



Titre: Definition and Analysis of Respondent Burden in Web-Based Travel
Title: Surveys

Auteur: Hedieh Keshavarz Mohammadian
Author:

Date: 2023

Type: Mémoire ou thèse / Dissertation or Thesis

Référence: Keshavarz Mohammadian, H. (2023). Definition and Analysis of Respondent
Citation: Burden in Web-Based Travel Surveys [Mémoire de maîtrise, Polytechnique
Montréal]. PolyPublie. <https://publications.polymtl.ca/54890/>

 **Document en libre accès dans PolyPublie**
Open Access document in PolyPublie

URL de PolyPublie: <https://publications.polymtl.ca/54890/>
PolyPublie URL:

**Directeurs de
recherche:** Catherine Morency
Advisors:

Programme: Génie civil
Program:

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Definition and analysis of respondent burden in web-based travel surveys

HEDIEH KESHAVARZ MOHAMMADIAN

Département des génies civil, géologique et des mines

Mémoire présenté en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*

Génie civil

Août 2023

© Hedieh Keshavarz Mohammadian, 2023.

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Ce mémoire intitulé:

Definition and analysis of respondent burden in web-based travel surveys

présenté par **Hedieh KESHAVARZ MOHAMMADIAN**

en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*

a été dûment accepté par le jury d'examen constitué de :

Owen WAYGOOD président

Catherine MORENCY membre et directrice de recherche

Matthew PALM, membre

DEDICATION

To Peyman, who always inspired and believed me during this journey,

To Mom, and Dad, for their boundless love and understanding,

To my brother for always being there for me,

ACKNOWLEDGEMENTS

First and foremost, I want to express my heartfelt gratitude to Professor Catherine Morency, my research supervisor at Polytechnique Montréal, for trusting me and giving me the opportunity to pursue my educational goals. Her unwavering support, guidance, patience, and encouragement throughout my studies have been a constant source of inspiration and a shining example of dedication. She has given me the support that I needed to keep pursuing my educational dreams and thanks to her, I have had the privilege to work with an exceptional research team. I will forever be thankful for your invaluable support and guidance throughout my studies.

I am deeply grateful to the project funding sources, Chaire Mobilité headed by Professor Catherine Morency and Mobilizing Justice project. I want to thank you my colleagues at Chaire Mobilté team, especially Pierre-Léo Bourbonnais for patiently addressing all my questions. Additionally, I want to express my appreciation to Brigitte Bouchard-Milord, and Hubert Verreault for their help during my project. Furthermore, I am thankful for the contributions of Howaida Hassan, Mathew Palm, and Steven Farber from the Mobilizing Justice project. Their collaboration has been very helpful in the success of my work.

Additionally, I would like to express my great appreciation to my colleague and friend, Hamed Malekzadeh, for his invaluable assistance throughout my studies, answering my questions with patience. Also, I would like to thank my colleague and friend, Kaveh Rezvani, for his help during my studies.

Moreover, I want to extend a special thank you to my friends Forouz Allahyari Fard, Mojde Sharafi who were the initial inspiration for embarking on this this journey. A heartfelt thank you goes out to my friends Leyla, Tannaz, Parastoo, Sadaf, Neda, and Mehnoush for their unwavering support over these past years. I am honored for your friendship and constant encouragement.

I would like to thank my Mom and Dad for trusting and encouraging me to follow my dreams miles away from home, from them. I am always grateful for your unwavering support and continuous love. I also want to extend my thanks to my only brother for his constant support in helping me perusing my goals. I am grateful and lucky to have you by my side.

Last but not least, I would like to thank my husband and best friend, Peyman, whom without his support I would have not been able to preserve this challenging journey. Thank you for being by

my side in each and every step of this way. Your understanding and support through all ups and downs meant a lot to me and has been always a source of power and motivations for me. Once again, thank you for always believing in me and trusting me, and for encouraging me to follow my dreams without fear.

RÉSUMÉ

Les enquêtes font partie des méthodes typiques de collecte de données dans le domaine des comportements de transport et de mobilité. En raison des avancements des technologies Web, les enquêtes en ligne sont devenues plus populaires et ont commencé à remplacer les méthodes traditionnelles de type stylo-et-papier. Bien que les enquêtes en ligne soient rapides et économiques, et capables d'atteindre un auditoire plus large, les problèmes sont le faible taux de participation et le taux élevé d'abandon qui sont devenus un grand défi. Le fardeau du répondant, c'est-à-dire l'effort nécessaire pour répondre à une enquête, est reconnu comme un facteur important affectant le taux de réponse. Le fardeau est un concept multidimensionnel qui englobe diverses dimensions, notamment les caractéristiques de l'enquête, les caractéristiques des répondants et la manière dont ces deux interagissent. Le fardeau réel est quantifié en s'appuyant sur les caractéristiques objectives du questionnaire, tandis que le fardeau perçu est mesuré par la perception que les répondants ont de l'enquête.

Cette étude vise principalement à identifier les éléments qui influencent le fardeau du répondant et à développer des modèles qui quantifient le fardeau perçu. Pour y parvenir, un ensemble de sept questions post-enquête sont posées à la fin de l'enquête liée à l'impact de la COVID sur le comportement de transport des résidents du Québec. Les questions invitent les répondants à auto-déclarer leur perception du fardeau de l'enquête, le niveau de difficulté, le niveau d'intérêt, la volonté de participer à de nouvelles enquêtes, la tendance à abandonner l'enquête et la durée perçue de l'enquête. Les données recueillies à partir de cette enquête ont servi à trois constructions latentes, à savoir la longueur perçue, l'effort perçu et l'intérêt perçu dans un modèle d'équations structurelles pour prédire le fardeau perçu comme variable de réponse. Selon les résultats, la longueur perçue et l'effort perçu sont significativement et positivement corrélés avec le fardeau, alors qu'une corrélation négative significative est trouvée entre l'intérêt perçu et le fardeau. Selon les résultats, bien que la durée de l'enquête soit classée comme l'un des facteurs les plus importants dans la création de fardeau, l'effort perçu requis pour répondre à une enquête a un impact plus important sur la formation du fardeau.

Le deuxième objectif de cette étude est d'examiner quelles variables et quels seuils sont les plus importants pour prédire la perception des répondants de la longueur de l'enquête, qu'elle soit perçue comme longue ou brève. Des arbres de décision et des forêts aléatoires sont utilisés pour classer

l'échantillon en deux classes: « durée perçue longue » et « durée perçue courte/correcte ». Les résultats ont révélé que les niveaux de fardeau supérieurs à 30-34 (sur une échelle de 0 à 100), la tendance à décrocher et la perception que l'enquête dure plus de 19-20 minutes augmentent la probabilité de percevoir une enquête comme longue.

Le troisième objectif de cette recherche consiste à étudier l'effet des types de questions sur la durée perçue et mesurée en utilisant deux modèles de régression multiple. Les résultats montrent que l'augmentation du nombre de questions à choix multiples et ouvertes augmentera significativement les durées perçue et réelle de l'enquête.

