EJTIR

Issue 17(3), 2017
pp. 360 – 383
ISSN:1567 – 7141
tlo.tbm.tudelft.nl/ejtir

Understanding future mode choice intentions of transit riders as a function of past experiences with travel quality

Andre Carrel[1]
Dept. of Civil, Environmental and Geodetic Engineering, The Ohio State University, Columbus, USA

Joan L. Walker [2]
Dept. of Civil and Environmental Engineering, University of California at Berkeley, Berkeley, USA

This paper empirically investigates the causes for transit use cessation, focusing on the influence of users' personal experiences, resulting levels of satisfaction, and subsequent behavioral intentions. It builds on a novel data set in which observed, objective measures of travel times are mapped to smartphone-based surveys where participants assess their travel experience. An integrated choice and latent variable model is developed to explain the influence of satisfaction with operations (travel times) and satisfaction with the travel environment (e.g., crowding) on behavioral intentions. Satisfaction is modeled as a latent variable, and the choice consists of participants' stated desire and intention to continue using public transportation. The results show how delays, in particular in-vehicle delays but also transfer times and being left behind at stops, contribute to passengers' intentions to cease transit use. Furthermore, a number of critical incidents, i.e., particularly memorable negative experiences, are found to have negative and significant effects on overall satisfaction and on willingness to continue using public transportation. The usefulness of the framework is demonstrated in a set of simulations in which the effect of three types of delays on passengers' willingness to remain transit riders is modeled. This work highlights the value and potential of using new data collection methods to gain insights on complex behavioral processes, and it is intended to form the basis for new modeling tools to understand the causes of transit use cessation and the impact of various strategies and service quality improvements to reduce ridership turnover.

Keywords: *Latent Variable Choice Model, Mode Choice, Public Transportation, Rider Loyalty, Satisfaction, Service Quality.*

1 Introduction and motivation

Public transit is a key element to efficient and sustainable urban transportation, and in the past decades, numerous public policies have been designed to increase its mode share in urban areas through subsidies, service expansions, and land-use zoning. Yet, as is noted by Perk, Flynn, and Volinski (2008), US transit agencies continue to see high levels of ridership turnover; in many cases, a steady influx of new users into the system is offset by similarly high rates of transit use cessation. On the individual level, these shifts are not trivial: As is explained by Vij, Carrel, and Walker (2013), travelers tend to build their lifestyle around the use of certain travel modes, and decisions between, for instance, an auto-oriented lifestyle and a transit-oriented lifestyle are relatively stable. In other words, users who quit using a transit system are often unlikely to return unless a major upgrade to the transit system is made.

[1]A: 470 Hitchcock Hall, 2070 Neil Avenue, Columbus, OH 43210, USA. T: +1.614.292.2771 E: carrel.20@osu.edu
[2]A: 111 McLaughlin Hall, Berkeley, CA 94720, USA. T: +1.510.642.6897 E: joanwalker@berkeley.edu

The fact that little is known about the reasons for which people shift away from transit-oriented lifestyles is primarily due to a lack of suitable data. Authors have generally identified changes in lifestyles associated with events such as marriage, or having children, as causes of transit use cessation. However, Carrel, Halvorsen, and Walker (2013) identified a variety of negative experiences with service quality (e.g., delays, high crowding levels) as further potential drivers. This paper aims to quantify the effects of negative experiences on transit users' future intentions of transit use. It is based on the San Francisco Travel Quality Study data set, presented in section 3.1, and uses a latent variable choice model to understand the link between their individual experiences, satisfaction and future intentions.

2 Literature review

There are very few publications that have investigated transit passenger loyalty. The only studies we are aware of that have done so in a generalizable fashion are by Trépanier, Habib, and Morency (2012) and Ma et al. (2013). Both made use of automatically collected passenger data from smart cards, but they did not link the observed usage patterns to riders' experiences with the service. This paper is concerned with making that link, and with describing the influence of individual experiences on transit rider satisfaction and on future behavior. The framework laid out in this paper has several components:

1. The link between individual experiences with transit service quality and reported levels of satisfaction on a daily level.
2. The link between satisfaction on a daily scale and overall satisfaction reported at the end of an extended period of time.
3. The link between overall satisfaction and future behavior.

The first item was the subject of a previous paper (Carrel et al., 2016) and will be expanded on in this paper. The pertinent literature is presented in more depth there and is only summarized in the paragraph below.

Satisfaction surveys are most valuable if a link can be made between satisfaction and objective service quality measures (Davis and Heineke, 1998). So far, this has not been done in transit satisfaction literature, with the exception of work by Friman and Fellesson (2009) and Carrel et al. (2016). Only in the latter was the link made on an individual rather than an aggregate level. In fact, customer satisfaction is a function of *personal use experience* (Woodruff, Cadotte, and Jenkins, 1983; Anderson and Sullivan, 1993), and in transportation, it is generally formed through multiple repeated experiences over time. To control for memory distortions, the analyst needs to be knowledgeable of the subject's usage history and needs to limit the time frame covered by the satisfaction survey (Fredrickson and Kahneman, 1993; Kahneman et al., 2004). In Carrel et al. (2016), several separate ordinal logit models were estimated, linking satisfaction with individual travel time components to observed travel times. It was found that the disutility of scheduled in-vehicle travel time was much lower than in-vehicle delay time, and that in-vehicle delays appear to be strong drivers of passenger dissatisfaction. Under certain circumstances, the latter might be perceived as more onerous than out-of-vehicle wait time. Furthermore, it was found that the baseline satisfaction with transit services and subjective well-being on the day of the survey were important covariates in the measurement of daily satisfaction.

The second item is the link between satisfaction on a daily level and satisfaction reported at the end of an extended period. It is recognized in the marketing literature that satisfaction is a dynamic phenomenon and can change over time (Mittal, Kumar, and Tsiros, 1999), and that this change is a function of personal experiences a decision-maker has made with the service or product in question (Anderson and Sullivan, 1993; Davis and Heineke, 1998). This is consistent with the findings of Kahneman et al. (1993), who found that subjects' ratings of a repeated experience were dependent on their history of previous experiences. Bates et al. (2001) extend this finding to the transportation realm and argue that personal experience is very important in the context of travel time variability, and that travelers will not choose a route based on average travel times on that route, but rather on travel times they have experienced in the past - i.e., their personal travel time distribution. That distribution is updated every time a person makes a trip, so a person's satisfaction reported at the end of an extended period should be a combination of their satisfaction at the beginning of that period and their satisfaction with experiences during the period. Abou-Zeid et al. (2012) based an experimental design aimed at capturing travelers' subjective well-being before and after a week of transit use on this notion, though satisfaction with daily experiences was not separately measured. Furthermore, Abou-Zeid et al. (2012) argue that travelers may be less cognizant of their well-being and satisfaction when they carry out routine behavior, such as choosing a car for commuting. This is consistent with findings by Lanken et al. (1994). We propose a further distinction. The effect described by Abou-Zeid et al. (2012) may be stronger when travel times and service quality are consistent, and less pronounced when the experienced travel times or other service quality aspects are variable, since delays or service failures can trigger negative emotions that influence future decision-making. In public transportation, negative emotions caused by service failures can be exacerbated by the passengers' feeling of not being in control of their experience (Anable and Gatersleben, 2005). Therefore, we postulate that even with habitual transit riders, satisfaction levels can change as a consequence of positive or negative experiences when service quality is variable. This is supported by work by Friman, Edvardsson, and Gärling (2001), who found a measurable impact of "critical incidents", i.e., memorable positive or negative experiences, on customer satisfaction with public transportation reported in a post-study survey.

The third item is the link between satisfaction with travel modes and future travel behavior. With specific regard to public transportation, Pedersen, Friman, and Kristensson (2011) have investigated the influence of satisfaction on mode choice with a statistical path analysis and found a positive association between remembered satisfaction and current choices. Satisfaction with travel modes has also been successfully included in discrete choice models of mode choice (Abou-Zeid and Ben-Akiva, 2010; Friman et al., 2013). In the application reported on in this paper, future choices were not directly observed, so to link satisfaction with behavior, an alternative set of variables was required. These variables were formulated in accordance with the Model of Goal-Directed Behavior (MGB), a theoretical model of behavior change that is grounded in psychology and the behavioral sciences. The MGB is a refinement of the widely used Theory of Planned Behavior (TPB) (Ajzen, 1991). In the TPB, satisfaction is considered part of the set of attitudes toward the behavior in question. Along with norms and beliefs, attitudes are linked to a person's intention to carry out a future behavior, which in turn leads to observed behavior. An example in which the TPB has been applied directly to mode choice is Bamberg, Ajzen, and Schmidt (2003). The MGB adds two elements to this framework: Anticipated emotions and behavioral desire. It postulates that there are three steps to behavior change, as shown in Figure 1: First, a person develops a desire to change behavior, followed by the formulation of an intention to change behavior. Lastly, the person will actually change behavior. The development of a behavioral desire and the transition to the two next steps are governed by a variety of factors. For more information on the model and influencing factors, see Perugini and Bagozzi (2001). A discussion of the difference between behavioral desire and intention can be found in Perugini and Bagozzi (2004). The Model of Goal-Directed Behavior was chosen over the Theory of Planned Behavior due to the explicit recognition of emotions, as the original study design included several questions on travelers' subjective well-being. Following the MGB, the outcome variables covered behavioral intentions as well as desire, in order to capture transit users in both stages of the decision-making process and to approximate the (unobserved) future choice as closely as possible.

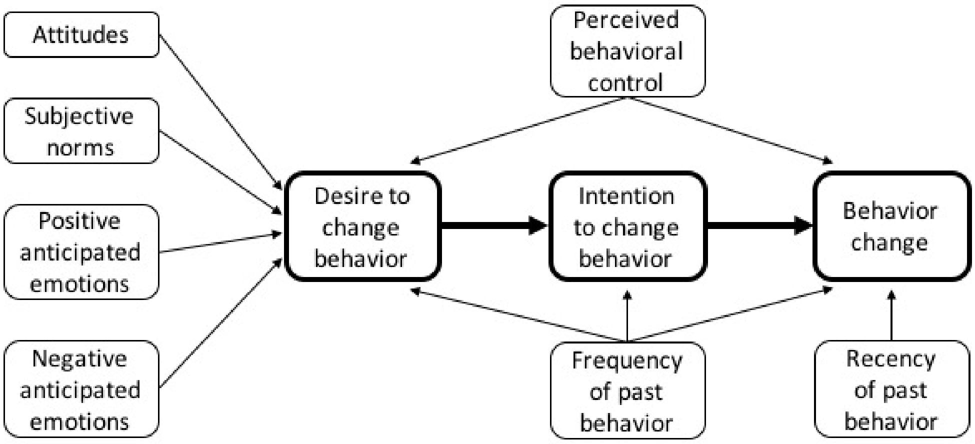


Figure 1. The Model of Goal-Directed Behavior (Perugini and Bagozzi, 2001).

