The impacts of varying survey design on reported trip satisfaction

Emily Grisé
School of Urban Planning, McGill University
E-mail: emily.grise@mail.mcgill.ca

Boer Cui
Department of Civil Engineering, McGill University
E-mail: boer.cui@mail.mcgill.ca

Anders Turim
Department of Geography, McGill University
E-mail: anders.turim@mail.mcgill.ca

Kevin Manaugh
Department of Geography and School of Environment, McGill University
E-mail: kevin.manaugh@mail.mcgill.ca

Ahmed El-Geneidy
School of Urban Planning, McGill University
E-mail: ahmed.elgeneidy@mcgill.ca

Word count: (6161 words + 5 tables) = 7411 words

December 2nd, 2018

ABSTRACT

Research has shown that the way in which a question is asked or the order in which it is received can influence the answer of respondents. Surveys are frequently used by transport professionals to better understand the psychology behind travelers' level of satisfaction and derive policies. An awareness of how survey design impacts individuals' responses to satisfaction-related questions is vital for any future policy analysis and development. This study uses travel data collected through a university travel survey, which was carried out during the fall and winter of 2017/2018, to examine the effects of seasonality, question wording, and question order on travel satisfaction. To test for question wording and seasonality, respondents were asked to rate their satisfaction with their last trip and satisfaction with their typical commute to work or school under either "warm and sunny" or "cold and snowy" conditions, depending on season of survey completion. To test the effect of question order, respondents were randomly assigned one of two survey versions in which the order of the two aforementioned trip satisfaction questions were switched. Our results confirm the presence of question order and wording effects on reported satisfaction levels, and to a lesser degree, differences in trip satisfaction due to seasonality. Results of this study highlight the need to practice caution when designing travel surveys, particularly with respect to how transport survey respondents are impacted by seasonality, question wording and order.

INTRODUCTION

Surveys are one of the most heavily utilized tools within the transport industry to gain insight into the psychology and behavior of travelers (Zmud, Lee-Gosselin, Munizaga, & Carrasco, 2013). A critical body of literature has developed around surveys, since the process by which a survey is designed is complex and can have varying biases and effects on results (Dillman, Smyth, & Christian, 2014). For example, questions that are asked first can provide a particular context that alters the way subsequent questions are answered (Schuman, 1992). Furthermore, question wording can impact respondents' answers (Schuldt, Konrath, & Schwartz, 2011). The purpose of this study is to address these survey design effects in response to travel satisfaction questions in a transport survey.

Trip satisfaction is one of the most important and commonly collected data from transport surveys as researchers have made the connection between an individual's satisfaction with their commute trips and their travel decisions such as the mode used (De Vos, Schwanen, van Acker, & Witlox, 2018). While there is a growing body of literature which aims to analyze how satisfied travelers are, there is little consensus in the literature regarding best practices for satisfaction survey design, or in the understanding of how various survey design techniques impact a respondent's reported satisfaction. To address this gap in the knowledge, we conducted a university-wide travel survey of commuters to McGill University located in Montreal, Canada during fall and winter of the 2017/2018 academic year. In this survey, we experimented with question order, question wording, and seasonality. In the first phase of the analysis, seasonality effects are examined and in the second phase we evaluate the impact of question order as well as wording. The experimental design of this survey presents important findings regarding the sensitivity of survey respondents to different survey design techniques.

LITERATURE REVIEW

Commute satisfaction

Satisfaction has been studied extensively in marketing as a way of measuring service quality and predicting customer behaviour (Cronin & Taylor, 1992; Fornell, Johnson, Anderson, Cha, & Bryant, 1996; Zeithaml, Berry, & Parasuraman, 1996). In the realm of transport, a rich body of literature explores the connection between satisfaction with travel and how it influences mode

loyalty (Minser & Webb, 2010; van Lierop, Badami, & El-Geneidy, 2017), life satisfaction (De Vos & Witlox, 2017) and overall well-being (De Vos & Witlox, 2017; Legrain, Eluru, & El-Geneidy, 2015).

Satisfaction is often defined as the response to a discrepancy between prior expectations and actual experience (Oliver, 1999), however, previous studies have also used the term "travel satisfaction" when referring to satisfaction with travel in general. It is important to distinguish the two types of commute satisfaction: trip satisfaction, which is influenced by individual's emotions during a trip, and satisfaction with daily travel, i.e. with their typical trips (De Vos & Witlox, 2017).

Attention has also been paid to the effects of weather conditions on commute satisfaction. Cold climate, sunshine, and higher temperatures increase travel satisfaction through the elicitation of positive emotions (Ettema, Friman, Olsson, & Gärling, 2017; Kööts, Realo, & Allik, 2011). These results demonstrate a strong correlation between weather conditions, particularly the presence of sunshine, on reported satisfaction. Montreal's climate comprises four distinct seasons with vastly different weather conditions. Local research has also examined the impact of seasons on reported commute satisfaction finding that cyclists in the summer are significantly more satisfied than those who cycled in the winter (Willis, Manaugh, & El-Geneidy, 2013), and that satisfaction on typical trips is lower in cold and snowy conditions for pedestrians, cyclists and bus users (St-Louis, Manaugh, van Lierop, and El-Geneidy (2014).