Ce travail souligne l'importance de mesurer et d'analyser le fardeau des répondants en tant qu'élément clé de prédiction du taux de participation à l'enquête et de la qualité des données recueillies. Les résultats ont révélé que des niveaux de fardeau plus élevés et une durée perçue plus longue ont amené les répondants à classer les enquêtes comme trop longues, ce qui est positivement associé à l'abandon des enquêtes. L'une des contributions de ce travail est associé à des questions qui ont servi comme les indicateurs pour entreprendre une analyse des facteurs latents. Les résultats montrent que des questions post-enquête supplémentaires doivent être élaborées pour améliorer la fiabilité et la validité des variables latentes grâce à une analyse factorielle confirmatoire et pour saisir avec succès les différentes dimensions du fardeau subjectif. De plus, les résultats de cette recherche seront utiles pour proposer des stratégies de conception d'enquête afin d'éviter que les enquêtes ne soient longues et que les répondants restent intéressés à les remplir.

Plusieurs limites sont présentes dans cette étude. La principale limite est la contrainte de poser un nombre limité de questions post-enquête, pour éviter de surcharger les répondants suffisamment motivés pour répondre à l'enquête au complet. Une autre limite est que les questions post-enquête sont posées aux répondants suffisamment intéressés pour répondre à l'enquête et atteindre la dernière section. Par conséquent, on s'attend à ce que leurs attitudes envers l'enquête soient relativement positives, ce qui peut entraîner un biais de sélection dans l'échantillon. Dans la section associée à l'analyse de régression, la principale limitation était associée à une faible variation de la fréquence des questions oui/non et des questions à curseur entraînant leur exclusion du modèle de régression. Il est important que leur rôle dans la variation temporelle réelle et perçue soit exploré dans d'autres enquêtes liées aux comportements de transport et de mobilité.

ABSTRACT

Surveys are among the typical methods to gather data in the field of transportation and mobility behaviour. Due to the advancement of web technologies, web surveys have become more popular and started to replace traditional pen-and-paper methods. Although web surveys are time- and cost-efficient, and able to reach a wider audience, low participation and increased drop-off rate have become a big concern. Respondent burden referred to as the effort needed to complete a survey, is recognized as an important factor affecting response rate. Burden is a multidimensional concept that encompasses various dimensions including survey features, respondents' characteristics, and how these two interact with each other. Actual burden is quantified relying on objective features of the questionnaire, while perceived burden is measured through perceptions of respondents towards the survey.

This study initially aims to identify the elements that influence respondent burden and to develop models which quantify subjective burden. To reach this, a set of seven post-survey questions are asked at the end of a survey related to the impact of COVID on transportation behaviour of Quebec residents. The questions prompt respondents to self-report their perception of the survey burden, difficulty, interest, willingness to participate in new surveys, inclination to abandon the survey, and perceived duration of the survey. The data gathered from this survey served as three latent constructs namely perceived length, perceived effort, and perceived interest in a Structural Equation Model to predict subjective burden as the response variable. According to the results, perceived length and perceived effort are significantly and positively correlated with burden, whereas a significant negative correlation is found between perceived interest and burden. According to the results, despite survey length being ranked as one of the most important factors in creating burden, perceived effort required to complete a survey has a greater impact on burden formation.

The second objective of this study is to examine which variables and what cut-off points are more important in predicting the respondent's perception of survey lengthiness, whether it is perceived as lengthy or brief by the respondents. Machine learning techniques namely Decision Trees and Random Forests are used to classify the sample into two classes: perceived "Long" and perceived "Correct/Short". The results revealed that burden levels greater than 30-34 (on a scale of 0 to

100), being inclined to drop-off, and perceiving the survey to be longer than 19-20 minutes increase the likelihood of perceiving a survey as lengthy.

The third objective of this research involves investigating the effect of question types on perceived and measured duration using multiple regression models. The results show that increasing the number of multiple-choice and open-ended questions will increase perceived and measured survey duration significantly.

This work highlights the importance of measuring and analyzing respondent burden as a key driver in survey participation rate and data quality. The results revealed that higher levels of burden and longer perceived duration led respondents to rank surveys as lengthy which is positively associated with dropping off from surveys. One of the contributions of this work is associated with questions that served as indicators to undertake latent factor analysis. The results show that additional post-survey questions are required to be developed with the aim to enhance the reliability and validity of latent constructs through confirmatory factor analysis and to successfully capture different dimensions of subjective burden. Additionally, the findings of this research are helpful to propose survey design strategies to prevent surveys from being long resulting in keeping respondents interested in completing them.

Several limitations are present in this study. One of the primary limitations of this study is the sampling technique. This sample is opportunistic, meaning it includes respondents who already showed interest in participating in surveys. As a result, some bias will be introduced into the results, and it will be challenging to draw inferences about a broader population. The second limitation is the constraint of asking a limited number of post-survey questions, to prevent overburdening respondents who were motivated enough to complete the survey. Another limitation is that post survey questions are asked from respondents that were interested enough to complete the survey and reach the last section. Hence, it is expected that their attitudes towards the survey to be relatively positive which may result in a selection bias in the sample. In the section associated to regression analysis, the limitation was associated with low variation in the frequency of yes/no and slider questions resulting in their exclusion in the regression model. Thus, it is important that their role in measured and perceived time variation be explored in other surveys related to transportation and mobility behaviour.

TABLE OF CONTENTS

DEDICATION	III
ACKNOWLEDGEMENTS	IV
RÉSUMÉ.....	VI
ABSTRACT	VIII
TABLE OF CONTENTS	X
LIST OF TABLES	XIII
LIST OF FIGURES.....	XIV
LIST OF SYMBOLS AND ABBREVIATIONS.....	XVI
CHAPTER 1 INTRODUCTION.....	1
1.1 Problem statement	2
1.2 Objectives.....	3
1.3 Thesis structure	4
CHAPTER 2 LITERATURE REVIEW	7
2.1 History of surveys	7
2.2 Definition of respondent burden	8
2.3 Factors affecting respondent burden	9
2.3.1 Factors associated with the survey	10
2.3.2 Factors associated with respondents	11
2.3.3 Factors associated with interaction among respondents and the survey	12
2.4 Analysis and modelling respondent burden	15
CHAPTER 3 DATA AND METHODOLOGY	27
3.1 Data	27
3.1.1 Sociodemographic and COVID related questions	28
3.1.2 Burden related questions	29
3.2 Latent Factor Analysis	30
3.2.1 Data Preparation.....	31
3.2.2 Structural Equation Modeling	31
3.2.3 Confirmatory Factor Analysis (Measurement Model)	32
3.3 Classification Analysis.....	34
3.3.1 Data Preparation.....	35

3.3.2	Decision Tree Modeling.....	35
3.3.3	Random Forest Modeling.....	38
3.3.4	Comparison Measures	39
3.4	Regression Analysis	42
3.4.1	Data Preparation	42
3.4.2	Multiple Regression Analysis	43
CHAPTER 4	RESULTS ON BURDEN AND SURVEY DURATION ANALYSIS	45
4.1	Descriptive Statistics	45
4.2	Latent Factor Analysis Results.....	50
4.2.1	Measurement Model.....	50
4.2.2	Structural Model.....	55
4.3	Classification Analysis Results and Comparisons	58
4.3.1	Classification and regression tree (CART)	58
4.3.2	Conditional Inference Tree (Ctree)	61
4.3.3	Random Forest results.....	64
4.3.4	Comparison among classification methods.....	68
4.4	Regression Analysis	69
4.4.1	Multiple Regressions.....	70
CHAPTER 5	DESIGN OF THE FIRST NATIONAL SURVEY ON TRANSPORT EQUITY	78
5.1	Mobilizing Justice Project and its Goals.....	78
5.2	Contribution to the Survey Design.....	78
5.3	Introduction to the Survey.....	79
5.4	Survey Design	81
5.4.1	Workshop on National Survey	81
5.4.2	CEAT Formation.....	83
5.4.3	Question pool	89
5.4.4	Mobilizing Justice Survey Test.....	94
CHAPTER 6	CONCLUSION	98
6.1	Summary and Contribution	98
6.2	Limitations	101