3 Methodology

3.1 Data source

The data were collected during the San Francisco Travel Quality Study, a large-scale study of transit service quality, which ran from October 21 to December 22, 2013. The design and organization of the study is described in detail in Carrel, Sengupta, and Walker (2016); what follows is a summary. The study involved an initial total of 856 participants recruited from the general public in San Francisco. It focused on usage of the San Francisco Municipal Transportation Agency network (commonly called "Muni" in San Francisco; this term is used in the remainder of the paper). As an incentive, participants received a free one-month transit pass, valid for unlimited travel on the Muni network. Participants were asked to complete an online entry survey, in which sociodemographic and mode access information was collected, as well as an exit survey at the end of the study. All participants received the entry survey on October 21. Depending on when participants received the transit pass, they were divided into two cohorts. Cohort 1 received the exit survey on December 8; cohort 2 received it on December 22. Response times to the surveys varied. After completing the entry survey, participants were asked to download a survey app for Android phones. They were instructed to keep location services enabled on their phones, and if they did so, the app collected location information from the phone every 30 seconds. Once per day at a time set by the participants, generally in the evenings, they received a survey prompt on their phones asking them whether they had used transit on that day. If they responded yes, they were presented with a survey on their phone (hereafter called "daily mobile survey") in which they were asked to rate their satisfaction with the transit service they had experienced that day. Participants were asked to use Muni on at least five days during the study period and fill out the corresponding daily mobile surveys. The time window for responding to the daily mobile surveys was between October 27 and December 1 for cohort 1 and between October 27 and December 15 for cohort 2. Since the daily mobile surveys were filled out only once per day, they referred to all transit trips made by the participant on that day, regardless of the number.

This paper focuses on the link between experienced travel times, satisfaction, and future transit use. The analysis builds on the following data collected during the study:

- Satisfaction with transit services: The online entry and exit surveys measured satisfaction with nine variables, each on a five-point Likert scale from 'very dissatisfied' to 'very satisfied': Overall reliability, in-vehicle travel time, wait time at the origin stop, transfer time (if applicable), crowding, cleanliness, safety, pleasantness of other passengers, and the accuracy of real-time information. The prompt in these surveys asked respondents to rate their overall satisfaction with Muni services. In the daily mobile surveys, respondents were asked to rate their satisfaction with the same nine variables, but only with respect to the transit service they had experienced that day. An exploratory factor analysis on the results confirmed that there were strong correlations within two groups of variables: The first four can be summarized as satisfaction with operations, whereas the following four can be thought of as satisfaction with the travel environment.

- Experienced travel times: During the study, the vehicle locations of all transit vehicles throughout the city were continuously recorded. The phone location data from the participants were then matched to the transit vehicle location data to automatically identify whether the participant had used transit on that day and to extract wait times, in-vehicle travel times and transfer times. The measured travel times were then compared to timetable information to identify deviations from the timetable. This methodology provided objective measurements of the transit travel times and delays experienced by the participants. An in-depth description of the methodology can be found in Carrel et al. (2015). For the purposes of the model presented in this paper, since the satisfaction surveys concerned an entire day, participants' travel times were aggregated on a daily level.
- Future transit use: In the entry and the exit surveys, the participants were asked a set of questions regarding their future transit use. In accordance with the MGB, they were asked about their behavioral intentions and their behavioral desires, with the following question prompts:
  - Question 1: "In 2014, do you intend to use [Muni] more or less than you do now, or the same way as you do now?"
  - Question 2: "Ideally (regardless of whether you intend to do so), do you want to be able to travel in San Francisco by [Muni] more or less than you do now, or about the same way as you do now?"

  After the entry survey had been distributed, it became apparent that the formulation of these questions was not optimal. Some respondents were confused by the difference between behavioral intention and desire, and others stated that they did not know in advance what their mode choices during an entire year would be. Therefore, five additional questions were added in the exit survey:
  - Question 3: "Compared to how often you used Muni in the month *before* the study, you anticipate using Muni in January...". Responses were on an 8-point Likert scale from "not at all anymore" to "much more". (This measured short-term intention compared to before the study)
  - Question 4: "Compared to how often you used Muni *during* the study, you anticipate using Muni in January... ". Responses were on an 8-point Likert scale from "not at all anymore" to "much more". (Short-term intention compared to during the study)
  - Question 5: "Compared to how often you used Muni during the study, you would prefer to use Muni in January...". Responses were on an 8-point Likert scale from "not at all anymore" to "much more". (Short-term desire)
  - Question 6: "As soon as my circumstances permit, I would like to use public transportation more". Responses were on a 5-point Likert scale from "strongly disagree" to "strongly agree". (Long-term desire)
  - Question 7: "As soon as my circumstances permit, I would like to use public transportation less". Responses were on a 5-point Likert scale from "strongly disagree" to "strongly agree". (Long-term desire)

  The five added questions incorporate three changes: First, the time frame for the intention questions was shortened from one year to one month, since it was assumed that participants would have a clearer sense of their mode use in the month following the study. Second, the word "intend" was removed, and instead, participants were simply asked how they were going to travel in January. Third, the cumbersome formulation "...do you want to be able to...", which was the original measure for desire, was replaced with "...would you prefer to...". After the exit survey was distributed, no emails were received asking for clarification on these questions, suggesting they were better understood than the original set. Participants were also presented with a list of possible reasons for their response to question 5 and were asked to rate the level of influence of every item on their behavioral desire. A descriptive analysis of the results is provided in section 4. Lastly, it should be noted that the question on short-term desire was only asked in comparison with the respondents' transit use during the study, but not in comparison with their transit use before the study. This was an unfortunate oversight on the part of the survey designer.

The questions on future transit use were intended as the set of outcome variables. In the original survey design, it was planned to use questions 1 and 2 to measure changes in behavioral desire and intentions during the study. The responses to the entry survey would have provided a baseline. When questions 3 through 5 were added, it was no longer possible to measure a baseline, so instead, they were formulated as comparisons such that the baseline was self-reported. The explicit distinction between behavioral desire and intention was introduced in the surveys to ensure consistency with the MGB. In total, the data set used for developing the model presented here included 449 respondents who had filled out the entry and exit surveys, and from whom at least one daily mobile survey response and relevant travel time data was recorded. More information on the characteristics of the study population can be found in Carrel, Sengupta, and Walker (2016).

3.2 Satisfaction response patterns

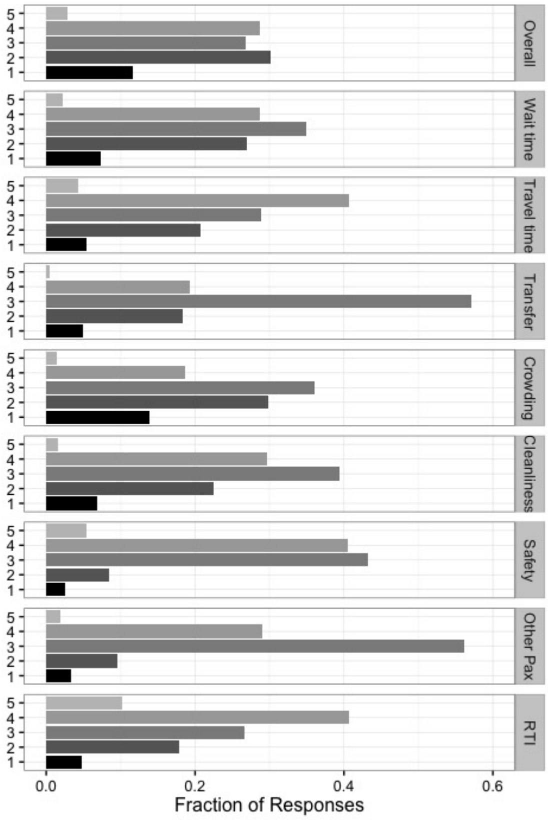
Before introducing the model, we will first consider an interesting observation with respect to the satisfaction reported in the exit survey. All participants were asked to complete the online exit survey, but in addition to that, an optional mobile exit survey was also distributed. Its formatting was identical to the daily mobile surveys, with the exception that the survey prompts asked respondents to indicate their satisfaction with their overall Muni experience during the study. Participants were asked to fill out the mobile exit survey in addition to the online exit survey, but it was made clear that that was not mandatory. The 5-point response scales in the daily mobile surveys were labeled only at the maximum and minimum with a frowny and a smiley face due to space constraints, where as the scale in the online survey was labeled with words since the online survey engine did not allow graphical labels. A distinct difference was noticed between users' responses with respect to their daily satisfaction and their overall satisfaction in the online and mobile exit surveys. In Figure 2c, the distributions of responses to the nine satisfaction items in the daily mobile survey are shown. It should be noted that this sample includes multiple responses per participant, and no correction has been made to account for correlation between responses from the same person. For each item, the darkest bar shows the proportion of "very dissatisfied" responses and the lightest bar shows the proportion of "very satisfied" responses, with the remaining bars showing the responses in between. Figures 2a and 2b show the distributions for the online exit survey and the mobile exit survey, respectively. These samples include only one observation per participant.

Whereas users were willing to state that they were "very satisfied" with their daily experiences on public transportation, it can be seen that:

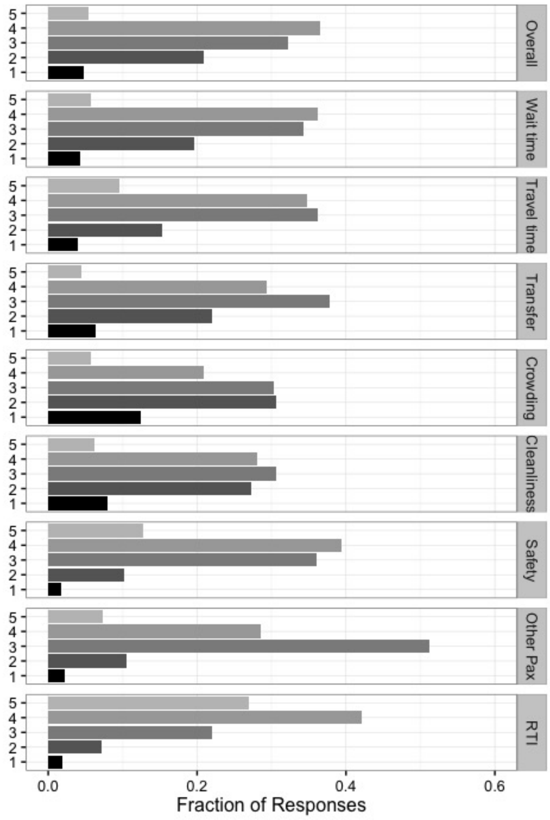
- Participants were less willing to state that they were "very satisfied" in the exit surveys
- This effect was more pronounced in the online exit survey than in the mobile exit survey.