Survey design

This section examines two of the important ways that measurement error could occur in a survey, First, the way that a particular question is worded and, second, the order in which questions are received.

Question wording

The bias generated from the wording in surveys can be attributed to the framing of a question, through the words and phrases used in the survey to relay information to the survey respondents (Druckman, 2001). Framing effects occur when the wording of a question affects the judgment and choice of the decision-maker (Iyengar, 1990). Early references to this effect concern the variance in an individual's decision when options are presented as potential gains (i.e. positively) or losses (i.e. negatively) (Kahneman & Tversky, 1984). Furthermore, framing effects can also be

present implicitly in the positive, neutral or negative connotations associated with controversial terms. For example, when "climate change" is used in place of "global warming", the percentage of people who agree with its occurrence decreases significantly for people who self-identify as conservative (Schuldt et al., 2011).

Furthermore, researchers studying student self-efficacy (van de Ridder, Peters, Stokking, de Ru, & ten Cate, 2014), medical training programs (Guyatt et al., 1999) and healthcare quality (Dunsch, Evans, Macis, & Wang, 2018) find that when questions are framed positively, reported satisfaction is higher. These researchers attribute this pattern to acquiescence bias where individuals tend to agree to the statement they are presented with, regardless of the content (Krosnick, 2000).

Question order

According to Dillman et al. (2014), a survey should not be viewed as a compilation of independent questions that have no effect on each other but rather each question and response should be evaluated with regard to the larger context of the survey. When earlier questions impact answers to later questions, this is referred to as a question order effect. The theory on survey priming helps to explain this effect, where answers to previous questions will influence responses to later questions (van de Walle & van Ryzin, 2011). There are two potential outcomes of question order effects, one where responses to later questions become more different (i.e. contrast effect) and the other where responses become more similar (i.e. consistency effect) (Schuman & Presser, 1981). Furthermore, Moore (2002) explains that in a set of related questions (e.g. satisfaction with last trip and typical trip), the question that appears first is answered in a non-comparative way but respondents will respond to the following question in comparison to their answer to the first question. He found that when a concept, or framework, is introduced in the non-comparative context, the response to the second question may be systematically different, either higher which demonstrates an additive effect or lower, demonstrating a subtractive effect.

Split-ballot testing has been used to test question order effects. Under this method, the study sample is randomly given a distinct survey version. van de Walle and van Ryzin (2011), conducted a citizen survey using a split-ballot methodology and found that the reported satisfaction changes depending on the order of the two sets of questions, one related to satisfaction with specific public services and another on overall satisfaction.

In terms of customer satisfaction with public transport, researchers have investigated the difference in the reported overall satisfaction with service prior and after introducing a set of questions pertaining to specific aspects of the service (dell'Olio, Ibeas, & Cecín, 2010). They found that a third of the respondents changed their overall satisfaction ratings when asked the second time around and the number of people whose scores improved was double those who reduced it. This indicates that the respondents were more critical when they did not think specifically about the various attributes of the public transport system.

DATA

Survey

The data used in this study were obtained from a commuter survey carried out at McGill University in Montreal, Canada. The survey targeted all staff and faculty in addition to a randomly selected sample of 30% of the student population. The online survey was active from September until November 2017 and January until March 2018, during which 16,930 email requests were distributed. A single reminder was sent to every person who did not respond to the original invitation after a week of non-completion, and prizes were offered as incentives for participation. The survey had a response rate of 33.4%.

Respondents were asked to recall their most recent trip to McGill, and specify the transport mode(s) used, personal and household characteristics, and their satisfaction with their last trip. Additionally, respondents were asked to rate their satisfaction with their "typical" trip to McGill during a specific weather condition. Those who took the survey in fall were asked to report their typical satisfaction on a "warm and sunny day" and those who took the survey in winter reported their typical satisfaction on a "cold and snowy day".

Using a split-ballot method, respondents were randomly assigned one of two survey orders. The design framework for this survey is illustrated in Figure 1. In the first order (referred to as scenario 1), respondents were asked to rate their satisfaction with their last trip, then report their satisfaction with specific trip characteristics relevant to the mode they took along with a host of other questions included in the survey before answering their satisfaction with their typical trip in that season. In the second question order (referred to as scenario 2), respondents were asked to

