment of evaluation is important to note, as this is not necessarily during or immediately after the trip – a trip can also be evaluated several days later. The analysis assumes that this does not affect the scores which are given.
- Respondents are asked to evaluate the last trip they made with HTM. All responses in which the values filled in can cast any doubt on which exact line used between what stops and at what time were removed. This is mostly due to missing values, respondents evaluating a trip

- not operated by HTM, filling in an incorrect line number with origin and destination allowing multiple lines or filling in 'temporary stop' as origin or destination.
- Buses replacing trams due to planned construction works were not considered in any analysis and thus removed from the dataset.
- If either 'overall trip satisfaction' or 'perceived crowding' was not graded or grade with 'I do not know' or 'does not apply for this trip' the response was deleted, for two reasons:
    1. These two marks are the ones this thesis is most interested in, so responses with these questions unanswered offer no usable information for analysis.
    2. It makes no sense that the respondent is unable to answer these questions or thinks they do not apply to his travels.

This left a dataset of 3738 trip evaluations which based on survey responses were considered to be usable for analysis. Next, these trip evaluations have to be combined with actual occupancy numbers. Up to present, this has not been done at HTM. An Excel tool has been built in order to accommodate converting the input data from a response on HTMs Klantenpanel to be coupled to the occupancy of the trip reviewed. The manual for this tool can be found in appendix F.1. The procedure followed to do so is seen in figure 3.8 and described below.

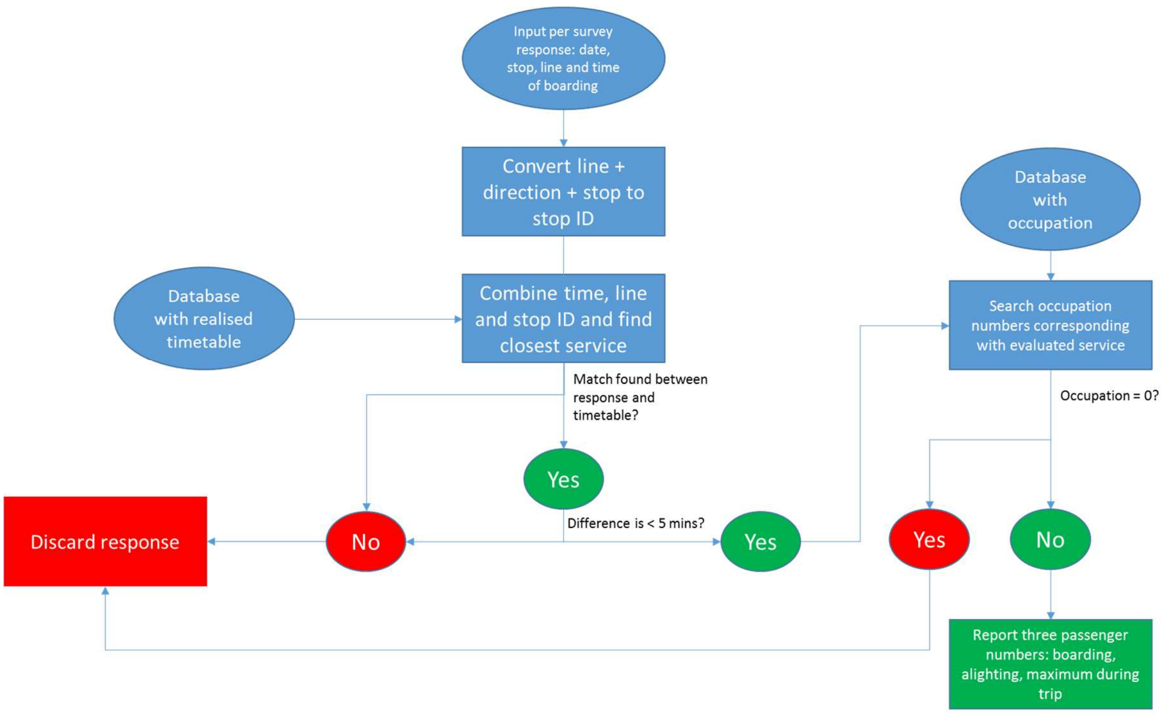


Figure 3.8: Process of combining survey responses to occupancy numbers

Firstly, the response values given were converted to values which HTMs SQL-databases can understand. This meant that all input on where a respondent boarded and alighted from a vehicle was converted into a corresponding stop ID. This was done in Excel, using tables containing all stop ID's for all stops on a line. Afterwards this table was put into the SQL-database to retrieve the corresponding service, using the following procedure:

- Given the line number, direction, time and stop at which the respondent reported to have boarded a vehicle the system reported all vehicles which left this stop **five minutes or less** from the time reported. This choice is made based on the <u>realised</u> timetable on the day reported. The choice for five minutes as threshold was made based on most tram lines driving each ten minutes during working days, resulting in at least one vehicle being at most five

minutes away under normal circumstances. The trip number of the service this vehicle was driving is then remembered by the system.
- If more than one vehicle of this specific line passed the stop less than five minutes from the time reported, the closest vehicle in time was taken. No difference was made between early or late vehicles.
- If no vehicle of the line reported passed the stop mentioned in the direction mentioned within five minutes of the time reported the time and stop reported are considered unreliable and the response was deleted by the database.

If a trip number was determined the database then requested the occupancy numbers for this service. As an output three occupancy numbers were generated:

- Vehicle occupancy at the stop of boarding
- Vehicle occupancy at the stop of alighting
- The maximal vehicle occupancy during the trip (including boarding and alighting stops).

The software carries out the matching process very precisely: if a response was unable to meet *any* of the requirements necessary for analysis errors were returned. As a result, any response containing any form of error was excluded from all further analyses. If no errors were found in the process above for a response this unfortunately did not mean usable output could be retrieved for two reasons. Firstly, sometimes the trip made by the customer did not match the trip made by the service the response was combined to.[7] Due to software and time restrictions it was too difficult to correct this mistake. Secondly, some services returned an in-vehicle occupancy of zero for all stops due to faulty smart card software. These services were identified by finding those which have a maximal occupancy of zero. Both categories of responses were discarded as well. After all these transactions, a dataset of 2858 down from 3738 responses (76.5%) remained for analysis.

Lastly, for each of these 2858 responses punctuality numbers were searched and reported. This was done by taking information used to get the occupancy numbers (service number, IDs of the stops at which the passenger boarded and alighted) and reporting both the planned and actual departure time of that specific service at these stops. By then subtracting the actual departure time from the planned departure time punctuality values could be retrieved in seconds, which is the unit used in HTMs planning systems.

3.3. Statistical model building

While exploratory data analysis can provide some general insight on the data, using only these techniques is insufficient in gaining more insight in the relationship between the different variables. The aim is to formulate and test a statistical model which explains causally how customer satisfaction is affected by crowding. The resulting model will answer sub question 5 as formulated in section 1.2: *How can crowding perception of a PT-traveller be predicted*? Several methods are able to formulate such a model. Table 3.9 below provides an overview of techniques which could be used to analyse relationships within a given dataset.

[7] This is caused by rush hour additions (in HTM terms: korttrajectritten), which are stored in the system as a service of a specific line but do not drive the whole length of its route. In case of a slightly incorrect boarding time results in the algorithm matching the response to a 'korttrajectrit', which does not stop at the alighting stop reported.

Table 3.9: overview of methods for data analysis (based on Kroesen, 2017)

Method	Advantages	Disadvantages
Basic statistic techniques (e.g. linear regression) (Montgomery et al., 2015)	- Easy to use and understand - Requires little computational power.	- Not suited for complex analysis between large sets of variables - Theoretical support necessary for conclusions on causality
Structural Equation Modelling (Nachtigall et al., 2003)	- Able to model complex multi-level relationships large numbers of variables - Able to incorporate latent and subjective variables	- Theoretical support necessary for conclusions on causality - Types of relations that can be tested is limited
Machine Learning (Kubat, 2017)	Able to use advanced mathematical functions, many iterations and algorithms to approach optimal solution	Risk of 'black-box' methodology in which the process leading up to results becomes unclear
Qualitative Analysis (Lapan et al., 2012)	- Highly flexible approach - Able to embed a phenomenon in its social context	- Limited possibilities to generalise conclusions - Limited options to draw quantitative conclusions

The method which will be used in building a statistical model in this thesis is Structural Equation Modelling (hereafter: SEM), as this offers the best trade-off between advantages and disadvantages given time and software availability combined with the authors' knowledge. SEM was developed for quantitative research in the social sciences. As a result, the technique is able to model complex relations between variables and to accommodate for latent variables. Several software packages exist for SEM. This thesis will use AMOS, a package developed by IBM (Byrne, 2013). AMOS was chosen as this was the software available to the author via an academic license, whereas other SEM software programs were not.

SEM models consist of two main components, a structural model and measurement model(s). One of the advantages of SEM is its ability to work with latent variables, variables which are derived from other variables rather than directly observed. Measurement models are used to examine the relation between latent variables and their measures. Section 3.4 will discuss how the latent variables seen in figure 2.7 will be measured. If a measurement model for latent variables fits data sufficiently, it can be assumed that this measurement model is properly able to measure the latent variables.

The structural model shows quantitative causal relationships between the latent variables. The measurement model ensures that latent variables in the structural model are properly represented before estimating the structural model. Thus, if the measurement model does not fit adequately it makes no sense to estimate the fit of the structural model. The structural model is seen as a system of pre-determined independent regression equations which provides the input for an optimization problem. Software then estimates for what values the set of equations fits the data best and how good exactly this set of equations approaches reality.

In estimating a Structural Equation Model the software is provided with an input which already defines which causal relations exist and which do not. SEM is thus primarily a confirmatory modelling technique: the model ultimately tests if a proposed model ultimately fits the dataset. It is thus crucial that the proposed model makes sense. Otherwise testing a model does not make sense, since correlation does not equal causation. When using statistical methods a theoretical framework is crucial to be able to translate observed correlation into causation. The framework presented in figure 2.7 is embedded in literature and will form the basis of the structural model, although slight adaptations are necessary: section 3.4 will discuss these changes.

SEM also has several possible issues that should be noted. The main remark to be made is that SEM estimates whether the *proposed* model fits the data or not. The software tests whether the suggested relations are statistically significant and can also suggest if adding or removing links improve model fit, but this is done empirically rather than theoretically. This means the software will advise any relationship containing a high correlation to be added to the model, while correlation, of course, does not equal causation. This means, in evaluating a model theory should be leading in estimating the exact model structure and not the extra fit achieved by adding one extra random correlation (Nachtigall et al, 2003). A poor theory will most likely result in an inadequate model.

Moreover, SEM is a large sample method. This means that the data sample analysed should be large, in general having a size of at least 200 samples, for SEM to be used properly (Nachtigall et al, 2003). Based on the available data described in section 3.1, the sample size does not pose a problem.

3.4. Practical framework

Section 2.3 has presented a framework to capture how customer satisfaction can be explained based on a combination of the service which is actually provided and how customers experience this service. The preceding sections have shown that quite some information is available on these topics. However, not all parameters and variables which are described in the framework presented in section 2.3 can be used for this specific research. Two main reasons can be identified for this:

1. Some variables are simply not known given the dataset used. This is a drawback of using existing survey data rather than collecting own data. However, due to time constraints it was chosen to use existing data.

2. Some variables could theoretically be retrieved or calculated from the data available without extra data collection. However, in some cases this was found to be extremely time-consuming. Given the limited time scope of this research it was sometimes decided not to invest this time.

The following section gives an overview of which variables drop out of the model as a result of these issues. Factors which have to be excluded are:

- Reliability is reduced to whether the service taken was punctual or not. Ideally reliability aspects which include other services than the one taken by the respondent are taken into account. For example, it might be useful to know whether the preceding service dropped out or was early. However, including this in the model requires a lot of work, as it is difficult to retrieve the punctuality of the preceding service of a given service using other means than handwork. Creating an algorithm which can use a given service as input and produce the punctuality of the preceding service(s) as output would require a lot of time. As a result, this is left out of the model.

- Secondly, the whole strategic variable 'PT Supply' was decided to be left out of the model completely. Again, this comes as a result of a trade-off between time investment given limited time and the added value of doing the research on the model. On a low level of detail, these variables have no additional explanatory power at all (as values will be the same for all of Den Haag), calculating them on a high level of detail will be extremely time-consuming. An additional problem in this situation is that line and stop densities can vary heavily between neighbourhoods of Den Haag and thus between sections of lines. Moreover, after discussion with experts at HTM it was concluded that if the effort would be done the conclusions would be of limited value to HTM: changing factors such as line density for the tram network would be extremely difficult.

- Lastly, the opinion (and thus perception) of customers on several aspects mentioned in literature are simply not asked in HTMs Klantenpanel trip evaluation list and thus not known.

Table 3.10 provides an overview of all variables which have to be left out.

Table 3.10: Overview of variables left out of analysis.

Dimension	Variable	Reason	Remark
PT Supply	Network structure, line density, stop density	Too time-consuming to determine, large differences between line segments, limited added value.	
Weather	Temperature, precipitation, wind speed	No knowledge is available on the exact impact of weather on customer satisfaction.	
Reliability	Cancellation	Too time-consuming to determine	
	Regularity	Too time-consuming to determine	
Traveller Characteristics	Previous PT Experiences	Not known	
	Income	Not known	
	Car access	Not known	
Perceived Service Quality	Perception of speed	Not known	Has been added in survey since January, 2019.
	Perception of waiting times	Not known	Could be added to survey.
	Perception of price	Not known	Could be added to survey.

The resulting framework, which will be used in analysis, can be seen in figure 3.11.

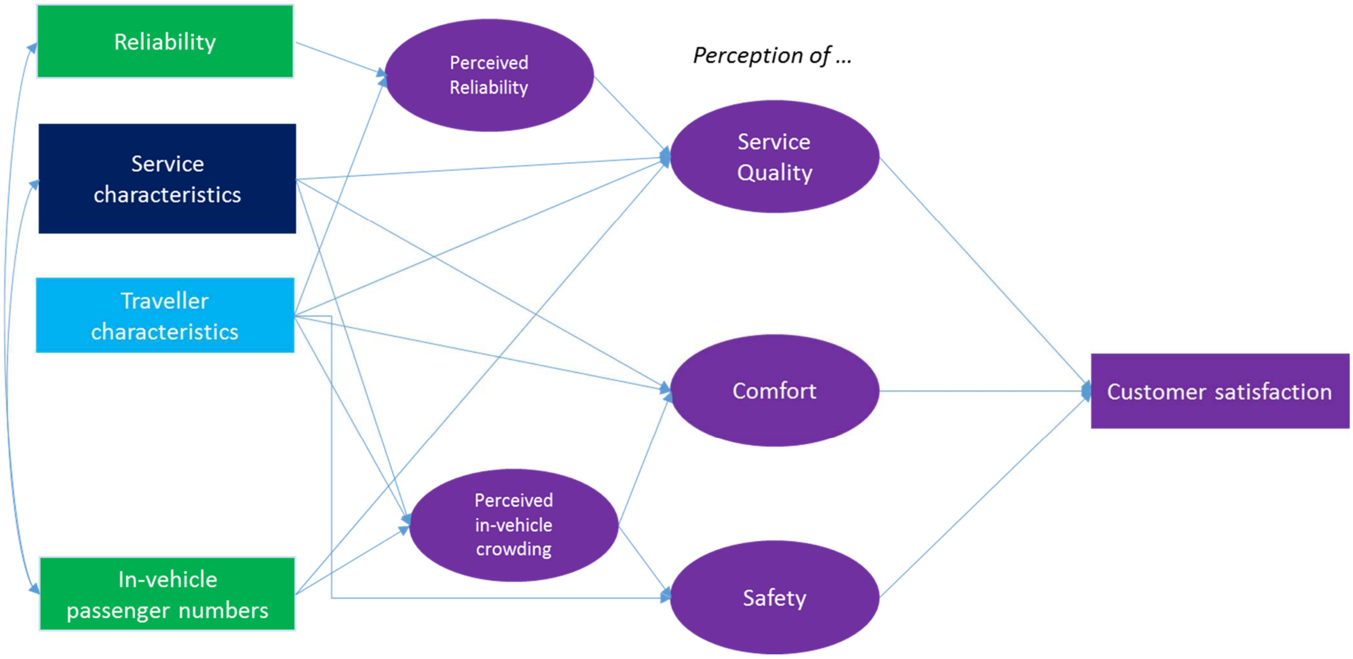


Figure 3.11: Framework used for analysis

All measurable variables mentioned in this framework need to be operationalised. Table 3.12 shows an overview of how each specific variable is measured and what unit and level of detail will be used in analysis.

Table 3.12: Operationalisation of variables

Category	Variable	Unit
Reliability	Departure delay	Seconds [s]
	Arrival delay	Seconds [s]
Service Characteristics	Frequency	Veh/h
	Vehicle Type	Type
	Moment of travel	Peak/off-peak
Traveller characteristics	Age	Year
	Gender	M/F
	PT use Frequency	Times per week
	Education	[low/middle/high]
	Travel Motive	[commute/business/education/leisure/other]
In-vehicle occupancy	Load Factor	$\% \left[\frac{\frac{pax}{veh}}{\frac{seats}{veh}}\right]$
	Standing passenger density	$\frac{\frac{pax}{veh}}{\frac{m^2}{veh}}$

An overview of subjective variables has already been presented in table 3.7. Table 3.13 shows how all evaluations by customers are allocated to the latent variables as shown in figure 3.11.

Table 3.13: Allocation of subjective evaluation aspects over factors

Factor	Attribute
Comfort	Comfort
	Cleanliness of vehicle
	Cleanliness of stop
	Friendliness of staff
	Ease of buying a ticket
	Driving style of driver
Customer Satisfaction	Overall customer satisfaction
Perceived Crowding	Probability of finding a seat
Perceived Reliability	Punctuality
Safety	Feeling of safety during this trip
	General feeling of safety in PT
Service Quality	Information supply on stop
	Information supply during delays or disruptions
	Frequency

During operationalisation, using a Structural Equation Model poses two challenges:

- SEM is unable to use choice rules. This means that the ideal measurement of crowding as described in section 2.2 (Load Factor for low occupancy levels, standing passenger density for high occupancy levels) cannot be used in the model.
- SEM offers a limited set of possible mathematical functions, only being able to incorporate linear and quadratic functions. This poses, of course, problems when investigating relationships which cannot be represented with linear or quadratic functions.

Regarding crowding and delay values several options exist. Chapter 4 will check if a linear or quadratic function approaches the relation between in-vehicle occupancy and perceived occupancy sufficiently and explore the advantages and disadvantages for each of the possible variables and make a choice as to what value can be used best.

3.5. Panel Effects

An assumption made by AMOS in estimating the model is the assumption that each data point is measured independently. However, for the HTM panel this is clearly not the case, as many respondents have evaluated more than one trip: these trips correlate with each other. This effect could be quite strong: one member made 79 responses in 2018 and six members out of 743 are responsible for 10% of responses. The model will be tested assuming that each response has been made independently, thus neglecting these panel effects.

To test whether this simplification affects conclusions, the dataset has been adapted to a situation in which the independence assumption of responses can reasonably be assumed, i.e. in which it is certain that only one response per respondent is considered and all others are deleted. This is not the best way to deal with panel data: ideally if more than one response per person exists these responses are combined to get an optimal impression of this respondent, as in such a case no data is lost. However, the software available is unable to do this. Software able to estimate Structural Equation Models using panel data does exist, however – in future comparable research it can be useful to use this software.

The dataset is thus reduced to one response per respondent to test for panel effects. The data as available to HTM provides no respondent ID so different measures are necessary. To identify different respondents the postal code (six digits) is a useful metric. Insight in the composition of the panel, which is known to HTM, shows the postal code to be unique for over 99.5% of panel members, which is sufficient for the goal of the analysis. All responses were sorted using a random variable (in this case: number of the day a trip was evaluated), and afterwards the dataset was tested for duplicates by SPSS, whereby an identical postal code was set to identify duplicates. Each last response (in the order responses were sorted in) by a certain postal code was coded on a new variable with a 1 and all other responses with a 0. Next, all responses which contained a 0 on this variable were deleted. After this analysis, a dataset of 743 responses out of 2858 remained for which independence between respondents can be assumed with certainty. The model will be estimated again using this dataset in section 5.3 and afterwards outcomes will be compared.

3.6. Conclusion

This chapter has presented the methodology which will be used to answer the research questions. This thesis will use a case study of the urban Public Transport Network of Den Haag to do so, using the full network of HTM, urban PT operator in Den Haag. The data used to get insight into customer satisfaction will be all responses over 2018 to HTMs own customer satisfaction response group (HTM Klantenpanel), for which passengers can voluntarily subscribe. This is done because of the high level of detail on the exact trip made which is available for this data. For each response, it is tried to retrieve occupancy and punctuality data on the service which a respondent to the HTM Klantenpanel evaluated. In section 3.4, the theoretical framework as presented in chapter 2 was converted to a practical framework based on limitations imposed by data or time restrictions.

Structural Equation Modelling (SEM) using AMOS will be used to estimate a quantitative model based on the framework. This method was chosen due to SEM being able to evaluate complex relations between many variables.

4. Exploratory Data Analysis

As has been described in chapter three, this research uses a variety of survey and smart card data to investigate the effects of in-vehicle crowding on customer satisfaction. Advanced statistical modelling techniques will be used to achieve this. However, to use these techniques to full strength and interpret results correctly, it is important to also have general insight in the data used. To achieve this, an exploratory data analysis has been carried out on the data used. This chapter will present the results of this analysis.

This chapter is structured as follows: section 4.1 will discuss the demographic characteristics of the respondents and assess to what extent the respondents accurately reflect the group they should represent. Afterwards the data used will be explored. Section 4.2 will discuss the survey data, section 4.3 will investigate the crowding data coupled to survey responses and section 4.4 will look into delay data.

4.1. Composition and representativeness

The analysis will use two main sources of customer satisfaction as described in chapter three: HTMs Klantenpanel, and the OV-Klantbarometer. The responses to these surveys will be analysed. To draw the correct conclusions from these analyses, it is important to know whether respondents to these surveys properly reflect the groups they are supposed to reflect. This section provides some information on the composition of the respondent groups and reflects on whether they accurately represent the group they should represent: HTM travellers.

To provide information on whether a sample is representative for a group as a whole, first the demographics of HTMs total customer pool should be determined. HTM does allowed keep personal information on each of its customers and research shows that not all people use Public Transport. Research has shown that only a small portion of the population uses Public Transport frequently (Zijlstra et al., 2018; Bussink & De Konink, 2015). A first task is therefore to find the demographic distribution of PT travellers.

Bussink and De Konink (2015) in their analysis compare demographic characteristics of responses in the OV-Klantbarometer to general demographics. They implicitly assume that responses to the OV-Klantbarometer accurately represent the characteristics of PT travellers. Based on the size of their data (92.500 responses over 2014) and the way this survey is carried out (travellers are randomly asked to respond) it is plausible that this conclusion is indeed correct. Moreover, as the Dutch Research Institute for Mobility (Kennisinstituut voor Mobiliteit, KiM) notes no better data on demographic characteristics of PT travellers is available (van der Loop et al., 2018). To conclude, this research assumes that the demographic characteristics as observed in the OV-Klantbarometer accurately reflect those of PT travellers in general, both as a result of assumptions made in the OV-Klantbarometer based on this conclusion, a lack of contrary evidence and a lack better information sources. Table 4.1 shows an overview of the personal characteristics which are known from each respondent in both the OV-Klantbarometer and HTM Klantenpanel.

To see whether the HTM Klantenpanel is representative, the demographics of the Klantenpanel are compared to the OV-Klantbarometer. Appendix C provides an overview of demographic characteristics of the HTM Klantenpanel and tries to compare this information to both the OV-Klantbarometer and, where OV-Klantbarometer data is not available, general demographic characteristics of the Den Haag region. It should be noted that the OV-Klantbarometer makes a difference between light-rail lines 3 and 4 and the other urban tram lines. Both are shown separately here and in appendix C but differences are quite small and not statistically significant.

Table 4.1: Demographic characteristics known in surveys

HTM Klantenpanel	OV-Klantbarometer
Age [year of birth]	Age [year of birth]
Gender [M/F]	Gender [M/F]
Education level [low/medium/high]	
Frequency of PT Travel	Frequency of PT Travel
Household composition	
Type of ticket	Type of ticket
Possession of smart phone	
Reason of travel	Reason of travel
Postal code of home address	

In the available demographics, two main differences can be observed between the composition of the OV-Klantbarometer and the HTM Klantenpanel. Firstly, a majority of responses on HTM Klantenpanel is done by males while both the OV-Klantbarometer as well as scientific studies (e.g. Bakker, 2018) show that a majority of PT travellers in the Netherlands is female. The distribution of the HTM Klantenpanel has 60% males while the OV-Klantbarometer has around 40% males. Figure 4.2 shows this distribution graphically.

When analysing age a second difference can be observed. In HTMs Klantenpanel, 85% of responses come from respondents aged above 40, with the group above 65 accounting for 41% of responses. The OV Klantbarometer shows that these groups are quite overrepresented, especially the 65+ group, while, on the other hand, young travellers are significantly underrepresented. Figure 4.3 shows this graphically. For other demographic characteristics measured differences are much smaller. As a result, it is important to note the overrepresentation of elderly and males in the dataset used in the upcoming analyses and, if possible, to correct conclusions to represent this correctly.