6.3	Research perspectives	102
CHAPTER 7	REFERENCES.....	104

LIST OF TABLES

Table 2.1 Factors affecting respondent burden and participation rate, replicated from (Keshavarz and Morency 2023)	13
Table 2.2 Details of effective factors on perceived burden and gratification, replicated from (Haraldsen 2004)	17
Table 2.3 Measurement model: indicators used for respondent burden dimensions, replicated from (Fricker, Kreisler et al. 2012)	19
Table 2.4 List of latent constructs and associated indicators in the SEM model, replicated from (Fricker, Yan et al. 2014)	21
Table 2.5 Indicators forming latent constructs and the source of data, replicated from (Yan, Fricker et al. 2020)	22
Table 2.6 Burden point based on type and complexity of questions, replicated from (Schmid and Axhausen 2019).....	23
Table 2.7 Summary of models and variables used for the analysis of burden in previous studies, replicated from (Keshavarz and Morency 2023).....	25
Table 4.1 Descriptive statistic of the sample sociodemographic (n=1507)	45
Table 4.3 Descriptive statistics of the observed variables	46
Table 4.4 Latent constructs and their indicators and survey questions	52
Table 4.5 Measurement model validity and reliability, square root of AVE in bold and correlation between constructs in off-diagonal	54
Table 4.6 Relationship among constructs based on SEM estimates.	58
Table 4.7 Performance metrics of the decision tree according to the confusion matrix	61
Table 4.8 Performance metrics of the CTree according to confusion matrix.	64
Table 4.9 Performance metrics of the Random Forest according to confusion matrix.....	68
Table 4.10 Model performance scores and comparison.....	68
Table 4.11 The results of multiple regression models	75
Table 5.1 Elements discussed in the workshop (replicated from brainstorm session report)	82
Table 5.2 Content of pre-meeting survey	84
Table 5.3 Feedbacks from CEAT on second component of pre-meeting survey	84

LIST OF FIGURES

Figure 1.1 Schematic representation of the study	6
Figure 2.1 Respondent burden drivers (Bradburn 1978).....	16
Figure 2.2 Theoretical model representing the relationship between response burden and effective elements, replicated from (Fisher and Kydonieffs 2001)	16
Figure 2.3 The interaction between respondent and the survey leading to response burden and gratification, replicated from (Haraldsen 2004)	17
Figure 2.4 PLS path model of burden, replicated from (Fricker, Kreisler et al. 2012).....	20
Figure 2.5 Structural model of respondent burden, replicated from (Fricker, Yan et al. 2014)	20
Figure 2.6 SEM formation of response burden, replicated from (Yan, Fricker et al. 2020).....	22
Figure 3.1 An overview of survey sections and questions layout (montage of different screens).	28
Figure 3.2 Graphical interface of survey tool and design of survey questions	30
Figure 3.3 Confusion matrix of a binary classification problem.....	40
Figure 3.4 Distribution of measured duration in minutes	43
Figure 4.1 Distribution of measured variables	48
Figure 4.2 Box plot presenting the distribution of observed variables.	48
Figure 4.3 Perception of survey duration	49
Figure 4.4 Correlation matrix among measured variables.	49
Figure 4.5 Path diagram of measurement model specification, a three-factor solution (CFA model).....	52
Figure 4.6 Path diagram of structural and measurement models specifications	56
Figure 4.7 Model performance with varying complexity parameter (cp)	59
Figure 4.8 Results of classification and regression tree (CART).....	60
Figure 4.9 Confusion matrix of classification and regression tree (CARET).....	61
Figure 4.10 Results of conditional inference tree (CTree).....	63
Figure 4.11 Confusion matrix of conditional inference tree (CTree)	63
Figure 4.12 Exploring model tuning: OOB Error vs. <i>mtry</i>	65
Figure 4.13 Exploring model tuning: OOB Error vs. number of trees in the forest.	65
Figure 4.14 Variable importance based on Mean Decrease Gini and Mean Decrease Accuracy metrics in Random Forest	67
Figure 4.15 Confusion matrix of Random Forest	67

Figure 4.16 Average number of different question types in each survey section	70
Figure 4.17 Minimum, mean, and maximum survey measured duration in each survey section ..	70
Figure 4.18 Correlation matrix between predictors and target variables	71
Figure 4.19 Box plot presenting the distribution of various question types	71
Figure 4.20 Residual analysis of measured duration model, predicted values vs. Real values (left), and residual plot (right)	76
Figure 4.21 Residual analysis of perceived duration model, predicted values vs. Real values (left), and residual plot (right)	77
Figure 5.1 CEAT member feedback question types and perceived complexity	85
Figure 5.2 CEAT members feedback on optimal survey duration.....	86
Figure 5.3 Schematic overview of pilot survey design	89
Figure 5.4 Elements associated with home location	92
Figure 5.5 Variables associated to mobility anchor point.....	93
Figure 5.6 Variables associated to well-being anchor point	94
Figure 5.7 An overview of Mobilizing Justice survey sections and questions layout (montage of different screens)	97

LIST OF SYMBOLS AND ABBREVIATIONS

AVE	Average Variance Extracted
CART	Classification and Regression Tree
CEAT	Community Equity Advisory Table
CFA	Confirmatory Factor Analysis
CTree	Conditional Inference Tree
CR	Construct Reliability
DT	Decision Tree
DWLS	Diagonally Weighted Least Square
EFA	Explanatory Factor Analysis
FN	False Negative
FP	False Positive
FP	False Positive
lavaan	Latent Factor Analysis
MDA	Mean Decrease Gini
MDG	Mean Decrease Accuracy
MGA	Multigroup Analysis
MJ	Mobilizing Justice
HMLM	Hierarchical Multilevel Linear Model
OOB	Out-of-Bag
OLS	Ordinary Least Square
PLS	Partial Least Square
PB	Perceived Burden
PE	Perceived Effort
PI	Perceived Interest
PL	Perceived Length
RF	Random Forest
SEM	Structural Equation Modelling
SSHRC	Social Science and Humanities Research Council
TN	True Negative
TP	True Positive

CHAPTER 1 INTRODUCTION

Data collection through surveys is a long established method in the field of transportation and mobility behaviour. Nowadays, by evolving web technologies, web-based surveys have gained more interest due to their ability to target a broader audience and by being less costly, and less time-consuming for survey administrators.