The sample size for the mobile exit survey was smaller than the online exit survey since the former was optional, but all participants who took the mobile exit survey also took the online exit survey. We propose four possible reasons for the observed discrepancies:

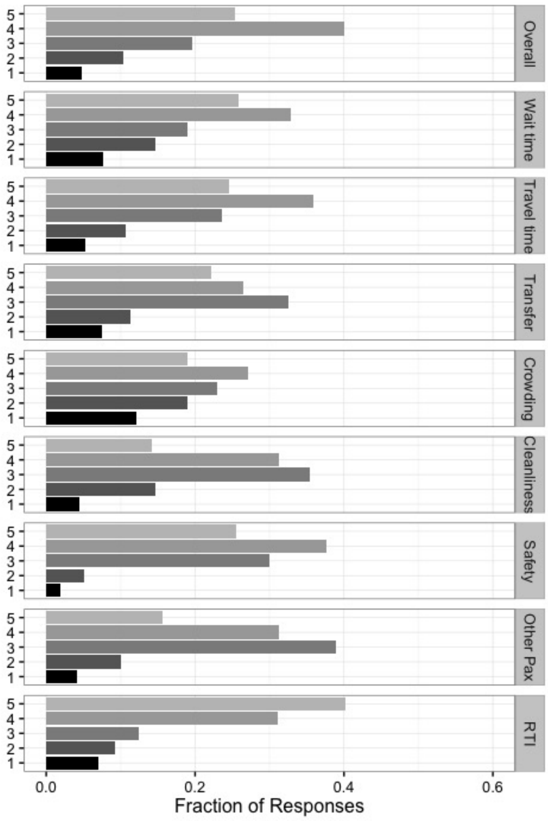
1. The service quality experienced by study participants between the end of the daily survey prompts and the time they filled out the exit surveys was markedly worse than the service quality on the days for which daily surveys were filled out. While possible, this explanation is not plausible.
2. The different time frames to which the questions are referring to: When asked about their



(a) Online exit survey (N = 482)



(b) Mobile exit survey (N = 353)



(c) Mobile survey (N = 6846)

Figure 2. Distribution of survey responses. RTI stands for Real-Time Information.

overall satisfaction, participants may recall negative events that occurred before the study period.

3. The different presentation of the questions, i.e., the fact that the labels differed between the mobile and the online survey versions. While this may account for some of the differences between the online and the daily mobile survey responses, the fact that the mobile exit survey response patterns also differ from the daily mobile survey response patterns indicates that this cannot be the only factor at play.

4. The different environments in which the surveys may have been filled out. It is more likely that the online survey was taken by people at home or at the office, whereas the daily mobile survey may have been taken anywhere. However, one could argue that the mobile exit survey may also have been taken anywhere.

The definitive reason for this discrepancy cannot be elicited without further investigation. Until that is possible, researchers designing future studies should be aware that the medium through which the survey is delivered (in this case, smartphone vs. online survey engine) can have an effect on the response patterns. The fact that the "very satisfied" category in the online exit survey had very few responses would have introduced nonlinearity in the latent variable measurement model and caused estimation problems since the measurement equations assume a linear relationship between the latent variable and the indicator variables. To avoid these problems, the satisfaction ratings for the online exit survey were re-scaled to a four-point scale where "satisfied" and "very satisfied" were included in one category.

3.3 Model development

Carrel et al. (2016) described the link between experiences and satisfaction in a static context and considering the individual travel time components separately. This paper extends that work in several ways and embeds it in a broader modeling framework. The primary purpose is to link personal experiences to future behavior via the intermediate construct of satisfaction. In this case, the overall satisfaction with travel times and operations was of interest, and not the satisfaction with the individual travel time components. Following the exploratory factor analysis described in section 3.1, we assume that there are two underlying and unobserved latent satisfaction variables - satisfaction with operations and with the travel environment. The observed variables, i.e., the recorded responses, are indicators of that underlying latent variable. To make the links between satisfaction in the entry survey, experiences during the study, reported daily satisfaction, satisfaction in the exit survey and future mode choice behavior, a latent variable choice model (Walker, 2001) was developed, as shown in Figure 3. In the figure, ellipses denote latent variables, rectangles denote observed variables, and the arrows show the directionality of effects being modeled.

The latent entry and exit satisfaction are shown as "entry satisfaction with operations" and "exit satisfaction with operations" with the respective indicator variables I_1 through I_6. The indicator variables were the reported satisfactions with the in-vehicle travel time, wait time and overall reliability. The daily satisfaction was modeled using the same latent variable constructs; those are shown as d_1 through d_r. The indicator variables for the daily satisfaction are omitted in the figure due to space limitation, but every daily satisfaction item had four indicator variables which were the aforementioned three satisfaction measurements plus satisfaction with transfer time. There were five variables for daily satisfaction: the four most recent responses for which travel time data were available, labeled d_1 through d_4, plus a fifth variable including the average of all remaining daily observations, labeled d_r. This structure was chosen since there were variable numbers of responses per participant. Every participant with at least one daily mobile survey response was included in this data set. The structural model for satisfaction with operations reflects the temporal dependencies between the individual surveys: The daily satisfaction ratings are influenced by satisfaction reported in the entry survey as well as by travel times experienced on that day and by the user's general feeling on that day. The daily responses in turn feed into the satisfaction reported in the exit survey. The assumption is made that the exit satisfaction depends on the entry satisfaction only by way of the daily satisfaction. All coefficients relating the daily latent satisfaction variables,

d_1 through d_4 and d_r, to the exit satisfaction are constrained to be the same. The coefficients of the measurement equations of d_1 through d_4 are also constrained to be the same, but the coefficients for the measurement equation of d_r are allowed to differ to reflect the fact that d_r is averaged over a number of days. In addition to satisfaction with operations, the exit satisfaction with the travel environment is included as a separate latent variable, labeled "exit satisfaction with environment". The indicator variables for the latter are the reported satisfaction with crowding, cleanliness, safety and other passengers in the exit survey. No objective measurements were available for the travel environment. Both latent variables feed into the utility for the choice model.

In addition, several variables on negative critical incidents during the study which were self-reported in the exit survey were included as explanatory variables affecting the exit satisfaction directly. These were:

- The number of times a participant arrived late at work or school ("Late arrival at work").
- The number of times a participant arrived late at a leisure activity ("Late arrival at leisure").
- The number of times a participant reported that he or she wanted to use Muni but was not able to because of a delay on the system ("Could not travel due to delay").
- The number of times a participant was left behind at a stop because the vehicle was full ("Left-behind").

In Figure 3, the Greek letters denote groups of coefficients corresponding to the notation in Table 1. Different letters are assigned to different groups of coefficients to improve readability of the model.

As explained at the beginning of this section, the original outcome variables on behavioral intention and desire proved to be problematic. Nonetheless, the exploratory data analysis and first model specifications used the responses to those questions. Since they were asked both in the entry and the exit survey, they would have provided a more objective measure of changes in participants' intended/desired future behavior. Unfortunately, we found that the data for these two questions contained a lot of noise and showed little correlation, both between the entry and exit responses and between the responses and other variables. Therefore, three of the alternative variables that had been added in the exit survey were used: Questions 3, 5, and 6. These three variables served as (imperfect) indicators of the latent choice that was of interest but that could not be observed (Bollen, 2014): Whether or not a person was going to use public transportation less in the future. Though the responses were recorded on five-point Likert scales, the indicator variables were reduced to a binary choice between desiring/intending to use Muni less and desiring/intending to use it the same or more. This was done to reduce model complexity.

Question 3 contains a self-reported baseline and was formulated to ask the respondent about differences between pre-study transit use and post-study transit use. It can therefore be considered a reasonable replacement for question 1. In the figure, it is labeled as "change in intentions". Question 5 also contains a baseline, though it is with respect to transit use during the study rather than before. However, as 93% of all survey respondents used Muni more than 2 days per week before the study (see Carrel, Sengupta, and Walker (2016) for details), negative responses such as "[I will] not use Muni at all anymore" or "[I will] use Muni much less" can nonetheless be seen as indicative of decreases compared to the pre-study baseline. Although the different baselines may reduce correlations between the two variables, it was decided that this was outweighed by the ability to have indicators for both short-term intention and desire. Question 6, on the other hand, was formulated as a statement with which respondents could disagree or agree. This did not require them to make a comparison with a baseline, which had the advantage that even if respondents did not know exactly how much they would use transit in the far future, they could express a general sentiment. Questions 4 and 7 were not included due to the strong similarity to questions 3 and 6.

The model specification assumes a set of causal relationships as shown by the arrows in Figure 3. Most importantly, it assumes that the participants' desire and intention to stop using transit is primarily a function of their satisfaction with transit services reported in the exit survey, which in

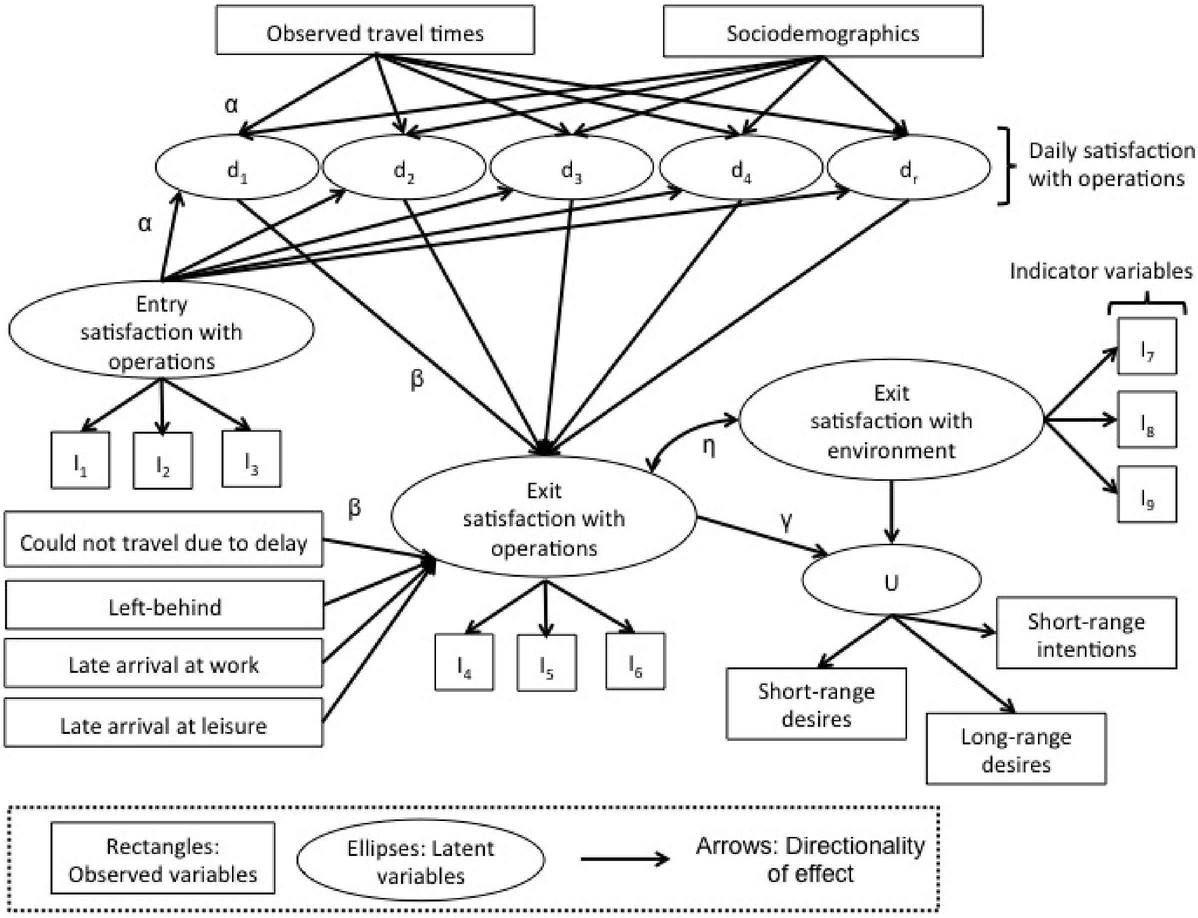


Figure 3. Model structure.