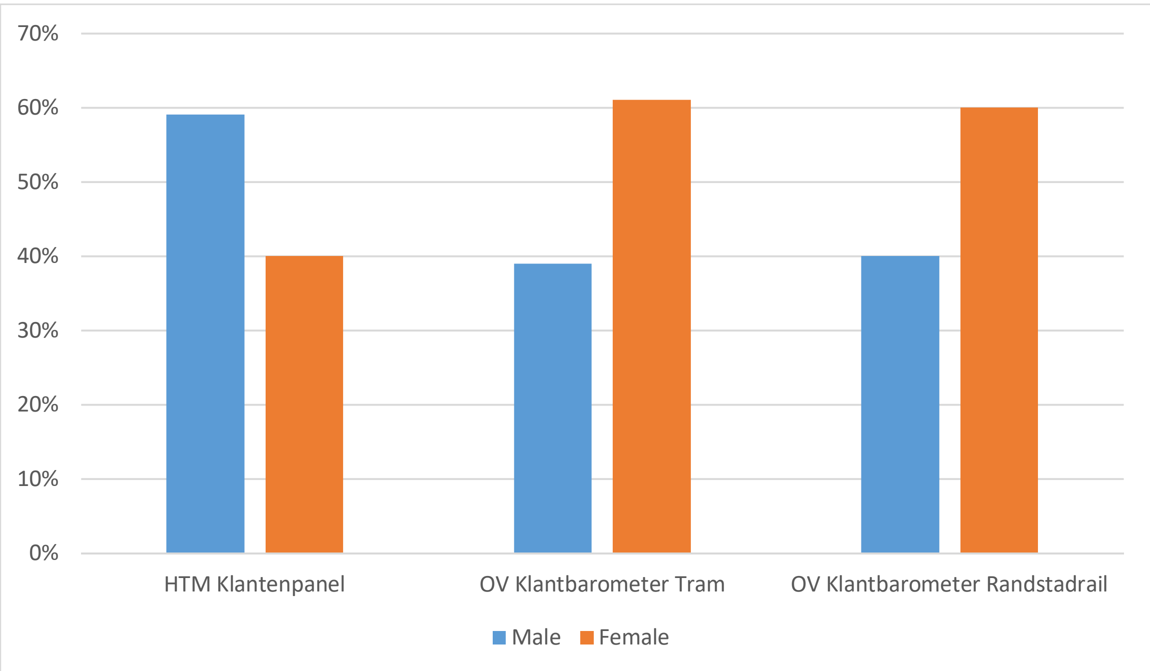


Figure 4.2: Gender distribution in different PT Satisfaction surveys

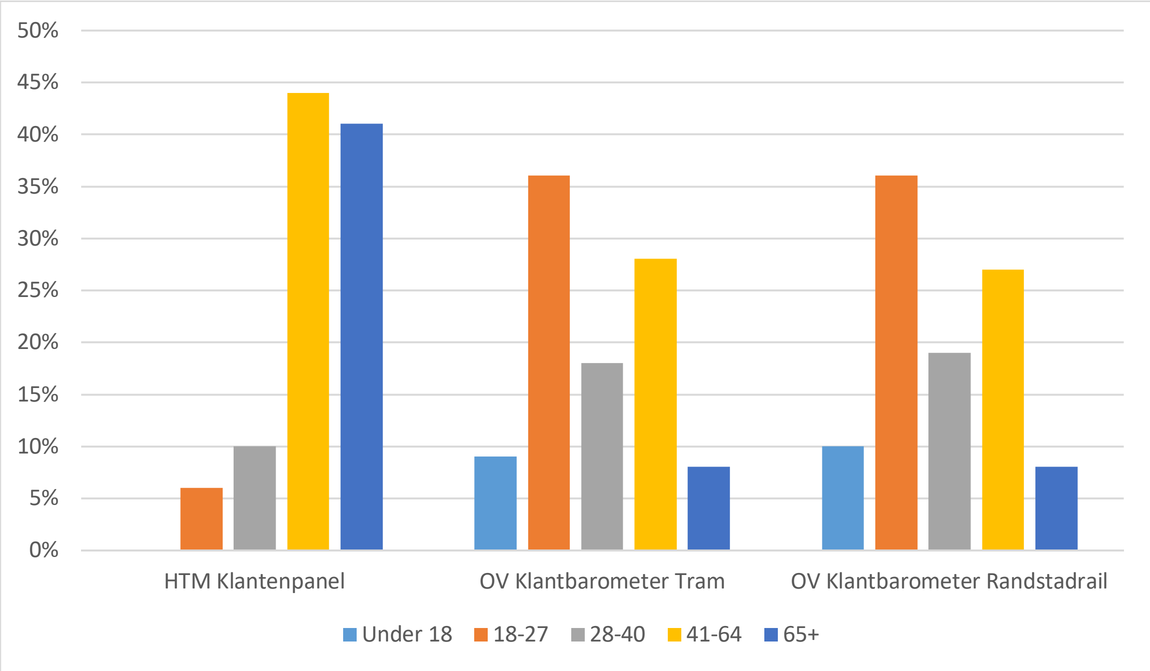


Figure 4.3: Age distribution in different PT Satisfaction surveys

Figures 4.2 and 4.3 show the distribution of demographics in the OV-Klantbarometer, these match the observations on the composition of PT travellers found in literature by Bussink and de Konink (2015) and Zijlstra et al. (2018): looking at age elderly (65+) and young (18-27) travel on average more often with Public Transport and women tend to travel more using PT than men. The model will be able to take these differences into account.

Next up the travel motive can be compared. Tables 4.4 and 4.5 show this comparison. As both surveys ask categorise the answers in a slightly different way two tables are used to present responses. Both tables show for what specific purpose the trip evaluated was made. HTM Klantenpanel asks explicitly that if a homebound trip is evaluated the motive of the tour should be noted, while the OV-Klantbarometer allows 'going home' to be filled in. For comparison, an extra column is added at the OV-Klantbarometer in which trips going home are distributed over the other categories – assuming that the motif of the return leg of a homebound trip follows the same distribution as observed. Nevertheless, comparison remains a bit difficult of course, due to the different answer options. It is, however, clear that education is severely underrepresented in HTM Klantenpanel.

Next, travel frequency can be compared. Again, the HTM Klantenpanel and OV-Klantbarometer ask these questions differently, so two tables are necessary to present the results. These can be found in tables 4.6 and 4.7. The differences in the way questions are asked (mainly usage being counted in days per week or times per week) makes comparison, again, difficult. It can be concluded, however, that HTMs Klantenpanel has very little responses from incidental PT users. Given the nature of the panel, however, this is logical.

Table 4.4: Travel Motif distribution Klantenpanel (trip specific travel motive 2018)

Motive	HTM Klantenpanel (%)
Work	29.4
Education	3.8
Business	3.2
Medical	6,4
Groceries	7,4
Shopping	8,7
Visiting friends/family	13.4
Informal care	1.2
Recreation	9.9
Eating&Drinking	2.8
Sport/Leisure	5.9
Different	8.0

Table 4.5: Destination Distribution OV Klantbarometer

Destination	OV Klantbarometer Tram (%)		OV Klantbarometer RR (%)	
Living	22	-	30	-
Working	25	32	28	41
Education	13	16	13	19
Shopping	12	15	6	9
Sport	3	4	2	3
Visiting friends/family	11	14	9	13
Different	15	19	10	15

Table 4.6: Frequency of travel HTM Klantenpanel

Frequency	HTM Klantenpanel
4+ days per week	37
1-3 days per week	44
1-3 days per month	16
6-11 days per year	2
5- days per year	1

Table 4.7: Frequency of Travel OV Klantbarometer

Frequency	OV Klantbarometer Tram	OV Klantbarometer RR
6+ times per week	25	27
5 times per week	13	17
4 times per week	14	14
3 times per week	11	11
2 times per week	13	13
<2 times per week	24	18

To conclude, from a demographic perspective in HTM Klantenpanel an overrepresentation can be seen of men and elderly. When looking at, for example, travel motive, a similar conclusion can be drawn, in which for example people travelling to or from education are underrepresented. Differences in the way HTM Klantenpanel and OV-Klantbarometer ask their respondents for travel purposes and travel frequencies the answers are not perfectly comparable. The Structural Equation Model is able to accommodate for this misbalance in responses as long as all possible answers have been given a sufficient number of times – the tables in this section suggest this is the case. Regarding the model which will be presented in chapter five thus no extra action is necessary. When converting the model to implications, however, it should be noted that all average values in the dataset are skewed.

4.2. Survey Data HTM Klantenpanel

This section will deal with the exploratory data analysis on the survey responses as filled in by respondents, after data preparation and coupling as has been described in section 3.2. To start with, it can be analysed what kind of trips are analysed. Figure 4.8 shows how the 2858 responses used are divided over the different lines which HTM operates and figure 4.9 shows how these responses are divided over the different vehicle types which HTM uses. The data behind these figures can be found in appendix C. The dataset as delivered incorrectly stated that line 2 is driven with GTL vehicles, while in fact this is done with a mix of Avenio (mostly) and Citadis. It would have been too time-consuming to derive which vehicle type was used exactly for each evaluated trip on line 2. As a result, it has been assumed that all services on line 2 are driven with Avenios. This is the assumption which approaches reality best though not perfectly.

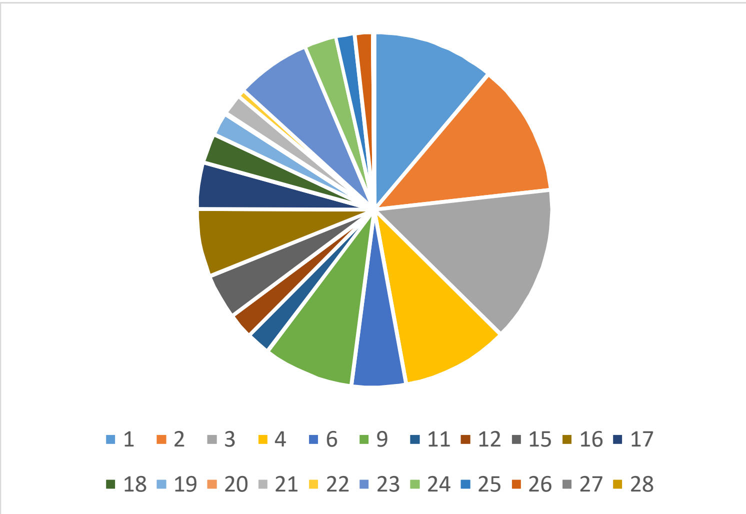


Figure 4.8: Distribution of responses over line

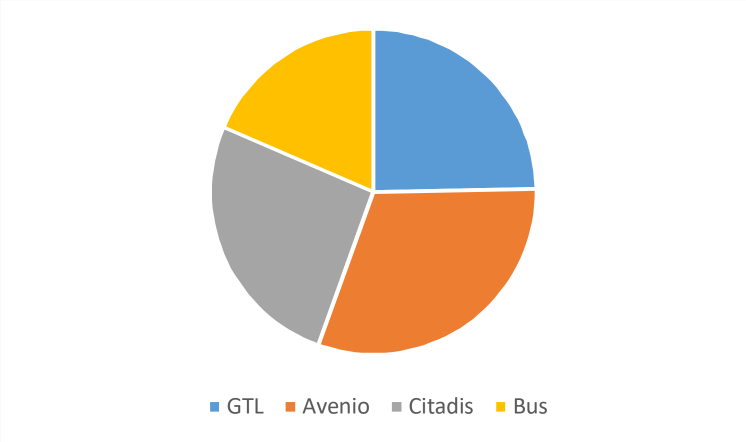


Figure 4.9: Distribution of responses over vehicle type

Next up, some general statistics can be derived on how respondents evaluated their travels. Table 4.10 provides an overview of how each of the questions asked was answered by respondents. It is also reported how often a question was answered with 'I do not know' or 'does not apply.' In estimating the structural equation model, these answers will be treated as missing values. As mentioned in section 3.2, if overall customer satisfaction or perceived crowding was not graded the response was deleted, hence no missing values are present for these two variables.

Table 4.10: Descriptive statistics on survey responses

Variable	N	'Do not know'/ 'Does not apply' (%)	Mean	Median	Mode	Std. Deviation
Perceived Crowding	2858	0.0	7.92	8	10	2.35
Comfort	2838	0.7	7.53	8	8	1.80
Cleanliness of vehicle	2797	2.1	7.67	8	8	1.43
Cleanliness of stop	2818	1.4	7.57	8	8	1.47
Information supply on stop	2749	3.8	7.96	8	8	1.86
Information supply during delays or disruptions	673	76.5	6.65	7	10	2.97
Friendliness of staff	1195	58.2	7.77	8	8	2.10
Driving style of driver	2810	1.7	7.79	8	8	1.67
Ease of buying a ticket	1232	56.9	8.75	10	10	1.78
Punctuality	2736	4.3	8.02	8	8	1.83
Frequency	2714	5.0	7.47	8	8	1.82
Overall satisfaction	2858	0.0	7.69	8	8	1.72
General feeling of safety in PT	2829	1.0	7.86	8	8	1.32
Feeling of safety during this trip	2827	1.1	8.15	8	8	1.43

It is interesting to note that some questions were quite often responded with 'I do not know' or 'does not apply'. This mainly applies to the factors 'information supply in-vehicle', 'friendliness staff' and 'ease of buying ticket'. It is reasonable to assume that this is legit:

- It is reasonable to think that most passengers know the trip they make and do not pay any attention to information in-vehicle unless disruptions occur.
- In Avenio and Citadis trams a passenger does not encounter any staff unless service members are present in-vehicle. In a GTL it is also perfectly reasonable to not encounter any staff when entering the vehicle in the back.
- A lot of passengers travel using a subscription or an OV-chipkaart with sufficient credit and thus they only need to tap in and out. As a result, they do not need to buy a ticket.

While preparing data for analysis, the values 'I do not know' and 'does not apply' were deleted – during the analyses they will be treated as missing values. A resulting question is whether this makes that these questions need to be discarded from the dataset altogether in analyses. Section 5.1 will test and discuss whether this is the case.

4.3. Crowding

Section 3.2.3 has described how crowding data is retrieved based on survey responses. For the service a respondent has used three variables have been retrieved:

- The number of passengers in the vehicle at the moment of boarding
- The number of passengers in the vehicle at the moment of leaving
- The maximal number of passengers in the vehicle between these two stops

As has been discussed in section 2.2, these values need to be converted to other metrics for proper analysis to be possible. Table 4.11 shows the values which were used for this conversion, based on calculations by Wieffering (2016). Regarding the number of available places for standing passengers, the norms used by HTM are used and not the norms defined by the manufacturer. Using these values, each passenger number was converted into three other metrics:

- Load Factor
- Standing Passenger Density
- If the passenger is able to sit (yes or no). This is, of course, directly coupled to the Load Factor.

Table 4.11: Values used to convert occupancy to Load Factor and Standing Passenger Density (Wieffering, 2016)

Number	Vehicle type	Number of seats (#)	Available room for standing passengers (m2)	Capacity for standing passengers (#)	Load Factor when full (%)
1	GTL	71	25.1	80	213
2	Avenio	70	33.8	113	261
3	Citadis	86	32	132	253
4	Bus (MAN Lion City)	31	12.6	46	248

These metrics can be analysed. First, we can have a look at the Load Factor. Table 4.12 gives some information of the statistics which were found regarding the Load Factor. All percentages are based on the dataset containing 2858 responses, in which the Load Factor is defined for all cases.

Next, the standing passenger density can be investigated. As for the majority of trips the Load Factor is below 100%, the standing passenger density is often zero. Table 4.13 provides an overview of the number of cases in which each passenger density statistic is larger than zero; table 4.14 then provides descriptive statistics on those cases in which the passenger density is larger than zero.

Table 4.12: Descriptive Statistics on Load Factor (n=2858)

Variable	Load Factor on Boarding (%)	Load Factor on alighting (%)	Maximal Load Factor (%)
Mean	40.5	32.6	64.4
Median	32.2	25.8	56.3
Standard Deviation	3.4	3.0	4.1
Minimum	0	0	1.2
Maximum	268.6	235.7	268.6

Table 4.13: Number of cases with lack of seat availability

Load Factor	Boarding (%)	Alighting (%)	Maximum (%)
> 100%	6.2	3.4	16.9
< 100%	93.8	96.6	83.1

Table 4.14: Descriptive statistics on standing passenger density (only cases with LF > 100%)

Variable	Standing Passenger Density on Boarding (pax/m2)	Standing Passenger Density on alighting (pax/m2)	Maximal Standing Passenger Density (pax/m2)
n	177	97	482
Mean	0.78	0.67	0.86
Median	0.50	0.41	0.63
Standard Deviation	0.7	0.7	0.8
Minimum	0.03	0.03	0.03
Maximum	3.49	3.38	3.49

This data shows that during most trips seating availability is not an issue. As a result, in this dataset the Load Factor seems to be a better variable to measure occupancy numbers than standing passenger density, as the amount of evaluated trips in which the load factor was over 100% is limited. The number of trips evaluated in which this was the case structurally is even more limited: only in 35 cases (1.2%) the Load Factor was over 100% at both the stop the respondent boarded and the stop the respondent alighted.

It can be concluded that for a clear majority of trips evaluated seating availability was no issue: even when looking at the maximal number of passengers on-board only in one out of six cases there were more passengers than seats as reported. In practice, this number will be slightly higher as the passenger numbers are based on the number of checked-in passengers and not corrected for fare evasion or tickets which do not require check-in. HTM currently uses a correction factor of 12% to take these passengers into account.

A next step in analysis can be to compare these passenger numbers to the perceived occupancy passengers have experienced. As section 3.3 has explained, the *probability of finding a seat* is used as an indicator for perceived crowding. Actual and perceived crowding can be shown visually using scatter/dot-diagrams. Figures 4.15 to 4.17 show these diagrams.

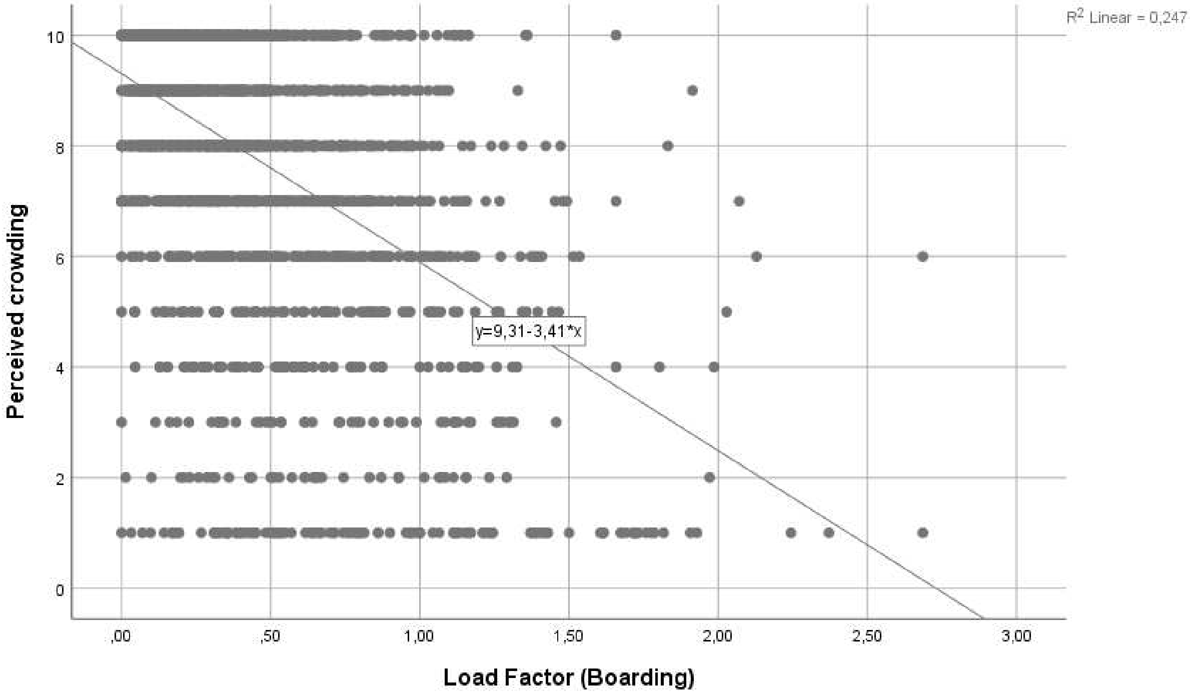


Figure 4.15: Scatter/dot-plot of perceived occupancy vs. Load Factor on boarding

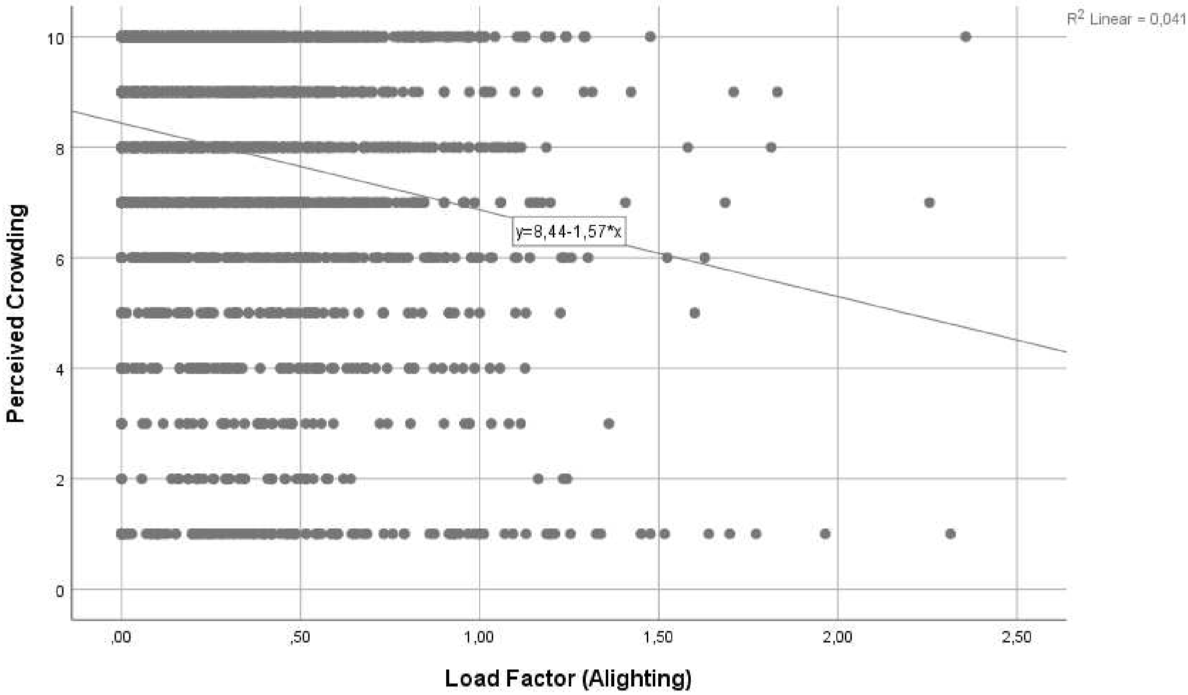


Figure 4.16: Scatter/dot-plot of perceived occupancy vs. Load Factor on alighting

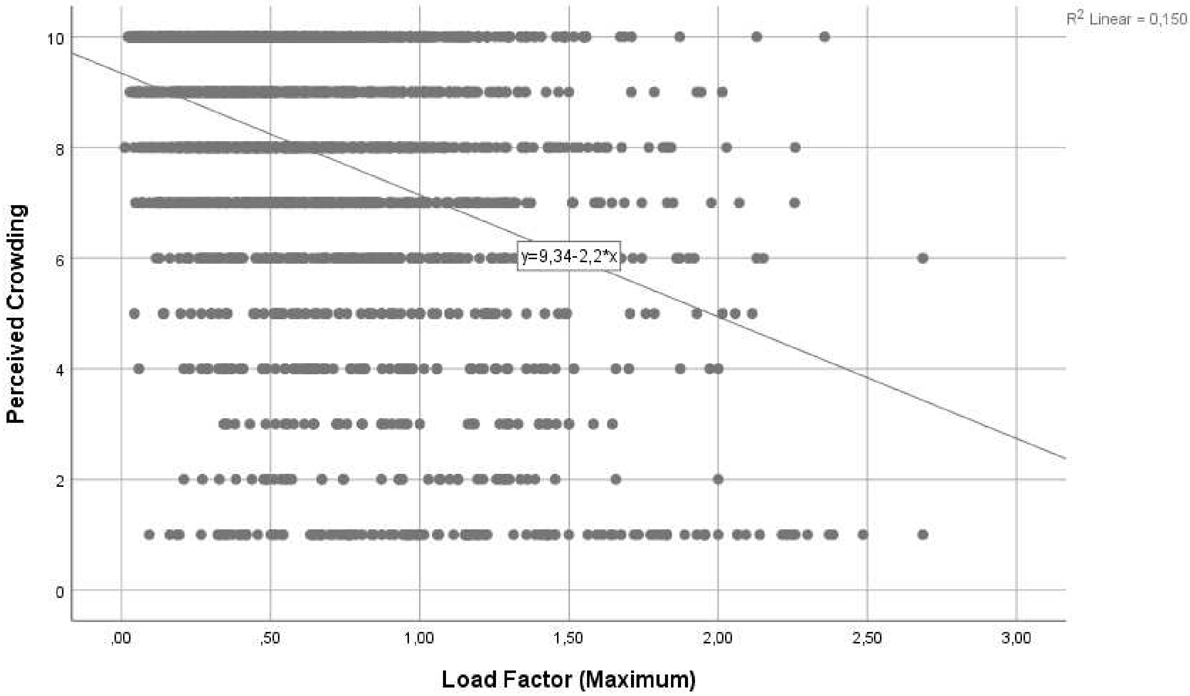


Figure 4.17: Scatter/dot-plot of perceived occupancy vs. maximal Load Factor

These figures show that there is no clear one-on-one relationship between objective and subjective crowding. It can, however, be seen that in general lower marks are given for a higher Load Factor. The most likely explanation for the lack of visibility of a clear relationship is the large amount of noise which is present on the marks given. This means that the mark given for perceived occupancy is affected by more factors than just actual occupancy. The model as presented in section 3.4 will try to explain this noise.