report their satisfaction with their typical trip first, then their last trip satisfaction, followed by their satisfaction with the mode-specific attributes.
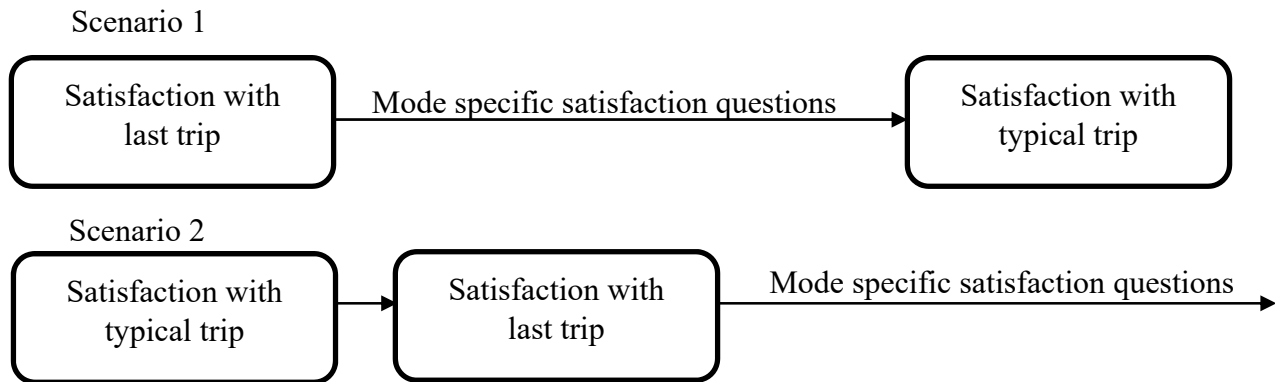
Scenario 1

| Satisfaction with last trip | → Mode specific satisfaction questions → | Satisfaction with typical trip |

Scenario 2

| Satisfaction with typical trip | → | Satisfaction with last trip | → Mode specific satisfaction questions → |

FIGURE 1 Question order design in the survey.

Study Sample

The final sample used for this study consists of 3,102 commuters, which were obtained after removing respondents that did not fully complete the survey, used different modes of transport (such as skateboarding, or the McGill Shuttle), or those who reported using a mode other than the mode used on their last trip as their typical mode of commuting for the sake of consistency. Cyclists were also removed from the analysis due to their small sample size in winter. The final sample for this study is made up of 59% public transport users, 28% pedestrians, and 13% motorized non-transit users consisting of mostly drivers (87%), but also includes those who carpool, use taxi, motorcycle, or scooter.

Using a 5-point Likert-scale, where 1 represents "very unsatisfied" and 5 "very satisfied", survey respondents reported their level of satisfaction with their typical and last commute to McGill. In fall, 731 respondents were asked for their last trip satisfaction before their typical trip satisfaction (scenario 1) and 659 respondents were asked for their typical trip satisfaction and then their last trip satisfaction (scenario 2). In winter 913, respondents were asked for their last trip satisfaction and then their typical trip satisfaction and 799 respondents were asked for their typical trip satisfaction and then their last trip satisfaction.

Weather conditions differed substantially between the two seasons in which the survey was active. Mean temperatures varied from 10.1°C in fall to -7.4°C in winter. Accumulated snow on the ground also changed by season with an average of less than one centimeter in fall, to an average

68 centimeters in winter. Precipitation occurred on the day of the respondents' last trip for 54% of the participants in fall, and 46% in winter.

METHODOLOGY

The analysis was carried out in two parts: part A, where the seasonality effects are examined, and part B, where the impact of question order as well as wording is explored. Respondents were segmented by their mode of transport for all analyses. We begin by analyzing the seasonal and implied seasonal effects on satisfaction. The types of analyses that will be done are outlined in Figure 2. Arrows in the figure labelled A, B, C, and D represent the comparisons between the two seasons, which will be referred to in the analysis section. First, we compare last trip satisfaction for fall and winter respondents. Then, we examine how differently individuals respond to the different wording of typical trip satisfaction that is asked according to the season of survey distribution. Independent sample t-tests are applied for both analyses to identify statistical differences, first between fall and winter respondents on their last trip and then between "warm and sunny day" and "cold and snowy day" respondents on their typical trip.
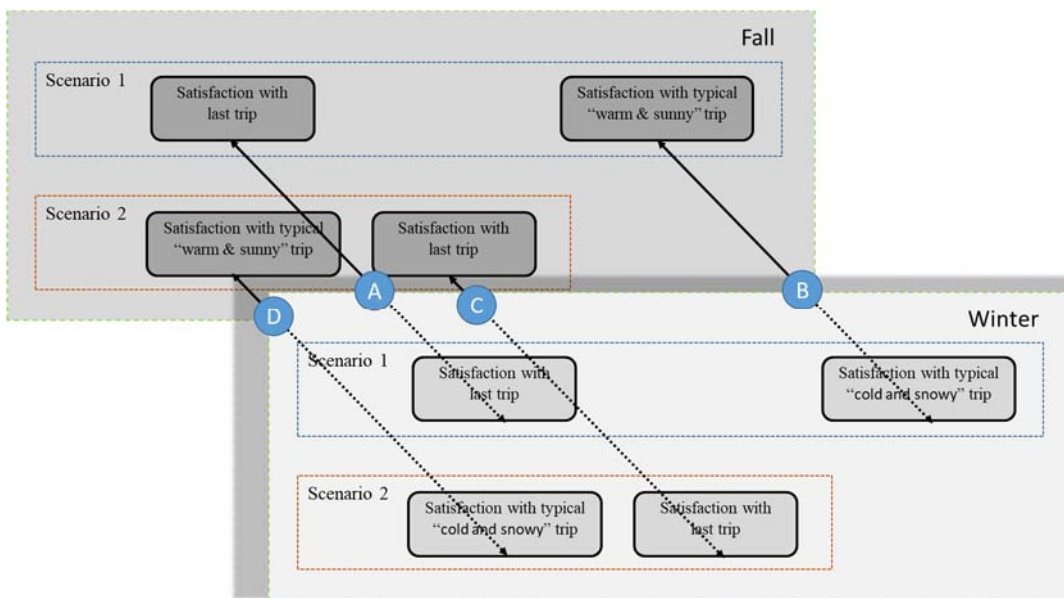

FIGURE 2 Evaluation framework – Part A.