turn depends on the quality of transit service experienced during the study and their reported satisfaction during the study. Their sociodemographic characteristics and entry satisfaction influence the satisfaction reported during the study, but not the exit satisfaction directly. In other words, it is assumed that experiences during the study were significantly more important to participants' behavioral intentions and desires than experiences before the study, such that any direct influence of the latter on the latent choice can be disregarded. In addition, because participants were not required to fill out a daily mobile survey after every day on which they used Muni (as long as they submitted the minimum number), the estimation results can only be interpreted properly if it is assumed that the experiences reported through the daily mobile surveys are representative of the participants' average experiences during the study. We recognize that these are limitations of the model and the data, as is further discussed in section 3.4.

In what follows, the model specification is presented. The interpretation of all coefficients used in the model specification is shown in Table 1. The structural equation for the entry satisfaction with operations was:

$$Sat_{entry,ops} = \gamma_{entrymean\_ops} + \eta_{entry} \cdot \omega_{entry} \tag{1}$$

The structural equation for the daily satisfaction with operations (denoted d_i in Figure 3) was:

$$Sat_{daily,ops} = \alpha_{age} \cdot age + \alpha_{income} \cdot income + \alpha_{unknownincome} \cdot unknown\_income + \alpha_{longuser} \cdot longuser + \alpha_{entrysat} \cdot Sat_{entry,ops} + \alpha_{mood} \cdot mood + \alpha_{ivtt} \cdot ivtt + \alpha_{early} \cdot early + \alpha_{delay} \cdot delay + \alpha_{waitover5} \cdot waitover5 + \alpha_{waitunder5} \cdot waitunder5 + \alpha_{nowait} \cdot no\_wait + \alpha_{transfertime} \cdot transfer + \alpha_{unobservedtransfer} \cdot unobserved\_transfer + \alpha_{notransfer} \cdot no\_transfer + \alpha_{leftbehind} \cdot left\_behind + \eta_{daily} \cdot \omega_{daily} \tag{2}$$

The structural equation for the exit satisfaction with operations was:

$$Sat_{exit,ops} = \sum \left(\beta_{dailysat} \cdot Sat_{daily,ops} \right) + \beta_{leftbehind\_1\_9} \cdot leftbehind\_1\_9 + \beta_{leftbehind\_10} \cdot leftbehind\_10 + \beta_{latework} \cdot latework + \beta_{lateleisure} \cdot lateleisure + \beta_{notravel} \cdot notravel + \eta_{exitops} \cdot \omega_{exitops} + \eta_{errorcorr} \cdot \omega_{errorcorr} \tag{3}$$

The summation term above is the summation over all latent daily satisfaction variables. The structural equation for the exit satisfaction with operations was:

$$Sat_{exit,env} = \eta_{exitenv} \cdot \omega_{exitenv} + \eta_{errorcorr} \cdot \omega_{errorcorr} \tag{4}$$

And finally, the choice model was:

$$V = (\mu_{shortint} + \mu_{shortdes} + \mu_{longdes}) \cdot (ASC_{shortint} + ASC_{shortdes} + ASC_{longdes} + \gamma_{ops} \cdot Sat_{exit,ops} + \gamma_{env} \cdot Sat_{exit,env}) \tag{5}$$

The μ and alternative-specific constant (ASC) terms above are specified such that they only enter into the equation if the choice being modeled relates to the respective outcome variable, and they are zero otherwise. The measurement equations all had the same functional form. For example, the conditional probability for the indicator "satisfaction with in-vehicle travel time" (IVTTSat_entry) of the entry satisfaction with operations is:

$$P(IVTTSat_{entry} | I_{IVTT}, \delta_{IVTT,entry}, \lambda_{IVTT,entry}) = \frac{1}{\sigma_{IVTT,entry}} \cdot \phi \left(\frac{(IVTTSat_{entry} - \delta_{IVTT,entry} - \lambda_{IVTT,entry} \cdot I_{IVTT}}{\sigma_{IVTT,entry}} \right) \tag{6}$$

Table 1. List of coefficients of the structural equations.

Coefficient	Meaning
η_{entry}	Error term (entry survey satisfaction with operations)
α_{age}	Age (in year brackets)
α_{income}	Income (10,000 USD brackets)
$\alpha_{unknownincome}$	Unknown income (Binary)
$\alpha_{longuser}$	Long-time user (Binary: System user > 2 years)
$\alpha_{entrysat}$	Entry satisfaction with operations (4 pt. Likert)
α_{mood}	General mood (5 pt. Likert)
α_{ivtt}	In-vehicle travel time (Minutes)
α_{early}	Early arrival at destination stop (Minutes)
α_{delay}	Late arrival at destination stop (Minutes)
$\alpha_{waitover5}$	Wait time greater than 5 minutes
$\alpha_{waitunder5}$	Wait time less than or equal to 5 minutes
α_{nowait}	No wait time inferred from location data (Binary)

Continued on next page

Coefficient	Meaning
$\alpha_{transfertime}$	Transfer time (Minutes)
$\alpha_{notransfer}$	No transfer inferred from location data (Binary)
$\alpha_{leftbehind}$	Denied boardings (Inferred from location data)
$\alpha_{unobservedtransfer}$	Transfer reported but not observed in location data (Binary)
η_{daily}	Error term (daily mobile satisfaction with operations)
$\beta_{dailysat}$	Daily satisfaction with operations (5 pt. Likert)
$\beta_{leftbehind\_1\_9}$	Between 1 and 9 denied boardings (Self-reported)
$\beta_{leftbehind\_10}$	10 or more denied boardings (Self-reported)
$\beta_{latework}$	Arrived late at work or school (Self-reported)
$\beta_{lateleisure}$	Arrived late at a leisure activity (Self-reported)
$\beta_{notravel}$	Wanted to use Muni but could not due to delay (Self-reported)
$\eta_{exitops}$	Error term (exit survey satisfaction with operations)
$\eta_{errorcorr}$	Error term correlation
$\eta_{exitenv}$	Error term (exit survey satisfaction with the travel environment)
$ASC_{shortint}$	ASC intended Muni use short-term
$ASC_{shortdes}$	ASC desired Muni use short-term
$ASC_{longdes}$	ASC desired Muni use long-term
γ_{env}	Coefficient for exit satisfaction with the travel environment
γ_{ops}	Coefficient for exit satisfaction with operations
$\mu_{shortint}$	Scale parameter intended Muni use short-term
$\mu_{shortdes}$	Scale parameter desired Muni use short-term
$\mu_{longdes}$	Scale parameter desired Muni use long-term

3.4 Limitations

The model is subject to a few limitations which are discussed here. First, the choice indicator variables were only measured in the exit survey. Two of the three indicator variables referred to self-reported baselines, one of which was relative to the traveler's behavior prior to the study and one of which was relative to behavior during the study. The three indicators covered different time scales and different stages of the decision-making process. This design was inspired by the MGB, a behavioral theory that is a theoretical model, in the hopes of developing indicators that were as correlated as possible with the future outcome. Nonetheless, the model cannot determine whether there is a causal relationship between the events during the study, the indicator variables, and the choice. The assumptions stated in section 3.3 are necessary to interpret the correlation observed in the data as a causal relationship, but any interpretation of the results should be done with this caveat in mind.

The uncertainty regarding whether the observed correlation truly represents causality could have been reduced by a control group. Even though a control group was not available, the assumption that other, external influences on the choice (aside from experiences with service quality) could largely be ignored still appears reasonable for several reasons: The study covered less than two months, and it was carried out in San Francisco, where winter weather is relatively stable and temperate. We are not aware of any notable events during the study that could have impacted transit use. Participants were asked about major lifestyle changes or moving plans, and those factors were controlled for. Finally, as will be shown in the results, the estimated coefficients were statistically significant.

The model estimation was done via maximum likelihood procedure. If the model were misspecified, i.e., if it omitted variables relevant to the choice or if the causal relationships were incorrect, this would lead to biases in the coefficient estimates, though it is not possible to quantify any potential biases, as this would require knowledge of the true model (Bollen et al., 2007). In a latent variable model, structural misspecification in one part of the model can cause biased coefficient estimates in correctly specified parts of the model as well. Furthermore, a misspecification would also impact the efficiency of the estimation technique and the accuracy of the hypothesis tests.

Table 2. Cross-tabulation of two questions regarding cessation of transit use.