For now, it is important to make assumptions on which basis the model will test the relationship between objective and subjective crowding. Haywood and al. (2017) posed the relationship between occupancy levels and customer satisfaction was linear and tested this for high levels. In section 2.2, some critical remarks were made to this conclusion. Table 4.18 shows the explained variance (R^2) obtained from fitting a linear, quadratic or cubic polynomial on data presented in figure 4.15 to 4.17. While the explained variance is quite low, all variables are nevertheless significant ($p = 0.000$). From this table two conclusions can thus be drawn. Firstly using a complex non-linear relationship offers very little extra explanatory power compared to a simple linear relation. Based on the academic convention of using the simplest sufficient type of relation and the conclusions of Haywood and al. (2017) it will therefore be assumed in the model that the relationship between objective and subjective crowding is linear. Secondly, the moment of boarding seems to be the most useful of the occupancy numbers in terms of explanatory power.

Table 4.18: Explained variance from fitting various polynomials on figures 4.15 to 4.17

Polynomial	Load Factor on Boarding	Load Factor on Alighting	Maximal Load Factor
Linear	0.247	0.041	0.150
Quadratic	0.247	0.042	0.152
Cubic	0.249	0.043	0.152

As section 3.3 has described, Structural Equation Modelling is unable to operationalise variables in more than one way, making it impossible to operationalise crowding in the optimal way defined in section 2.2. As a result, choices have to be made how to operationalise crowding. Based on this exploratory analysis the following modelling choices have been made:

- Load Factor only will be used as a metric for measuring in-vehicle occupancy. As table 4.13 shows, standing passenger density is only usable in a small number of cases, as in most trips evaluated free seats were available.
- The moment of boarding will be used for measuring the number of passengers. Two reasons support this choice. As table 4.18 shows boarding seems to relate most to the measurement of perceived occupancy (seat availability).

4.4. Delays

As has been mentioned in section 3.3, for each response used in the HTM Klantenpanel data on the punctuality of the service was also retrieved. This section will present and discuss punctuality in the dataset.

Two numbers were retrieved per response: the delay at the moment of boarding, and the delay at the moment of alighting. Figure 4.19 shows the resulting values visually in a boxplot. All delays are measured in seconds.

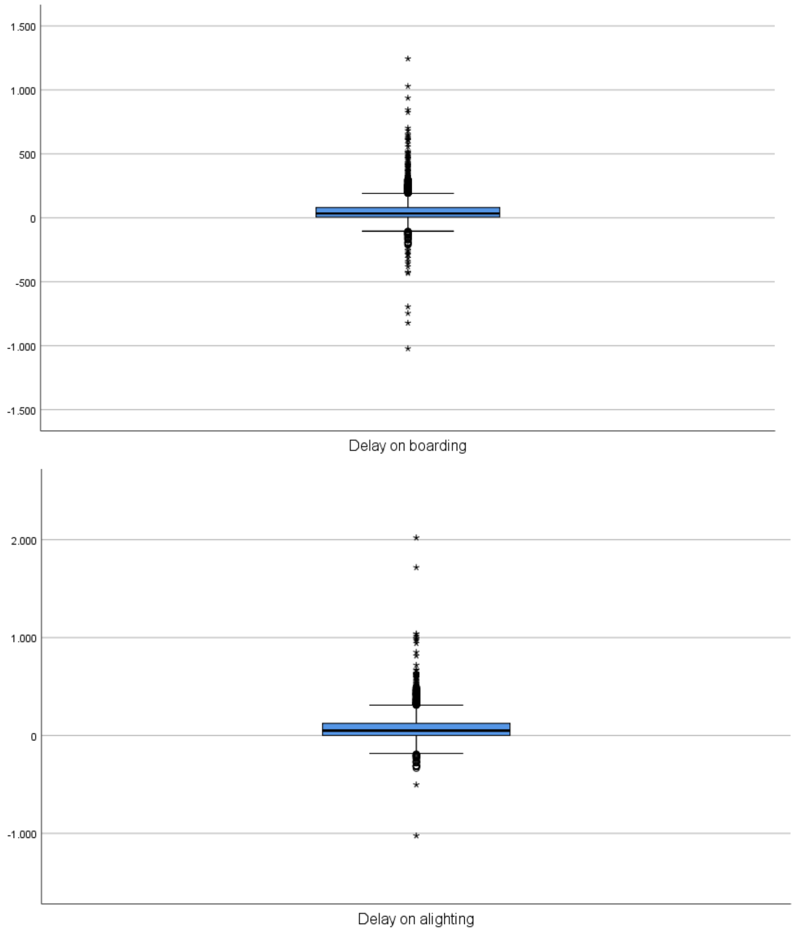


Figure 4.19: Boxplot of Delays on boarding (above) and delays on alighting (below)

The data shows that average departure delays are slightly positive, which is logical as departing early is prohibited by MRDH. It should be noted that these scheduled times are accurate to the second, while actual times communicated to passengers are rounded to minutes. A delay of 120 seconds or more is considered 'late' by MRDH in the HTM concession.

The range of delays is high, reporting from vehicles driving more than 15 minutes (900s) early to vehicles being more than 30 minutes (1800s) late. On the upper side these delays can be explained by, for example, traffic accidents or technical problems with a vehicle. Finding an explanation for the early vehicles is much less intuitive.

Regardless of the delays as reported, it should be noted that actual delay does not mean the passenger experiences a delay: if all vehicles are 15 minutes late and frequencies are 4 services per hour all vehicles seem on time to the passenger, so a delay value does not tell everything. Luckily, in the HTM Klantenpanel trip evaluation questionnaire a respondent is asked whether he has experienced a delay or disruption in the trip evaluated and he is asked to give the punctuality of the service a mark. This provides insight in perceived reliability and punctuality. These variables will be combined with actual delay to get perceived reliability.

Lastly, as is the case for occupancy numbers, SEM only allows for one measurement of delay. As a result, a choice between these two variables has to be made. Section 4.3 has explained that for occupancy numbers the occupancy at the moment of boarding will be used. Accordingly, it seems fair to use the delay at the same moment. Two other reasons can be given to support using boarding instead of alighting delays:

- Using the delays at alighting poses a problem in that some passengers evaluate their trip while making it.
- Passengers experience waiting for a (delayed) vehicle at a stop worse than waiting inside a (delayed) vehicle (van Hagen, 2011). It is thus reasonable to assume that departure delays affect passenger perception more than delays which grow while travelling.

The model which will be presented in chapter 5 will thus use the delay at the moment of boarding.

4.5 Conclusion

This chapter has explored the data which will be used in building the Structural Equation Model. From this analysis, a variety of conclusions can be drawn. Firstly, regarding representativeness it should be noted that the HTM Klantenpanel has an overrepresentation of males and elderly compared to PT travellers in general. As a result other variables such as travel purpose also have a different distribution in HTM Klantenpanel than in OV Klantbarometer. The dataset is large enough to correct for this in the model.

The dataset was then investigated for three factors: the responses given, the corresponding occupancy numbers and the corresponding delays. In general no oddities were found in doing so: high delays and/or occupancy levels found do indeed occur sometimes. Regarding the relation between occupancy and perceived occupancy, the data gave no reason to reject the assumption of linearity which is used in SEM. In estimating the model, occupancy and delays will be measured at the moment when the respondent boarded a service, as these values seemed to represent customer perception best.

5. Results

Chapter three has presented the conceptual model which will be used for analysis has been presented. This chapter will deal with the quantitative estimation of this model. Firstly, a factor analysis will be presented to analyse whether the match between factors and indicators as presented in table 3.11 can be justified based on data. Afterwards, the estimated model will be presented and its outcomes will be discussed. In this chapter only the most relevant quantitative insights will be presented: appendix E provides an overview of all values and relations which have been estimated.

5.1. Measurement model

Firstly it needs to be checked whether the factors defined in section 2.3 are sufficiently approximated by the indicators used based on the questions asked in the survey. A confirmatory factor analysis (CFA) has been carried out to test whether this model fits the actual data or not.

To start with, the validity of this model can be tested using a variety of fit indices. The Comparative Fit Index (CFI) is 0.916, being above the threshold of 0.9 indicating good model fit. The Root Mean Square Error of Approximation (RMSEA) is 0.087, just above the threshold of 0.08 which indicates the upper bound for good model fit. Lastly, the chi-square statistic is insignificant. While this is a widely used metric of model fit, it is well-known that the chi-square value is often insignificant for models which are based on large datasets even in cases of good model fit. With 2858 cases the dataset used in this analysis is large enough to conclude that the conclusion of the chi-square test is no reason to reject good model fit. (Byrne, 2013).

Table 5.1 shows how each indicator load on the factor it is coupled to, as well as the Average Variance Extracted (AVE) per factor. An AVE of 0.5 is considered to be an indicator of good fit. The weight of some indicators is fixed to one, these indicators are marked with an asterisk (*). This has to be done if a factor is explained by only a single indicator. In such a case, the variance extracted automatically becomes one as well. In Structural Equation Modelling, an indicator is commonly said to load sufficiently to a factor if its weight is larger than 0.5 and good if this value is higher than 0.7. The only indicator which does not meet this threshold is the ease of buying a ticket, which has a factor loading of 0.492 – only just below 0.5. Deleting this indicator and running the analysis again improves the AVE of perceived comfort to above 0.5 but results in poorer other fit indices (CFI, RMSEA). As a result the decision is made to include the variable in the structural model. In conclusion, the measurement model offers an acceptable fit, and all indicators as proposed in section 3.4 will remain in the causal model when estimated.

5.2. Structural model

Having concluded that the indicators which have been selected can properly be used, we can next go on to test the framework as presented in section 3.4. This section will analyse the estimated model. Firstly, it will be checked whether this model fits the data sufficiently to be able to draw conclusions from the model. Afterwards a general overview of model results will be presented showing the effects which the different variables have on each other. Lastly, interesting conclusions and relations found will be presented by analysing the model layer by layer.

Before the model results can be analysed it is necessary to first analyse whether the model fits the data sufficiently enough to make drawing conclusions viable. If this is not the case, interpreting results would make no sense. Table 5.2 shows the commonly used statistics to measure this, along with conventional values which indicate good model fit and the actual values for the model.

Table 5.1: Results of Confirmatory Factor Analysis (CFA)

Factor	**Indicator**	***Weight***	**Average Variance Extracted (AVE)**
Perceived Safety	General feeling of safety in PT	0,739	0.716
	Feeling of safety during this trip	0,942	
Perceived Service Quality	Frequency	0,708	0.496
	Information supply during delays or disruptions	0,710	
	Information supply on stop	0,696	
Perceived Comfort	Cleanliness of vehicle	0,751	0.477
	Comfort during trip	0,780	
	Driving style of driver	0,728	
	Cleanliness of stop	0,634	
	Friendliness of staff	0,721	
	Ease of buying a ticket	0,492	
Perceived Reliability	Punctuality	1*	1*
Perceived Crowding	Probability of finding a seat	1*	1*

Table 5.2: Fit indices of structural model

Statistic	**Value for good model fit**	**Model values**
χ^2-statistic	< 0.05	0.000
χ^2/df	< 5	11.059
Goodness of Fit (GFI)	> 0.90	0.910
Root Mean Square Error of Approximation (RMSEA)	< 0.08	0.059
Standardized Root Mean Square Residual (SRMR)	< 0.06	Not available

These values can be compared to the model values. The χ^2-statistic is significant (0.000), moreover the ratio of the χ^2-statistic to the degrees of freedom (d.f.) is higher than could be expected from a good model fit. Poor model fit based on the χ^2-statistic, however, is common in SEM-models in which a large sample is used. As a result, poor model fit based on the χ^2-statistic is tolerated if other indices indicate a good model fit (Byrne, 2013). With 2858 observations the sample size used here is large for a SEM-model. Other fit indices show that indeed this is the cause of the insignificant result of the χ^2-test. The Goodness-of-Fit is 0.910, above the minimum threshold of 0.9 which is used to indicate a good model fit. The Root Mean Square Error of Approximation (RMSEA) is 0.059, below the common acceptance level of 0.08, which again indicates a good model fit. The SRMR is unavailable in this model as a result of using Full Information Maximum Likelihood as a method of dealing with missing values. It can thus be concluded that the model offers an acceptable explanation of the observed covariance between the variables of interest. `

Lastly, we can analyse to what extent the model is able to explain observed variance in customer satisfaction. The squared multiple correlation for customer satisfaction is 0.789, which means that the model is able to explain 78.9% of the variance of the customer satisfaction as defined by respondents – a reasonable large number.

Having concluded that the model models the situation well and can be used, the model itself can be analysed. Figure 5.3 shows the structural model, along with the estimated relations. Two methods of presenting these relations exist:

- Standardised coefficients, which show by what number of standard deviations a variable increases if the other variable increases by one standard deviation. This coefficient is useful to show the relative strength of a relation in comparison to the rest of the model.
- Unstandardised coefficients, which show the absolute increase of a variable if the other variable increases by one. This coefficient is useful to calculate the actual impact of variables on each other.

As the standardised coefficients provide a much better insight into the relative strength and relevance of relations, all relations presented and explained in this chapter will use standardised coefficients unless explicitly mentioned otherwise. In chapter 6, where the model will be used to estimate scenarios and practical consequences, unstandardised estimates will be used.

Table 5.4 shows the total standardised effects between all variables. These effects are calculated by summing up all effects found between two variables via all possible routes. Effects which were found to be significant at the 0.05 level are marked in **bold**. These values can be analysed step by step to find and describe the most interesting relations.

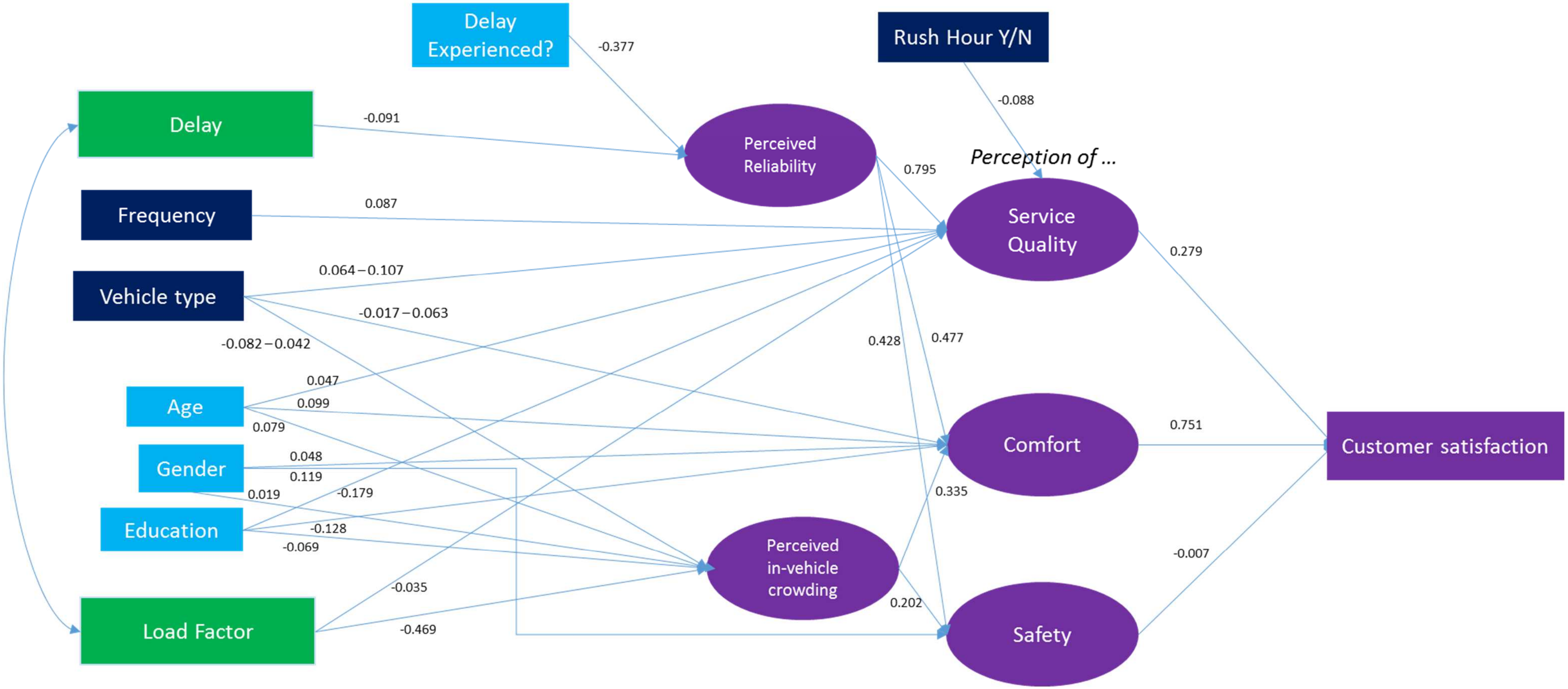


Figure 5.3: Estimated structural model with standardised regression weights

*Table 5.4: standardised total effects between all variables in the structural model (**bold** is significant at the 0.05 level).*

	Delay Experienced Y/N	PT Travel Frequency	Rush Hour?	Purpose_work	Purpose_business	Purpose_school	Purpose_visit	Avenio?	Citadis?	GTL?	Education	Age	Gender	Delay	Load Factor	Frequency	Perc Reliability	Perc Crowding	Perc Comfort	Perc Service Quality	Perc Safety
Perceived Reliability	**-,377**	,000	,000	,000	,000	,000	,000	,000	,000	,0000	,000	,000	,000	**-,091**	,000	,000	,000	,000	,000	,000	,000
Perceived Crowding	,000	-,018	-,032	-,002	,012	-,030	,008	**-,040**	**-,082**	**,042**	**-,069**	**,079**	,019	,000	**-,469**	,000	,000	,000	,000	,000	,000
Perceived Comfort	**-,180**	**-,048**	-,036	-,032	,005	-,012	,020	**,063**	-,017	,045	**-,151**	**,126**	**,054**	**-,043**	-,157	,000	**,477**	,335	,000	,000	,000
Perceived Service Quality	**-,299**	**-,045**	**-,088**	**-,056**	,008	-,013	,009	,107	,067	,064	**-,179**	**,047**	,020	**-,072**	-,035	**,087**	**,795**	,000	,000	,000	,000
Perceived Safety	**-,161**	-,004	-,007	,000	,002	-,006	,002	-,008	-,017	,008	-,014	,016	**,122**	-,039	-,095	,000	**,428**	**,202**	,000	,000	,000
Overall Satisfaction	**-,199**	**-,044**	**-,048**	**-,036**	,006	-,012	,015	**,071**	**,008**	**,047**	**-,148**	**,094**	**,040**	**-,048**	**-,111**	**,024**	**,528**	**,215**	**,647**	**,279**	-,007

To start with, the relation between occupancy and customer satisfaction, the main goal of the model and this thesis, can be investigated. Figure 5.5 below shows the part of the model which links the load factor to customer satisfaction, including the effects variables have on each other. The indirect relation between occupancy and customer satisfaction is indeed present. However, the relationship between load factor and perceived service quality much less strong than the relation between load factor and perceived comfort. As a result, the effect of occupancy levels on customer satisfaction goes via perceived occupancy and perceived comfort, which is in line with literature and the hypothesis as formulated in section 2.5. The correlation between occupancy and customer satisfaction is -0.111. The effect of the Load Factor on perceived occupancy is strong at -0.469.

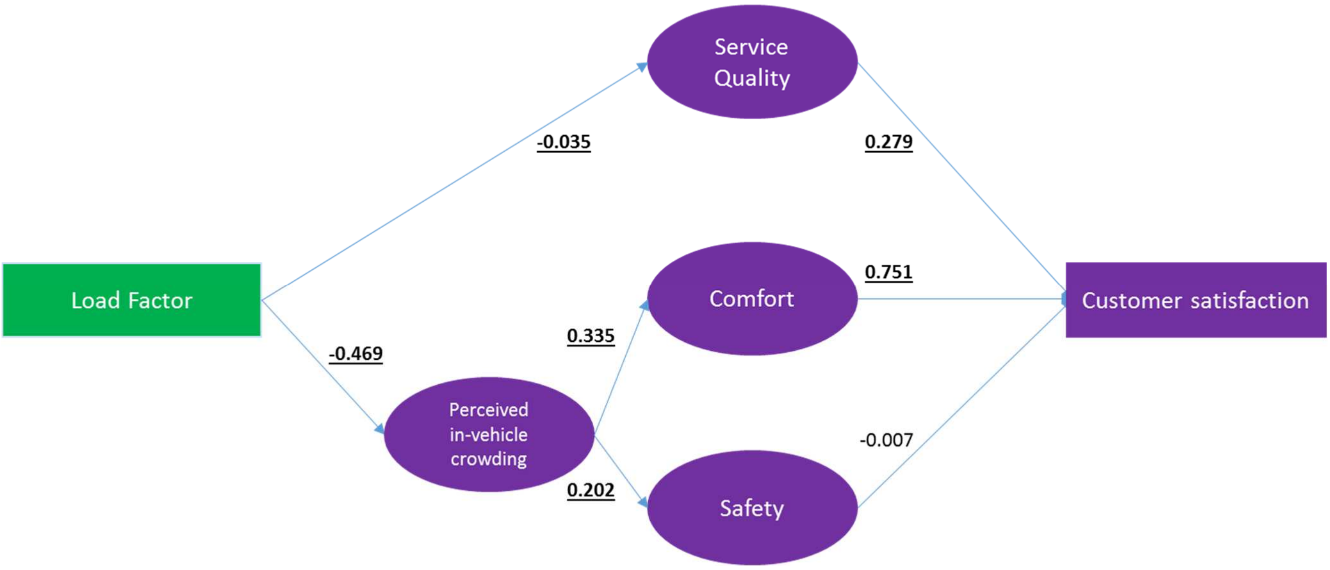


Figure 5.5: Link between occupancy and customer satisfaction

However, besides the relation between occupancy and customer satisfaction a lot of other different interesting conclusions can be drawn from the model. To start with the construction of customer satisfaction based on three factors can be looked upon. Figure 5.6 shows this level of the model, including the correlations between the three factors identified.

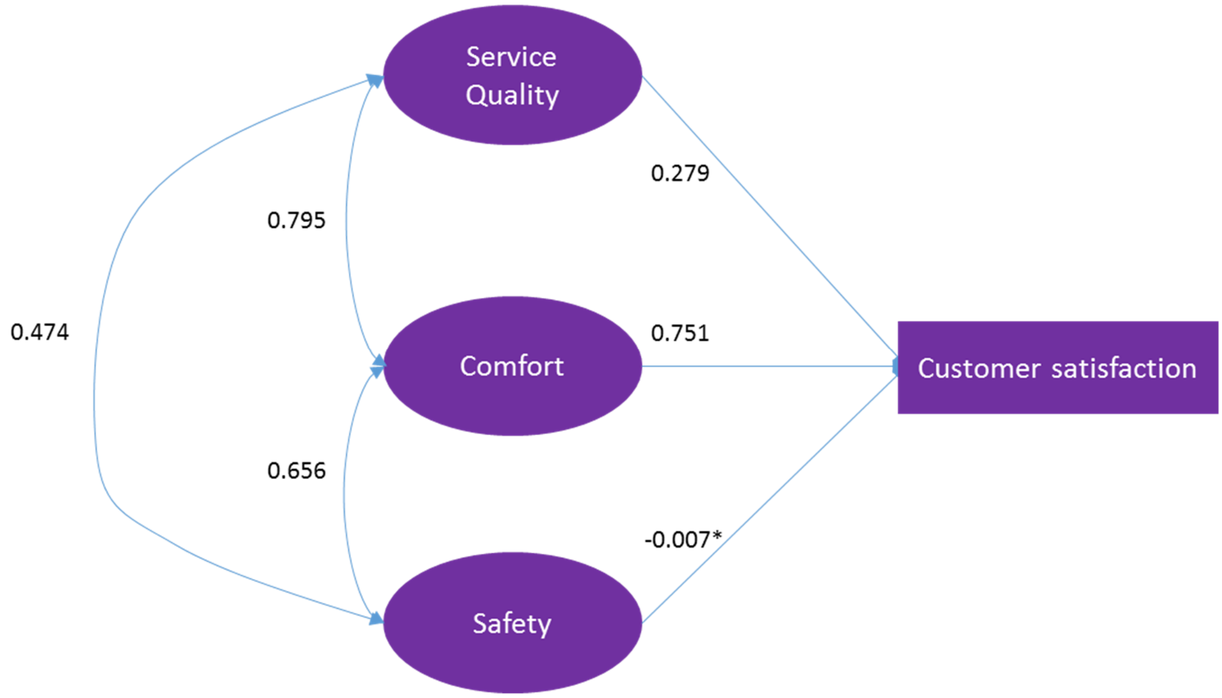


Figure 5.6: Factors determining customer satisfaction and their correlations

It is interesting to see that comfort seems to be the dominating factor in determining customer satisfaction, having a much higher effect than the other two factors. On the other hand the relation between perceived safety and customer satisfaction is almost equal to zero and not significant. This does not mean, however, that safety plays no role in determining customer satisfaction. As figure 5.6 shows the correlations between the three dimensions of customer satisfaction are high, between 0.47 and 0.79.