In the second part of the analysis (Part B), we evaluate the impact that question wording as well as question order has on the reported average satisfaction level of respondents for their last

and typical trips. This comparison is conducted for each season separately. The evaluation framework is presented in Figure 3, the four comparisons shown on the diagram are described below:

1. "First question asked" – a comparison of average satisfaction of the first question asked (last trip or typical trip) between the two scenarios;

2. "Second question asked" – a comparison of reported average satisfaction of the second question asked (last trip or typical trip) between the two scenarios;

3. "Last trip order" – a comparison of average last trip satisfaction when asked first in scenario 1 compared to asked second in scenario 2; and

4. "Typical trip order" – a comparison of average typical trip satisfaction when asked second in scenario 1 compared to asked first in scenario 2.
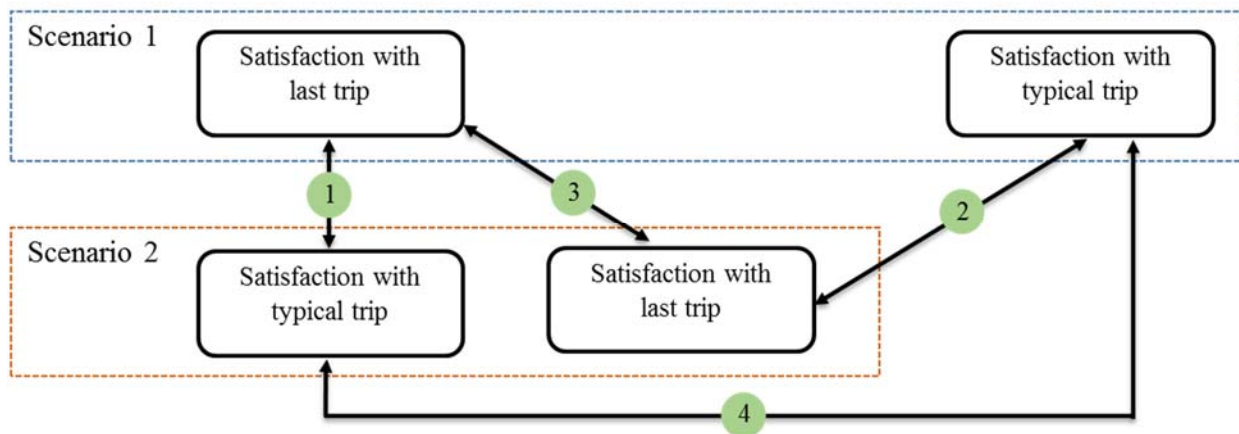

FIGURE 3 Evaluation framework – Part B.

To evaluate the impact that question order and question wording has on reported last and typical trip satisfaction, we conduct independent t-tests to examine statistical differences between reported satisfaction levels to of the "first question asked" between in the two scenarios and then again for the second question. Following the t-tests, we undertake two series of binary logistic regression analyses to determine the predictors of reported satisfaction, and to capture the impact of question wording on satisfaction responses. The dependent variable in the first model is the satisfaction rating for the question that respondents were asked first (Comparison 1), and in the second model the dependent variable is the rating of the satisfaction question asked second (Comparison 2). We transformed respondent's ordinal satisfaction responses into binary variables

by recoding 1, 2 and 3 to represent unsatisfied, and 4 and 5 as satisfied. While this practice is common and simplifies the interpretation of the results, we acknowledge that it may impact the model estimates (Manor, Mathews, & Power, 2000). To investigate the impact of question order and wording, we include a dummy variable to account for when the question answered was satisfaction with a typical trip. For example, when modeling satisfaction with the "first question asked", the typical trip satisfaction dummy variable indicates whether the typical trip satisfaction question asked first is more likely to result in a satisfied response compared to the satisfaction with last trip question asked first. In the second model, we control for the response to the first question asked. A model is developed for each of the three modes and segmented by fall and winter respondents. In both models we control for gender, age, income, and travel time. Additionally, weather-related variables corresponding to the day the trip occurred, including average temperature, precipitation, and snow on ground are included.

The third and fourth comparisons directly evaluate the impact of question order. Comparison three evaluates reported satisfaction with the last trip when the question is asked first compared to the same question when it is asked second (split-ballot testing) and likewise for satisfaction with a typical trip in comparison four. Independent t-tests are carried out on the reported satisfaction ratings that we are comparing in each case. Then, logistic regression models are used to capture the effect of question order on reported satisfaction. The dependent variable in this set of models is the transformed binary satisfaction rating for the question being tested (either last trip or typical trip) in both question order scenarios. The key independent variable is the question order scenario, where the reference question order that is compared to is scenario 1 (in which the last trip question was asked before typical trip). Additionally, both models have the same control variables as the initial two series of logistic regressions.