		In January, I would like to use transit...			
		less	the same	more	Sum
"As soon as my circumstances permit, I want to use transit less"	Agree	96	81	13	190
	Neutral	42	80	32	154
	Disagree	49	210	84	343
	Sum	187	371	129	687

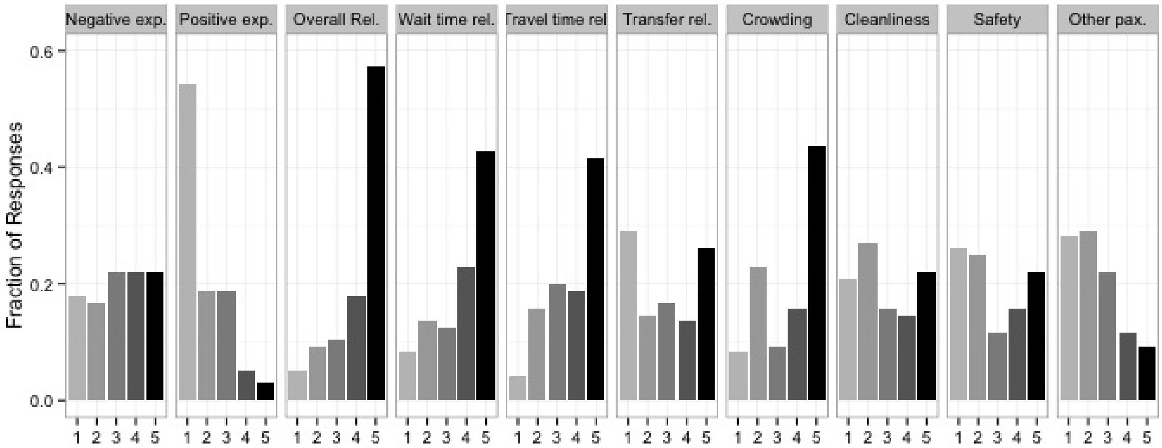
A second limitation is that the coefficients of the structural model relating the daily satisfaction to the exit satisfaction needed to be constrained to be equal, due to estimation difficulties if they were unconstrained. As a consequence, all daily satisfaction ratings were weighted equally. One possible cause is that participants' satisfaction was not measured at the same times with respect to the time of the exit survey, as they were only required to give five responses over the course of the study. By specifying separate coefficients for each of the past satisfaction ratings, it was hoped that a time dependency would be observable, for example, that the most recent experience would have a stronger influence than more distant experiences. The failure to observe such a time dependency in this model does not mean it is not present, but is more likely due to the data limitations. In future research, satisfaction should be measured at the same time intervals between the daily mobile surveys and the exit surveys for all participants, and the model should be re-estimated with those data to discover potential time dependencies. In addition, if it is possible to sample satisfaction for all participants on consecutive days, the data would allow the observation of serial dependency between measurements of different days.

4 Results

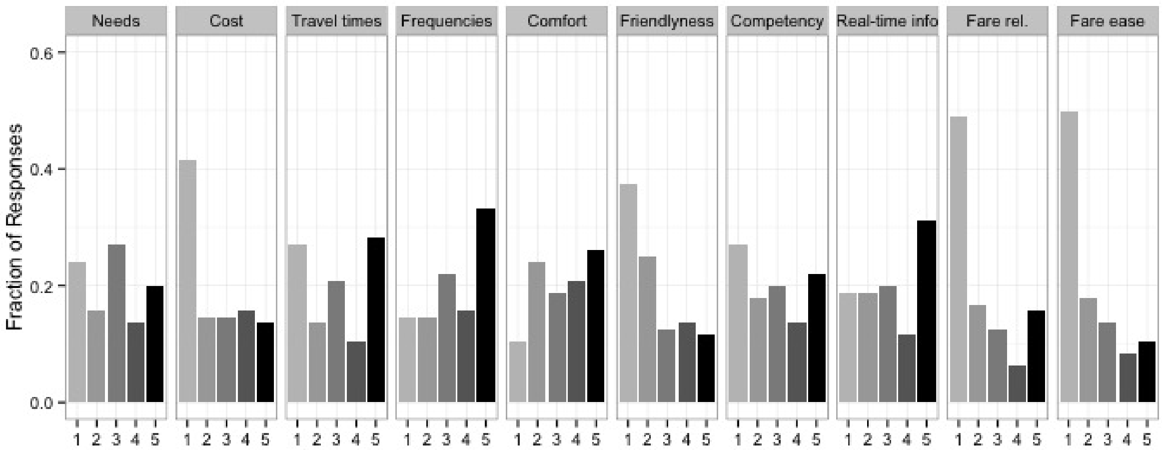
4.1 Descriptive analysis

Table 2 shows a cross-tabulation of the responses to two questions: "Would you prefer to use Muni less/the same/more in January 2014" and "As soon as my circumstances permit, I would like to use public transportation less" (responses to the latter were levels of agreement). Out of the 687 participants who filled out the exit survey, 187 stated that they would prefer to use public transportation less in January, and they were subsequently asked for the reasons for that statement. Out of those 187, 96 also agreed with the statement that they would like to use public transportation less as soon as their circumstances permitted. Figure 4 shows the stated reasons on a scale from "not at all influential" (1) to "very influential" (5). It can be seen that only 17 out of the 96 participants stated that negative experiences during the study did not influence their desire to reduce their use of transit, and 16 said such experiences were slightly influential. The remaining 63 were split evenly between "somewhat influential", "moderately influential" and "very influential". Among the specific reasons mentioned, the most important ones were overall unreliability, crowding levels, wait time unreliability and unreliability of in-vehicle travel times. Unreliability of transfer times was mentioned less frequently, but that was in part due to the fact that not all participants transferred. Unreliability and crowding levels are of course linked due to bus bunching (Daganzo and Pilachowski, 2011). It can also be seen that out of the other environmental variables, cleanliness, safety, comfort, the friendliness and competence of staff and the pleasantness of other passengers were reported to be much less influential than crowding. Travel times and service frequencies when there are no delays were asked about separately, and as can be seen, were reported to be less influential by participants than travel time reliability variables. Lastly, the least influential variables were related to the cost of travel and the fare payment system.

For participants who responded to the question "Compared to how much you used Muni during the study, you anticipate using Muni in January..." either by saying that they were going to increase or decrease their use of Muni, a follow-up question was asked regarding their anticipated mode shifts. Participants who said they anticipated decreasing their Muni use were asked what modes they were going to shift their trips to, and participants who said they anticipated increasing their



(a) In order from left to right: Negative experiences during the study, positive experiences during the study, overall reliability, wait time reliability, travel time reliability, transfer time reliability, crowding, cleanliness, safety, pleasantness of other passengers.



(b) In order from left to right: Ability of Muni to meet daily travel needs, cost, on-board travel times when there are no delays, frequencies of service, comfort, friendliness of staff, competence of staff, accuracy of real-time information, reliability of fare payment system, ease of use of fare payment system.

Figure 4. Stated reasons for wanting to use transit less or not at all anymore

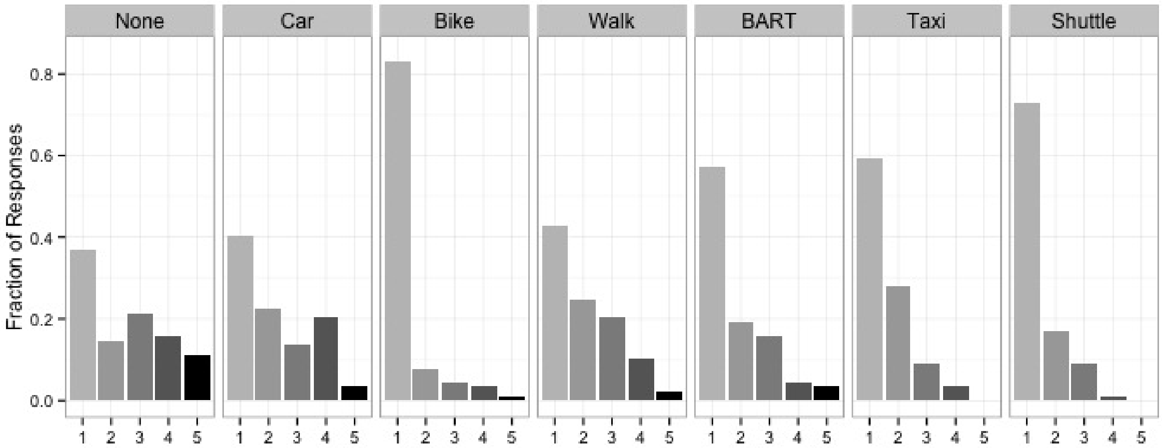


Figure 5. Decrease in Muni use: Travel modes substituted for Muni trips. (1 - applies to none of the trips, 5 - applies to all trips.)

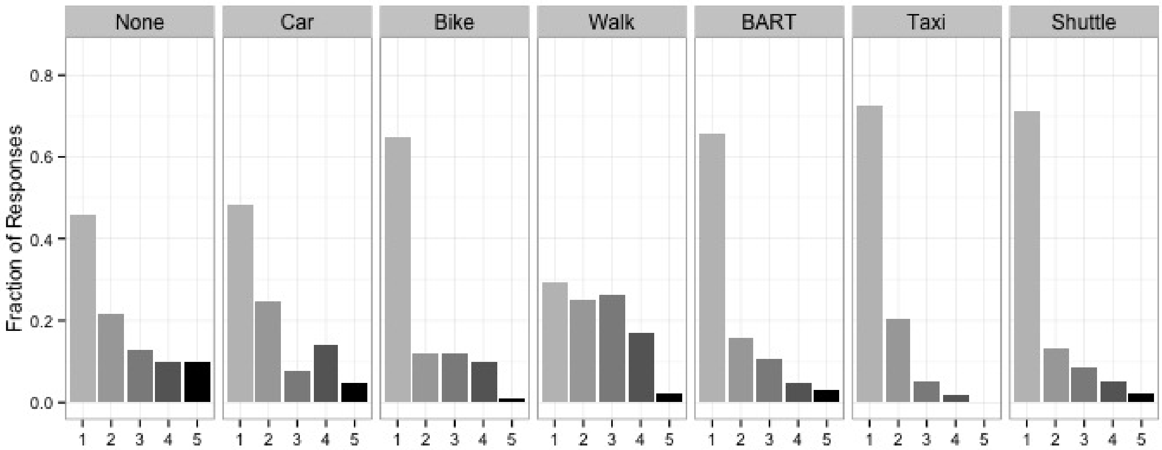


Figure 6. Increase in Muni use: Travel modes substituted with Muni trips. (1 - applies to none of the trips, 5 - applies to all trips.)

Muni use where asked what modes they were shifting their trips from. The results are shown in Figures 5 and 6. "None" refers to trips that the participant did not make before or would cease making when the shift to or from Muni occurred. It can be seen that increased Muni use drew mostly from walk trips, followed by trips that were not made before and auto trips. On the other hand, people who decreased their Muni use primarily either ceased making those trips, began using the automobile or walking.

4.2 *Modeling results*

The parameters of the structural model are explained in Table 1. The estimation results for the structural model are found in Table 3, and the estimation results for the measurement models are in Table 5 in the appendix. Based on a mix of theoretical considerations regarding model identifiability (Bollen, 2014; Ben-Akiva et al., 2002) and empirical work with the data, it was determined that several of the coefficients needed to be constrained in order to ensure identification. Generally speaking, the constrained value can be freely chosen by the modeler, but a value of 1 is a common approach to support the interpretability of the remaining coefficients. This concerned two of the error terms (η_{daily}, η_{entry}). Furthermore, one of the scale parameters of the choice model and one of the coefficients of each measurement model for latent satisfaction in the exit survey had to be constrained to set the scale for the remaining coefficients. $\mu_{longdes}$ and two of the λ coefficients were chosen for this and set to 1. These coefficients are marked by asterisks in Tables 3 and 5. While the estimation was conducted, it was observed that the likelihood function along the dimension of the ASC for short-term intention appeared to be very flat, which caused erratic coefficient estimates for that ASC. Due to this empirical identification issue, the ASC for short-term intention was constrained to 1 as well. As in all cases where coefficients were constrained, we verified that the constraint did not impact the fit of the model, i.e., that the final value of the likelihood function remained approximately the same when the constraint was introduced. This ensured that the normalizations stabilized the parameter estimates but did not impact the fit and interpretation of the model.