Via these effects safety does impact customer satisfaction to some extent. The correct way to interpret the insignificant relationship between perceived safety and customer satisfaction is thus that, in this dataset, the factor 'perceived safety' adds no extra information that the other two factors do not already offer. A possible explanation for this result might be that safety, according to van Hagen (2011) the most basic requirement of a good PT service, is simply not an issue experienced in Den Haag and thus the difference between safe and unsafe situations is not captured by responses. The number of responses in which safety was evaluated as being insufficient is limited (3.8% for safety during the specific trip evaluated and 5.3% for safety in general), which can support this conclusion.

Descending one layer in the model it can be analysed what factors affect perceived service quality, comfort and safety. The impact of perceived occupancy and load factor has already been discussed above. Perceived reliability has an effect both on perceived comfort (t = 25.78[8]) and perceived service quality (t = 37.368). It is interesting to see that personal and service characteristics also impact perceptions, in the following ways:

- Travelling in an Avenio has a clear positive effect on perceived comfort (t = 4.407) when compared to the other vehicle types HTM offers, between which no significant differences exist.

[8] The t-value is used to determine whether a relationship is significant. If the t-value is larger than 1.96 or smaller than -1.96 a relation is significant at the 0.05 level. For the 0.01 level the critical t-value is 2.25 or -2.25. The corresponding size of the effects can be seen in figure 5.1 and table 5.2.

- Perceived service quality is significantly lower (t = -4.082) if a trip is made during rush hours, and significantly higher (t = 4.304) if the frequency (in trips/hour) on the line on which the trip is made increases. The effect of travelling on or off-peak on perceived comfort is just insignificant (t = -1.791).

- Age has a significant impact on both perceived service quality (t = 2.629) and perceived comfort (t = 6.915). The relationship is positive, meaning that elderly evaluate their trips more positive than young people do. This is in line with earlier research into customer satisfaction at HTM.

- Gender has a significant impact on perceived comfort (t = 3.123) and perceived safety (t = 7.066). In both cases men are slightly more satisfied than women.

- Education level affects perceptions negatively: both regarding perceived service quality (t = -10.361) and perceived comfort (t = -9.233) a clear relation can be found. Highly educated travellers are thus significantly more negative than low educated travellers. This relation seems much stronger than the relation between gender or age and these perceptions.

- Frequent travellers experience both comfort (t = -2.883) and service quality (t = -2.472) slightly lower than infrequent travellers, albeit the relationship is not extremely strong and only visible when comparing daily travellers to very infrequent travellers.

- Travel purpose often has little effect on traveller perception: the only significant relation found was that commuters perceive service quality (t = -2.827) and possibly comfort (t = -1.915) to be lower.

At the lowest level, we can have a look at how perceived occupancy and perceived reliability are influenced. It is logical that perceived occupancy is affected heavily by the actual occupancy (t = -9.823) and that perceived reliability is affected heavily by delays (t = -5.517). The latter relation is much weaker than the first. This might be because a delay is not always perceived as a delay[9] by a traveller, which weakens the relation between delay and perceived reliability. The model confirms this, as the relation between whether the passenger has experienced a delay or disruption and perceived reliability is much stronger (t = -23.041). While the relation between objective and subjective occupancy is significant and standardized coefficient is -0.469, which is far from one-on-one. It is interesting to ask why this is the case. The most likely reason is that passengers are unable to perfectly see the exact level of crowding in a vehicle.

Looking at service characteristics, it is interesting to see the effects of vehicle type on perceived crowding. Occupancy in a GTL is perceived significantly better different from the base scenario (buses) (t = 2.072), whereas Citadis (t = -4.071) and Avenio (t = -1.973) is rated much more negatively. Perhaps this could be explained by the way perceived occupancy is measured: it is measured as the probability to find a seat. Both newer trams are built to accommodate relatively more standing passengers, while a GTL has a relatively large number of seats compared to its total capacity, which reflects it performing better: the Load Factor of a fully loaded Avenio or Citadis tram is almost 50% higher than the Load Factor of a full GTL Tram (Wieffering, 2016).

When determining the net effect of vehicle type on overall customer satisfaction the effect of

[9] Examples in which a delay is not perceived as a delay include situations in which:

1. Customers do not check any timetable before travelling to a stop.
2. A delay makes that a passenger can board an earlier service than during normal operations.
3. Delays are (almost) equal to the headway of the line.

vehicle type on perceived comfort dominates: Avenio trams score significantly better than all other vehicles, between which no significant differences exist.

Lastly, some exogenous variables correlate as well. These connections are often logical (e.g. younger people travel more often for educational or commuting purposes than elderly). One interesting large correlation is the one between load factor and departure delay (t = 13.547). Fuller trams are thus significantly more delayed and delayed trams are fuller. This follows logic as well, but it is good to note that expectations and outcomes are in line.

5.3. Panel Effects

As section 3.5 has described, the data used in estimating the model above fails to meet one assumption which can affect results: the assumption of independence between responses. Therefore, the same model has been estimated again, this time using only one observation per respondent.

With 743 responses compared to 2858, the dataset used is, of course, much smaller. This means effects have to be much larger for the model to be significant. Nevertheless the general conclusions from this smaller model are the same. The fit indices drop with less than 0.01, and the relations found remain comparable. As a result, while the exact numeric outcome of this smaller model differs slightly qualitative conclusions remain the same. While some differences can be observed between both models, these seem to be mostly caused by differences in the sample size. If significance between variables changes, this is often for variables whose effects on customer satisfaction were very small anyway.

However, some differences between the two models are found as well. Some relations are found to be insignificant at the 0.05 level as a result of this. It is found that some personal characteristics, mainly travel frequency and travel purpose, are found to be insignificant in this smaller model. In general, this concerns relations which were very weak when quantified in the larger model. Moreover, the impact of frequency on perceived service quality and vehicle type (Avenio) on comfort just fails to meet the 0.05 threshold. As a result, in this smaller model these effects cannot be called significant. On the other hand, the effect of perceived safety on customer satisfaction is significant, albeit not large (0.074). The exact outcomes can be found in appendix E.

To conclude not being able to properly incorporate panel effects does not seem to cause too much harm. As a result, the model using the large dataset will be used as a basis in further analyses and policy implications.

5.4. Conclusion

To sum up, the model as defined in section 3.4 fits the dataset defined in section 3.5 sufficiently to be able to draw conclusions. The results show that the indirect relationship between occupancy levels and customer satisfaction exists and is significant, with comfort being the intermediate factor. Higher occupancy levels thus lead to lower customer satisfaction because of the discomfort high passenger numbers cause.

Looking at factors which further affect customer satisfaction indirectly, some interesting conclusions can be found. Firstly, vehicle type seems to have an impact on perceptions: the newer Avenio trams are evaluated as significantly more comfortable than other tram types and buses. An explanation for this relationship can be that the newer trams are more comfortable in general due to their young age, but additionally the Avenio can accommodate standing passengers better, offering them a more pleasant journey while standing.

Looking at the effect of personal characteristics, model results are in line with existing knowledge at HTM: elderly are more satisfied than younger travellers, females are slightly more satisfied than males and level of education and satisfaction correlate negatively. Travel purpose does not seem to have a huge effect on customer satisfaction, besides business travellers being slightly more negative than other passengers. Lastly, when analysing correlations between the exogenous variables in the model a number of significant correlations can be found, which can and have all be easily explained. Earlier HTM research on customer satisfaction is mostly confirmed by model results.

To conclude not being able to properly incorporate panel effects does not seem to cause too much harm. While some differences can be observed between both models, these seem to be mostly caused by differences in the sample size. If significance between variables changes, this is often for variables whose effects on customer satisfaction were very small anyway. As a result, the model using the large dataset will be used as a basis in further analyses and policy implications. These will be described in chapter six.

6. Implications

Chapter five has presented a model which provides quantitative insight on the impact which in-vehicle occupancy levels have on crowding perception and, ultimately, customer satisfaction. Having gained these insights, they can be used to help analyse and improve HTMs performance. Subquestions six to eight as described in chapter one relate to the outcome of the model, being:

- How does the predicted growth of travellers relate to the perception of crowding?
- How does the prediction of scores of seating opportunities affect HTMs predicted performance?
- How can HTM use this knowledge to improve its services?

This chapter will aim to answer these questions. Section 6.1 will present an excel calculation model which converts the model presented in chapter 5 into a calculation tool. Section 6.2 will give several examples of how this knowledge can be used to estimate the impact of several changes in HTMs services. Section 6.3 will generalise these insights to a context in which passenger numbers keep on growing.

6.1. Calculation model

The model presented in chapter 5 gives insight in the relations which affect customer satisfaction. The resulting model is in essence a set of linear regression models, and like linear regression models, besides its explanatory power it can be used to some extent to predict behaviour as well. In order to be able to use the model easily for calculations and estimations, the variables found in the structural equation model have been converted into an Excel model. A detailed manual of this model, how to use and how to adapt it can be found in appendix F.

Table 6.1 provides an overview of the input/output screen of the model. The user is asked to provide data on the service evaluated, and moreover personal characteristics can be provided. The calculation model will, as a result, estimate the average marks which would be given on all questions for that specific service. The unstandardized total effects as estimated by the Structural Equation Model serve as a means to make this calculation. An overview of all unstandardized total effects can be found in appendix E in table E.2. In order to turn these relations into a predictive model, it is necessary to assume that all means correspond to each other. This means that if a service has exactly the average characteristics and is evaluated by exactly an average person this gives the mean value of all indicators as an output. 'Average' in this case means a respondent with average personal characteristics based on the responses to the HTM Klantenpanel. Contrary to the model as estimated in chapter five in this case it is relevant that the composition of this panel is not equal to the composition of actual PT Travellers, as described in section 4.1. Given the conclusions from section 5.2 and the differences as observed in section 4.1, the overrepresentation of males and elderly will result in the model having a slight positive bias.

Better performing services will result in an above-average result, poorer services in a below-average result. As a result, caution should be taken when interpreting results literally: the change in a value is much more reliable than the actual value.

Table 6.1: Input/output screen of the model.

Input		
Trip characteristics		*Unit*
Vehicle type	Avenio	
Frequency	6	trip/h
Rush hour?	Y	Y/N
Number of passengers in vehicle on boarding	20	passengers
Delay on boarding	0	
Does passenger experience delay?	N	Y/N
Personal characteristics		
Age	30	Year
Gender	Male	
Education	Higher	
PT Travel Frequency	300	Trip/year

Output		
Load Factor	28.6% [pax/seats]	
Expected Marks	***Trip average***	***Specific person***
Overall satisfaction	**8.0**	**7**
Perceived crowding	7.9	7
General feeling of safety in PT	7.9	8
Comfort during trip	7.8	7
Cleanliness of vehicle	7.8	7
Cleanliness of stop	7.7	7
Information supply on stop	8.2	8
Information supply during delays or disruptions	6.9	6
Friendliness of staff	8.1	7
Driving style of driver	8.0	7
Ease of buying a ticket	8.9	8
Punctuality	8.3	8
Frequency	7.7	7
Feeling of safety during this trip	8.2	8

In general, the calculation is done as follows:

- The value entered is compared to the average values of the dataset used for the model for that specific variable. The absolute difference between the value entered and the average of the dataset is then calculated.
- The absolute difference is then multiplied by the total effect that this variable has on the variable which is being calculated. This results in the impact that this specific variable has on the score.
- This process is repeated for all variables shown in dark red in table 6.1. All values found are then summed to find the delta that this specific trip has compared to the average situation.

- This delta is then added to (or subtracted from) the average score to determine the expected score for this trip.

Table 6.2 gives an example of what this calculation looks like for the input data as provided in table 6.1 for general customer satisfaction. From top to bottom, it can be seen that the calculation starts with the average value for a variable in the dataset and for each input variable the model calculates what number needs to be added to or subtracted from this average to reach the expected value for this specific trip.

Table 6.2: Example of calculation in Excel Model

	Overall Satisfaction
Average in dataset	7.69
Trip variables	*Change compared to data average*
Vehicle type	0.26
Passenger numbers	0.06
Objective delay	0.05
Rush Hour?	-0.13
Frequency	0.00
Experienced Delay	0.10
Expected trip change to dataset average	0.29
Expected average mark for service	7.98
Personal characteristics	*Change compared to trip average*
Age	-0.28
Gender	0.05
Education	-0.23
PT Travel Frequency	-0.20
Expected personal change to trip average	-0.65
Expected mark given by person	7.33

By playing with input values the impact of changes can be seen: how does an increase of passenger numbers convert to a change in satisfaction values? How does a change in the vehicle type used affect perceptions? Using the tool, we can investigate this type of changes.

6.2. Impact of changes on customer satisfaction

The excel model presented in section 6.1 can be used to describe the effects of (small) changes in occupancy levels on expected customer satisfaction. This section will present four possible (theoretical) situations for which this can be done.

6.2.1. Occupancy levels

It can be calculated how many passengers need to be convinced to take an earlier or later tram to have an impact on expected customer satisfaction. Table 6.3 provides an overview of the impact of changes in (relative) occupancy on the different aspects of customer satisfaction. This is assuming all other factors besides occupancy such as delays and personal characteristics do not change.

Table 6.3: changes in average evaluation if the Load Factor changes

Increase in Load Factor (% point)	***5***	***10***	***20***
Overall satisfaction	-0.03	-0.05	-0.11
Perceived crowding	-0.16	-0.32	-0.64

It can be seen that while the score of crowding quickly changes when the vehicle becomes more crowded general satisfaction drops much slower. A change of almost 20 percentage point in the Load Factor (which corresponds to six passengers in a bus, fourteen passengers in an Avenio or GTL and seventeen in a Citadis) is necessary to change general customer satisfaction by 0.1. In order to change customer satisfaction in this way it is thus necessary to persuade quite large groups of travellers to take a different service. Of course, persuading passengers to take a different service also leads to a reduction in service quality on the service which sees increasing passenger numbers. It can be computed if the net effect nevertheless is positive.

This can be done as follows. In the current situation, the expected average customer satisfaction can be calculated as follows:

$$Satisfaction = \frac{\sum_{i=1}^{n} pax_i * satisfaction_i}{\sum_{i=1}^{n} pax_i}$$

In this formula Satisfaction is the expected customer satisfaction weighted over all services and passengers in a given time frame, n is the number of services in this time frame, pax_i is the number of passengers on service i and $satisfaction_i$ is the calculated expected customer satisfaction according to the model for service i.

In the ideal situation passengers are evenly distributed over services. The formula can then be used again to compute the expected overall satisfaction in the optimal situation. The difference between the two outcomes shows how much can be gained by distributing passengers better over services. The net effect of redistributing passengers more evenly over services on customer satisfaction is always positive, as in all cases more passengers profit from the 'busy' service being less crowded than suffer from the 'quiet' service being more crowded. This means that if passengers are more evenly distributed among services this leads to a structural improvement in customer satisfaction.

This theory can also be converted into practice. Many lines of HTMs network have strong passenger flows during rush hour, and as a result manly lines have extra services at specific moments during rush hour to travel. These high frequencies are necessary to transport all passengers, but services are often not used optimally. Two types of imbalances can be seen:

- Large differences in passenger numbers between two time periods.
- Large differences between passenger numbers between individual services.

As described above these differences result in satisfaction being lower than could be. For both cases an example will be discussed of how much can be gained.

In the first case (large differences between two adjacent time periods) lines 3 and 4 can be considered. These lines share a long line segment, between HMC Westeinde and Seghwaert, and this segment is among the busiest of HTMs network. Both lines run at frequencies of 6 vehicles per hour and are reinforced with line 4k during rush hour, which also has a frequency of 6 vehicles per hour.

To retrieve passenger numbers, all weekdays between January 1st, 2019 and June 30th, 2019 were used. All passenger numbers per line were retrieved. As an output the 80th percentile of this dataset was used. This value offers the most accurate representation of passenger numbers on busy day without irregularities and is the same way of determining occupancy numbers as HTM uses in own analyses. Table 6.4 shows occupancy numbers at Leidschenveen station in the direction of Den Haag during the morning peak aggregated at a 30 minute level. This aggregation was chosen as differences in passenger numbers between vehicles within this 30 minute level are negligible. The colour represents the occupancy rate of the vehicle, which can vary from dark green for an empty vehicle via yellow to dark red for a fully loaded vehicle.

Table 6.4: Passenger numbers at Leidschenveen (direction Den Haag Centraal). Services with an asterisk have an average Load Factor of over 100%.

Time	Line 3	Line 4	Line 4k
06:00			
06:30		*	
07:00			
07:30	*	*	*
08:00	*	*	*
08:30	*	*	*
09:00		*	
09:30			

It can be seen that on some moments differences in occupancy between lines are quite large, though it is likely that a large share of passengers board and alight at a stop at which all three lines stop – these three lines share 16 stops between the city centres of Zoetermeer and Den Haag. This data can be analysed both horizontally and vertically, the most notable things being:

- The large differences in passenger numbers between lines 3 and 4 between 06:00 and 06:30
- The large differences in passenger numbers between lines 3, 4 and 4k between 06:30 and 07:00
- The sudden large drop in passenger numbers seen between the intervals 09:00 – 09:30 and 09:30 – 10:00.

Table 6.5 shows the calculated evaluations for current occupancy numbers as well as the expected results if all passengers would be evenly distributed along vehicles, using the formula as described earlier. This shows evaluations can be improved by up to 0.05 on average for overall customer satisfaction and up to 0.30 on the probability of finding a seat by convincing passengers to take a different, less crowded vehicle.

Table 6.5: effects of evenly distributing passengers in table 6.4 on customer satisfaction

Interval	Current average overall satisfaction	Current perceived crowding	Expected average overall satisfaction	Expected perceived crowding
06:00 – 06:30	8.28	7.95	8.33	8.23
06:30 – 07:00	8.22	7.44	8.26	7.60
09:00 – 10:00	8.35	7.16	8.39	7.30

Secondly, at some lines passenger numbers during rush hour are quite imbalanced: between services, which often go every 10 minutes or even more frequent, large differences in occupancy can be seen. The main explanation according to HTM is connections: connecting trains and buses often go every 15 or 30 minutes and if trams drive more often, more people will travel in the service which connects best. Logically, these services are more crowded as a result. Not all travellers need to connect though. If some travellers can be convinced that travelling 10 minutes earlier or later results in a more pleasant journey this can lead to an improved experience. Of course, it needs to be ensured that the less busy service does not become overcrowded as a result.

As an example, table 6.6 shows departures of line 9 towards Vrederust at station HS between 17:30 and 19:00. Again, the 80th percentile was used based on occupancy data over the first six months of 2019. All vehicles drive towards the same terminus via the same route and stop at all stops, using Avenio tram types. In spite of this large differences in occupancy can be observed over a short period of time – the 18:27 service, for example, carries almost 2.4 times as much passengers as the 18:12 service. The largest differences in occupancy can be seen at the edges of peak hours when the frequency drops from 12 services/h to 6 services/h.

Table 6.6: Crowding of line 9 at Station Hollands Spoor (direction Vrederust) plus expected customer satisfaction. Services with an asterisk have an average Load Factor of over 100%.

Departure time	Number of passengers	Expected average Overall satisfaction	Expected average Perceived crowding
17:32	*	8.36	6.49
17:37	*	8.25	5.82
17:42		8.51	7.35
17:47	*	8.28	5.97
17:52	*	8.46	7.04
17:57	*	8.45	7.02
18:02		8.57	7.68
18:07	*	8.45	6.99
18:12		8.64	8.11
18:17	*	8.44	6.94
18:27	*	7.99	4.89

Again, it can be calculated how changes in passenger numbers affect overall customer satisfaction. The formula to do so remains the same, only the time span differs. Three pairs of services have been taken as an example, table 6.7 shows the effect of evenly distributing passengers travelling on these services on customer satisfaction.

As in the first example, the positive impact of making the busiest services quieter slightly outweighs the negative impact of making the quieter services busier. In the case considered, the gains are quite small: on the services with the largest differences up to 0.1 can be won on overall customer satisfaction on average. The impact on the rating of perceived crowding is much larger, as this evaluation drops much quicker in case of crowded vehicles. Here gains up to 0.3 on average can be made for the busiest service.

Table 6.7: effects of evenly distributing passengers in table 6.6 on customer satisfaction

Pair of Services	Current average overall satisfaction	Current perceived crowding	Expected average overall satisfaction	Expected perceived crowding
17:37 – 17:42	8.35	6.43	8.38	6.58
17:47 – 17:52	8.35	6.43	8.37	6.50
18:17 – 18:27	8.16	5.68	8.21	5.92

To conclude, by distributing passengers more evenly over services small gains can be made in customer satisfaction. Overall customer satisfaction can be improved by up to 0.05 point and perceived crowding can be improved by up 0.3 point on a scale of one to ten. While not enormous this could certainly help in improving passenger experience and customer satisfaction, albeit slightly. It is thus advisable to put effort in trying to distribute passengers more evenly among trams during rush hour.

6.2.2. Vehicle Types

At the moment, HTM uses four different vehicle types. Each of these vehicle types has its own characteristics and capacity. If the vehicle type changes. The model can be used to determine the effect

HTM uses two metrics to determine the usage of a service compared to its capacity:

- The 'inzetnorm', which is defined as the capacity of a vehicle during regular operations.
- The crush capacity, which is the maximum number of passengers which can reasonably be carried in a vehicle.

These norms are defined per vehicle type. Table 6.8 provides an overview of these norms for all vehicles HTM uses, both expressed in absolute passenger numbers and in percentage of the Load Factor.

Table 6.8: Inzet- en volnormen (Wieffering, 2016)

Vehicle Type	Inzetnorm (pax)	Inzetnorm (% LF)	Crush Capacity (pax)	Crush Capacity (% LF)
GTL-8	125	176	151	213
Regio Citadis	165	191	218	253
Avenio	150	214	183	261
MAN Lion City	50	161	77	248

However, as different vehicles have different capacities the change in Load Factor as result of using a different vehicle needs to be considered as well. Table 6.9 gives an overview of the number of seats in all vehicle types, both in absolute values and relative to each other. The difference in number of seats between an Avenio and a GTL tram is only one, the impact of a vehicle change between those two on the Load Factor is negligible. However, a Citadis has many more seats which, of course, does impact the Load Factor if the number of passengers remains the same. The replacement of trams with buses and vice versa is not considered here.

Table 6.9: (Relative) number of seats in all vehicle types

Vehicle Type	Number of seats (#)	Relative capacity (Bus = 1)
GTL-8	71	2.29
Regio Citadis	86	2.77
Avenio	70	2.26
MAN Lion City	31	1

Knowing all these values the effect of a change in vehicle type used on a service can be calculated. This is dependent on the level of crowding in a service before the change, of course. The change in overall satisfaction and perceived crowding can be calculated using the following formula:

$$\Delta_b = \left(c_{o,b} - c_{n,b}\right) + \frac{s_o - s_n}{s_o} * u_b * LF$$

In this equation:

- b is the attribute for which the change is calculated
- n is the new vehicle type
- o is the old vehicle type
- $c_{a,b}$ is the unstandardized total effect of vehicle type a on an attribute b
- s_a is the number of seats in vehicle type a
- u_b is the unstandardized total effect of in-vehicle passenger numbers on attribute b
- LF is the current Load Factor of the service for which the change is estimated.

Given a vehicle type and attribute, all values besides LF can be directly found in the model outcomes. For example, the effects of a vehicle type change on overall customer satisfaction and perceived crowding can be calculated. The unstandardized effect of Load Factor on overall customer satisfaction was found to be -0.539, which means overall customer satisfaction drops by 0.539 on a ten point scale if the load factor increases by 1. For perceived crowding this effect was found to be -3.188. The resulting formulas are then as follows:

$$\Delta_{overall\ satisfaction} = (c_o - c_n) + \frac{s_o - s_n}{s_o} * -0.539 * LF$$

$$\Delta_{perceived\ crowding} = (c_o - c_n) + \frac{s_o - s_n}{s_o} * -3.188 * LF$$

Regarding c_a chapter five found only the positive impact of Avenio trams to be significant for overall satisfaction and the positive impact of GTL trams to be significant for perceived crowding. This conclusion will be used here as well, and thus in this context only a distinction will be made between Avenio trams and all other vehicle types.

A future development which is to be expected in which this information is useful is the replacement of the remaining GTL-8 trams with new trams, which will be gradually done between the moment of writing (2019) and 2025. The newer trams generally score better, so an increase in overall satisfaction as a result of this replacement can be expected. In 2018 35% of passengers travelled on lines driven by a GTL (lines 1, 6, 12, 16) (HTM, 2019). Assuming the average Load Factor of 40.6% found in the dataset used for analysis, replacing all GTL trams with Avenio trams will lead to an expected increase in overall satisfaction of:

- Circa 0.25 for the lines on which currently GTL trams are used
- Circa 0.09 for the HTM network as a whole.