Despite this, surveys themselves and the platforms on which surveys are launched are not flawless. Survey drop-outs and low response rates are among the most pressing challenges researchers and surveying institutions face in this regard. Response rate is influenced by various factors; therefore, identifying these factors is crucial for making changes. The respondent burden refers to “the effort required by a respondent to answer a questionnaire” (Briz-Redón. Á., 2021). It is negatively associated with completion rate and the quality of data gathered by the survey. Furthermore, the completion rate is a very important aspect of any survey, as without an appropriate level of completion rate, the results and the conclusions derived from the survey can be deemed unreliable. In general, a rate around 15%-20% is considered appropriate. However, this value can be higher when the sample size is small or less when the sample size is truly big (Lund, B., 2021). In designing surveys, one of the objectives is to minimize the respondent burden to achieve the highest possible completion rate and to collect data of the highest quality. To meet this objective, the very first step is to precisely understand the concept of respondent burden and to recognize the factors affecting it. Previous works have defined and measured burden in different ways. Some studies only considered the objective measures of the survey such as number of questions and their type to measure burden. Some used proxy measures such as survey duration to quantify burden. Other research believed that burden should be defined in a subjective context and the way a survey is perceived by respondents. Hoogendoorn (2004) made a distinction among actual and perceived burden and demonstrated that the former can be measured by considering the actual survey features and the latter by accounting for how people feel about the survey.

The most precise definition is the one that considers burden as a multidimensional concept reflecting the potential effect of survey features, characteristics of respondents, and how these two interact with each other. Survey length, number of questions, question types, graphical interface, and sponsorship are regarded as factors affecting burden associated with the survey. Actual burden can be measured through these observed survey characteristics. Subjective burden can be

quantified by post-survey questions that directly investigate how respondents perceive the surveying task. Interest in the topic, difficulty, cognitive complexity of the task, survey frequency, and perceived value can be assessed for their impact on burden by post-survey questions. Among the factors associated with respondents influencing burden, age, gender, race, household size, and educational level are found to be affecting burden. Following the identification of factors that affect burden, metrics can be defined to measure and quantify respondent burden. The next step includes developing models to investigate the importance and the power of different indicators in burden formation. Once a concise and accurate definition and measurement of burden is established, the relationship between burden and response rate can be explored. The methodology and models could then be generalized to suit other survey methods.

1.1 Problem statement

To the authors knowledge, few studies were carried out measuring burden as a subjective phenomenon and the research was limited to quantifying actual burden. This can be due to lack of methodology development capable of assessing perceived burden. Moreover, limited studies examined perception of burden in web-based surveys and research is more focused on telephone and face-to-face interviews. Also, a measure of burden has never been used in a transport-related survey, nor have any transportation-related indicators been developed in this regard. Since web surveys are used quite frequently to gather data associated with transport and mobility behaviour, measuring burden in these survey types is very constructive. Despite survey duration often being used as a proxy or indicator of burden assessment, no research is done to investigate what leads respondents to perceive a survey as either lengthy or short. In fact, the objective is to explore different predictors and different thresholds to determine at what point respondents call a survey short or long. A survey lasting 20 minutes, for example, may be perceived as lengthy by respondents. Or the joint effect of survey duration and other influential factors such as perceived difficulty and perceived interest contribute to its lengthiness. This is something that has never been studied in previous works. Lastly, with advancement of graphical design and surveys interface, new question types such as choosing on maps to declare locations (associated to travel surveys) and sliders to gather continuous data instead of Likert scale questions have been developed. Investigating the influence of different question types on survey duration, both measured and perceived duration, is a valuable area of research. The findings are helpful in developing survey

design strategies to maintain a high response rate and to increase response quality by keeping respondents interested in the surveys.

The main research questions to be answered in this study are as followed:

1. How to define and measure subjective burden in a web-based transportation related survey using respondents self-reported assessment?
2. How to identify different predictors and thresholds which determine perceived survey duration?
3. How do survey features such as question types, and respondents' characteristics affect actual and perceived duration of a survey?
4. How to use the knowledge gained in this research to design the first national survey on transport equity and social exclusion?

1.2 Objectives

The main goal of this research is to develop methods to define and measure subjective burden in web-surveys related to transportation and mobility behaviour. Identifying the important factors contributing to burden is vital to propose strategies by keeping respondent burden level to the lowest possible level. The following includes the principal objectives of this research project:

1. Identify elements influencing respondent burden.
2. Develop a model to quantify subjective burden to evaluate the strength of contribution of each factor in burden creation.
3. Investigate different thresholds and joint effects of various predictors in respondent's perception of survey duration.
4. Provide strategies to reduce survey duration based on objective survey characteristics such as type and complexity of questions.
5. Incorporate the knowledge gained from the beforementioned objectives in the design of the first Canadian national survey on transport equity and social exclusion.

This work aims to satisfy the research objectives using data coming from a survey associated with changes of mobility behaviour of Quebec residents before, during, and after (anticipated behaviour) the COVID era.

1.3 Thesis structure

The first chapter of this document provides an introduction on the importance of surveys in gathering data in different fields of science and transportation domain. Respondent burden is then introduced as a primary driver of survey drop-out and low participation rate. Research gaps are recognised in this chapter which helps defining the research objectives and developing appropriate strategies to carry out this analysis.

The second chapter provides a complete review of previous work carried out to conceptualise burden and its different types. The section also includes an overview of different drivers of respondent burden including the ones associated to the questionnaire, ones related to respondents' characteristics, and ones related to the interaction between survey takers and the survey. The section ends with an outline of different models developed whether to measure burden as the target variable or as a predictor variable resulting in an outcome.

Chapter three starts with an exploration of data used to conduct this research followed by an overview of different methods employed. The data is derived from a survey of Quebec residents mobility behaviour in pre- and post-COVID era. The methodology is divided in three sections:

1. Latent factor analysis to measure subjective burden using Structural Equation Modelling (SEM).
2. Classification analysis to assess respondents' perception of survey length using multiple machine learning algorithms including Decision Trees and Random Forest.
3. Regression analysis to investigate the relationship among observed survey features such as type and complexity of questions and actual and perceived duration using two multiple regression models.

In section four, in the first part, a thorough discussion of the results and outcomes retrieved from employing latent factor, classification, and regression analysis is provided.

Section five is considered a case study in survey design. In this part, the steps taken to design the first pilot of a national survey concerning transportation equity and social exclusion are discussed. The survey is designed with a focus on safety, suppressed, and excess travel, and access to opportunities targeting the low-income population. The objective is to shed some light to overcome mobility related challenges in future urban and transportation planning in Canadian cities. In the process of the survey design, the goal is to implement the results and learnings from section four, findings from previous literature, as well as the knowledge and expertise of different teams involved in the Mobilizing Justice project in order to better meet the survey objectives. In the following design iterations, along with fulfilling the project primary objectives, different survey scenarios will be examined in terms of burden and attempts will be made to decrease the burden level of the final version.

Lastly, section six summarizes the results of the analysis and provides a conclusion of the work conducted in this research. It continues by discussing the contribution of this research, and also touches on the limitations encountered during the course of the study.

A schematic representation of the work done in this analysis is provided in Figure 1.1.