RESULTS

Season of Survey Distribution (Analysis Part A)

Table 1 summarizes the results of the independent t-tests examining the seasonal differences in average satisfaction levels with the last trip as well as typical trip by mode and by question order scenario. Please note that letters A, B, C, and D in the Table 1 correspond to the arrows presented in Figure 2 corresponding to which comparison is being conducted.

TABLE 1 Seasonal differences in average trip satisfaction and typical trip satisfaction

Scenario			Walk	Transit	Drive
1	**A**	**Season of survey distribution**	Average satisfaction level		
		Last trip in Fall	4.32	3.87	3.46
		Last trip in Winter	4.02	3.85	3.7
		$\Delta_{\text{fall-winter}}$	0.3*	0.02	-0.24
	B	**Question wording**	Average satisfaction level		
		Typical warm and sunny trip (fall)	4.61	3.83	3.91
		Typical cold and snowy trip (winter)	3.04	3.26	3.01
		$\Delta_{\text{typical fall-typical winter}}$	1.57*	0.57*	0.9*
2	**C**	**Season of survey distribution**	Average satisfaction level		
		Last trip in Fall	4.63	4.15	3.84
		Last trip in Winter	4.03	4.04	4.04
		$\Delta_{\text{fall-winter}}$	0.6*	0.11	-0.2
	D	**Question wording**	Average satisfaction level		
		Typical warm and sunny trip (fall)	4.68	4.14	3.74
		Typical cold and snowy trip (winter)	3.6	3.69	3.4
		$\Delta_{\text{typical fall-typical winter}}$	1.08*	0.45*	0.3*

* 95% significance level

With regards to the satisfaction with last trip (Comparisons A and C), only pedestrians reported significantly lower satisfaction levels in winter compared to fall, irrespective of the scenario. This result indicated that reported satisfaction from active mode users are more likely to be impacted by harsh winter conditions. However, for transit users the difference between fall and winter is not significant. Interestingly, drivers were more satisfied in the winter than fall but the difference is not significant at the 95% level.

Regarding the differences between satisfaction on a typical "warm and sunny day" compared to typical "cold and snowy day" across scenarios (Comparisons B and D), the difference between satisfaction on a typical "warm and sunny day" compared to typical "cold and snowy day" (Comparisons B and D) is much larger than the differences across last trip in fall and winter (Comparisons A and C) for pedestrians. Also, we see a statistically significant difference among public transport users and drivers, reporting a higher satisfaction during a typical "warm and sunny day" compared to "cold and snowy day" (Comparisons B and D). In addition, we observe a question order effect, the difference between seasons is consistently higher in scenario 1, where typical trip satisfaction was asked second (Comparison B).

Question Order and Wording (Analysis Part B)

T-test Results

Table 2 presents the average satisfaction levels with last and typical trip, using t-tests to assess statistical significance between the two questions. Please note that Comparisons 1, 2, 3, and 4 mentioned in the text refer to relations presented in Figure 3. A horizontal comparison (i.e. $\Delta_{typical-last}$) in the table examines statistical differences in average satisfaction in the "first question asked" (Comparisons 1) and "second question asked" (Comparison 2) analysis. A vertical comparison (i.e. $\Delta_{second-first}$) in the table represents the t-test conducted to assess differences in "last trip order" (Comparison 3) and "typical trip order" (Comparison 4).

TABLE 2 Results of independent t-tests for part B of analysis

	Fall Average			Winter Average		
	Last trip	Typical trip	$\Delta_{typical-last}$	Last trip	Typical trip	$\Delta_{typical-last}$
Walk						
First question asked	4.32	4.68	0.36*	4.02	3.6	-0.42*
Second question asked	4.63	4.61	-0.02	4.03	3.04	-0.99*
$\Delta_{second-first}$	+0.31*	-0.07		+0.01	-0.56*	
Transit						
First question asked	3.87	4.14	0.27*	3.85	3.69	-0.16*
Second question asked	4.15	3.83	-0.32*	4.04	3.26	-0.78*
$\Delta_{second-first}$	+0.28*	-0.31*		+0.19*	-0.43*	
Drive						
First question asked	3.46	3.74	0.28	3.7	3.4	-0.3*
Second question asked	3.84	3.91	0.07	4.04	3.01	-1.03*
$\Delta_{second-first}$	+0.38*	+0.17		+0.34*	-0.39*	

* 95% significance level

Our results show that changing the wording of a satisfaction question appears to impact reported satisfaction levels. For the "first question asked" analysis (Comparison 1) satisfaction with a typical trip is consistently answered higher than last trip satisfaction in the fall with the exception of drivers. Whereas, in the winter we see that satisfaction with a typical trip is answered lower than last trip satisfaction across all modes. Since respondents answer the first question in the survey in a non-comparative manner, this pattern may be attributed to the framing of the typical

trip satisfaction question. Asking satisfaction with a typical trip may have exaggerated people's level of satisfaction positively when the words "warm and sunny day" were used, evoking positive emotions; and negatively when the idea of a "cold and snowy day" was introduced. Accordingly, the wording of typical trip satisfaction in our survey appears to impact reported satisfaction.