6.2.3. Frequency changes

A next case which is interesting to consider is a change in frequency on a line. The effect this has on travellers has two aspects: firstly, travellers experience a higher or lower frequency which affects their perceived service quality. However, especially on short term it also means that the same number of passengers is divided over a different number of vehicles. This strengthens the effect of increasing frequencies on the short term and worsens the effect of decreasing frequencies. For example, let us take a tram line with a frequency of 6 vehicles per direction per hour, which is standard on most HTM lines. If the frequency changes the same number of passengers need to be distributed over a different number of vehicles. This changes the occupancy of vehicles and thus impacts customer satisfaction. Table 6.10 provides an overview to give an impression of these effects on the general satisfaction with a trip, given a specific Load Factor per vehicle before the change. The change which includes the change of frequency and the corresponding change of occupancy in all vehicles. This includes:

- Frequency change
- Changes in passenger numbers per vehicle
- Change of expected delay as a result of the change in passenger numbers, defined using the correlation between Load Factor and delays

Table 6.10: Impact of frequency changes

Frequency change	**6 -> 5**	**6 -> 5**	**6 -> 7**	**6 -> 7**
Load Factor per vehicle before change (%)	*25*	*50*	*175*	*250*
Change in overall satisfaction	-0.045	-0.054	0.153	0.211

The most important conclusion to be drawn is that the positive impact of increasing frequencies on busy lines is much larger than the negative impact of decreasing frequencies on quiet services. The size of this effect becomes even larger when this difference is multiplied by the number of passengers: the number of passengers which experiences the smaller discomfort is much smaller than the number of passengers who profit from the larger increase in comfort.

It can be concluded that increasing frequencies on crowded lines results in significant increases in customer satisfaction for passengers on those lines, up to 0.2 point. If services have to be cut in order to make this possible, customer satisfaction suffers little if this is done by lowering frequencies on quiet lines.

6.2.4. Disruptions

Both for changes in passenger numbers and frequency changes it can be seen that effects on customer satisfaction are certainly present though not necessarily large. A recurring theme in literature has, however, been that perception plays an important role in customer satisfaction. The model includes a variable ‘delay experienced?’ This is a binary variable which represents the question: ‘Did you experience a delay or disruption during your journey?’ The model shows that the effect of a perceived disruption is much larger than the effect of an actual delay: each minute of actual delay leads to a drop of 0.06 in overall satisfaction while answering the question ‘did you experience a delay or disruption?’

with ‘Yes’ leads to a drop of 1.20 in overall satisfaction, equal to 20 minutes of delay. Aside from major disruptions in which running a service is impossible delays that size are not likely given the frequencies on HTMs network, in which vehicles often drive every 10 or 15 minutes. As the effects of the perception of a delay are much higher than the effects of actual delays, focus should be given on decreasing the extent to which passengers experience delays in particular to increase satisfaction.

6.3. Growing passenger numbers and customer satisfaction

Sections 6.1 has presented a tool which is able to estimate the impact of changes in the service provided by HTM and section 6.2 has discussed some possible interesting changes which could exist. However, as mentioned in chapter 1 HTM aims at transporting more passengers over the upcoming years and development in and around Den Haag will likely also result in more and more demand for Public Transport in the region. As a result, it can be concluded that passenger numbers on the HTM network will continue to increase over the upcoming years – between 2015 and 2018 passenger numbers at HTM grew by 1.4% per year on average, with the growth in Ypenburg, Rijswijk and Delft being as high as 3.9% during weekdays and 6.1% on weekends (HTM, 2019). Naturally this results in higher pressure on the network.

Subquestion 6 as mentioned in section 1.2 was formulated because of this:

How does the predicted growth of travellers relate to the perception of crowding?

If it is assumed the number of passengers grows with a constant rate, whether it is for a specific service, a line or the network as a whole, the number of passengers in a specific vehicle in x year from now can be calculated using the formula:

$$pax_{future} = pax_{now} * (1 + \frac{n}{100})^x$$

An estimation can then be made on how long it takes before passenger growth has a noticeable negative impact on customer satisfaction levels. This differs, of course, based on current passenger numbers and vehicle types. Table 6.11 provides an overview for a variety of occupancy levels in the current situation for two values of n: 1% and 3%. These values are the lower and upper bound for the HTM network as a whole, according to Janiek de Kruijff, business analyst responsible for estimating passenger growth at HTM (personal communication, July 23, 2019). As long as all other input variables (frequency, delay, personal characteristics) remain equal these values do not change. Figure 6.12 shows table 6.11 visually: the relation between passenger growth and its effect on customer satisfaction seems to be exponential.

The main conclusion to be drawn that in case of an equal passenger growth of 1% on all lines and services of the network the expected drop in customer satisfaction cause purely by increasing passenger numbers is negligible in most cases on the short term. If passenger growth is increased to 3% a year, however, customer satisfaction can be expected to drop significantly on short notice if no action is taken, especially on these network sections which are already quite crowded. It should thus be noted that these conclusions are extremely dependent on the expected passenger growth due to the exponential nature of a growth based on x % per year. The Load Factors used in table 6.9 are all widely seen in HTMs network: Load Factors of 200% are seen during rush hours on the busiest parts of the network, for example in the Tram Tunnel Grote Markt – more quiet services are seen all over the place outside peak travel moments.

Table 6.11: Number of years before overall customer satisfaction drops by 0.1 point, given current Load Factor and constant yearly growth

Current occupancy (LF) (%)	Yearly growth	
	1%	3%
50	31	11
75	22	7
100	17	6
125	13	5
150	11	4
175 (≈ inzetnorm GTL/Bus)	10	3.5
200 (≈ inzetnorm Citadis/Avenio)	9	3
250 (≈ crush capacity all vehicles)	7	2.5

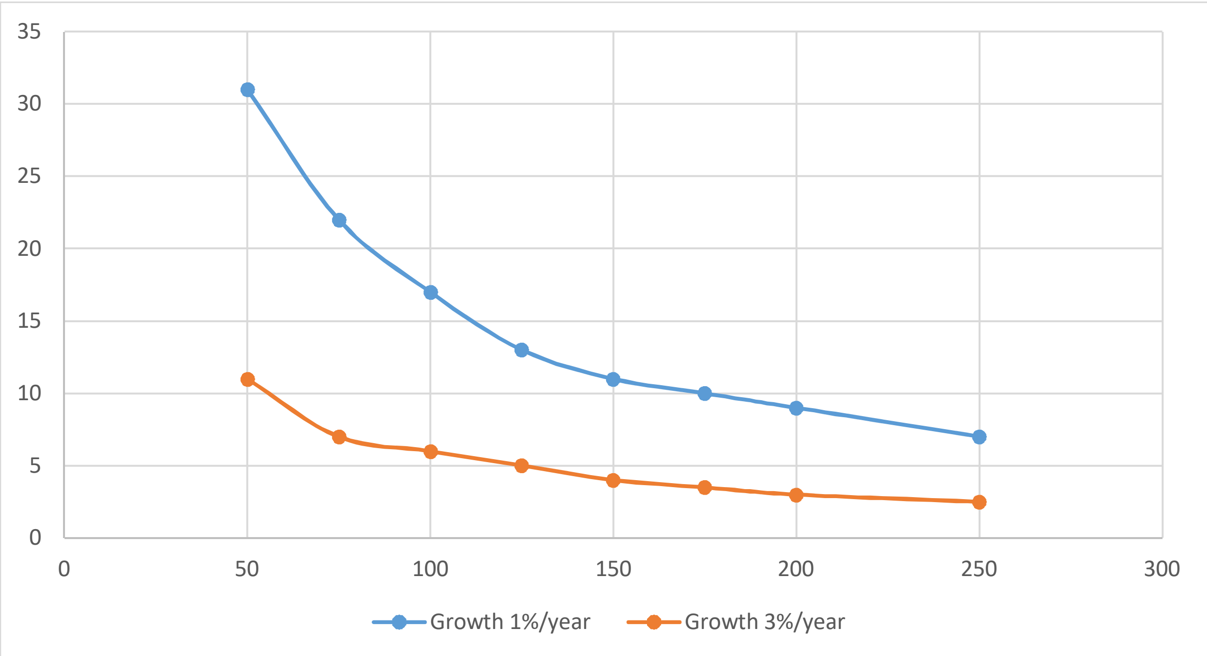


Figure 6.12: Number of years before overall customer satisfaction drops by 0.1 point, given current Load Factor and a fixed yearly growth

To conclude, the size of the effect of passenger growth on overall customer satisfaction depends heavily on how large growth is. In case of a relatively small growth of 1% per year it will take more than five years before effects become noticeable even for the most crowded sections. If the growth is larger (e.g. 3%) effects can be seen on most services within seven years. As section 6.2.1 has shown ensuring this growth is evenly distributed is crucial to minimize the negative impact on customer satisfaction.

6.4. Conclusion

From these analyses, a variety of conclusions can be drawn. It can be seen that changes in the in-vehicle occupancy or frequency of a service have an impact on customer satisfaction when evaluating such a service. Based on the model these effects are not very large but certainly noticeable. It is shown that in order to improve overall customer satisfaction by more than 0.1 point in-vehicle passenger numbers have to be changed with more than 20 percentage point of the Load Factor. However, small improvements on the scale of one or two tenths of a point can certainly be achieved by tweaking the current service level... Analysis of lines 3, 4 and 9 showed that without changing the timetable and with equal passenger numbers overall customer satisfaction can be improved by up to 0.1 point and the evaluation of seating availability can be increased by up to 0.5 point just by ensuring passengers are more evenly distributed over existing services.

Most importantly, the perception of a disruption or delay on a trip has an extremely heavy impact on the evaluation of a trip, much more than in the case of actual delays. An important lesson to learn here is that it needs to be evaluated whether a disruption or delay actually hinders passengers. If a passenger does not experience a delay it might be better to not inform him even if there is an actual disruption, as satisfaction is heavily affected by the perception of a delay or disruption.

An analysis of the effect of passenger growth on customer satisfaction showed that except for large increases on already crowded sections short term effects of increasing passenger numbers are relatively small. It should be noted that on the long term negative effects on customer satisfaction can be seen due to increased crowding. The time span of this effect does mean that there is time for adequate changes in service supply to accommodate these extra passengers.

7. Conclusion and discussion

This chapter wraps up all insights: section 7.1 answers the main research question as stated in chapter one and provides an overview of conclusions. Section 7.2 provides resulting policy recommendations for HTM. In section 7.3 the author reflects on this thesis and provides some academic and scientific recommendations.

7.1. Conclusion

While the effect of in-vehicle crowding on customer experience and customer satisfaction intuitively is logical, research on this topic has been very limited. This thesis tried to investigate this topic, resulting in the main research question for this thesis:

What is the relation between objective and subjective in-vehicle crowding in Public Transport and customer satisfaction?

In literature, customer satisfaction in Public Transport is mostly defined as the difference between experience and expectation. Both are influenced by two main factors:

- The performance of the system. This includes both the lay-out of the system, which includes the route and timetable of the line, vehicle type used, etc., as well as the performance of the specific service which is used by the respondent, which includes, for example, in-vehicle occupancy and punctuality of that service.
- A variety of personal characteristics, such as age or past PT experiences, which vary per customer which define how the customer perceives the experience of the system.

In an urban public transport network, crowding occurs when too many passengers travel with a certain vehicle compared to the capacity of this vehicle. The optimal way to measure crowding uses a metric which represents the ratio between the number of passengers and vehicle capacity, as the capacity. Literature is divided on the exact moment when crowding effects start to come into play. It is certain, however, that travelling standing or sitting has an effect on how a customer experiences crowding.

Based on literature and available data, a conceptual framework was developed which is shown in figure 7.1. This framework explains the relation between occupancy, in-vehicle crowding experience and customer satisfaction and was mathematically analysed using a Structural Equation Model.

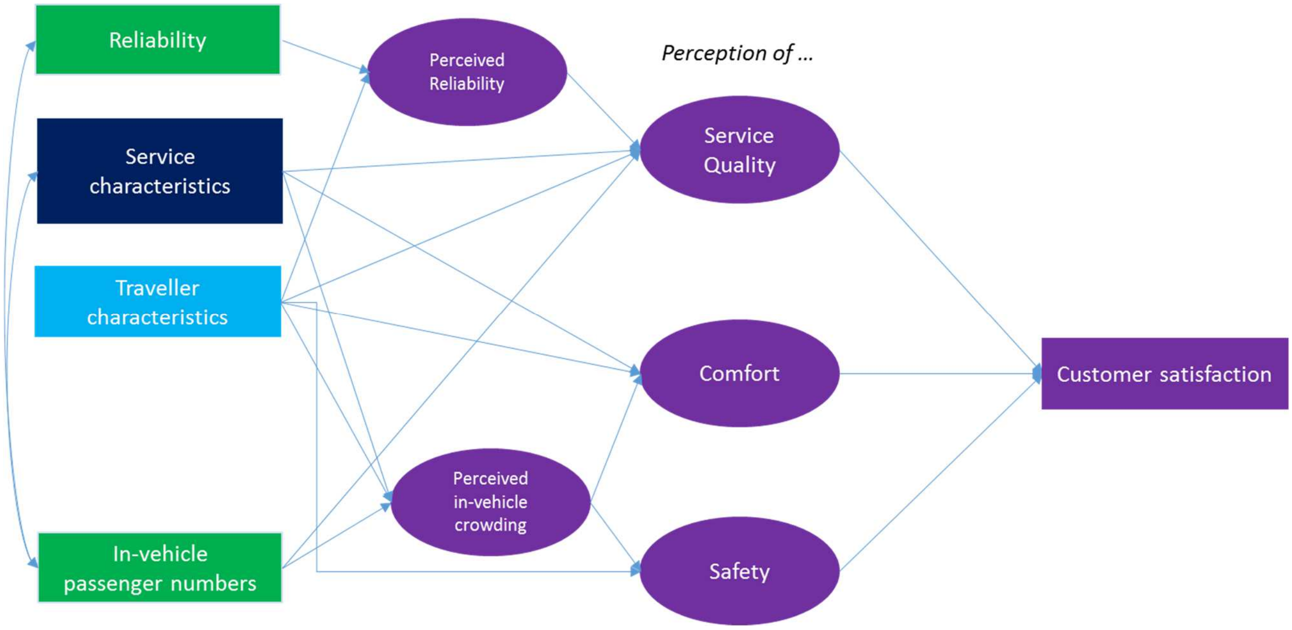


Figure 7.1: Framework used for analysis

No reason was found to reject linearity in the relation between objective and subjective in-vehicle crowding, with the subjective evaluation of in-vehicle crowding dropping linear if in-vehicle passenger numbers rise. Literature found the relation not to be one-on-one and analysis using just these two variables showed a lot of other noise which affected the perception of crowding.

In the estimated Structural Equation Model, as presented in figure 5.3, the effect of in-vehicle crowding on customer satisfaction in Public Transport was found to be significant: in the model estimated the total correlation between objective occupancy and customer satisfaction was estimated at -0.1 and between subjective occupancy and customer satisfaction to be 0.2. The difference in sign can be explained by the method of measurement: higher values on objective occupancy correspond with more passengers and thus less satisfaction and higher values on subjective crowding correspond with fewer passengers and thus a higher level of satisfaction. The results thus showed that occupancy levels had a significant impact on customer satisfaction. The results confirmed the hypothesis that the effect of crowding of customer satisfaction is indirect: crowding affects perceived comfort and service quality which in turn affects customer satisfaction.

An Excel calculation tool was developed to be able to quantify the model results for practical situations and to be able to calculate the expected effect of changes in HTMs services. From calculations made with this tool several conclusions can be drawn:

- Given service and passenger numbers, optimal customer satisfaction is reached when passengers are distributed evenly among services.
- The effect of lowering frequencies on quiet lines is smaller than increasing frequencies on busy lines on the level of an individual passenger: lowering the frequency by one service on a line with 6 vehicles per hour and a Load Factor of 50% results in a drop of overall satisfaction with 0.05, while increasing the frequency by one service on a line with 6 vehicles per hour and a Load Factor of 175% results in an increase of overall satisfaction by 0.15. As much more passengers travel on busy lines compared to quiet lines, this effect becomes much larger when all passengers are considered instead of one.
- Growth estimates for the HTM network vary between 1% and 3% per year on average. Analysis showed that in case of a growth near the lower end of this estimate increase of passenger numbers will have little effect on customer satisfaction even over a time span of a decade,

while a growth near the upper end requires action within five years to prevent measurable lower customer satisfaction as a result of crowding.
- The analysis of the impact of a (perceived) disruption illustrates that the effect of perceiving a disruption or delay by a passenger is much larger than the effect of an actual delay or disruption. An experienced delay results in a drop of overall satisfaction by 1.20 on a ten point scale, while each minute of actual delay only accounts for a drop of 0.06 point.

7.2. Policy Recommendations

Given the conclusions, a number of recommendations can be made to HTM. These recommendations can be classified into two categories: advice given on the analysis done and model outcomes, and recommendations for further internal research at HTM. First the recommendations based on conclusions can be named:

- Small improvements in overall customer satisfaction can be made by using existing capacity in a more efficient way. Analysis of lines 3, 4 and 9 showed that without changing the timetable and with equal passenger numbers overall customer satisfaction can be improved by up to 0.1 point and the evaluation of seating availability can be increased by up to 0.5 point just by ensuring passengers are perfectly evenly distributed over existing services, regardless of actual passenger numbers.
- From a customer satisfaction perspective it can be recommended to lower frequencies on quiet lines (LF < 30% of the inzetnorm) to allow for more services on busy (LF > inzetnorm) lines if possible. Even on an individual level the decrease in overall satisfaction from lowering frequencies slightly is offset by a higher increase in overall satisfaction – multiplied by the number of passengers experiencing this change this effect becomes much stronger.
- The effect of a perceived delay is much larger than the effect of an objective delay. A recommendation is thus to ensure passengers are not disappointed when travelling, and to not point them to disruptions which would otherwise have gone by unnoticed.
- The impact of passenger growth on customer satisfaction is highly dependent on the size of growth of passenger numbers. An increase of the growth by one percentage point can make a large difference. It is advisable to closely monitor realised growth – if the growth is 3% or higher per year on busy lines action on short notice is to be advised.
- The new Avenio trams have a clear positive impact on customer satisfaction compared to the older trams. Replacing the remaining GTL-8 trams with a comparable tram type is likely to further improve satisfaction on the lines which are operated by GTL-8 trams today.
- The coupling of evaluations of the HTM Klantenpanel to occupancy numbers proved to be time-consuming because of a variety of issues. Most eye-catching was the fact that stops in the Klantenpanel were sometimes named differently from their names in HTMs databases and that changes in line routing were not immediately changed in the panel stop input data. This has led to unnecessary data loss in this research and makes repetition of the analysis carried out time-consuming.

7.3. Academic Discussion

During the research resulting in this thesis, assumptions have been made in order to come to the results that have been reached. Sometimes, these assumptions were enforced by data or time limitations and sometimes they were made simply because choices have to be made. This section will discuss assumptions made and results achieved and note how these could be altered and improved to gain further knowledge.

Firstly, the framework used as a basis for the statistical model can be evaluated. Structural Equation Modelling is a confirmatory analysis which requires a theoretical framework to be developed and then tests whether this framework is correct. Literature has used a variety of different categorisations to explain customer satisfaction in Public Transport. Some studies have opted to use a number of factors which together define customer satisfaction while others directly identify relevant, measurable attributes to explain it. This thesis has chosen for a structure in which customer satisfaction is explained by three dimensions: perceived service quality, perceived comfort and perceived safety. The model presented in chapter 5 found very high correlations between these factors. This suggests that a different construction of customer satisfaction might fit this data better. This could either be done using less dimensions (one or two), or by not using dimensions all together and have all indicators correlate directly with customer satisfaction. This thesis did not consider this type of construction for the Den Haag case. In later research, an exploratory analysis testing different set-ups of how customer satisfaction can be defined model wise is useful to get to know what the best construction is.

The model could also be made more accurate by expanding the factors which were left out: several factors which are known to have an impact on how a passengers experience his journey were not included. This includes, amongst others, the effect which the weather has on passenger numbers (on a rainy day people will tend to use PT more, on a sunny day they might prefer to bike) as well as passenger experience. This was not included because there is insufficient knowledge on how weather affects perceptions exactly: which factors are relevant and what does this relation look like? Quantifying the effect of weather on customer perception in PT might be an interesting topic for research.

Regarding data some comments can be made as well. Using existing data from the HTM Klantenpanel had the advantage of having a large dataset ready for analysis and thus saving a lot of time in data collection and preparation. On the other hand it meant that some insights which were identified in literature, such as customer perception of speed, were not known in analysis. In an optimal situation with enough time, a dataset is collected which does include these insights, as this allows for questions which can be specified based on the exact knowledge gap. This could for example be useful to gain more insight in customer expectations. It is interesting to see what effects this might have on customer satisfaction.

The choice for a linear relation between crowding and customer satisfaction was in line with the research by Haywood et al. (2017). The discussion of their research in section 2.2 showed scepticism about their conclusions, as it feels counterintuitive that the difference between a Load Factor of 0% and 50% affects customer satisfaction as much as the difference between a Load Factor of 75% and 125%. The conclusion found still feels to be counterintuitive and a different mathematical function might model the actual relationship better.

The choice for the Load Factor as measurement for crowding can also be discussed. As Yap et al. (2018) showed, using the Load Factor yields a slightly different crowding multiplier for different vehicle types due to the differences between vehicles. The Load Factor was chosen as a measurement due to its ease of interpretation and the relatively low occupancy levels in the dataset. Moreover the model takes the differences in perception between different vehicle types into account.

Regarding the effect of frequency changes on customer satisfaction it was not considered whether passengers change their route or mode choice because of a frequency change. Lowering frequencies can lead to a decrease in passenger numbers if passengers decide to take a different route or modality because of a lower frequency.

However, as no quantitative model of the effect of in-vehicle crowding on customer satisfaction in Public Transport has been made up to this thesis, it is logical that improvements and drawbacks were found. The thesis has shown that it is indeed possible to quantify the impact of in-vehicle crowding on customer satisfaction in a highly complex environment with subjective factors, and has successfully done so. The results provide a stable foundation for further research and assist HTM in decision making using quantitative arguments.

Bibliography

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

Abenoza, R., Cats, O., & Susilo, Y. (2018). How does travel satisfaction sum up? An exploratory analysis in decomposing the door-to-door experience for multimodal trips. *Transportation*, 1-28.

Bakker, P. (2018). *De Keuze van de Reiziger.* Kennisinstituut voor Mobiliteitsbeleid, Den Haag.

Bakker, P., Van Der Loop, H., & Savelberg, F. (2015). *Uitwisseling gebruikersgroepen 'auto-ov' Kennisinstituut voor Mobiliteitsbeleid.*

Björklund, G., & Swärdh, J. (2017). Estimating policy values for in-vehicle comfort and crowding reduction in local public transport☆. *Transportation Research Part A: Policy and Practice, 106*, 453-472.

Bussink, B., & De Konink, S. (2015). *Het vergelijken van reizigerstevredenheid tussen gebieden: voor welke factoren corrigeren?* Den Haag.

Byrne, B. (2013). *Structural Equation Modelling with AMOS: Basic concepts, applications and programming.* Routledge.

Cantwell, M., Caulfield, B., & O'Mahony, M. (2009). Examining the Factors that Impact upon Public Transport Commuting Stress. *Journal of Public Transportation 12, 2*, 1-21.

Cats, O., Abenoza, R., Liu, C., & Susilo, Y. (2015). Identifying priority areas based on a thirteen years evolution of satisfaction with public transport and its determinants. *Transportation Research Record, 2538*, 86-95.

Cats, O., West, J., & Eliasson, J. (2016). A dynamic stochastic model for evaluating congestion and crowding effects in transit systems. *Transportation Research Part B: Methodological, 89*, 43-57.

CBS. (2018). Trends in Nederland 2018 - Onderwijs. Den Haag.

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*(3), 605-622.

de Palma, A., Kilani, M., & Proost, S. (2015). Discomfort in mass transit and its implication for scheduling and pricing. *Transportation Research Part B: Methodological, 71*, 1-18.

Del Castillo, J., & Benitez, F. (2012). A Methodology for Modeling and Identifying Users Satisfaction Issues in Public Transport Systems Based on Users Surveys. *Procedia - Social and Behavioral Sciences, 54*, 1104-1114.

Diana, M. (2012). Measuring the satisfaction of multimodal travelers for local transit services in different urban contexts. *Transportation Research Part A: Policy and Practice, 46*, 1-11.

Eboli, L., & Mazzulla, G. (2007). Service Quality Attributes Affecting Customer Satisfaction for Bus Transit. *Journal of Public Transportation, 10*(3), 21-34.

European Union RTD Programme - Project QUATTRO. (1998). *Quality Approach in Tendering Urban Public Transport Operations.* European Union, Belgium.

Fellesson, M., & Friman, M. (2008). Perceived Satisfaction with Public Transport Service in Nine European Cities. *Journal of the Transportation Research Forum, 47*(3), 93-103.

Friman, M., & Fellesson, M. (2009). Service Supply and Customer Satisfaction in Public Transportation Service Supply and Customer Satisfaction in Public Transportation: The Quality Paradox. *Journal of Public Transportation, 12*(4), 57-69.

Friman, M., Edvardsson, B., & Gärling, T. (2001). Frequency of negative critical incidents and satisfaction with public transport services. I. *Journal of Retailing and Consumer Services, 8*, 95-101.

Gemeente Den Haag. (2019). *Bevolkingsmonitor gemeente Den Haag*. Opgehaald van https://denhaag.buurtmonitor.nl/dashboard/Bevolking-en-wonen/Maandmonitor-bevolking/?regionlevel=gemeente

Gemeente Den Haag. (2019). Den Haag in Cijfers. Den Haag.

Guirao, B., García-Pastor, A., & López-Lambas, M. (2016). The importance of service quality attributes in public transportation: Narrowing the gap between scientific research and practitioners' needs. *Transport Policy, 49*, 68-77.