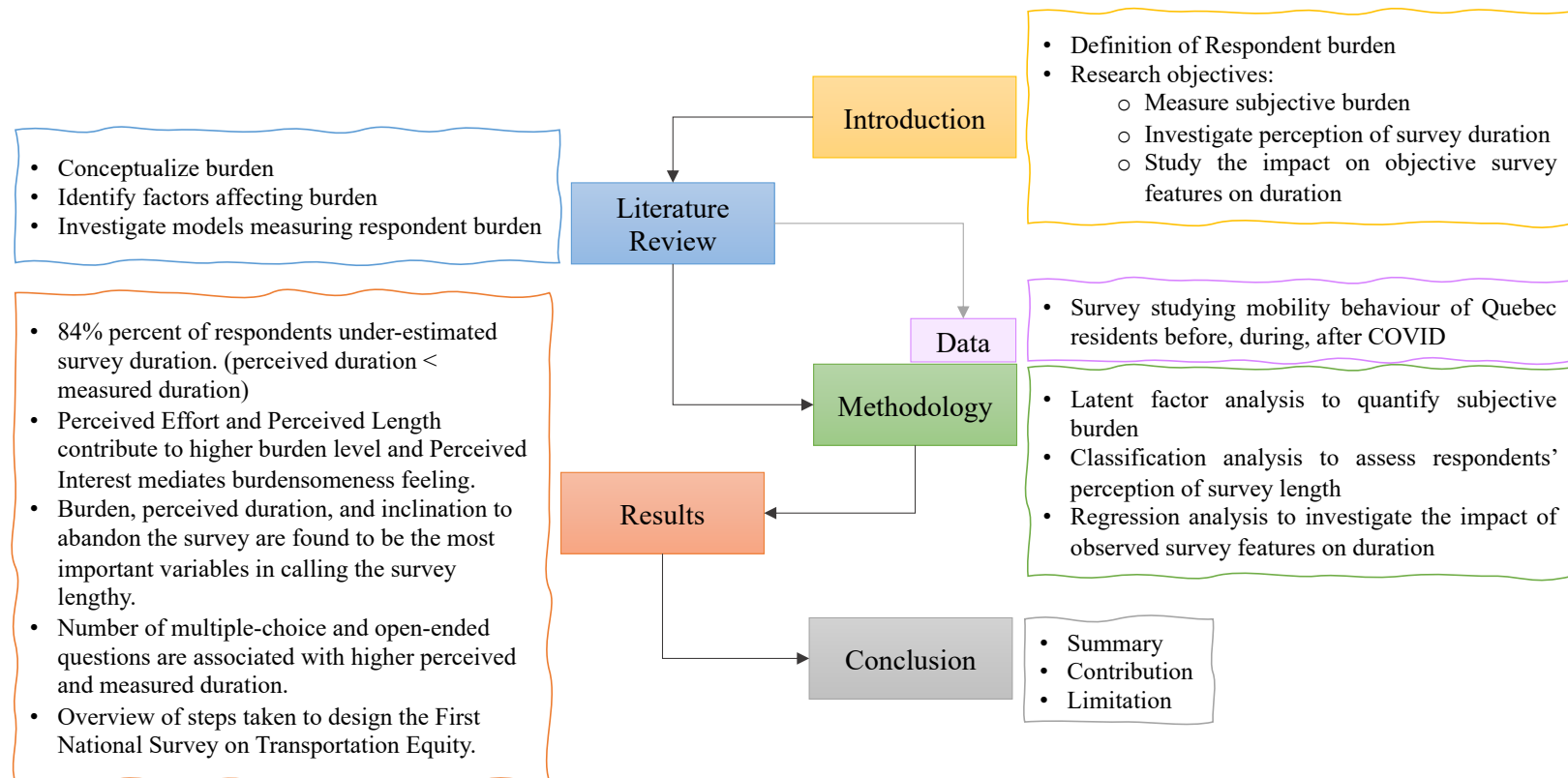


Figure 1.1 Schematic representation of the study

CHAPTER 2 LITERATURE REVIEW

2.1 History of surveys

Surveys are very important tools for collecting data in different fields of science; namely, sociology, medicine, and engineering. From face-to-face interviews to pen-and-paper surveys, web-based surveys have become increasingly popular with the advancement of web technologies. Traditionally, surveys have been used to gather data related to transport and mobility behaviour in the field of transportation (Zhang, Wang et al. 2011, Qureshi and Abdullah 2013). Using survey data helps shaping future urban planning, enhance public transportation development, improve active transit infrastructure, inform policy decisions, and enable responsible organizations to undertake effective interventions. The importance of using surveys to gather high-quality, accurate, and unbiased data becomes apparent here.

Prior to survey design, it is crucial to decide on the survey method. Research objectives, research topic, population under study, research timeline, budget, and in-hand technology all play a role in the decision to use one survey method over another (Wright 2017). Surveys conducted online are becoming increasingly popular due to their low cost and time-saving nature for survey administrators, as well as the ability to reach a broader audience, including those who are difficult/impossible to reach through other modes (Saleh and Bista 2017). Various factors including the ease of sending invitations via email, posts on social media, and links shared on the news, easy access to collected data, and being prone to errors in coding and data entry makes web-based surveys useful (Haraldsen 2004, Wright 2017). In spite of this, online surveys may present a challenge to reach marginalized populations that cannot afford expensive data plans and tools to freely participate in online surveys.

While web-based surveys are cost-effective and time-efficient, survey drop-outs and low response rates are becoming increasingly prevalent (Dykema, Jones et al. 2013). Low response rates are one of the major challenges researchers face. Therefore, it is imperative to identify the reasons why respondents abandon surveys and come up with solutions to keep them engaged. Specifically, Crawford, Couper et al. (2001) argued that web users are becoming impatient and less likely to

complete surveys because of the heavy burden of web interactions. To this point, the concept of respondent burden can be introduced.

2.2 Definition of respondent burden

Respondent burden has been defined and measured in a variety of ways in previous studies. Some research has used objective elements of surveys, such as number and type of questions or survey frequency to quantify burden (Bradburn 1978, Schmid and Axhausen 2019) and some used proxies such as survey duration to measure burden (Singer, Van Hoewyk et al. 1999, Groves, Presser et al. 2004, Hoogendoorn 2004, Bergman and Brage 2008, Giesen 2012). To be able to measure burden, the very first step is to define and conceptualize it. Researchers define burden in various ways. Frankel (1980) defined self-perceived burden as “factors associated with the presence of negative feelings such as annoyance, frustration or inconvenience which may be experienced by survey participants”. Respondent burden was defined by Sharp and Frankel (1983) as “the hardship involved in participating in a survey”. According to Lavrakas (2008), burden is “the degree to which a survey respondent perceives participation in a survey research project as difficult, time-consuming, or emotionally stressful”. A more recent study described burden as “the effort required by a respondent to answer a questionnaire” (Briz-Redón 2021).

Bradburn was the pioneer researcher who highlighted that “burdensomeness” is a subjective phenomenon related to individual’s perception of the task, rather than being solely an inherent objective feature of the survey (Bradburn 1978). Giesen and Bradburn suggested ways to measure response burden that include subjective and objective elements, implying that psychological aspects should not be underestimated (Bradburn 1978, Giesen 2012). In this regard, Hoogendoorn (2004) distinguished *actual burden* from *perceived burden*. Objective measures of survey complexity, like the number, length, and type of questions, and the frequency of interviews, used to determine the actual burden (Bradburn 1978, Hoogendoorn 2004, Yan, Fricker et al. 2014, Bottone, Modugno et al. 2021). Respondents' subjective perceptions of task burdensomeness, whether they feel burdened while completing a survey or not, are used as determinants of subjective burden. To measure subjective burden, post-survey questions can be asked to directly inquire about respondent’s self-perception of survey burden. In some cases, target groups can be defined to study respondent interaction with the survey and relative burden and effort associated with it among a smaller group of respondents. In addition, the cognitive response burden was found to be more

influential in the data quality compared to observed factors such as time (Haraldsen, Jones et al. 2013).