With regards to the "second question asked" analysis (Comparison 2), we see no trend in the difference between reported typical trip and last trip satisfaction across all mode users in fall. While in winter we observed that satisfaction with a typical trip is consistently answered lower than satisfaction with last trip across all mode users. We also see that in winter, the difference between the two questions asked second (Comparison 2) is greater than the difference between the two questions asked first (Comparison 1), implying a contrasting effect arising from question order where individuals may have answered more differently to the typical "cold and snowy" question given their response to their last trip. It is possible that a mental comparison of their satisfaction with their last trip and the loaded negative connotation associated with the wording of typical "cold and snowy" day caused them to provide a lower rating for the second question.

With respect to question order, we see that average satisfaction with the last trip (Comparison 3) is always higher when asked second than first for fall respondents. The same trend is observable in winter but is not significant for pedestrians. On the other hand, average satisfaction with the typical trip (Comparison 4) is always lower when asked second than first for winter respondents only.

Statistical Modeling

Table 3 presents model results with satisfaction with the "first question asked" as the dependent variable (Comparison 1), while Table 4 presents satisfaction with the "second question asked", (Comparison 2). Both models show that respondents answer the typical satisfaction question significantly different depending on the season, mode, and question order scenario. In Table 3, we see that fall respondents are significantly more likely to answer their first satisfaction question with a satisfied response when the question asked was regarding their typical commute rather than their last trip. With the exception of transit users, winter respondents answered typical satisfaction, when it was asked first, lower than last trip satisfaction. For winter transit riders, response to their first satisfaction question appears to be driven by actual weather experienced on that day. Likely their trip took place on a cold and snowy day, therefore little difference between the responses to the two satisfaction questions was observed. In Table 4, all winter respondents are significantly

less likely to report a satisfied response when typical trip satisfaction is asked second. In fall, only pedestrians are more likely to say they were satisfied with the "second question asked" when the second question was typical trip satisfaction. In addition, we can see that being satisfied with the first question, regardless of what the question was, increases the likelihood of having a satisfied response to the second question for all modes, especially in fall.

With respect to the remaining independent variables in our model, none of the personal characteristics are consistently significant across modes and seasons. Being male is only significant in the transit model which may be due to the insecurity experienced by females in transportation settings as described in previous studies (Loukaitou-Sideris & Fink, 2008). Age is significant for all modes in winter and only transit and walking in fall. A middle or high-income driver is four times more likely to report being satisfied with their first question answered in fall, compared to a low-income driver. In winter, a middle or high-income transit user or pedestrian is more likely to report being satisfied as they may be choosing to use these modes out of choice instead of necessity. Weather conditions are significant for transit users in the winter, drivers in the fall and not at all for pedestrians. Likewise, travel time is not significant for pedestrians perhaps due to predictability of the mode. However, for transit users, an increase in travel time decreases the likelihood of them reporting being satisfied in both seasons. The same result is seen for fall drivers. In the second set of models, we observe that the number of significant control variables is reduced and no trends can be observed. The contributing factor may be the question order effects that were discussed previously.

TABLE 3 Results of logistic regression modeling satisfaction with first question answered (Comparison 1)

Variables	Walk Fall OR	Sig.	Walk Winter OR	Sig.	Transit Fall OR	Sig.	Transit Winter OR	Sig.	Drive Fall OR	Sig.	Drive Winter OR	Sig.
Satisfaction with typical trip	3.997	***	0.647	**	2.252	***	1.051		2.566	***	0.837	**
Personal characteristics												
Gender: Male	1.077		0.922		1.682	**	1.41	**	0.576		1.128	
Age	1.05	*	1.031	*	1.02	*	1.015	**	0.999		1.048	*
High income *(ref = low income)*	0.786		2.733	**	1.883	*	2.224	***	3.876	**	0.585	**
Middle income *(ref = low income)*	1.658		2.317	***	0.963		1.669	***	3.858	**	0.524	***
Trip characteristics												
Average temperature (Celsius)	1.043		1.007		1.015		1.043	***	0.891	**	1.047	
Precipitation (cm)	0.974		0.967		1.002		0.944	*	1.05	**	1.007	
Snow on ground dummy	0.997		0.998		1.05		1.012	*	0.872	*	1.005	
Travel Time (min)	0.997		1.001		0.971	***	0.976	***	0.961	***	0.998	
Number of respondents	400		463		770		1055		220		194	

*** 99% significance level, ** 95% significance level, * 90% significance level

TABLE 4 Binary Logistic Regression modeling satisfaction with second question answered (Comparison 2)