Haywood, L., & Koning, M. (2015). The distribution of crowding costs in public transport: New evidence from Paris. *Transportation Research Part A: Policy and Practice, 77*, 182-201.

Haywood, L., Koning, M., & Monchambert, G. (2017). Crowding in public transport: Who cares and why? *Transportation Research Part A: Policy and Practice, 100*, 215-227.

Hong, S., Min, Y., Park, M., Kim, K., & Oh, S. (2016). Precise estimation of connections of metro passengers from Smart Card data. *Transportation, 43*, 749-769.

Hörcher, D., Graham, D., & Anderson, R. (2017). Crowding cost estimation with large scale smart card and vehicle location data. *Transportation Research Part B: Methodological, 95*, 105-125.

Hörcher, D., Graham, D., & Anderson, R. (2018). The economics of seat provision in public transport. *Transportation Research Part E: Logistics and Transportation Review, 109*, 277-292.

HTM. (2018). *Onze missie & visie*. Opgehaald van https://www.overhtm.nl/nl/over-ons/onze-missie-visie/

HTM Personenvervoer N.V. (2018). *Jaarverslag 2017.* Den Haag.

HTM Personenvervoer N.V. (2019). *Jaarverslag 2018.* Den Haag.

HTM Personenvervoer NV. (2017). *Jaarverslag 2017.* HTM Personenvervoer NV, Den Haag.

Jonkeren, O., Harms, L., Jorritsma, P., Huibregtse, O., & Bakker, P. (sd). *Kennisinstituut voor Mobiliteitsbeleid | Waar zouden we zijn zonder de fiets en de trein?*

Kahneman, D. (2012). *Thinking, Fast and Slow.* London: Penguin Books Ltd.

Kennisinstituut voor Mobiliteitsbeleid. (2013). *De maatschappelijke waarde van kortere en betrouwbaardere reistijden.* Kennisinstituut voor Mobiliteitsbeleid, Den Haag.

Kroes, E., Kouwenhoven, M., Debrincat, L., & Pauget, N. (2013). *On the value of crowding in public transport for île-de-France.* OECD.

Kroesen, M. (2017). Travel Behaviour Research. Delft.

Kruijff, J. d. (2019, July 23). Personal communication. (M. Seerden, Interviewer) Den Haag.

Kubat, M. (2017). *An Introduction to Machine Learning.* Springer International Publishing.

Lapan, S., Quartaroli, M., & Riemer, F. (2012). *Qualitative Research: An Introduction to Methods and Designs* (1st ed.). San Francisco: Jossey-Bass.

Li, Z., & Hensher, D. (2011). Crowding and public transport: A review of willingness to pay evidence and its relevance in project appraisal. *Transport Policy, 18*, 880-887.

Li, Z., & Hensher, D. (2013). Crowding in Public Transport: A Review of Objective and Subjective Measures. *Journal of Public Transportation, 16*(2), 107-134.

Maister, D. (1985). *The Psychology of Waiting Lines.*

Maslow, A. (1943). A theory of human motivation. *Psychological Review, 50*(4), 370-396.

Ministerie van I&M. (2014). *Concessie voor het hoofdrailnet 2015-2025.* Ministerie van Infrastructuur & Milieu, Den Haag.

Mohd Mahudin, N., Cox, T., & Griffiths, A. (2012). Measuring rail passenger crowding: Scale development and psychometric properties. *Transportation Research Part F: Traffic Psychology and Behaviour, 15*, 38-51.

Montgomery, D., Peck, E., & Vining, G. (2015). *Introduction to Linear Regression Analysis* (5 ed.). Hoboken: John Wiley & Sons Inc.

Morfoulaki, M., Tyrinopoulos, Y., & Aifadopoulou, G. (2010). Estimation of Satisfied Customers in Public Transport Systems: A New Methodological Approach. *Journal of the Transportation Research Forum.*

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.

Mouwen, A. (2015). Drivers of customer satisfaction with public transport services. *Transportation Research Part A: Policy and Practice, 78*, 1-20.

MRDH. (2017). *Concessiemonitor MRDH 2017.* Metropoolregio Rotterdam-Den Haag.

MRDH. (2019). *Kerngetallen MRDH*. Opgehaald van https://mrdh.nl/gemeenten

Nachtigall, C., Kroehne, U., Funke, F., & Steyer, R. (2003). (Why) Should we use SEM? Pros and cons of Structural Equation Modelling. *Methods of Psychological Research (8)*, pp. 1-22.

Olsson, L., Friman, M., Pareigis, J., & Edvardsson, B. (2012). Measuring service experience: Applying the satisfaction with travel scale in public transport. *Journal of Retailing and Consumer Services, 19*, 413-418.

Prud'homme, R., Koning, M., Lenormand, L., & Fehr, A. (2012). Public transport congestion costs: The case of the Paris subway. *Transport Policy*.

Qin, F. (2014). Investigating the in-vehicle crowding cost functions for public transit modes. *Mathematical Problems in Engineering, 2014*, 13.

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.

Şimşekoğlu, Ö., Nordfjærn, T., & Rundmo, T. (2015). The role of attitudes, transport priorities, and car use habit for travel mode use and intentions to use public transportation in an urban Norwegian public. *Transport Policy, 42*, 113-120.

Susilo, Y., Lyons, G., Jain, J., & Atkins, S. (2012). Rail Passengers' Time Use and Utility Assessment. *Transportation Research Record: Journal of the Transportation Research Board, 2323*, 99-109.

Theler, B., & Axhausen, K. (2013). *When is a bus full? A study of perception.* ETH Zürich.

Tirachini, A., Hensher, D., & Rose, J. (2013). Crowding in public transport systems: Effects on users, operation and implications for the estimation of demand. *Transportation Research Part A: Policy and Practice, 53*, 36-52.

Tirachini, A., Hensher, D., & Rose, J. (2014). Multimodal pricing and optimal design of urban public transport: The interplay between traffic congestion and bus crowding. *Transportation Research Part B: Methodological, 61*, 33-54.

Tirachini, A., Hurtubia, R., Dekker, T., & Daziano, R. (2017). Estimation of crowding discomfort in public transport: Results from Santiago de Chile. *Transportation Research Part A: Policy and Practice, 103*, 311-326.

Tirachini, A., Sun, L., Erath, A., & Chakirov, A. (2016). Valuation of sitting and standing in metro trains using revealed preferences. *Transport Policy, 47*, 94-104.

van de Wiel, C. (2017, 1 26). NS voor de rechter om overvolle treinen. *NRC Handelsblad*.

van der Loop, H., Bakker, P., Savelberg, F., Kouwenhoven, M., & Helder, E. (2018). *Verklaring van de ontwikkeling van het ov-gebruik in Nederland over 2005-2016.* Kennisinstituut voor Mobiliteitsbeleid, Den Haag.

van Hagen, M. (2011). *Waiting Experience at Train Station.* Delft: Eburon Academic Publishers.

van Hagen, M., & Bron, P. (2014). Enhancing the Experience of the Train Journey: Changing the Focus from Satisfaction to Emotional Experience of Customers. *Transportation Research Procedia*, (pp. 253-263). Frankfurt.

van Hagen, M., & Sauren, J. (2014). Influencing the Train Experience: Using a Successful Measurement Instrument. *Transportation Research Procedia*, (pp. 264-275). Frankfurt.

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.

van Oort, N., Brands, T., de Romph, E., & Yap, M. (2016). Ridership Evaluation and Prediction in Public Transport by Processing Smart Card Data:A Dutch Approach and Example. In N. van Oort, T. Brands, E. de Romph, M. Yap, F. Kurauchi, & J. Schmöcker (Red.), *Public Transport Planning with Smart Card Data.* CRC Press.

van 't Hart, J. (2012). *Increasing customer satisfaction with public transport.* Delft.

Wardman, M., & Whelan, G. (2011). Twenty years of rail crowding valuation studies: Evidence and lessons from British experience. *Transport Reviews*.

Wieffering, M. (2016). *Inzet- en volnormen HTM voertuigen.* NHTV, Den Haag.

Yap, M., Cats, O., & van Arem, B. (2018). Crowding valuation in urban tram and bus transportation based on smart card data. *Transportmetrica A: Transport Science*.

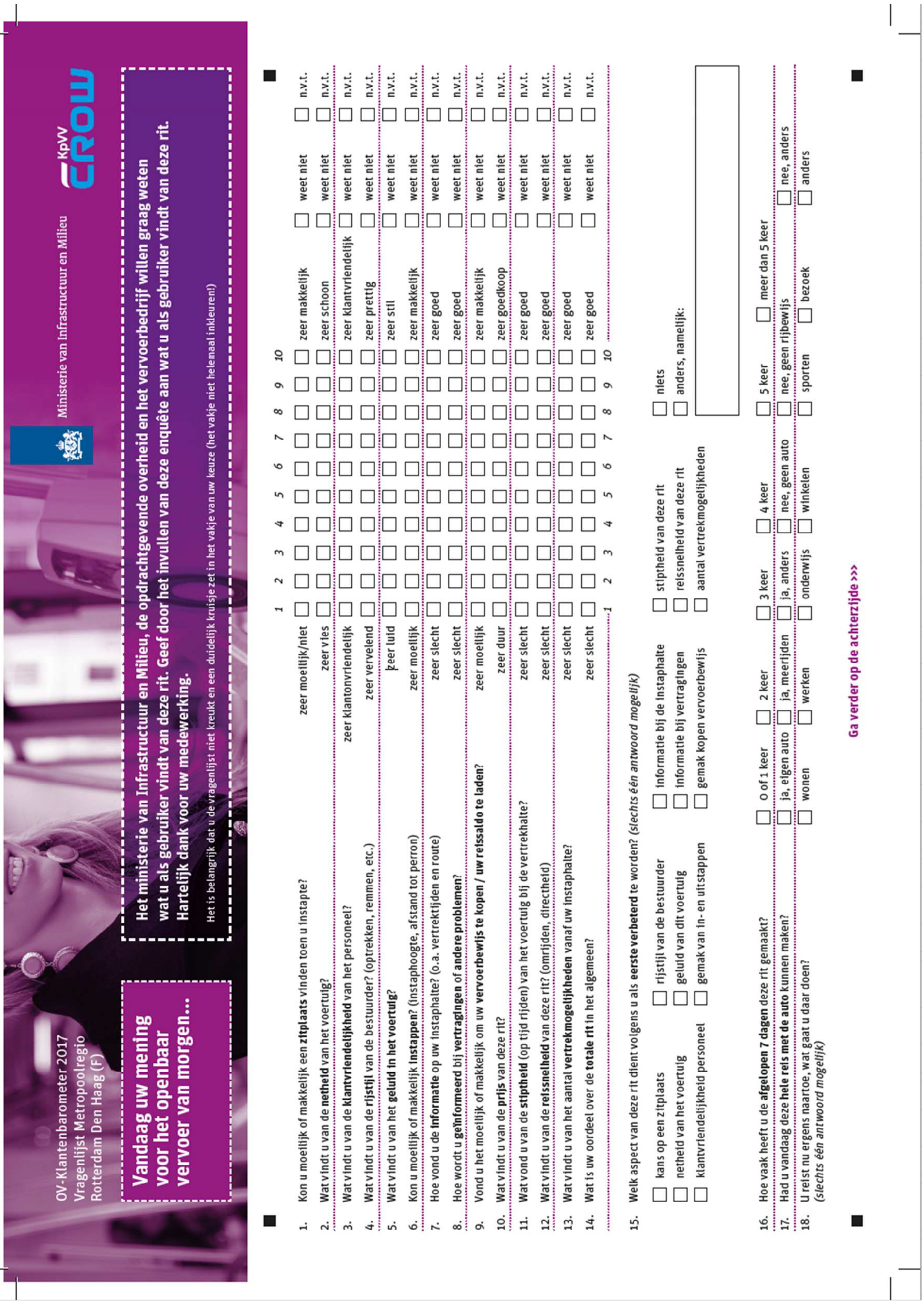
Yaya, L., Fortià, M., Canals, C., & Marimon, F. (2015). Service quality assessment of public transport and the implication role of demographic characteristics. *Public Transport, 7*, 409-428.

Zijlstra, T., Bakker, P., Harms, L., Durand, A., & Wüst, H. (2018). *Busgebruikers door dik en dun.* Kennisinstituut voor Mobiliteitsbeleid, Den Haag.

Appendices

Appendix A: Questionnaires

- OV Klantbarometer



OV-Klantenbarometer 2017
Vragenlijst Metropoolregio
Rotterdam Den Haag (F)

Ministerie van Infrastructuur en Milieu

KpVV CROW

Vandaag uw mening voor het openbaar vervoer van morgen...

Het ministerie van Infrastructuur en Milieu, de opdrachtgevende overheid en het vervoerbedrijf willen graag weten wat u als gebruiker vindt van deze rit. Geef door het invullen van deze enquête aan wat u als gebruiker vindt van deze rit. Hartelijk dank voor uw medewerking.

Het is belangrijk dat u de vragenlijst niet kreukt en een duidelijk kruisje zet in het vakje van uw keuze (het vakje niet helemaal inkleuren!)

			1	2	3	4	5	6	7	8	9	10			
1.	Kon u moeilijk of makkelijk een **zitplaats** vinden toen u instapte?	zeer moeilijk/niet	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	zeer makkelijk	☐ weet niet	☐ n.v.t.
2.	Wat vindt u van de **netheid** van het voertuig?	zeer vies	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	zeer schoon	☐ weet niet	☐ n.v.t.
3.	Wat vindt u van de **klantvriendelijkheid** van het personeel?	zeer klantonvriendelijk	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	zeer klantvriendelijk	☐ weet niet	☐ n.v.t.
4.	Wat vindt u van de **rijstijl** van de bestuurder? (optrekken, remmen, etc.)	zeer vervelend	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	zeer prettig	☐ weet niet	☐ n.v.t.
5.	Wat vindt u van het **geluid in het voertuig**?	zeer luid	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	zeer stil	☐ weet niet	☐ n.v.t.
6.	Kon u moeilijk of makkelijk **instappen**? (instaphoogte, afstand tot perron)	zeer moeilijk	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	zeer makkelijk	☐ weet niet	☐ n.v.t.
7.	Hoe vond u de **informatie** op uw instaphalte? (o.a. vertrektijden en route)	zeer slecht	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	zeer goed	☐ weet niet	☐ n.v.t.
8.	Hoe wordt u **geïnformeerd** bij **vertragingen** of **andere problemen**?	zeer slecht	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	zeer goed	☐ weet niet	☐ n.v.t.
9.	Vond u het moeilijk of makkelijk om uw **vervoerbewijs te kopen / uw reissaldo te laden**?	zeer moeilijk	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	zeer makkelijk	☐ weet niet	☐ n.v.t.
10.	Wat vindt u van de **prijs** van deze rit?	zeer duur	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	zeer goedkoop	☐ weet niet	☐ n.v.t.
11.	Wat vond u van de **stiptheid** (op tijd rijden) van het voertuig bij de vertrekhalte?	zeer slecht	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	zeer goed	☐ weet niet	☐ n.v.t.
12.	Wat vindt u van de **reissnelheid** van deze rit? (omrijden, directheid)	zeer slecht	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	zeer goed	☐ weet niet	☐ n.v.t.
13.	Wat vindt u van het aantal **vertrekmogelijkheden** vanaf uw instaphalte?	zeer slecht	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	zeer goed	☐ weet niet	☐ n.v.t.
14.	Wat is uw oordeel over de **totale rit** in het algemeen?	zeer slecht	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	zeer goed	☐ weet niet	☐ n.v.t.
			1	2	3	4	5	6	7	8	9	10			

15. Welk aspect van deze rit dient volgens u als **eerste verbeterd** te worden? *(slechts één antwoord mogelijk)*

☐ kans op een zitplaats	☐ rijstijl van de bestuurder	☐ informatie bij de instaphalte	☐ stiptheid van deze rit	☐ niets
☐ netheid van het voertuig	☐ geluid van dit voertuig	☐ informatie bij vertragingen	☐ reissnelheid van deze rit	☐ anders, namelijk:
☐ klantvriendelijkheid personeel	☐ gemak van in- en uitstappen	☐ gemak kopen vervoerbewijs	☐ aantal vertrekmogelijkheden	

16.	Hoe vaak heeft u de **afgelopen 7 dagen** deze rit gemaakt?	☐ 0 of 1 keer	☐ 2 keer	☐ 3 keer	☐ 4 keer	☐ 5 keer	☐ meer dan 5 keer	
17.	Had u vandaag deze **hele reis met de auto** kunnen maken?	☐ ja, eigen auto	☐ ja, meerijden	☐ ja, anders	☐ nee, geen auto	☐ nee, geen rijbewijs		☐ nee, anders
18.	U reist nu ergens naartoe, wat gaat u daar doen? *(slechts één antwoord mogelijk)*	☐ wonen	☐ werken	☐ onderwijs	☐ winkelen	☐ sporten	☐ bezoek	☐ anders

Ga verder op de achterzijde >>>

19. Met welk **vervoerbewijs** reist u?

OV-chipkaart, u moet in- en uitchecken

- ☐ op saldo met OV-chipkaart
- ☐ regio-(ster)abonnement op OV-chipkaart
- ☐ OV-chipkaart gekocht in voertuig of op halte/station (bijv. wegwerp OV-chipkaart)
- ☐ studenten OV-kaart
- ☐ zakelijke kaart (NS Business Card, Mobility Mixx, Xximo etc.)
- ☐ OV-chipkaart anders, incl. acties dagkaarten winkelketens

Andere kaartsoort, zonder in- en uitchecken

- ☐ papieren kaartje gekocht in voertuig of op halte/station
- ☐ regio-(ster)abonnement zonder chip
- ☐ anders (niet OV-chipkaart), zoals gratis probeerkaartje

		1	2	3	4	5	6	7	8	9	10			
20. Als u met een **OV-chipkaart** reist, hoe beoordeelt u het **gebruiksgemak** van de OV-chipkaart?	zeer moeilijk	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	zeer makkelijk	☐ weet niet	☐ n.v.t.
21. Als u straks overstapt of bent overgestapt op ander openbaar vervoer, hoe beoordeelt u de **overstaptijd**?	zeer onprettig	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	zeer prettig	☐ weet niet	☐ n.v.t.
22. Hoe **veilig** voelt u zich in het openbaar vervoer?	zeer onveilig	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	zeer veilig	☐ weet niet	☐ n.v.t.
23. Hoe **veilig** voelt u zich (meestal) tijdens deze **rit**?	zeer onveilig	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	zeer veilig	☐ weet niet	☐ n.v.t.
24. Hoe **veilig** voelt u zich (meestal) op de **halte** waar u bent ingestapt?	zeer onveilig	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	zeer veilig	☐ weet niet	☐ n.v.t.

25. Bent u in de afgelopen 12 maanden **zelf slachtoffer** geweest van één of meer van de volgende incidenten in het openbaar vervoer of op de halte?

	niet	1 keer	2 keer	3 keer	4 keer of vaker
a. lastig gevallen	☐	☐	☐	☐	☐
b. bedreigd	☐	☐	☐	☐	☐
c. bestolen	☐	☐	☐	☐	☐
d. mishandeld	☐	☐	☐	☐	☐
e. anders	☐	☐	☐	☐	☐

26. Bent u in de afgelopen 12 maanden **ooggetuige** geweest van één of meer incidenten in het openbaar vervoer of op de halte? ☐ ja ☐ nee

27. Kunt u aangeven hoe u **deze rit beleeft**?

	1	2	3	4	5	6	7	8	9	10			
a. ongezellig	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	gezellig	☐ weet niet	☐ n.v.t.
b. kil	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	sfeervol	☐ weet niet	☐ n.v.t.
c. druk	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	rustig	☐ weet niet	☐ n.v.t.
d. inefficiënt	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	efficiënt	☐ weet niet	☐ n.v.t.
e. oncomfortabel	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	comfortabel	☐ weet niet	☐ n.v.t.

28. Wat vindt u van het comfort in dit voertuig?	zeer slecht	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	zeer goed	☐ weet niet	☐ n.v.t.
29. Wat vindt u van de netheid van de instaphalte?	zeer vies	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	zeer schoon	☐ weet niet	☐ n.v.t.

30. Als u bij vraag 23 een cijfer van 6 of lager heeft gegeven, kunt u dan aangeven waarom? (meerdere antwoorden mogelijk)

☐ te druk ☐ agressie ☐ slechte verlichting ☐ vandalisme
☐ te rustig ☐ rondhangende personen ☐ gebrek aan toezicht ☐ anders

31. Wanneer zou u vaker met bus of tram van HTM reizen? (meerdere antwoorden mogelijk)

☐ als u 's avonds voor een vast, goedkoop tarief kunt reizen
☐ als u in het weekend voor een vast, goedkoop tarief kunt reizen
☐ als u in het voertuig met uw pinpas een kaartje kan kopen
☐ als u in het voertuig uw OV-chipkaart kan opladen

32. Hoe bent u bij de vertrekhalte gekomen? ☐ te voet ☐ per fiets ☐ per scooter ☐ per auto ☐ anders, namelijk

33. Wat is uw **geboortejaar**? ☐☐☐☐

34. Wat is uw **geslacht**? ☐ man ☐ vrouw

Hartelijk dank voor uw medewerking. U kunt de enquête weer inleveren bij de enquêteur.

Hier niets invullen! ritcode (uniek) 6

0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9

- **HTM Klantenpanel [NB: dit is een pdf-import in word, ga deze op een later moment nog mooi maken]**

HTM-

Ritbeoordeling 2018

Instructie:

Op dit formulier kun je je oordeel geven over de gemaakte rit in de vorm van rapportcijfers van 1 t/m 10. Als een bepaald aspect niet van toepassing is voor de rit die je beoordeelt, kun je 'n.v.t.' aanvinken.

N.B. Je beoordeling dient één enkele rit met bus, tram of RandstadRail van HTM te betreffen. Als je bent overgestapt, kies dan een deeltraject. Je kunt deze vragenlijst het beste invullen na afloop van je rit.

Lees zo nodig de instructie nog eens door.

[VRAGENLIJST MET 1 VRAAG PER PAGINA]

1.1. Welke rit ga je beoordelen?
Vul hieronder de gegevens van die rit in.
Als je niet precies meer weet wat je exacte instaptijdstip was, geef dan een zo goed mogelijke benadering.

datum: (dd/mm/yyyy)

instaptijd (uren / minuten):

lijn (dropdown): ○overzicht lijnen met richting ○ tram ...
○RandstadRail ...
○bus ...
○[99] andere lijn [als lijn=99] namelijk lijn: [als lijn=99] richting:

instaphalte: ... [als lijn=99] naam of locatie instaphalte:

uitstaphalte: ... [als lijn=99] naam of locatie uitstaphalte:

wagennummer [mouse-over plaatje Wagennummers]:

1.2. [ZELFDE PAGINA][V] Was er tijdens de rit sprake van een vertraging of andere problemen?

○ ja, er was vertraging

○ ja, er was een ander probleem. Namelijk:

○ ja, er was vertraging en er was een ander probleem. Namelijk:

○ nee
○ dat weet ik niet meer

1.3. [ZELFDE PAGINA] Met welk doel maakte je deze rit?
Als je een rit naar huis beoordeelt, kies hier dan het doel van je rit op de heenweg.
○ van/naar mijn werk
○ van/naar school, studie, opleiding, stage ○ zaken-/dienstreis, bezoek congres e.d.
○ medisch: bezoek aan huisarts, tandarts, specialist, ziekenhuis e.d.
○ boodschappen doen ○ winkelen ○ bezoek aan familie, vrienden, kennissen ○ kinderen halen/brengen of andere zorgtaken ○ uitstapje naar het strand, museum, theater, bioscoop, e.d. ○ horecabezoek (restaurant, kroeg, cafÃ©)
○ naar sportclub, koor, vereniging of andere vrijetijdsbesteding ○ anders, namelijk:

1.4. Kon je moeilijk of makkelijk een zitplaats vinden toen je instapte?

zeer moeilijk/niet 1	2	3	4	5	6	7	8	9	zeer makkelijk 10	weet niet	n.v.t.

1.5. Wat vond je van het comfort van het voertuig tijdens je rit?

ze er sle cht 1	2	3	4	5	6	7	8	9	ze er goed 10	w:et ni et	n.v.t.

1.6. Wat vond je bij deze rit van:

	zeer vies 1	2	3	4	5	6	7	8	9	zeer schoon 10	weet niet	n.v.t.
de netheid van het voertuig	○	○	○	○	○	○	○	○	○	○	○	○
de netheid van de instaphalte	○	○	○	○	○	○	○	○	○	○	○	○

1.7. Wat vond je bij deze rit van:

	zeer slecht 1	2	3	4	5	6	7	8	9	zeer goed 10	weet niet	n.v.t.
de informatie op de instaphalte (vertrektijden, route, etc.)	○	○	○	○	○	○	○	○	○	○	○	○
De reisinformatie in het voertuig	○	○	○	○	○	○	○	○	○	○	○	○

1.8. Wat vond je van de klantvriendelijkheid van het personeel?

zeer klant-onvriendelijk 1	2	3	4	5	6	7	8	9	zeer klant-vriendelijk 10	weet niet	n.v.t.

1.9. Vond je het moeilijk of makkelijk om voor je rit een vervoerbewijs te kopen / reissaldo te laden

Zeer moeilijk 1	2	3	4	5	6	7	8	9	zeer makkelijk 10	weet niet	n.v.t.