In regard to burden measurement, Yan, Fricker et al. (2020) considered burden as a vaguely conceptualised phenomenon that is rarely directly measured. According to them, only two studies measured burden directly, and the rest relied on its indirect measurement. First is a study that attempted to directly measure burden by periodically asking respondents about their interest and perceived burden throughout different sections of a questionnaire. The findings showed a decrease in interest and increase in burden as respondents proceed with filling out the survey (Galesic 2006). And a second study, which included a set of post-survey questions after the Consumer Expenditure Interview Survey to directly determine response burden (Yan, Fricker et al. 2014). According to Yan and Fricker, burden has been quantified using three approaches. The first approach is associated with surveys' objective burdensome features, such as survey duration, number of pages, or number of survey items to be responded (Singer, Van Hoewyk et al. 1999, Groves, Presser et al. 2004, Hoogendoorn 2004, Bergman and Brage 2008, Giesen 2012). The second approach involved assessing respondents' attitudes towards the survey by examining if the survey was useful or interesting to them. This approach measured burden not in a direct way but as a mediator of perceived burden (Sharp and Frankel 1983, Hoogendoorn 2004). In the third approach, survey burden was estimated by observing undesirable outcomes of the task, such as unwillingness to participate in future surveys and instances of survey drop-outs (Stocké and Langfeldt 2004).

Accordingly, the best way to define burden is as a multidimensional concept that takes into account respondents' characteristics, survey features, and their interactions with the survey content. It is crucial to understand the magnitude of different factors that contribute to respondent burden and develop measures to quantify these factors in order to improve data quality and reduce drop-outs. In the next section, different factors associated with response burden, response rate, and survey drop-out will be discussed.

2.3 Factors affecting respondent burden

A variety of factors contribute to response burden that can be classified in three categories: those associated with the questionnaire, those associated with respondents, and those associated with the respondent's interaction with survey through surveying task. A summary of all factors discussed above is presented in Table 2.1.

2.3.1 Factors associated with the survey

Among survey related elements, the *time* required to complete the survey is considered the most effective factor at determining burden. Previous researchers revealed a positive correlation between survey duration and drop-out rate, meaning that longer surveys are imposing higher burden on respondents (Yan, Conrad et al. 2010, Göritz 2014, Revilla and Ochoa 2017, Rudžionienė, Grigas et al. 2018). Nevertheless, a study of postal surveys demonstrated that sometimes shorter surveys with complex questions may require more effort to complete (Rolstad, Adler et al. 2011). A distinction has been made between measured and perceived survey durations and how they impact burden. The impact of perceived duration though is discussed in the third category: “factors associated with interaction between respondents and the survey”. A previous study also investigated the effect of the *announced questionnaire duration* and found that longer anticipated questionnaire duration results in lower participation rates (Galesic and Bosnjak 2009).

Revilla (2017) argued that *question format* is an important factor affecting online survey completion rates. Their study found grid format, open-ended questions, and multiple-response questions with multiple options more burdensome. A suggestion made by Smyth, Dillman et al. (2006) was to avoid multiple response questions and instead asking respondents to provide a yes or no answer on each item in a web-based survey. In line with this finding, a recent study developed a point system based on questions type and complexity, in which closed yes/no questions were assigned a point equal to 1, while open-ended questions were assigned a point equal to 6 which is much higher than the former question type (Schmid and Axhausen 2019). Concerning survey features, researchers found that there was a negative correlation between *total number of questions* and number of open questions within each block of questions in momentary interest of the survey which leads to higher burden (Galesic 2006, Backor, Golde et al. 2007). Furthermore, a study by Krosnick and Berent (1993) on different surveys including face-to-face, telephone, and self-administered questionnaires found that verbal response scales as opposed to numerical response scales discourage satisficing behaviour and lead to more accurate response. According to Hoogendoorn (2004), satisficing behaviour referred to the situation when respondents attempt to simplify their task by not changing their responses and trying to only provide acceptable answers (i.e, answer a set of Likert scale questions by choosing the neutral answer).

Regarding the survey platform, *graphical interface* is one of the most significant factors affecting web-based survey response burden. Study by Chung, Sriukenthiran et al. (2021) showed that using an appealing interface increases the likelihood of completing an online survey. Verzosa (2020) revealed the importance of gamification in respondents' interest and increasing response quality. Bourbonnais and Morency (2018) demonstrated that completing a survey on a small screen takes longer, h requires more effort. In this regard, Rudžionienė, Grigas et al. (2018) claimed that a mobile and tablet friendly design for surveys can keep respondents interested in completing surveys.

Lastly, the reputation of the survey provider and *survey sponsorship* have also been found to influence response rate. Based on multiple studies on web surveys, government- and university-sponsored surveys have higher participation rates than those coming from other sources (Walston, Lissitz et al. 2006, Fang, Wen et al. 2012).

2.3.2 Factors associated with respondents

Previous research has investigated how respondent's sociodemographic characteristics, including *age, gender, household size, and education level, race, and marital status* influence completion rate and inclination towards participation in new surveys.

In the work by Bourbonnais and Morency (2018) regarding interview durations, it was found that older respondents and those with lower education levels would require more time to complete surveys. This was in line with the results of a recent study which found that older respondents exhibit higher non-response rate in telephone and face-to-face surveys. This work also revealed that higher academic achievement correlates with lower non-response rate (Briz-Redón 2021).

In regard to respondent's *gender*, studies by Keusch (2015) and Patrick, Singer et al. (2013) uncovered a higher participation rate among female respondents in web surveys compared to men.

Concerning respondent's *race*, a study made by Patrick, Singer et al. (2013) on college students' participation in a survey related to substance use found lower inclination towards survey participation among black students compared to non-blacks. Contrary results were found in a recent research revealing no significant association between race and survey completion among respondents (Busby and Yoshida 2015).

Regarding *marital status* and survey completion, not much research is done. A recent study by Busby and Yoshida (2015) found that respondents' marital status had no significant impact on survey completion.

About *household size*, previous research found a positive correlation among respondent's household size and survey burden. This is a case in surveys that are gathered in the household scale when a survey taker acts as a proxy and declares activities on behalf of other household members (Fricker, Kreisler et al. 2012, Bourbonnais and Morency 2018).

2.3.3 Factors associated with interaction among respondents and the survey

Among factors associated with the way respondents interact with the survey, previous research has shown that perceived burden is significantly influenced by *perceived survey duration* (Fricker, Kreisler et al. 2012).

With respect to survey subject, research on telephone and online surveys found that survey participation and drop-out rate are noticeably affected by *interest* in the survey topic (Groves, Presser et al. 2004, Zillmann, Schmitz et al. 2014, Rudžionienė, Grigas et al. 2018).