Table 3. Estimation results for the structural equations of the latent variable choice model.

Coefficient	Value	Robust Std. Error	Robust t-stat	p-value
η_{entry}	1.00	*	*	*
α_{age}	0.01	0.00	1.64	0.10
α_{income}	-0.01	0.01	-0.80	0.42
$\alpha_{unknownincome}$	-0.02	0.65	-0.04	0.97
$\alpha_{longuser}$	-0.25	0.17	-1.47	0.14
$\alpha_{entrysat}$	1.02	0.16	6.54	0.00
α_{mood}	0.38	0.08	5.02	0.00
α_{ivtt}	0.00	0.01	0.18	0.85
α_{early}	0.01	0.06	0.24	0.81
α_{delay}	-0.20	0.04	-4.88	0.00
$\alpha_{waitover5}$	-0.02	0.03	-0.74	0.46
$\alpha_{waitunder5}$	0.01	0.03	0.55	0.58
α_{nowait}	-0.28	0.09	-3.04	0.00
$\alpha_{transfertime}$	-0.05	0.01	-3.28	0.00
$\alpha_{notransfer}$	-0.21	0.12	-1.83	0.07
$\alpha_{leftbehind}$	-0.12	0.10	-1.20	0.23
$\alpha_{unobservedtransfer}$	-0.51	0.19	-2.70	0.01
η_{daily}	1.00	*	*	*
$\beta_{dailysat}$	0.10	0.01	8.54	0.00
$\beta_{leftbehind\_1\_9}$	-0.10	0.06	-1.85	0.06
$\beta_{leftbehind\_10}$	-0.29	0.12	-2.45	0.01
$\beta_{latework}$	-0.06	0.01	-4.48	0.00
$\beta_{lateleisure}$	-0.04	0.02	-2.75	0.01
$\beta_{notravel}$	-0.07	0.03	-2.59	0.01
$\eta_{exitops}$	0.42	0.07	6.51	0.00
$\eta_{errorcorr}$	0.16	0.18	0.88	0.38
$\eta_{exitenv}$	0.57	0.05	12.56	0.00
$ASC_{shortdes}$	1.07	0.12	9.06	0.00
$ASC_{longdes}$	0.51	0.14	3.68	0.00
$ASC_{shortint}$	1.00	*	*	*
γ_{env}	0.23	0.10	2.23	0.03
γ_{ops}	0.39	0.10	3.89	0.00
$\mu_{shortint}$	1.78	0.18	9.95	0.00
$\mu_{shortdes}$	2.33	0.69	3.38	0.00
$\mu_{longdes}$	1.00	*	*	*

In what follows, the model estimation results are presented and discussed. For all variables that influence satisfaction, positive coefficient estimates mean that an increase in the respective variable leads to increased satisfaction, whereas negative coefficient estimates mean that an increase in the variable leads to decreased satisfaction. When comparing these results to those in Carrel et al. (2016), one should be cognizant of the fact that the assumptions underlying this model are different from those underlying the models in Carrel et al. (2016). In the latter, the link between the individual travel time components and satisfaction with those components were modeled. No assumptions were made about the relationship between the models, and it was assumed that the reported satisfaction ratings were true measures of the participant's satisfaction. On the other hand, the model results presented here assume that there is an underlying, latent satisfaction with operations (and thus, travel times). The individual satisfactions with travel time components serve as indicators of that latent underlying satisfaction.

4.2.1 Effects of sociodemographics, baseline satisfaction and mood

First, we investigate the effect of sociodemographic attributes on people's reported daily satisfaction. It can be seen in Table 3 that age has a positive effect (0.01 per year of age) and is significant, whereas having been a long-time user ($\alpha_{longuser}$) has a negative effect (-0.25). The latter, however, is not significant. The effect of income on reported daily satisfaction is also negative (-0.01 per $10,000), as is the non-response variable to income. The results with respect to income are intuitive, as higher income is associated with a higher value of time. Both effects, however, are not significant at $p = 0.42$ and $p = 0.97$, respectively. The effect of the entry satisfaction ($\alpha_{entrysat} = 1.02$), which again is a latent variable, and of the participant's general mood on the day of the survey ($\alpha_{mood} = 0.38$) are positive and significant at $p < 0.01$. In other words, these two variables have a markedly stronger effect on satisfaction than age, income and the length of Muni use. Of course we would expect the entry satisfaction to depend on sociodemographic variables as well, but it was not possible to include the mood and entry satisfaction in both the daily and the entry survey as this caused estimation problems. These results are in line with findings from Carrel et al. (2016), though the difference between the age and income variables and the mood and entry satisfaction variables is more pronounced in the latent variable model.

4.2.2 Travel time variables

The scheduled in-vehicle travel time (IVTT) is found to have virtually no effect on overall satisfaction with operations (α_{ivtt}). Delays with respect to the scheduled IVTT have a significant negative effect ($\alpha_{delay} = -0.20$ per minute, $p < 0.01$) and earlier arrivals have an insignificant effect on satisfaction ($\alpha_{delay} = 0.01$ per minute, $p = 0.81$). Unfortunately, this model was not able to capture the effect of wait times on a joint, latent satisfaction variable, as both coefficients for wait times below 5 minutes ($\alpha_{waitunder5}$) and wait times above 5 minutes ($\alpha_{waitover5}$) are insignificant. The reason for this merits further investigation, but as discussed in Carrel et al. (2016), it might be linked to the fact that the majority of wait times was short, with an average around 2 minutes. It must be noted that the observed wait times used in the model estimation only capture time actually spent standing at the origin stop. They exclude additional schedule delay times (i.e., the time between when a person wanted to leave and the time of a transit vehicle departure), which some participants may have chosen to spend elsewhere, such as in their homes. While these may have been perceived as wait times by some participants, it was not possible to automatically identify them with location tracking data from the phones, and therefore, the link between the latent satisfaction with travel times and the wait times could not be established in this case. As passengers rely more on real-time information, the strategy of spending wait time at locations other than the stop and of going to the stop only when an arrival is predicted will most likely become more prevalent.

Wait times could not be identified from the tracking data alone for approximately 50% of observations in the data set. The observations with missing wait times are denoted by a binary variable, the coefficient of which (α_{nowait}) is negative and significant. To the best of our knowledge, a missing wait time observation could be due to one or more of the following causes:

1. If there was insufficient location tracking data available for that portion of the trip. This includes both smartphone and vehicle location data.
2. If the tracking data showed the participant getting on a vehicle immediately, with no time spent at the stop.
3. if the participant was carrying out an activity near the stop (e.g. work, shopping) which made it impossible to distinguish activity time from wait time. This distinction was particularly in parts of San Francisco where the transit network is very dense.

4. If the wait time was incurred when the participant transferred from BART (regional rapid transit) to a local metro train inside an underground metro station.

It is not known whether the missing observations skewed the distribution of observed wait times in any particular way. The insignificance of the wait time coefficients suggests that the observed wait times at the stop were generally in a range that did not significantly affect riders' overall satisfaction with operations. Together with the low sensitivity toward wait times observed by Carrel et al. (2016), the results of this model suggest that in future research, a different approach should be taken to identifying wait times. There are three possible avenues:

- Adding other sensor data such as accelerometer in order to better identify when a person actually walked to a transit stop.
- Directly asking a participant about the perceived wait time and where it was spent.
- Tracking the use of real-time arrival information on the phone in order to determine when a participant first looked at upcoming departures. This could serve as an indicator of the beginning of a wait.

Unlike the coefficient of the wait time at the origin stop, the transfer time coefficient ($\alpha_{transfertime}$) is negative and significant at $p < 0.01$. A comparison of the transfer time coefficient and the IVTT coefficient shows that according to the model, one minute of in-vehicle delay causes as much dissatisfaction as four minutes of transfer time. The model further includes two binary variables related to transfer time: $\alpha_{notransfer}$ captures cases where no transfer was identified from the location tracking data and the participant did not report a transfer, and $\alpha_{unobservedtransfer}$ captures cases where the participant reported having transferred but the transfer could not be identified from the location tracking data. Both are negative and significant at $p < 0.10$. While this result is intuitive for the latter coefficient, it is not intuitive for the former, as it suggests that in general, passengers who transfer tend to report a higher satisfaction than passengers who do not transfer. This merits further investigation with a different data set.

4.2.3 *Effect on exit satisfaction and critical incidents*

The coefficient $\beta_{dailysat}$ in Table 3 links daily satisfaction with operations to the exit satisfaction with operations. As expected, it is positive (0.1), and it is also significant at $p < 0.01$, showing a positive correlation between daily satisfaction and exit satisfaction.

Interestingly, all five coefficients related to self-reported critical incidents have negative and significant estimates at $p < 0.1$. The first two are the number of times a person arrived late to work or school ($\beta_{latework}$) due to a transit delay and the number of times a person arrived late to a leisure activity due to a transit delay ($\beta_{lateleisure}$). Both were self-reported in the exit survey. $\beta_{notravel}$ captures cases where participants reported on their daily mobile surveys that they had wanted to use public transportation that day but could not due to a delay and were forced to choose a different mode. However, there was no obligation to report these incidents, so it must be assumed that the reported numbers are a lower bound. Therefore, the estimated coefficient is an upper bound on the impact of such incidents. A special case of critical incidents were denied boardings: These were captured both through self-reports in the exit survey and through automated detection. The latter affect the daily satisfaction with operations via $\alpha_{leftbehind}$. The automated detection was only based on location data: If a participant was observed to be at a stop and not board a passing vehicle but board the following one, it was recorded as a denied boarding. There is a risk of misclassification, as the participant may have been carrying out an activity and may not have intended to board the first vehicle. Therefore, the number of such automatically detected incidents in the data set is an upper bound, and the coefficient estimate is a lower bound for the impact. In Table 3, $\alpha_{leftbehind}$ is negative but not significant. On the other hand, the effect of self-reported denied boardings was found to exhibit some nonlinearity. There were 11 possible answers to the self-reported question: From 0 to 9, and then "10 or more". The model was found to produce the best fit if these two categories were separated, as in Table 3. Both coefficients, $\beta_{leftbehind\_1\_9}$ and $\beta_{leftbehind\_10}$, are negative and significant at $p < 0.10$, but the latter is approximately three times larger than the former. It is possible (and plausible) that the larger coefficient for 10 or more denied boardings is capturing protest responses.