1.10. Wat zijn bij deze rit je rapportcijfers voor:

	zeer slecht 1	2	3	4	5	6	7	8	9	zeer goed 10	weet niet	n.v.t.
de stiptheid van het voertuig (op tijd rijden)	○	○	○	○	○	○	○	○	○	○	○	○

bij je instaphalte												
de frequentie (aantal ritten per uur)	○	○	○	○	○	○	○	○	○	○	○	○

1.11. Wat is je totaaloordeel over deze rit?

zeer slecht 1	2	3	4	5	6	7	8	9	zeer goed 10	weet niet	n.v.t.

1.12. .

	zeer onveilig 1	2	3	4	5	6	7	8	9	zeer veilig 10	weet niet	n.v.t.
Hoe veilig voel je je over het algemeen in het openbaar vervoer?	○	○	○	○	○	○	○	○	○	○	○	○
Hoe veilig voelde je je tijdens deze rit?	○	○	○	○	○	○	○	○	○	○	○	○

1.13. Welke van de zojuist beoordeelde aspecten wil je nog toelichten?

Vink alle aspecten aan waarover je een toelichting wilt geven. Tussen haakjes zie je het door jou gegeven rapportcijfer staan. [als geen antwoord, weet niet of n.v.t. dan geen vermelding] □ beschikbaarheid zitplaats toen je instapte (rapportcijfer) □ comfort in het voertuig (rapportcijfer) □ netheid van het voertuig (rapportcijfer) □ de netheid van de instaphalte □ informatie op de instaphalte (rapportcijfer) □ reisinformatie in het voertuig (rapportcijfer)

□ informatie bij vertragingen of andere problemen (rapportcijfer) □ klantvriendelijkheid personeel (rapportcijfer) □ rijstijl bestuurder (rapportcijfer)

□ gemak kopen vervoerbewijs / laden reissaldo (rapportcijfer) □ de stiptheid (op tijd rijden) van het voertuig (rapportcijfer) □ frequentie, aantal ritten per uur (rapportcijfer) □ oordeel totale rit in het algemeen (rapportcijfer) □ veiligheid in het OV (rapportcijfer) □ veiligheid tijdens de rit (rapportcijfer) □ ik heb (ook nog) andere opmerkingen

1.14. Je toelichtingen: [tekstvakken voor toelichtingen bij de vragen die bij 1.13 aangevinkt zijn]

1.15. [alleen stellen als 1.2 = ja] [1.11a] Wat vond je bij deze rit van de informatie bij vertragingen of andere problemen?

zeer slecht 1	2	3	4	5	6	7	8	9	zeer goed 10	weet niet	n.v.t.

1.16. [alleen als 1.11a = <6] [1.11a.1] Wat had beter gekund aan de informatie van HTM bij vertraging of andere problemen bij deze rit?

1.17. [q.1.11b] Wat vond je van de drukte in het voertuig?

zeer druk 1	2	3	4	5	6	7	8	9	zeer rustig 10	weet niet	n.v.t.

1.18. [1.8a] Wat vond je van de rijstijl van de bestuurder? (optrekken, remmen, etc.)

zeer onprettig 1	2	3	4	5	6	7	8	9	zeer prettig 10	weet niet	n.v.t.

1.19. [1.14a] Mag HTM naar aanleiding van je antwoorden in deze vragenlijst contact met je opnemen?

Je contactgegevens die je hieronder invult worden uitsluitend gebruikt om naar aanleiding van deze ritbeoordeling eventueel contact met je op te nemen.

○ ja, bij voorkeur telefonisch, op telefoonnummer:

○ ja, bij voorkeur via e-mail, op e-mailadres:

○ nee

1.20. [q0; vanaf 25 mei; alleen panelleden die nog geen toestemming hebben gegeven]

Met ingang van 25 mei is de privacywetgeving gewijzigd. Daarom leggen wij graag uit met welk doel wij je gegevens verzamelen en verwerken en vragen wij je nogmaals toestemming om dat te doen.

Het doel van het verzamelen en verwerken van je gegevens is het peilen van de mening van reizigers over de dienstverlening van HTM. De resultaten van de onderzoeken worden uitsluitend gedeeld met HTM. HTM krijgt geen beschikking over gegevens die zijn te herleiden naar individuele personen. Lees hier meer over ons privacybeleid.

Geef je toestemming voor het verzamelen en verwerken van je gegevens?

○ ja, ik geef Citisens Marktonderzoek hiervoor toestemming en blijf lid van het HTM-panel

○ nee, ik geef geen toestemming en meld me hierbij af van het HTM-panel [--> AFMELDEN]

[ALS NOG NIET EERDER INGEVULD --> Variabel deel]

Nu volgt een ander onderwerp, namelijk: ....

Als je het volgende deel van de vragenlijst ook helemaal invult, hoef je dat bij de volgende Ritbeoordeling niet nog een keer te doen. De volgende ritbeoordeling die je dit kwartaal invult, bestaat dan alleen uit het eerste deel van de vragenlijst.

2.1. ...

Hartelijk dank voor je deelname

Je antwoorden zijn ontvangen.

Wil je nog een rit beoordelen? Bijvoorbeeld een ander deel van je reis als je overgestapt bent of de terugreis of een andere recente rit?

- ja, nu direct: klik dan door naar een nieuwe vragenlijst.
- ja, maar op een ander moment: gebruik dan de link in de uitnodiging opnieuw.
- nee, ik ben klaar met invullen: sluit dit onderzoek af.

Je vragen of klachten kun je kwijt aan HTM via het Reactieformulier.

Hartelijk dank voor je deelname

Deze reeks beoordelingen is afgerond. Je kunt binnenkort weer meedoen met een nieuwe reeks.

Appendix B: Search terms literature Review.

The literature review presented in chapter 2 was written based on an extensive search. This appendix provides a short overview of what search terms were used.

Literature research was done on the topics described in sections 2.1 and 2.2: customer satisfaction in Public Transport and Crowding in Public Transport. Terms used to find papers in these topics included:

Regarding Customer Satisfaction

- Comfort
- Service Quality
- Customer Satisfaction
- Passenger Experience

Regarding Crowding

- Smart Card
- Crowding
- Sitting/Standing
- Full Public Transport
- Crowding multiplier
- Crowding costs

If necessary the term 'Public Transport' was often added.

Papers advised by experts were also considered. Lastly, indirect searching was done:

- while reading literature which had been found, if an interesting paper was found in references
- By clicking on the suggestions provided by the database used, if the title seemed interesting

Naturally, the reference list shown does not include all papers found and read – some papers were found to be of little added value for this thesis.

Appendix C: Traveller characteristics and composition of HTM Klantenpanel

This appendix provides some information on the composition of HTMs Klantenpanel. The main conclusions have also been presented in chapter 3.

As subscription for and participation in the Klantenpanel Is fully voluntary, the background of the panel does not necessarily reflect the background of both PT travellers in Den Haag and the population of the city as a whole. It is however important to note which differences exist, as these affect the possibilities of generalizing results of the analyses made in this thesis. This appendix presents descriptive statistics on HTMs Klantenpanel and tries to make conclusions how representative its respondents are.

In order to be able to draw conclusions on representativeness, first it has to be determined what exactly is representative. For this, it is assumed that the responses as filled in in the OV Klantbarometer form an accurate representation of which people travel with HTM for what purposes. Due to the random nature of selection and size of the sample of this survey it is plausible to make this conclusion. Moreover, no more accurate data exists. As a basis, the HTM panel data on 2018 is compared with the OV Klantbarometer data of 2017.

To start with, one can have a look on age and gender. Table C.1Table 3 provides an overview of age of respondents. It can be seen that the HTM panel the elderly are overrepresented, ages 65 and over being responsible for 41% of all Klantenpanel responses while only accounting for 8% of all travellers. As a result, younger groups are underrepresented.

Table 3C.1: Age distribution of survey responses

Age group	HTM Klantenpanel	OV Klantbarometer Tram	OV Klantbarometer RR
< 18	0	9	10
18-27	6	36	36
28-40	10	18	19
41-64	44	28	27
65+	41	8	8

Next, Table C.2 shows how responses are distributed over gender. From this, it can be concluded that the distribution in the HTM Klantenpanel is mirrored to the one in the OV Klantbarometer. While gender in the population is obviously distributed circa 50/50, it is known that women tend to use PT more often, making the OV Klantbarometer distribution seem plausible. To conclude, the HTM Klantenpanel men are overrepresented.

Table C.2: Gender distribution of survey responses

Gender	HTM Klantenpanel	OV Klantbarometer Tram	OV Klantbarometer RR
Male	59	39	40
Female	40	61	60

Next up the travel motive can be compared. Tables C.3 and C.4 show this comparison. As both surveys ask categorise the answers in a slightly different way two tables are used to present responses. Both surveys ask the purpose the current trip was made for. The HTM Klantenpanel survey asks respondents to mark the motive of their outbound journey if they are travelling home, the OV-Klantbarometer does not. For comparison, an extra column is added at the OV Klantbarometer in which trips going home are distributed over the other categories – assuming that the motif of the return leg of a homebound trip follows the same distribution as observed. Nevertheless, comparison remains a bit difficult of course, mostly due to the difficulties in interpreting the 'varying' answer option in the Klantenpanel survey.

Table C.3: Travel Motif distribution Klantenpanel

Motive	HTM Klantenpanel
Work	29.4
Education	3.8
Business	3.2
Medical	6,4
Groceries	7,4
Shopping	8,7
Visiting friends/family	13.4
Informal care	1.2
Recreation	9.9
Eating&Drinking	2.8
Sport/Leisure	5.9
Different	8.0

Table C.4: Destination Distribution OV Klantbarometer

Destination	OV Klantbarometer Tram		OV Klantbarometer RR	
Living	22	-	30	-
Working	25	32	28	41
Education	13	16	13	19
Shopping	12	15	6	9
Sport	3	4	2	3
Visitation	11	14	9	13
Different	15	19	10	15

Next, travel frequency can be compared. Again, the HTM Panel and OV Klantenpanel ask these questions differently, so two tables are necessary to present the results. These can be found in tables C.5 and C.6. The differences in the way questions are asked (mainly usage being counted in days per week or times per week) makes comparison, again, difficult. It can be concluded, however, that HTMs Klantenpanel has very little responses from incidental PT users. Given the nature of the panel, however, this is logical.

Table 4C.5: Frequency of travel HTM Klantenpanel

Frequency	HTM Klantenpanel
4+ days per week	37
1-3 days per week	44

1-3 days per month	16
6-11 days per year	2
5- days per year	1

Table C.6: Frequency of Travel OV Klantbarometer

Frequency	**OV Klantbarometer Tram**	**OV Klantbarometer RR**
6+ times per week	25	27
5 times per week	13	17
4 times per week	14	14
3 times per week	11	11
2 times per week	13	13
<2 times per week	24	18

HTM also asks its panel respondents to provide their education level. This level of education can be compared to general education levels in the Netherlands and Den Haag. Table C.7 shows these numbers. It can be seen that the Klantenpanel resembles the total population of Den Haag quite okay, with lower education being a tad underrepresented. The population of Den Haag and the panel is on average higher educated than the Dutch average.

Table C.7: Level of education HTM Klantenpanel

Education Level	**HTM Klantenpanel**	**Den Haag Population (**Den Haag, 2018**)**	**Dutch Population (**CBS, 2018**)**
Elementary Education	1	7	9
Low (VMBO/MBO-1)	10	15	20
Middle (HAVO/VWO/MBO 2-4)	40	33	40
High (HBO/WO)	45	45	30
Different	4	-	1

Lastly, tables C.8 and C.9 show how respondents were distributed over HTMs network.

Table C.8: Number of responses per line

Line number	Frequency	Percent
1	317	11.1
2	346	12.1
3	405	14.2
4	276	9.7
6	143	5.0
9	235	8.2
11	62	2.2

12	67	2.3
15	118	4.1
16	178	6.2
17	120	4.2
18	76	2.7
19	61	2.1
20	2	0.1
21	53	1.9
22	19	0.7
23	194	6.8
24	83	2.9
25	50	1.7
26	49	1.7
27	1	0.0
28	3	0.1
Total	2858	100

Table C.9: Number of responses per vehicle type

Vehicle type	Frequency	Percent (%)
GTL	705	24.7
Avenio	881	30.8
Citadis	742	26.0
Bus	530	18.5
Total	2858	100

Appendix D: Data preparation

In building the Structural Equation Model, a lot of data is used. The data sources used need to be adapted slightly in order to be used correctly. This appendix explains how the data is prepared for analysis and what assumptions are made in the process. It is divided into three parts: first the preparation of the customer satisfaction data is done, then how crowding data is retrieved as accurately as possible from the survey data and last how crowding data is used.

To start with, customer satisfaction data has to be prepared. The basic data file used contains all filled-in surveys of the HTM Klantenpanel. The nature of the questions asked in the survey, however, makes that some changes are necessary before they can be used for proper analysis. The following list provides an overview of the changes and assumptions made:

- Respondents fill in their year of birth. Their age as of January 1st, 2019 can be calculated based on this. Age was used in the model; it was treated as a ratio variable (i.e. no categorisation or such was done).
- The moment of evaluation is important to note, as this is not necessarily during or immediately after the trip – a trip can also be evaluated several days later. The analysis assumes that this does not affect the scores which are given.
- Respondents are asked to evaluate the last trip they made with HTM. However, this is not always done as intended:
    - Some respondents evaluated services which were not operated by HTM but by Connexxion or Arriva, the regional bus operators. As these services fall outside the scope of this research, these responses were removed.
    - Some respondents made a trip which involved travelling on multiple HTM lines and services. Often, this results in responses in which only one of the lines is filled in and destination does not match with that specific line. All these responses were removed from the data set, as the stop at which was changed and the follow-up services are unclear.
- In general, all responses in which the values filled in can give any doubt on which exact line used between what stops and at what time were removed. Besides reasons mentioned above this is mostly due to missing values, filling in an incorrect line number with origin and destination allowing multiple lines (e.g. Station Hollands Spoor <-> Centraal Station) or filling in 'temporary stops' as origin or destination.
- Buses replacing trams due to construction works were not considered in any analysis.
- Lastly, the survey allows people to score each aspect of their trip on a scale of 1 to 10. Next to these, values 11 (no opinion) and 12 (inapplicable for this trip). For most questions, these answers are given less than 1% of the time *after* removing all previous cases. During the exploratory data analysis, these values are presented as filled in. In estimating statistical models, these answers are treated as missing values, as they pose no statistical value.

This left 3738 responses which were considered in the next step. In order to be able to thoroughly investigate the relationship between customer satisfaction and crowding, each of these responses has to be coupled to its respective crowding levels. When a traveller who is a member of the HTM Klantenpanel evaluates his or her trip, he or she provides information on the trip: date and time of boarding, as well as at which stop he or she boarded and alighted the vehicle. The analysis couples this data to occupancy rates based on smart card data. This is done by noting the boarding date, time, line, stop and direction and searching for the corresponding service in the realised exploitation for 2018. It is assumed the respondent does not make any error in filling in het boarding moment. If the time of boarding reported does not match exactly with any service, the algorithm automatically rounds the time to the nearest trip. This rounding down is due to the way the algorithm works and might not be the most accurate representation, which should be noted. After this coupling, data on this exact service can be searched for to get insight in how crowded that specific service was. If no service can be found (description of the algorithm can be found in section 3.3) the response was discarded. Moreover, if either the **general score** of the trip (the outcome of the framework) or the **probability of finding a seat** (which is used as a proxy of perceived occupancy) was scored with 11 or 12, the response was deleted as well. The reason for this is that the main goal of the framework and this thesis is to investigate the relationship between actual occupancy numbers, perceived occupancy and customer satisfaction. The availability of at least these variables is necessary for proper analysis. Moreover, it is impossible to evaluate customer satisfaction with 'does not apply' or 'does not know' for a trip.

After the correct service corresponding to a response has been found, data can be retrieved on the occupancy of this service. These occupancy numbers firstly and generally correspond to the number of passengers in the vehicle who have checked in after departure at a given stop. The values used for this research are not corrected afterwards. If a service was found to have a maximal occupancy of zero, the response was discarded. This left a set of 2858 responses, more than sufficient for model building.

The next question asked is what the 'correct' value of crowding is, from the perception of the respondent, i.e. which indicator of crowding best fits the perception of crowing by the respondent. This question has not yet been fully answered by science and consequently a variety of indicators will be used in the analysis to see which one fits best. Indicators which come to mind include:

- The occupancy at the moment of boarding the vehicle.
- The occupancy at the moment of leaving the vehicle.
- The maximum occupancy during the ride

It should be noted, however, that these occupancy numbers are just numbers and need to be converted for proper usage. An in-vehicle occupancy of 100 passengers is quite low for long trains but quite high for a passenger car, to use some extreme examples. Therefore these occupancy numbers will be converted to some variables which tell more about actual in-vehicle crowding. Extra information necessary to make this conversion is twofold: firstly, it has to be known what type vehicle provided the service and secondly the characteristics of this vehicle have to be known. Afterwards, occupancy numbers can be converted to crowding variables. As has been explained in chapter 2, several ways of measuring in-vehicle crowding exist. The following will be done for this analysis:

- Firstly, the occupancy numbers will be compared with the seat capacity of the vehicle. From this comparison a dummy variable will be derived, which states whether the respondent was able to sit or not based on this occupancy rate. It has the form:

  $$D_s = \begin{cases} 1 & if\ n_{pax} > n_{seats} \\ 0 & otherwise \end{cases}$$

  An important assumption which is made in this analysis is that each traveller only occupies one seat and that if seats are available a passenger always chooses to sit, even if this requires walking through the whole vehicle or travelling backwards. It has been empirically proven that this in reality is not the case.
- Secondly, the Load Factor (LF) will be calculated. The load factor compares the number of passengers to the number of seats and is defined as:

  $$Load\ Factor = \frac{number\ of\ passengers}{number\ of\ seats}$$

  The load factor is expressed as a percentage. The definition means that a Load Factor of more than 100% is perfectly possible – this means standing passengers are present in the vehicle.

- As research has shown, in crowded conditions (LF > 100%) the Load Factor often does not adequately represent crowding conditions (e.g. (Yap, Cats, & van Arem, 2018)), partly due to differences between vehicles – one vehicle is better adapted for allowing standing travellers comfortably than another. The composition of HTMs fleet, however, requires different vehicle types to be compared. A value which accounts better for this difference is (standing) passenger density. The passenger density is defined as:

$$Passenger\ Density = \frac{number\ of\ passengers}{space\ available\ for\ passengers}$$

  The passenger density has passengers per square meter (pax/m^2/veh) as a unit. Two measures of passenger density exist. Either the total passenger density can be used, or seated passengers (as well as the number of seats and the space these occupy) can be neglected, taking only into account standing passengers. The latter has the advantage of more adequately representing how much space standing passengers have at the cost of not representing situations without standing passengers at all.

To conclude, Load Factor is an optimal variable for measuring crowding in situations in which the number of passengers is lower than the number of seats and passenger density is better suited for situations in which passengers have to stand. As the calculation of both variables follows directly and easily from vehicle characteristics and occupancy numbers, both values will be calculated for the analysis. Furthermore the inclusion of a dummy variable which tells if the respondent was able to sit can help on choosing one of these measurements. Analyses will be carried out using all these variables to see which offers the most accurate representation: the results can be seen in section 4.3.

When retrieving the occupancy data from HTMs database, data on punctuality is retrieved as well. This was done for the stop at which the customer boarded the vehicle and the stop at which the customer alighted from the vehicle. Only the punctuality of the service boarded was considered: data restrictions make that it is impossible to put this punctuality in the broader situation on that specific moment. This is done by retrieving two variables:

- The nominal departure time at that specific stop, according to the timetable (NomDep)
- The actual departure time at that specific stop on that specific day (ActDep)

Both variables are measured to the second. The punctuality for the service is retrieved by

$$Delay = ActDep - NomDep$$

Delays are thus measured in seconds.

Having performed all these steps we now have all data necessary to start analysing. However, data preparation is not fully done yet at that moment: some last changes have to be made. This mainly has to do with the different types of variables that exist. For the context of Structural Equation Modelling, three different types of variables exist:

- Ratio or interval variables, which are continuous variables in which a hierarchy between coded values is present and it can be quantified with certainty how large differences between different values are (e.g. the difference in temperature between 20 and 30 degrees Celsius is as large as the difference between 60 and 70 degrees Celsius).
- Ordinal variables, in which a hierarchy is present but it is impossible to quantify how large differences between values actually are (e.g. education level: a college degree is certainly higher than elementary education, but how much higher?)

- Nominal or categorical variables, in which differences between values are clear but even no hierarchy can be made (e.g. gender: males are certainly different from females but none of the two is any better or worse than the other).

SEM uses linear regression to explain relations between variables, which is the technique made for analysing interval/ratio variables. For correctly including nominal and ordinal variables, the following needs to be done:

- According to Byrne et al (2013) ordinal variables in SEM may be treated as an interval/ratio variable provided there are at least four different possible values for the variable and these are coded in the correct way (from low to high)
- Each possible category of a nominal variable need to be recoded into dummy variables (0/1 variables). For all but one categories a dummy needs to be made which becomes one if a response belongs to that category. If all dummies are 0 then the response belongs to the last category.

Luckily, most ordinal variables have at least four possible values and thus can be treated as an interval variable. However, several variables to be recoded for a proper estimation of the Structural Equation Model. Table D.1 below provides an overview of which variables were recoded in what way.

Variable	Reason for recoding	Resulting variables	Coding resulting variables
Vehicle type	Nominal variable	Is_GTL, Is_Citadis, Is_Avenio	1 if the trip is made in that vehicle. If all are 0 then the vehicle is a bus
Gender	Nominal variable, useless answer options ('does not want to state gender')	Gender_Binary	0 for males, 1 for females, all other values replaced by missing values
Travel purpose	Nominal variable, many possible options for answer	goal_commute, goal_business, goal_leisure, goal_education	1 if the goal is the goal mentioned in the name. If all are 0 than purpose is 'different'
Education level	Ordinal variable, useless answer options ('other')	Education_ratio	Same as first, but with 'other' replaced by missing values
Travel time	Values coded (boarding moment) is no interval variable	Rush Hour?	1 if trip is made during rush hour (07-09 and 16-19 on weekdays), 0 else
Delay	Nominal variable	Delay_YN	1 if some delay or disruption was experienced, 0 else

Appendix E: Structural Equation Model

This appendix explains how the dataset considered for analysis was converted into a structural equation model (SEM). Section E.1 shows the result of the Confirmatory Factor Analysis (CFA), after which section E.2 shows the full estimated structural Equation Model itself.

E.1 Confirmatory Factor Analysis

To check whether the proposed model structure is correct, a confirmatory factor analysis was carried out. The outcome of the CFA is discussed in section 5.1. Figure E.1 shows the AMOS output of the CFA showing standardized values.

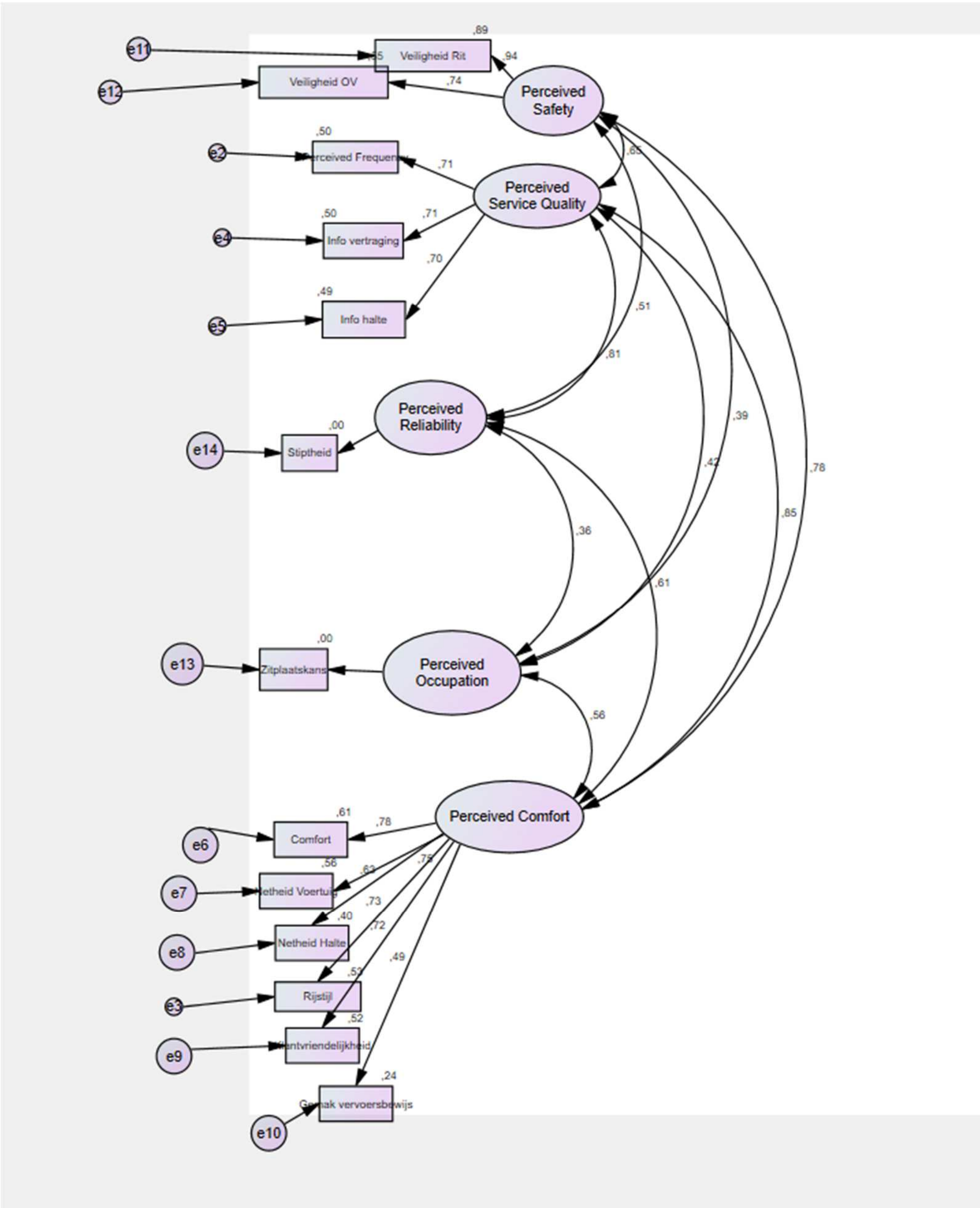


Figure E.1: CFA Model

E.2. Structural Equation Model

Chapter five has presented the main results of the estimated Structural Equation Model. This appendix will deal with those model results that were not included in chapter five. Table E.2 provides an overview of all unstandardized total effects between variables, table E.3 provides an overview of all standardized total effects between variables and figure E.4 shows the full estimated structural model.