Variables	Walk				Transit				Drive			
	Fall		Winter		Fall		Winter		Fall		Winter	
	OR	Sig.	OR	Sig.	OR	Sig.	OR	Sig.	OR	Sig.	OR	Sig.
Satisfaction with typical trip	3.128	**	0.142	***	0.872		0.304	***	1.799		0.188	***
Satisfaction with first question	16.199	***	8.272	***	9.102	***	6.666	***	19.489	***	3.953	***
Personal Characteristics												
Gender: Male	0.679		1.213		0.803		0.938		0.776		1.166	
Age	1.024		1.025		1.001		1.020	**	1.015		1.017	
High income *(ref = low income)*	1.088		1.628		2.253	**	1.334		0.555		1.127	
Middle income *(ref = low income)*	1.013		1.879	**	1.493		1.218		0.352		0.853	
Trip Characteristics												
Average temperature (Celsius)	1.061		1.000		0.978		1.020		1.089	*	0.983	
Precipitation (cm)	1.008		0.924		1.008		0.976		1.036		1.013	
Snow on ground dummy	0.826	**	0.992		0.949		1.003		1.055		0.988	
Travel Time (min)	0.998		0.999		0.993	*	0.987	***	1.001		1.000	
Number of respondents	400		463		770		1055		220		194	

*** 99% significance level, ** 95% significance level, * 90% significance level

Table 5 presents the results of the logistic regression modeling the effect of question order on both satisfaction with a typical trip (Comparison 4) and last trip (Comparison 3). In this table we present the odds ratio and statistical significance of the effect of receiving the survey under question order scenario 1 (last trip first and typical satisfaction second). Please note all models included personal and trip characteristics variables similar to Tables 3 and 4; for simplicity we only report the coefficient of the dummy variable which captures whether the respondent had the scenario 1 version of the survey. In general, we see that in both seasons, the likelihood of being satisfied with a typical trip is lower (except for winter pedestrians) when this question is asked second. This implies a subtractive effect (Moore, 2002), as respondents, regardless of the wording of the typical trip question (i.e. warm and sunny or cold and sunny), are reporting lower satisfaction when asked second. Perhaps respondents, after answering a series of questions in between the last trip and typical trip satisfaction questions, are growing tired of the survey and therefore responding with lower satisfaction. In addition, it seems that transit users are less likely to be satisfied with both typical trip and last trip in scenario 1 compared to scenario 2 regardless of season. For the last trip question, the trend is not as visible across seasons and modes.

Table 5 Question order effects

Independent variable	Dependent variable: Satisfaction with last trip (Comparison 3)											
	Walk				Transit				Drive			
	Fall		Winter		Fall		Winter		Fall		Winter	
	OR	Sig.	OR	Sig.	OR	Sig.	OR	Sig.	OR	Sig.	OR	Sig.
Scenario 1 (*ref = scenario 2*)	1.023		0.444	***	0.529	***	0.492	***	0.833		0.630	

Independent variable	Dependent variable: Satisfaction with typical trip (Comparison 4)											
	Walk				Transit				Drive			
	Fall		Winter		Fall		Winter		Fall		Winter	
	OR	Sig.	OR	Sig.	OR	Sig.	OR	Sig.	OR	Sig.	OR	Sig.
Scenario 1 (*ref = scenario 2*)	0.343	***	0.852		0.550	**	0.686	**	0.427	**	0.431	**

*** 99% significance level, ** 95% significance level, * 90% significance level

CONCLUSIONS

We began our analysis by examining seasonality effects in satisfaction responses (Analysis A). Comparing trip satisfaction in fall and winter, pedestrians were the only group of users that were significantly less satisfied in the winter compared to fall. This finding is not surprising given the reality of walking in winter conditions . Regarding the differently worded typical satisfaction question, all respondents reported significantly higher satisfaction with a typical "warm and sunny" trip compared to a typical "cold and snowy" trip. Such stark differences in responses of satisfaction to these typical satisfaction questions are likely exaggerated due to the positive or negative emotions evoked from the question wording. Future research trying to capture seasonal variation in satisfaction levels with commute should use the same wording when collecting the data in both seasons.

The second part of analysis (Analysis B) contributes to the literature by examining how respondents answered a commute satisfaction question depending on which question was asked first and second. We find that the likelihood of being satisfied is higher when satisfaction with a typical trip is the "first question asked" in fall for all mode users. In contrast, the likelihood of being satisfied is lower when the "second question asked" is satisfaction with typical trip in winter for all mode users. The most consistent observation in these models is how likely a respondent is satisfied with their second question when they are satisfied with their first, regardless of question wording. We conclude that the effect of question wording as well as question order is present in transport surveys and will impact the answers of commute satisfaction questions.

In summary, this research highlights the need for caution when designing a travel survey as survey distributers may have vested interests in the results and therefore frame the questions to improve reported satisfaction and perception of the respondents. We confirm the presence of question order and wording effects, and, to a lesser degree, differences due to season of distribution. van de Walle and van Ryzin (2011) suggest that randomization of questions for the survey sample should offset some of the impacts of question order. In research such as the present study, we cannot determine which question order is objectively better, and therefore implementing a split-ballot methodology is recommended for future travel surveys, while using the same wording to capture seasonal effects.

REFERENCES

Cronin, J., & Taylor, S. (1992). Measuring service quality: A reexamination and extension. *Journal of Marketing, 56*(3), 55-68.

De Vos, J., Schwanen, T., van Acker, V., & Witlox, F. (2018). Do satisfying walking and cycling trips result in more future trips with active travel modes? An exploratory study. *International Journal of Sustainable Transportation, 0*(0), 1-17.