4.2.4 *Effects on future behavior*

The final component is the choice model. It is specified with only two inputs: The participant's exit satisfaction with operations and the participant's exit satisfaction with the travel environment. All other variables in the model, including the daily satisfaction with operations, the critical incidents, the travel time experiences and the entry satisfaction with operations, affect future behavior through the exit satisfaction with operations. The indicator variables for the satisfaction in the exit survey are on a four-point Likert scale, whereas the choice indicators are binary. Given that the indicator variables had different reference points, one has to make the assumption that the participants did not significantly change their pre-study frequency of transit use during the study in order to interpret the results. Given that assumption, the choice is between (a) continuing to use public transportation at the same frequency as before and during the study or using it more frequently, and (b) using public transportation less frequently or discontinuing it altogether. The choice indicator for the former is 1, and for the latter it is 0. Thus, positive coefficients mean there is a positive correlation between the input variable and the participant's willingness to use public transportation the same or more in the future. As can be seen in Table 3, the effects of both satisfaction with operations and satisfaction with the travel environment in the exit survey are positive (0.39 and 0.23, respectively) and significant at $p < 0.05$. This is intuitive, as higher satisfaction leads to a higher willingness to continue using transit in the future.

Of particular interest here is the relative difference between the two coefficients. The coefficient of the latent satisfaction with operations is approximately 1.7 times the coefficient of the latent satisfaction with the travel environment. Since these are latent variables, their exact values cannot be calculated, but the comparison can be made with the help of two of the indicator variables: A change in the latent satisfaction with operations variable that causes a one-point increase in satisfaction with overall reliability has 1.46 times the effect of a change in the satisfaction with travel environment variable that causes a one-point increase in satisfaction with crowding. This confirms that for the present group, overall satisfaction with operations has a stronger influence on future mode choice decisions than overall satisfaction with the travel environment. This is consistent with the results of the descriptive analysis in section 4.1. In future research, it would be interesting to add objective measurements to the travel environment variables. Crowding would be of particular interest, given the importance reported by participants in section 4.1; this could be calculated with data from automatic passenger counting or fare payment systems.

With the help of the final choice model, it is now possible to calculate the relative influence of various experiences on passengers' willingness to remain transit riders in the future. As a calculation example, the effect on the choice utility of one incident of not being able to travel due to a delay on the transit network is $\beta_{notravel} \cdot \gamma_{ops} = -0.07 * 0.39 = -0.027$.

4.2.5 *Calculation of trade-offs*

If the coefficient for wait times were negative and significant, it would be possible to calculate the impact of negative critical incidents in terms of the equivalent amount of dissatisfaction caused by wait times. For instance, if $\alpha_{waitover5}$ and $\alpha_{waitunder5}$ were -0.02 and significant and $\alpha_{leftbehind}$ were -0.12 and significant, one could state that the dissatisfaction caused by one instance of a denied boarding would be equivalent to the dissatisfaction caused by approximately 6 minutes of wait time at the origin stop. Such trade-offs provide good rules of thumb for transit planning professionals, as is illustrated by the popularity of the rule of thumb that a minute of out-of-vehicle travel time is twice as onerous as a minute of in-vehicle travel time (Wardman, 2004). Therefore, a goal of future research should be to derive significant wait time coefficients in order to calculate such trade-offs.

Table 4. Results of the simulation.

	10 min delay on board	10 min transfer, late to work	Left behind, 10 min wait
Change in probability per person	-0.006	-0.003	-0.002
Potential systemwide change	-1628	-745	-504
% of SFMTA yearly turnover	5.4	2.5	1.7

4.3 Simulation

To illustrate a potential use of this type of model, a simulation was conducted. Three hypothetical scenarios were simulated, and the group of subjects from whom the data had been collected served as a convenience sample. Three hypothetical simulation scenarios were designed:

1. Impact of every participant experiencing one additional ten-minute on-board delay.
2. Impact of every participant having one additional experience of being left behind at a stop with corresponding ten minutes of additional wait time.
3. Impact of every participant experiencing one additional 10-minute transfer wait time and arriving late at work.

The scenarios measured the impact of one additional event since the baseline was the choice probability calculated from the set of experiences that the participants had had during the study. The output of each simulation run was an average probability of remaining a transit rider in the future, which was contrasted with the baseline probability. First, we discuss the assumptions of the simulation scenarios in the context of the San Francisco transit system. The first simulation scenario is on the high end of typical in-vehicle delay times on Muni. The maximum delay observed in the data set was 9:16 minutes. Therefore, this simulation scenario captures the effect of a major, system-wide disruption. The second simulation scenario is also on the pessimistic side; out of 449 participants, 90 (20%) were observed to have been left behind at a stop at least once during the study. This scenario assumes that it happened to each participant one additional time. The third simulation scenario is more typical of day-to-day operations on Muni. The average transfer time experienced by participants who transferred was 7:21 minutes, and on average, participants reported being late to work or school due to difficulties on transit 2.43 times during the study.

Table 4 shows the simulation results. In the base case, on average, participants had a 0.77 probability of remaining transit riders in the future. The first data row shows the change in probability due to the simulated incident. In the second row, the change in probability was extrapolated to the SFMTA's entire ridership of 280,000, and the potential loss of riders due to the simulated event was calculated. To put the numbers into context, the third row shows what percentage of the SFMTA's yearly turnover (approximately 30,000 passengers) the simulated losses would represent. Since the scenarios are limited to single events that differ in the likelihood of occurrence and severity, this is not intended to model actual events observed on the SFMTA's network. Rather, this is intended to demonstrate how this type of model can help analysts understand the potential impact of various operating strategies and capital investment programs on ridership turnover. To analyze a proposed operational or infrastructure change, two pieces of input are required: First, the analyst needs to know the anticipated frequencies of delays and critical incidents before and after the changes. Second, an approximate knowledge of the rider population that will be impacted by the change can help in constructing the sample used for the simulation.

5 Discussion of results

The results show the link between service quality problems and loss of ridership from two different angles. First, in section 4.1, we selected participants who reported a behavioral desire or intention to reduce their use of public transportation, and investigated the self-reported reasons for which they wanted to do so. It was seen that travel time reliability was mentioned as the overall most important factor. In terms of the number of "very influential" responses, crowding came in second, but in terms of the average of all responses, the second-most important factors were wait time reliability and travel time reliability. Both had an average response of 3.78, compared to 4.13 for overall reliability and 3.64 for crowding. In a broader sense, even crowding can be considered a reliability variable since the crowding of vehicles is related to vehicle bunching, and passengers do not know ahead of time whether they will be able to find a seat (Polydoropoulou and Ben-Akiva, 2001). Overall, travel time variables were more influential than travel environment variables. However, it should be noted that this study concerned users who were already regular transit users, and therefore, it may be a self-selected group that might be less concerned with the travel environment than, for example, a comparison group of auto users.

Second, in section 4.2, we presented model estimation results and applied them to a simulation in section 4.3. The model results are generally in line with previous findings presented in Carrel et al. (2016), showing that neither the scheduled in-vehicle travel time nor early arrivals at the destination have a significant effect on satisfaction, but that in-vehicle delays are an important driver of dissatisfaction. By extension, it is shown that in-vehicle delays also have a strong impact on overall satisfaction after an extended time period and on passengers' desire to stop using public transportation. The transfer time coefficient was also negative and significant, but the origin wait time coefficients were not significant. It is assumed that this may be partly related to the difficulties associated with properly identifying origin wait times, as explained in section 4.2, and partly to the fact that participants appeared to be choosing to spend their wait times at locations other than the stop and rely on real-time information to go to the stop when an arrival was predicted. Therefore, the wait times detected from location data may not necessarily have corresponded to the wait times as defined by the user. It is also interesting that the coefficient of the binary variable for missing wait time observations was estimated to have a negative and significant impact on satisfaction. Unfortunately, since it is unknown what reasons led to a missing wait time measurement in any specific case, this result is difficult to interpret. Based on the reasons for missing wait time data discussed in section 4.2, the following may be occurring:

- It is possible that the wait times that were not observed due to missing or insufficient data were on average significantly longer than the observed wait times. However, we are currently not aware of any plausible reasons why this may be the case.
- It is possible that if there was a large time gap between participants' desired departure times and the next departures, and participants chose to spend that time carrying out other activities nearby, they still perceived it as wait time, leading to lower satisfaction.

There may be other explanations which we were not aware of at the time of writing. In future research, these data shortcomings should be addressed in order to solidify our understanding of the impact of various delay times on passengers. The model presented in this paper goes beyond previous satisfaction models by explicitly linking critical incidents and personal experiences with travel times to future behavioral intentions by way of customer satisfaction, and the significance of the relevant coefficients in Table 3 demonstrates that this link is present. It is shown that for the present group of participants, which consisted mostly of regular transit users, satisfaction with travel times and operational aspects is more important in determining their willingness to remain transit riders than satisfaction with the travel environment. Furthermore, even though the influence of wait times relative to in-vehicle delay times and transfer times requires further research, the results clearly demonstrate the value of developing models using participants' *personal experiences* with service quality as a means of understanding future mode choice intentions and the influence of various factors related to service quality. Besides travel times, we find that several types of critical incidents have measurable negative effects on participants' overall satisfaction in the exit survey and thus on their willingness to remain transit riders.

The value of the latent variable modeling framework used in this paper was that it permitted us to summarize an individual's overall satisfaction with operations in one variable and to determine the influence of a variety of experiences with travel times on that variable. It is flexible, and its specification can accommodate variables collected on different time scales, such as the daily satisfaction with operations and the entry and exit satisfaction. Most importantly, it allowed us to account for correlation between a participant's satisfaction ratings with respect to different travel time components.

6 Conclusions

In this paper, we presented an analysis and model results to understand the link between service quality, satisfaction, and transit ridership loss. This work emphasizes the importance of riders' personal experiences; an innovative procedure was used to map location data from users' mobile phones to vehicle location data in order to automatically identify personal experiences and use them in the estimation of the model. This demonstrates the value and potential of such new data collection methods in answering complex questions and observing phenomena that require panel data. The insights gained from these data help establish the link between travel time variability/critical incidents, satisfaction, and transit ridership loss. For the first time, this framework makes it possible to directly model the effect of negative personal experiences on future mode choice decisions and thus on ridership loss due to delays and system management strategies. In the future, it can be further refined with additional data in order to form the basis for new operational tools that would enable a move from system-based to person-based performance metrics for transit agencies.

References

Abou-Zeid, Maya and Moshe Ben-Akiva (2010). A model of travel happiness and mode switching. In: *Choice modelling: The state-of-the-art and the state-of-practice (proceedings from the inaugural international choice modelling conference)*, pp. 289–306.

Abou-Zeid, Maya et al. (2012). Happiness and travel mode switching: Findings from a Swiss public transportation experiment. In: *Transport Policy* 19.1, pp. 93–104.

Ajzen, Icek (1991). The theory of planned behavior. In: *Organizational Behavior and Human Decision Processes* 50.2, pp. 179–211.

Anable, Jillian and Birgitta Gatersleben (2005). All work and no play? The role of instrumental and affective factors in work and leisure journeys by different travel modes. In: *Transportation Research Part A: Policy and Practice* 39.2, pp. 163–181.