Table E.2: unstandardized total effects between all variables

	Delay Experienced Y/N	PT Travel Frequency	Rush?	Purpose _work	Purpose _business	Purpose _school	Purpose _visit	Avenio?	Citadis?	GTL?	Education	Age	Gender	Delay	Load Factor	Frequency	Perc reliability	Perc occupancy	Perc comfort	Perc servqual	Perc safety
Perc reliability	-2,497	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	-,002	,000	,000	,000	,000	,000	,000	,000
Perc occupancy	,000	-,001	-,166	-,011	,154	-,352	,052	-,202	-,436	,225	-,229	,012	,089	,000	-3,188	,000	,000	,000	,000	,000	,000
Perc comfort	-,656	-,001	-,080	-,069	,028	-,060	,057	,138	-,039	,104	-,214	,008	,112	,000	-,461	,000	,263	,145	,000	,000	,000
Perc servqual	-1,361	-,001	-,243	-,153	,057	-,084	,034	,290	,190	,186	-,317	,004	,050	-,001	-,128	,048	,545	,000	,000	,000	,000
Perc safety	-,804	,000	-,020	-,001	,018	-,042	,006	-,024	-,052	,027	-,027	,001	,343	,000	-,381	,000	,322	,119	,000	,000	,000
Punctuality	-2,497	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	-,002	,000	,000	1,000	,000	,000	,000	,000
Zitplaatskans	,000	-,001	-,166	-,011	,154	-,352	,052	-,202	-,436	,225	-,229	,012	,089	,000	-3,188	,000	,000	1,000	,000	,000	,000
Vervoersbewijs	-,545	-,001	-,066	-,058	,024	-,050	,048	,115	-,033	,087	-,178	,007	,093	,000	-,383	,000	,218	,120	,832	,000	,000
Comfort	-,908	-,001	-,110	-,096	,039	-,083	,080	,191	-,055	,144	-,297	,011	,155	-,001	-,638	,000	,364	,200	1,385	,000	,000
Klantvriendelijkheid	-,936	-,001	-,114	-,099	,041	-,086	,082	,197	-,056	,149	-,306	,011	,159	-,001	-,658	,000	,375	,206	1,428	,000	,000
Netheid_halte	-,559	-,001	-,068	-,059	,024	-,051	,049	,117	-,034	,089	-,183	,007	,095	,000	-,393	,000	,224	,123	,853	,000	,000
Netheid_voertuig	-,656	-,001	-,080	-,069	,028	-,060	,057	,138	-,039	,104	-,214	,008	,112	,000	-,461	,000	,263	,145	1,000	,000	,000
Info_halte	-1,315	-,001	-,234	-,148	,055	-,081	,033	,280	,184	,180	-,307	,004	,049	-,001	-,124	,047	,527	,000	,000	,966	,000
Info_voertuig	-1,814	-,001	-,323	-,204	,075	-,112	,046	,387	,253	,248	-,423	,005	,067	-,001	-,171	,064	,726	,000	,000	1,333	,000
Rijstijl	-,786	-,001	-,096	-,083	,034	-,072	,069	,165	-,047	,125	-,257	,010	,134	,000	-,552	,000	,315	,173	1,198	,000	,000
Frequentie	-1,361	-,001	-,243	-,153	,057	-,084	,034	,290	,190	,186	-,317	,004	,050	-,001	-,128	,048	,545	,000	,000	1,000	,000
Totaaloordeel	-1,203	-,001	-,176	-,131	,051	-,096	,074	,256	,029	,181	-,348	,010	,136	-,001	-,539	,018	,482	,154	1,074	,372	-,008
Veiligheid_rit	-,804	,000	-,020	-,001	,018	-,042	,006	-,024	-,052	,027	-,027	,001	,343	,000	-,381	,000	,322	,119	,000	,000	1,000
Veiligheid_OV	-,557	,000	-,014	-,001	,013	-,029	,004	-,017	-,036	,019	-,019	,001	,238	,000	-,264	,000	,223	,083	,000	,000	,693

Table E.3: Standardized total effects between all variables

	Delay Experienced Y/N	PT Travel Frequency	Rush Hour?	Purpose _work	Purpose _business	Purpose _school	Purpose _visit	Avenio?	Citadis?	GTL?	Education	Age	Gender	Delay	Load Factor	Frequency	Perc reliability	Perc crowding	Perc comfort	Perc servqual	Perc safety
Perc reliability	-,377	,000	,000	,000	,000	,000	,000	,000	,000	,0000	,000	,000	,000	-,091	,000	,000	,000	,000	,000	,000	,000
Perc crowding	,000	-,018	-,032	-,002	,012	-,030	,008	-,040	-,082	,042	-,069	,079	,019	,000	-,469	,000	,000	,000	,000	,000	,000
Perc comfort	-,180	-,048	-,036	-,032	,005	-,012	,020	,063	-,017	,045	-,151	,126	,054	-,043	-,157	,000	,477	,335	,000	,000	,000
Perc servqual	-,299	-,045	-,088	-,056	,008	-,013	,009	,107	,067	,064	-,179	,047	,020	-,072	-,035	,087	,795	,000	,000	,000	,000
Perc safety	-,161	-,004	-,007	,000	,002	-,006	,002	-,008	-,017	,008	-,014	,016	,122	-,039	-,095	,000	,428	,202	,000	,000	,000
Punctuality	-,377	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	-,091	,000	,000	1,000	,000	,000	,000	,000
Zitplaatskans	,000	-,018	-,032	-,002	,012	-,030	,008	-,040	-,082	,042	-,069	,079	,019	,000	-,469	,000	,000	1,000	,000	,000	,000
Vervoersbewijs	-,086	-,023	-,017	-,015	,002	-,006	,009	,030	-,008	,021	-,072	,060	,026	-,021	-,075	,000	,228	,160	,477	,000	,000
Comfort	-,141	-,038	-,028	-,025	,004	-,009	,015	,050	-,014	,035	-,119	,099	,043	-,034	-,124	,000	,376	,264	,787	,000	,000
Klantvriendelijkheid	-,135	-,036	-,027	-,024	,004	-,009	,015	,048	-,013	,034	-,113	,094	,041	-,033	-,118	,000	,358	,251	,751	,000	,000
Netheid_halte	-,105	-,028	-,021	-,018	,003	-,007	,011	,037	-,010	,026	-,088	,074	,032	-,025	-,092	,000	,279	,196	,585	,000	,000
Netheid_voertuig	-,127	-,034	-,025	-,022	,004	-,009	,014	,045	-,012	,032	-,106	,089	,038	-,031	-,111	,000	,337	,236	,706	,000	,000
Info_halte	-,196	-,030	-,058	-,037	,005	-,009	,006	,070	,044	,042	-,117	,031	,013	-,047	-,023	,057	,520	,000	,000	,655	,000
Info_vertraging	-,210	-,032	-,062	-,039	,006	-,009	,007	,075	,047	,045	-,126	,033	,014	-,051	-,025	,061	,559	,000	,000	,702	,000
Rijstijl	-,132	-,035	-,026	-,023	,004	-,009	,014	,047	-,013	,033	-,111	,092	,040	-,032	-,116	,000	,351	,246	,735	,000	,000
Frequentie	-,208	-,031	-,061	-,039	,006	-,009	,006	,074	,046	,045	-,124	,033	,014	-,050	-,024	,060	,551	,000	,000	,693	,000
Totaaloordeel	-,199	-,044	-,048	-,036	,006	-,012	,015	,071	,008	,047	-,148	,094	,040	-,048	-,111	,024	,528	,215	,647	,279	-,007
Veiligheid_rit	-,155	-,004	-,006	,000	,002	-,006	,001	-,008	-,016	,008	-,013	,015	,117	-,037	-,091	,000	,410	,194	,000	,000	,959
Veiligheid_OV	-,116	-,003	-,005	,000	,002	-,004	,001	-,006	-,012	,006	-,010	,012	,088	-,028	-,068	,000	,309	,146	,000	,000	,722

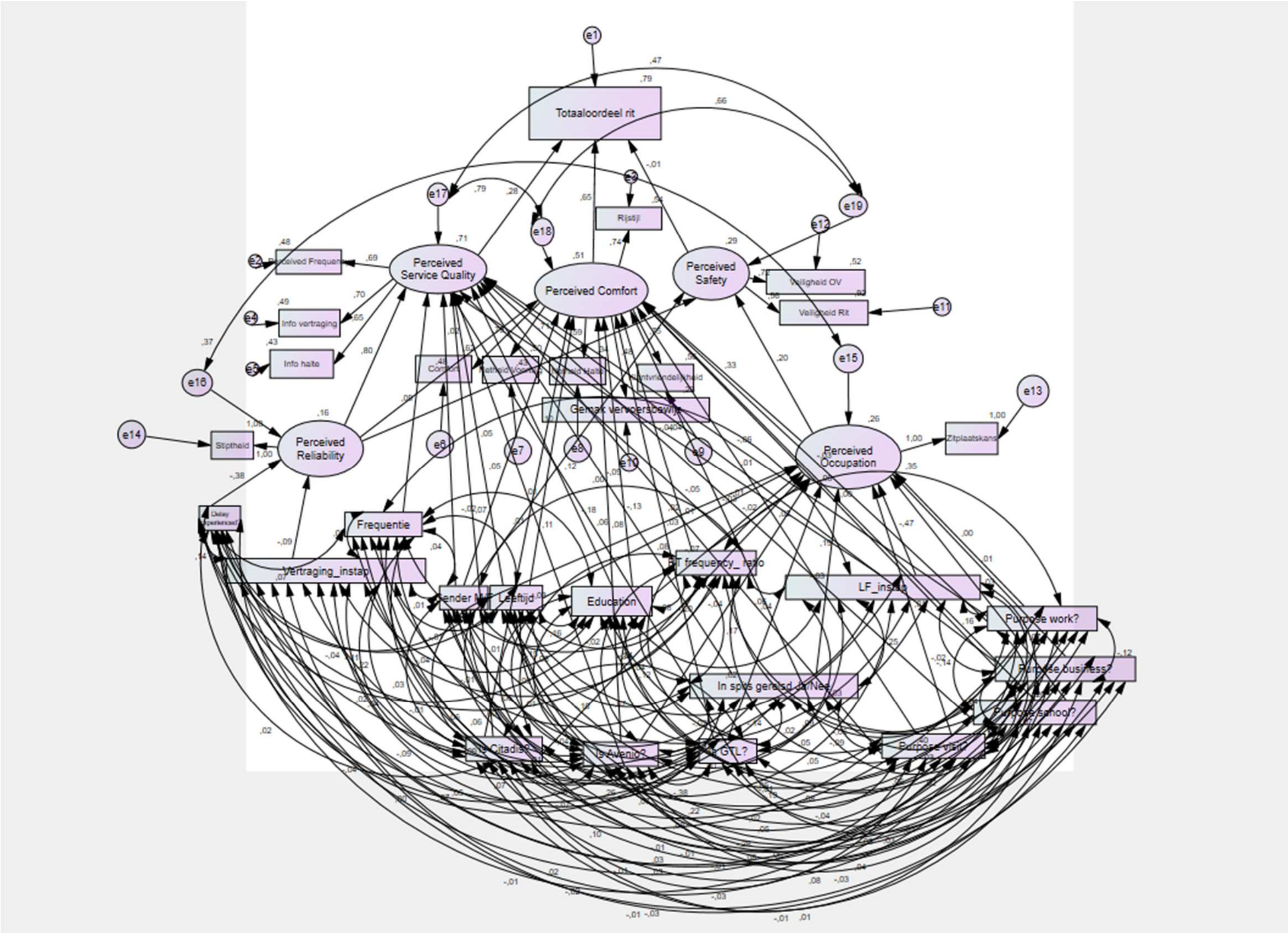


Figure E.4: estimated structural model

Besides all conclusions as reported in chapter five, the model has tested for all covariances between exogenous variables, which are at the same level in the model structure. Covariances which were found to be significant include the relationships between:

- Load Factor and Delay
- Rush Hour and Delay
- Load Factor and travel purpose
- Frequency and travel purpose
- Frequency and load factor
- Frequency and Rush Hour
- Delay and vehicle type
- Vehicle types
- Gender and education
- Age and education
- Age and Rush Hour
- Education and Travel Purpose
- Age and Travel Purpose
- Age, Gender and Vehicle Type

Most of these covariances are logical. For example, it is logical that age and travel purposes relate: younger people will travel more for educational purposes and elderly, who are retired, will not travel for work purposes. Other covariances can be explained mathematically: if a trip is made with an Avenio then that vehicle is neither a GTL nor a Citadis and as a result these vehicle types correlate as well.

However, some significant covariances need some more in-depth analysis for conclusions. The relation between vehicle types and delays comes to mind. Analysing the covariances the conclusion found was:

- GTLs have a significantly higher delay compared to buses ($t = 2.779$; $p = 0.005$)
- Avenios have significantly lower delays compared to buses ($t = -4.495$, $p < 0.001$)
- Citadis has no significant different delays compared to buses ($t = 0.793$; $p = 0.369$)

These coefficients can be explained as follows:

- Avenio lines often have excellent infrastructure where tram traffic and road traffic are handled separately. This means these lines are relatively safe for delays as a result of external factors.
- Buses often suffer from delays because of traffic jams and incidents. The Citadis lines (especially lines 3 and 4) suffer from delays as a result of capacity issues between Laan van NOI and Leidschenveen, where they share the tracks with RET line E.
- The lines on which GTLs provide services are often long and have long stretches in which the tram shares its route with road traffic. This makes them vulnerable for traffic jams (just like buses), but without the possibility to drive around potential obstacles.

Some covariances are difficult to explain based on data available. For example, the covariance between age and vehicle type and between gender and vehicle type. These might be explained by the demographic composition of Den Haag (e.g. lines with a certain vehicle type are used more in a part of the city which on average is older). On the other hand, while significant the impact of these covariances on the model output is negligible, the resulting effect on customer satisfaction being smaller than 0.05.

Panel effects

As section 3.5 has described, the data used in estimating the model above fails to meet one assumption which can affect results: the assumption of independence between variables. As has been described there, the same model has been estimated again, this time using a smaller dataset in which independence between responses can be assumed with certainty.

Looking at model fit, values comparable to the larger model can be seen for this smaller model. Table E. shows these values. The smaller model performs slightly worse than the large model. Due to different sample sizes comparing the chi-square statistics makes no sense as this statistic is sensitive to sample size. The p-value of the chi-square statistic remains 0.000 for the smaller model.

Indicator	**All-data ('larger') model**	**One per respondent ('smaller') model**
CFI	0.906	0.897
RMSEA	0.058	0.060

With 743 responses compared to 2858, the dataset used is, of course, much smaller. This means effects have to be much larger for the model to be significant. Some relations are found to be insignificant at the 0.05 level as a result of this. Relations which are not significant in this smaller model are marked with an asterisk. It is found that some personal characteristics, mainly travel frequency and travel purpose, are found to be insignificant in this smaller model. In general, this concerns relations which were very weak when quantified in the larger model.

Moreover, the impact of frequency on perceived service quality and vehicle type (Avenio) on comfort just fail to meet the 0.05 threshold. On the other hand, the effect of perceived safety on customer satisfaction is significant, albeit not large (0.074). Table XX shows relations in which a difference in significance on the 0.05 threshold can be seen between the two estimations. The exact outcomes can, again, be found in appendix E.

Relation	**Large dataset**	**Small dataset**
Perceived safety -> Customer satisfaction	Insignificant	Significant
Frequency -> Perceived Service Quality	Significant	Insignificant
Travel purpose -> all	Some significant	Insignificant
Vehicle Type -> Perceived Comfort	Avenio Significant, other insignificant	Insignificant

Vehicle type -> Perceived Occupancy	Significant	Citadis significant, other insignificant
Gender -> Perceived Comfort	Significant	Insignificant
Age -> Perceived Service Quality	Significant	Insignificant

The exact estimated values for other relations also vary as a result of using a different, smaller dataset. However, except for the effect of perceived safety no variable switches sign and the standardized coefficients remain comparable. The only major exception is the effect of travelling during rush hour, a coefficient which drops from -0.25 to -0.06 (albeit the effect remains statistically significant).

Having described the differences between both model estimations, it is important to interpret these differences: why do some conclusions differ? Moreover: what do these conclusions mean for the conclusions drawn? Several possible explanations exist:

- One possible explanation is that some relations found using the larger dataset only seem to exist due to the assumption of independence. Again, this assumption is clearly incorrect: Some respondents have provided more than 50 responses and the effect of their personal opinions and perceptions is overestimated by the model compared to respondents who only provided one response. Should this be the dominant explanation, the smaller dataset approaches the true situation better.
- A different, contrasting explanation is that the smaller dataset used in the smaller model lacks the predictive power that the larger dataset offers. As a result, relations which were found in the larger dataset are not found when using the smaller set due to a lack of a large enough number of responses. This for example can be the case for the different vehicles. This suggests the larger dataset is more usable.

As often is the case, I think both arguments are valid to some extent, while looking at the numbers the first reasons seems dominant. The best solution, of course, would be to test the model again using software which is able to incorporate multiple responses per respondent. While this software exists it was unavailable for the author during this thesis.

Appendix F: Manuals of Excel tools

During this thesis' research, two tools have been developed which assisted the author in carrying out the research described. This appendix consists of the manuals to these two tools. Appendix F.1 explains the tool used to enable the coupling of customer satisfaction survey data to occupancy and punctuality databases, appendix F.2 the tool which transforms the results from the structural equation model to a calculation model from which estimates can be made.

F.1. Data coupling

As mentioned in chapter 3, almost 4000 trip evaluations were made for the HTM Klantenpanel in 2018. Finding occupancy and punctuality data for these responses manually is an extremely time-consuming task, and so an Excel tool was developed. The aim of this tool is to interpret the relevant data from each response and convert it to a format which allows for easy searching in HTMs databases for occupancy and punctuality numbers.

The file, named 'converter HTMpanel to SQLdatabases.xlsx', has three visible sheets:

- A sheet named 'instructies and FAQ', which contains the Dutch manual for the tool.
- A sheet named 'Output' which contains the output which can then be used to search for occupancy and punctuality data in the corresponding databases at HTM.
- A sheet named 'Data_Ritbeoordelingen_HTMpanel' in which the data which should be converted needs to be inserted.

In using the tool, the first step is to find the corresponding HTM Klantenpaneldata. This data is delivered to HTM by Citisens as a .sav (SPSS) file. SPSS allows for easy conversion of .sav files to Excel (.xlsx/.csv)-files. After conversion, the full sheet can be copy-pasted into the 'Data_ritbeoordelingen_HTMpanel' tab. The sheet 'Output' will then contain the necessary output in the required format, being:

- Date and time of boarding according to the response
- The line which was evaluated
- The direction in which the passenger was travelling on this line
- ID of the boarding stop
- ID of the alighting stop

If the 'output' sheet (just this sheet) is then saved as a .csv-file this can be used as input for searching in HTMs databases which work based on SQL. Section 3.3 describes how this search is done. The corresponding SQL search code which searches the databases as described in section 3.3 has been saved in the database, which means only the corresponding input file as generated by this tool has to be inserted to start the search process.

While the tool automatically converts all data to the correct format, the searching for line, direction and stop ID's is done in a list of IDs stored in hidden sheets in the model. This is necessary, as each platform has its own ID and almost every stop thus has at least two IDs (one for the platform in one direction and a different ID for the stop in the other direction). Moreover, directions of lines (coded with 1 and 2 in HTMs databases) do not always match. As a result, for each line and direction stop IDs can be different and no 'general' algorithm can be used. The coupling of stop IDs to each line and direction was done manually by the author.

As a result of the way coding has been done, the model produces an error if:

- A service called at a stop on which this line normally does not call.
- A service halted at a different platform than a specific line normally does.

In case of the rerouting of lines, the sheet containing all stops for this line needs to be adapted so that for each line all correct stops and IDs per direction are up to date. In the case of the introduction of a new line, a new sheet needs to be added (named exactly "HalteIDs lijn <line number>) with all stops and stop IDs per direction of that specific line.

This tool was built with all stops for each line included as for the regular timetable of 2018, as well as those used during long-time construction works in 2018.

F.2. Calculation Model

Chapter 6 presented a calculation tool which can be used to estimate satisfaction values based on the results obtained from the structural equation model in chapter 5. The file, named 'Excelmodel SEM.xlsx' has four sheets:

- Input/Output, which is shown in table 6.1. The user needs to fill in values for all input variables and will then estimate the evaluations based on the input.
- Rekenwerk, which shows the calculations which are done to estimate the output values in the Input/Output sheet. The basis of the calculations is that the mean value for each aspect is searched and given. Next, based on for each input aspect provided in the Input/Output sheet the difference with the average for that input aspect is calculated. This difference is then multiplied with the unstandardized total effect the two variables have on each other to find the effect that this aspect has on evaluation for this trip. This is done for all Input variables, the outcomes are then summed with the estimated mean to obtain the estimated evaluation for the service based on the Input given.
- Modelcoëfficienten, which contains data on the unstandardized total effects all variables have on each other and the intercepts (expected means) of all relevant variables.
- Keuzelijsten, which contains a few lists which are used on the Input/Output sheet to limit variables such as vehicle type, gender and education level to the values used and estimated in the model.

Should the model at some point be re-estimated using new data, all that has to be done is that all values on the sheet Modelcoëfficienten have to be replaced with their new, updated values.

In using the model, the model parameters seem to underestimate satisfaction in case of extremely good performances and overestimate satisfaction in case of poor performance. Table 6.3 provides an example of both these situations.

Table 6.3: overview of expected evaluation of a good and a poor service provided.

Input	*Good performance*	*Poor performance*	*Unit*
Vehicle type	Avenio	Avenio	
Frequency	20	1	Veh/h
Rush hour?	N	Y	Y/N
Occupancy on boarding	0	180	passenger
Delay on boarding	0	600	s
Does passenger experience delay?	N	Y	Y/N

Output			
Load Factor	0,0%	257,1%	

Expected grades	*Trip average*	*Personal average*	
Overall satisfaction	**8,6**	**4,9**	
Perceived crowding	9,0	0,6	
General feeling of safety in PT	8,0	6,7	
Comfort during trip	8,1	4,8	
Cleanliness of vehicle	8,0	6,1	
Cleanliness of stop	7,9	6,3	
Information supply on stop	9,1	5,7	
Information supply during delays or disruptions	8,2	3,8	
Friendliness of staff	8,4	5,0	
Driving style of driver	8,2	5,9	
Ease of buying a ticket	9,1	7,5	
Punctuality	8,3	4,6	
Frequency	8,6	5,2	
Feeling of safety during this trip	8,3	6,5	

It can be seen that while in the 'good' case performance could not be better than given the marks given still remain between 8 and 9 while in case of extremely poor performance still some okay (> 6) marks are expected, even though the general satisfaction drops quite heavily. It thus seems that this tool, using all model parameters, overestimates the evaluation of very poor services and underestimates the evaluation of very good services. This can be explained by Structural Equation Modelling being unable to incorporate interaction effects. These interaction effects can be explained as follows: if a customer is extremely satisfied with certain parts of this travels he will tend to score other aspects higher as well. In the case of very poor performance the same effects occur the other way around: a customer dissatisfied with the crowding of a service will tend to grade other aspects which do not necessarily have to do with crowding lower.

The method used provides no simple solution for this problem: SEM is unable to properly incorporate these interaction effects. Solving these issues requires more advanced algorithms and techniques. For now, the conclusion needs to be that this calculation model seems to be a tad more inaccurate in extreme situations than it is for an average trip.

Appendix G: Translation of trip evaluation questions in HTM Klantenpanel

This appendix provides an overview of the Dutch names of travel aspects which are evaluated by respondents of the HTM Klantenpanel. These Dutch names are used in this thesis, here their English translations are shown.

Item	Translation
Totaaloordeel	Overall satisfaction
Zitplaatskans	Perceived crowding
Veiligheid OV	General feeling of safety in PT
Comfort	Comfort during trip
Netheid Voertuig	Cleanliness of vehicle
Netheid Halte	Cleanliness of stop
Info halte	Information supply on stop
Info Vertraging	Information supply during delays or disruptions
Klantvriendelijkheid	Friendliness of staff
Rijstijl	Driving style of driver
Gemak vervoersbewijs	Ease of buying a ticket
Stiptheid	Punctuality
Frequentie	Frequency
Veiligheid Rit	Feeling of safety during this trip