De Vos, J., & Witlox, F. (2017). Travel satisfaction revisited. On the pivotal role of travel satisfaction in conceptualising a travel behaviour process. *Transportation Research Part A: Policy and Practice, 106*, 364-373.

dell'Olio, L., Ibeas, A., & Cecín, P. (2010). Modelling user perception of bus transit quality. *Transport Policy, 17*(6), 388-397.

Dillman, D., Smyth, J., & Christian, L. (2014). *Internet, phone, mail and mixed-mode surveys: The tailored design method*. Hoboken: Wiley.

Druckman, J. (2001). The implications of framing effects for citizen competence. *Political Behavior, 23*(3), 225-256.

Dunsch, F., Evans, D., Macis, M., & Wang, Q. (2018). Bias in patient satisfaction surveys: A threat to measuring healthcare quality. *BMJ Global Health, 3*(2).

Ettema, D., Friman, M., Olsson, L., & Gärling, T. (2017). Season and weather effects on travel-related mood and travel satisfaction. *Frontiers in Psychology, 6*(140).

Fornell, C., Johnson, M., Anderson, E., Cha, J., & Bryant, B. (1996). The american customer satisfaction index: Nature, purpose, and findings. *Journal of Marketing, 60*(4), 7-18.

Guyatt, G., Cook, D., King, D., Norman, G., Kane, S., & van Ineveld, C. (1999). Effect of the framing of questionnaire items regarding satisfaction with training on residents' responses. *Academic Medicine, 74*(2), 192-194.

Iyengar, S. (1990). Framing responsibility for political issues: The case of poverty. *Political Behavior, 12*(1), 19-40.

Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist, 39*(4), 341-350.

Kööts, L., Realo, A., & Allik, J. (2011). The influence of the weather on affective experience. *Journal of Individual Differences, 32*, 74-84.

Krosnick, J. (2000). The threat of satisficing in surveys: The shortcuts respondents take in answering questions. *Survey Methods Newsletter, 20*(1), 4-8.

Legrain, A., Eluru, N., & El-Geneidy, A. (2015). Am stressed, must travel: The relationship between mode choice and commuting stress. *Transportation Research Part F: Traffic Psychology and Behaviour, 34*, 141-151.

Loukaitou-Sideris, A., & Fink, C. (2008). Addressing women's fear of victimization in transportation settings: A survey of U.S. transit agencies. *Urban Affairs Review, 44*(4), 554-587.

Manor, O., Mathews, S., & Power, C. (2000). Dichotomous or categorical response? Analysing self-rated health and lifetime social class. *Internation Journal of Epidemiology, 29*(1), 149 - 157.

Minser, J., & Webb, V. (2010). Quantifiying the benefits: Application of customer loyality modeling in public transportation context *Journal of the Transportation Research Board, 2144*, 111-120.

Moore, D. (2002). Measuring new types of question-order effects: additive and subtractive. *Public Opinion Quarterly, 66*(1), 80-91.

Oliver, R. (1999). Whence consumer loyalty? *Journal of Marketing, 63*, 33-44.

Schuldt, J., Konrath, S., & Schwartz, N. (2011). "Global warming" or "climate change"?: Whether the planet is warming depends on question wording *Public Opinion Quarterly, 75*(11), 115-124.

Schuman, H. (1992). *Context effects: State of the past/state of the art*. New York, US: Springer-Verlag.

Schuman, H., & Presser, S. (1981). *Questions and answers in attitude surveys: Experiments on question form, wording and context*. New York, NY.: Academic Press.

St-Louis, E., Manaugh, K., van Lierop, D., & El-Geneidy, A. (2014). The happy commuter: A comparison of commuter satisfaction across modes. *Transportation Research Part F: Traffic Psychology and Behaviour, 26*, 160-170.

van de Ridder, J., Peters, C., Stokking, K., de Ru, J., & ten Cate, O. (2014). Framing of feedback impacts student's satisfaction, self-efficacy and performance. *Advances in Health Sciences Education, 20*(3), 803-816.

van de Walle, S., & van Ryzin, G. (2011). The order of questions in a survey on citizen satisfaction with public services: Lessons from a split-ballot experiment. *Public Administration, 89*(4), 1436-1450.

van Lierop, D., Badami, M., & El-Geneidy, A. (2017). *Nothing but a good ride: The influences of satisfaction and loyalty in public transport*. Paper presented at the Transportation Research Board's 96th Annual Meeting Washington D.C.

Willis, D., Manaugh, K., & El-Geneidy, A. (2013). Uniquely satisfied: Exploring cyclist satisfaction. *Transportation Research Part F: Traffic Psychology and Behaviour, 18*, 136-147.

Zeithaml, V., Berry, L., & Parasuraman, A. (1996). The behavioral consequences of service quality. *Journal of Marketing, 60*(2), 31-46.

Zmud, J., Lee-Gosselin, M., Munizaga, M., & Carrasco, J. (2013). Transport surveys: Considerations for decision makers and decision making In J. Zmud, M. Lee-Gosselin, M. Munizaga, & J. Carrasco (Eds.), *Transport Survey Methods: Best Practice for Decision Making* Bingley, UK: Emerald Group Publishing Limited