Later, a set of experiments carried out by Sakshaug, Vicari et al. (2019) to examine the effect of *contact mode* on response rate. According to their study, paper invitations and paper reminders result in a higher response rate and fewer non-response items in a web-based survey.

Furthermore, the impact of *embedding a survey item in the invitation letter* and sending a *reminder message* in web-based surveys were examined. Results indicate that embedding a survey item in the follow up messages and sending funny reminders are associated with an increased likelihood of starting the survey (Moore and Varghese 2021). In line with this, in a recent work, Cobanoglu, Nanu et al. (2022) found respondents more encouraged to participate when a survey item was included in the invitation message. According to a study by Rudžionienė, Grigas et al. (2018), personalizing invitations also keeps respondents more engaged in a web-based survey.

In a recent study, Schmid and Axhausen (2019) found that *monetary incentives* were effective in increasing completion rate and reducing response burden in case of mail and mail back surveys. Moreover, according to Bottone, Modugno et al. (2021), respondent burden will be negatively affected by *cognitive complexity of the task* and *survey frequency*. This finding supports the conclusion of a previous study that found a significant effect of number of survey contacts and

perceived appropriateness of number of contacts on perceived burden (Fricker, Kreisler et al. 2012).

Lastly, respondents attitudes toward surveys, such as *enjoyment* and *perceived value*, were found to be positively correlated with response rate (Krosnick and Berent 1993, Rogelberg, Luong et al. 2000).

Table 2.1 Factors affecting respondent burden and participation rate, replicated from (Keshavarz and Morency 2023)

Factor category	Factor	Source	Findings
Associated with the survey	Survey duration	(Yan, Conrad et al. 2010, Göritz 2014, Revilla and Ochoa 2017, Rudžionienė, Grigas et al. 2018)	Longer surveys are associated with higher burden level.
	Announced duration	(Galesic and Bosnjak 2009)	Higher estimated questionnaire duration is positively related to a lower participation rate.
	Question format	(Krosnick and Berent 1993, Smyth, Dillman et al. 2006, Puleston and Sleep 2008, Grandmont, Graff et al. 2010, Revilla 2017, Schmid and Axhausen 2019)	Grid format, open-ended questions, and multiple-response questions are more burdensome. And yes/no questions impose less burden on respondents. Compared to numerical scales, verbal scales result in less satisfying behaviours.
	Number of questions	(Galesic 2006, Backor, Golde et al. 2007)	A negative relationship exists between the number of questions and the burden.
	Graphical interface and survey platform	(Bourbonnais and Morency 2018, Rudžionienė, Grigas et al. 2018, Verzosa 2020, Chung, Srikukenthiran et al. 2021)	Using innovative survey design and gamification elements make respondents more inclined to participate. A responsive design that is compatible with mobile and tablet devices increases response rates.
Associated with the respondent	Sponsorship	(Walston, Lissitz et al. 2006, Fang, Wen et al. 2012)	Higher participation rates for government- and university-sponsored surveys than those coming from other sources.
	Age	(Bourbonnais and Morency 2018, Briz-Redón 2021)	The burden is likely to be greater for older respondents.
	Gender	(Patrick, Singer et al. 2013, Keusch 2015)	Survey participation is higher among women than among men.

Table 2.1 Factors affecting respondent burden and participation rate, replicated from (Keshavarz and Morency 2023) (cont'd)

Factor category	Factor	Source	Findings
	Race/Ethnicity	(Patrick, Singer et al. 2013, Busby and Yoshida 2015)	Contrary findings on association between respondent's race and response rate.
	Marital status	(Busby and Yoshida 2015)	No significant relationship among marital status of respondents and response rate.
	Household size	(Fricker, Kreisler et al. 2012, Bourbonnais and Morency 2018, Briz-Redón 2021)	Increasing household size is associated with greater burden for respondent.
	Educational level	(Bourbonnais and Morency 2018)	Higher academic attainment is related to a higher response rate.
Associated with interaction between the respondent and the survey	Perceived duration	(Fricker, Kreisler et al. 2012)	Perceiving the survey to be long is leading to higher burden level.
	Interest in the survey	(Groves, Presser et al. 2004, Zillmann, Schmitz et al. 2014, Rudžionienė, Grigas et al. 2018)	Higher interest is associated with lower burdensome feeling.
	Contact mode	(Sakshaug, Vicari et al. 2019, Moore and Varghese 2021, Cobanoglu, Nanu et al. 2022)	Paper invitations and paper reminders led to a higher response rate and fewer non-response items in a web-based survey.
	Invitation letter and reminder	(Rudžionienė, Grigas et al. 2018, Moore and Varghese 2021, Cobanoglu, Nanu et al. 2022)	Embedding a survey item in both invitation and reminder message, personalizing the invitation letter and sending funny reminders will increase the probability of initiating the survey.
	Incentives	(Schmid and Axhausen 2019)	Positive effect of monetary incentives in completion rate.
	Complexity of the task	(Bottone, Modugno et al. 2021)	Complexity of the task negatively affected the burden of the respondent.

Table 2.1 Factors affecting respondent burden and participation rate, replicated from (Keshavarz and Morency 2023) (cont'd)

Factor category	Factor	Source	Findings
	Survey frequency (survey fatigue)	(Porter, Whitcomb et al. 2004, Backor, Golde et al. 2007, Fricker, Kreisler et al. 2012, Bottone, Modugno et al. 2021)	Frequency of survey requests causes survey fatigue and negatively affects burden.
	Enjoyment and perceived value	(Krosnick and Alwin 1987, Rogelberg, Luong et al. 2000)	Enjoyment and perceived value were found to be associated with higher response rate.

2.4 Analysis and modelling respondent burden

Having identified the effective elements of respondent burden, the next step is to determine their importance in creating burden. This section provides an overview of various models that have been used to estimate response burden both as a dependent variable and as an independent variable.

A study by Bradburn (1978) outlined four factors that contribute to burden: survey duration, respondents' effort, interview frequency, and respondents' stress and discomfort during interview (Figure 2.1). In this work, psychological dimension of response burden referred as the effort and stress involved in responding to a survey, was examined for the first time. According to this study, survey duration is not the sole driver of burden, and it is the correlation among survey duration and the required effort that leads to burden which further defines completion rate. The effort is determined based on respondents' familiarity with the subject and whether the interview includes detailed components or requires gathering data from various sources. Stress is defined as the level of discomfort experienced by respondents during an interview for example, when asked embarrassing or ego-threatening questions. This study introduced the burdensomeness feeling not as an objective feature of the survey task but as a result of the interaction between the survey content and the way it is perceived by respondents. For instance, a boring questionnaire may take longer to complete rather than a long but interesting one. In other words, interest in the survey topic mediates the burdensomeness feeling. Bradburn did not explore the contribution of these elements to response burden using any models.