Anderson, Eugene and Mary Sullivan (1993). The antecedents and consequences of customer satisfaction for firms. In: *Marketing Science* 12.2, pp. 125–143.

Bamberg, Sebastian, Icek Ajzen, and Peter Schmidt (2003). Choice of travel mode in the theory of planned behavior: The roles of past behavior, habit, and reasoned action. In: *Basic and Applied Social Psychology* 25.3, pp. 175–187.

Bates, John et al. (2001). The valuation of reliability for personal travel. In: *Transportation Research Part E: Logistics and Transportation Review* 37.2, pp. 191–229.

Ben-Akiva, Moshe et al. (2002). Integration of choice and latent variable models. In: *Perpetual motion: Travel behaviour research opportunities and application challenges*. Elsevier Science, pp. 431–470.

Bollen, Kenneth (2014). *Structural equations with latent variables*. John Wiley & Sons.

Bollen, Kenneth et al. (2007). Latent variable models under misspecification: Two-stage least squares (2SLS) and maximum likelihood (ML) estimators. In: *Sociological Methods & Research* 36.1, pp. 48–86.

Carrel, Andre, Anne Halvorsen, and Joan Walker (2013). Passengers' perception of and behavioral adaptation to unreliability in public transportation. In: *Transportation Research Record* 2351, pp. 153–162.

Carrel, Andre, Raja Sengupta, and Joan Walker (2016). The San Francisco Travel Quality Study: Tracking the trials and tribulations of a transit taker. In: *Transportation. Advance Online Publication.*

Carrel, Andre et al. (2015). Quantifying transit travel experiences from the users' perspective with high-resolution smartphone and vehicle location data: Methodologies, validation, and example analyses. In: *Transportation Research Part C: Emerging Technologies* 58, pp. 224–239.

Carrel, Andre et al. (2016). In pursuit of the happy transit rider: Dissecting satisfaction using daily surveys and tracking data. In: *Journal of Intelligent Transportation Systems* 20 (4), pp. 345–362.

Daganzo, Carlos and Josh Pilachowski (2011). Reducing bunching with bus-to-bus cooperation. In: *Transportation Research Part B: Methodological* 45.1, pp. 267–277.

Davis, Mark and Janelle Heineke (1998). How disconfirmation, perception and actual waiting times impact customer satisfaction. In: *International Journal of Service Industry Management* 9.1, pp. 64–73.

Fredrickson, Barbara and Daniel Kahneman (1993). Duration neglect in retrospective evaluations of affective episodes. In: *Journal of Personality and Social Psychology* 65.1, p. 45.

Friman, Margareta, Bo Edvardsson, and Tommy Gärling (2001). Frequency of negative critical incidents and satisfaction with public transport services. In: *Journal of Retailing and Consumer Services* 8.2, pp. 95–104.

Friman, Margareta and Markus Fellesson (2009). Service supply and customer satisfaction in public transportation: The quality paradox. In: *Journal of Public Transportation* 12.4, pp. 57–69.

Friman, Margareta et al. (2013). Psychometric analysis of the satisfaction with travel scale. In: *Transportation Research Part A: Policy and Practice* 48, pp. 132–145.

Kahneman, Daniel et al. (1993). When more pain is preferred to less: Adding a better end. In: *Psychological Science* 4.6, pp. 401–405.

Kahneman, Daniel et al. (2004). A survey method for characterizing daily life experience: The day reconstruction method. In: *Science* 306.5702, pp. 1776–1780.

Lanken, Bas et al. (1994). Attitude versus general habit: Antecedents of travel mode choice. In: *Journal of Applied Social Psychology* 24.4, pp. 285–300.

Ma, Xiaolei et al. (2013). Mining smart card data for transit riders' travel patterns. In: *Transportation Research Part C: Emerging Technologies* 36, pp. 1–12.

Mittal, Vikas, Pankaj Kumar, and Michael Tsiros (1999). Attribute-level performance, satisfaction, and behavioral intentions over time: A consumption-system approach. In: *The Journal of Marketing* 63.2, pp. 88–101.

Pedersen, Tore, Margareta Friman, and Per Kristensson (2011). The role of predicted, on-line experienced and remembered satisfaction in current choice to use public transport services. In: *Journal of Retailing and Consumer Services* 18.5, pp. 471–475.

Perk, Victoria, Jennifer Flynn, and Joel Volinski (2008). *Transit ridership, reliability and retention*. Tech. rep. FDOT BD 549-32. Tampa, FL: National Center For Transit Research.

Perugini, Marco and Richard Bagozzi (2001). The role of desires and anticipated emotions in goal-directed behaviours: Broadening and deepening the theory of planned behaviour. In: *British Journal of Social Psychology* 40.1, pp. 79–98.

– (2004). The distinction between desires and intentions. In: *European Journal of Social Psychology* 34.1, pp. 69–84.

Polydoropoulou, Amalia and Moshe Ben-Akiva (2001). Combined revealed and stated preference nested logit access and mode choice model for multiple mass transit technologies. In: *Transportation Research Record* 1771, pp. 38–45.

Trépanier, Martin, Khandker Habib, and Catherine Morency (2012). Are transit users loyal? Revelations from a hazard model based on smart card data. In: *Canadian Journal of Civil Engineering* 39.6, pp. 610–618.

Vij, Akshay, Andre Carrel, and Joan Walker (2013). Incorporating the influence of latent modal preferences on travel mode choice behavior. In: *Transportation Research Part A: Policy and Practice* 54, pp. 164–178.

Walker, Joan (2001). Extended discrete choice models: Integrated framework, flexible error structures, and latent variables. Massachusetts Institute of Technology.

Wardman, Mark (2004). Public transport values of time. In: *Transport Policy* 11.4, pp. 363–377.

Woodruff, Robert, Ernest Cadotte, and Roger Jenkins (1983). Modeling consumer satisfaction processes using experience-based norms. In: *Journal of Marketing Research* 20.3, pp. 296–304.

Appendix A.

Table 5. Estimation results for the measurement equations of the latent variable choice model. The abbreviations are explained at the end of the table.

	Latent variable	Indicator variable	Value	Robust Std err	Robust t-test	p-value
δ	Entry sat. ops.	Overall reliability	2.90	0.06	47.39	0
δ	Entry sat. ops.	IVTT	2.87	0.05	52.22	0
δ	Entry sat. ops.	Wait time	3.08	0.06	55.68	0
δ	Daily sat. ops.	Overall reliability	2.34	0.12	19.93	0
δ	Daily sat. ops.	IVTT	2.31	0.10	22.21	0
δ	Daily sat. ops.	Wait time	2.24	0.11	19.87	0
δ	Daily sat. ops.	Transfer time	2.22	0.12	19.33	0
δ	Rem. daily sat. ops.	Overall reliability	2.44	0.09	27.87	0
δ	Rem. daily sat. ops.	IVTT	2.23	0.09	24.01	0
δ	Rem. daily sat. ops.	Wait time	2.30	0.08	28.44	0
δ	Rem. daily sat. ops.	Transfer time	2.06	0.11	18.26	0
δ	Exit sat. ops.	Overall reliability	1.87	0.11	17.56	0
δ	Exit sat. ops.	IVTT	2.17	0.08	27.64	0
δ	Exit sat. ops.	Wait time	1.99	0.09	22.83	0
δ	Exit sat. env.	Crowding	1.59	0.04	41.79	0
δ	Exit sat. env.	Safety	2.29	0.03	69.36	0
δ	Exit sat. env.	Other Pax	2.15	0.03	69.20	0
δ	Exit sat. env.	Cleanliness	1.91	0.04	48.37	0
λ	Entry sat. ops.	Overall reliability	0.74	0.06	11.99	0
λ	Entry sat. ops.	IVTT	0.63	0.06	10.70	0
λ	Entry sat. ops.	Wait time	0.64	0.05	13.86	0
λ	Daily sat. ops.	Overall reliability	0.57	0.04	15.85	0
λ	Daily sat. ops.	IVTT	0.51	0.03	18.87	0
λ	Daily sat. ops.	Wait time	0.52	0.03	17.21	0
λ	Daily sat. ops.	Transfer time	0.53	0.04	15.05	0
λ	Rem. daily sat. ops.	Overall reliability	0.37	0.05	7.42	0
λ	Rem. daily sat. ops.	IVTT	0.41	0.04	9.65	0
λ	Rem. daily sat. ops.	Wait time	0.35	0.05	7.07	0
λ	Rem. daily sat. ops.	Transfer time	0.25	0.08	3.34	0
λ	Exit sat. ops.	Overall reliability	1.00	*	*	*
λ	Exit sat. ops.	IVTT	0.73	0.05	14.02	0
λ	Exit sat. ops.	Wait time	0.82	0.03	24.34	0
λ	Exit sat. env.	Crowding	1.00	*	*	
λ	Exit sat. env.	Safety	0.94	0.11	8.94	0
λ	Exit sat. env.	Other Pax	0.81	0.12	6.96	0
λ	Exit sat. env.	Cleanliness	1.12	0.24	4.61	0
σ	Entry sat. ops.	Overall reliability	0.74	0.03	24.55	0
σ	Entry sat. ops.	IVTT	0.75	0.02	30.87	0
σ	Entry sat. ops.	Wait time	0.78	0.02	41.99	0
σ	Daily sat. ops.	Overall reliability	0.76	0.03	29.10	0
σ	Daily sat. ops.	IVTT	0.88	0.02	47.85	0
σ	Daily sat. ops.	Wait time	0.99	0.02	66.03	0
σ	Daily sat. ops.	Transfer time	0.94	0.04	25.35	0
σ	Rem. daily sat. ops.	Overall reliability	0.58	0.03	18.01	0
σ	Rem. daily sat. ops.	IVTT	0.63	0.03	23.57	0
σ	Rem. daily sat. ops.	Wait time	0.73	0.03	21.25	0
σ	Rem. daily sat. ops.	Transfer time	0.88	0.09	9.81	0
σ	Exit sat. ops.	Overall reliability	0.75	0.05	16.05	0
σ	Exit sat. ops.	IVTT	0.76	0.02	35.25	0
σ	Exit sat. ops.	Wait time	0.72	0.03	21.66	0
σ	Exit sat. env.	Crowding	0.81	0.03	27.19	0
σ	Exit sat. env.	Safety	0.56	0.02	30.34	0
σ	Exit sat. env.	Other Pax	0.58	0.03	21.43	0
σ	Exit sat. env.	Cleanliness	0.68	0.08	8.54	0

Continued on next page

Latent variable	Indicator variable	Value	Robust Std err	Robust t-test	p-value

Legend:

Entry sat. ops.:	Satisfaction with operations in entry survey.
Exit sat. ops.:	Satisfaction with operations in exit survey.
Exit sat. env.:	Satisfaction with travel environment in exit survey.
Daily sat. ops.:	Satisfaction with operations in daily surveys d_1 through d_4.
Rem. daily sat. ops.:	Satisfaction with operations in daily survey d_r.