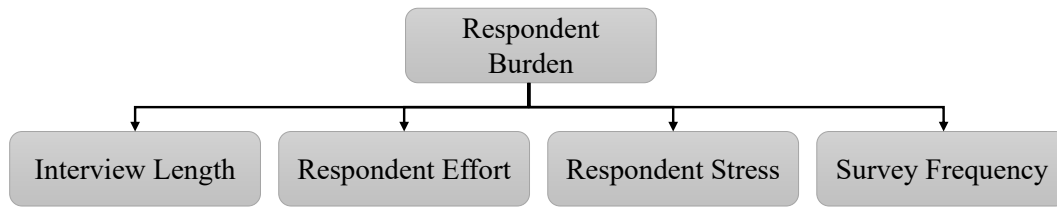


Figure 2.1 Respondent burden drivers (Bradburn 1978)

Fisher and Kydonieffs (2001) proposed a theoretical model for response burden. According to this work, burden is influenced by 32 elements separated into three categories: respondent burden, design burden, and interaction burden (Figure 2.2). Respondent burden is solely concerned with the respondents' attitudes, regardless of the survey design. Design burden, on the other hand, refers to the way in which the survey is constructed, the presentation of its content, and the methods employed for data collection. Lastly, interaction burden involves the way respondents with specific characteristics interact with the survey.

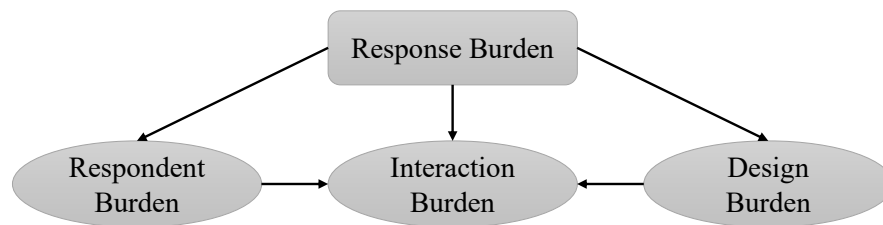


Figure 2.2 Theoretical model representing the relationship between response burden and effective elements, replicated from (Fisher and Kydonieffs 2001)

The study by Haraldsen (2004) argued that respondent burden and design burden should be considered as elements causing burden and should not be distinguished from the concept of perceived burden. In his work, he emphasizes that instead of focusing on time and cost reduction, response quality must be given priority and should not be mixed up with those causing burden. His definition of response burden is based on the balance between response burden and response quality. Moreover, according to him, being a survey respondent can also produce a sense of gratification, and the experience shouldn't necessarily be negative. He implemented his findings in a theoretical model illustrating how survey design, response burden, and response quality are related together (Figure 2.3). According to this model, survey properties are explained by four aspects of the questionnaire itself, three aspects of survey procedure, and four aspects of the data collection procedure and web survey toolbox. In terms of respondent characteristics, respondents' availability, interest, and competence were considered. The availability of the respondent is

measured by the amount of time and concentration they are willing to devote to the questionnaire. Participants' response strategies and satisfaction behaviours were observed and rated by the interviewer to determine their interest in and competence with the questionnaire by the interviewer. Details associated with the proposed theoretical model is presented in Table 2.2.

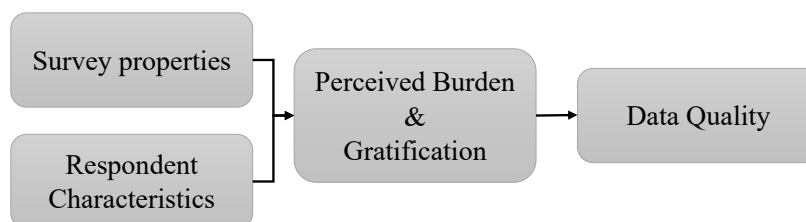


Figure 2.3 The interaction between respondent and the survey leading to response burden and gratification, replicated from (Haraldsen 2004)

Table 2.2 Details of effective factors on perceived burden and gratification, replicated from (Haraldsen 2004)

Effective factors on respondent burden	
Survey Properties	Survey Instrument: Number of Questions Question content Question flow Questionnaire layout Survey procedures: Content strategy Administrative tasks Confidentiality measures Web survey toolbox: Presentation format Navigational information Modes of communication Tailoring tools Interactive tools
Respondent Characteristics	Availability Interest Competence

Later, Galesic (2006) examined the reasons for abandoning web surveys and their aftereffects by taking the burden and interest levels into account as causes of dropout. In this work, respondents were asked to self-report their perceived burden and perceived interest at the end of each section of the survey. The questions were arranged differently in this study to examine whether dropouts are caused by specific questions or by the cumulative effect of earlier questions. To study effective factors on drop-out risk, a Cox proportional hazards regression model was used. The incentive type, the announced duration of the survey, sociodemographic characteristics of respondents (age and gender, educational level), work obligation, interest in the survey, and perceived burden level significantly impacted the dropout rate. As a result of the survey, the momentary interest dropped while respondents completed different sections and the burden was augmented at the dropout point. The findings were in line with decision field theory that predicted an increased tendency to abandon a survey prior to the dropout point. Additionally, this study examined some predictions from decision field theory showing how the quality of answers gets lower as one approaches the dropout point. In addition, Hierarchical Multilevel Linear Models (HMLM) were used to study the changes in momentary interest and burden considering the features of the question block for analysis at the first level, and respondents' characteristics at the second level.

Fricker, Kreisler et al. (2012) used Partial Least Square (PLS) modeling to measure burden associated with a Consumer Expenditure Interview Survey. Data associated with predictive variables were gathered by a set of post survey questions including respondents' perception of survey burden, duration, difficulty, interest, and frequency. These data served as indicators associated with four latent constructs. Perception of survey length (PL), perception of effort (PE), and perception of frequency of the interview requests (PF) were implemented into the PLS model as latent variables to predict perceived burden (PB). Details regarding indicators associated with each latent construct are presented in Table 2.3. The results show that PE, PB, and PL are significantly associated with perceived burden. Figure 2.4 shows the impact and significance level of independent constructs on burden formation (Fricker, Kreisler et al. 2012).

Table 2.3 Measurement model: indicators used for respondent burden dimensions, replicated from (Fricker, Kreisler et al. 2012)

Item	Question & response options
Dimension: perceived length (PL)	
Perceived survey length (final wave)	Do you feel that the length of today's interview was too long, too short, or about right? <i>1= too short; 2= about right; 3= too long</i>
Household size	Number of members in the household
Estimated duration (final wave)	How long do you think today's interview took? <i>Response in minutes</i>
Dimension: perceived effort (PE)	
Interest in survey topic	How interesting was this survey to you? <i>1=very interesting; 2=somewhat; 3=not very; 4=not at all</i>
Perceived ease in responding to survey questions	How difficult or easy was it for you to answer the questions in this survey? <i>1=easy; 2=somewhat easy; 3=some difficult; 4=very difficult</i>
Dimension: perceived frequency (PF)	
Number of contact attempts	Number of contact attempts recorded in the contact history for the sample unit
Perceived appropriateness of the number of contact attempts	Thinking about the number of phone calls you received before each interview, would you say that it was too many, or did it seem like a reasonable number? <i>1=Reasonable; 2=Too many</i>
Perceived appropriateness of the number of survey requests in the survey panel	Throughout the survey, you were asked to participate in several interviews. Would you say that this was too many interviews, or did it seem like a reasonable number? <i>1=Reasonable; 2=Too many</i>
Dimension: perceived burden (PB)	
Perceived burden (final wave)	How burdensome was this survey to you? <i>1= not at all burdensome; 2= not very; 3= somewhat; 4= very</i>

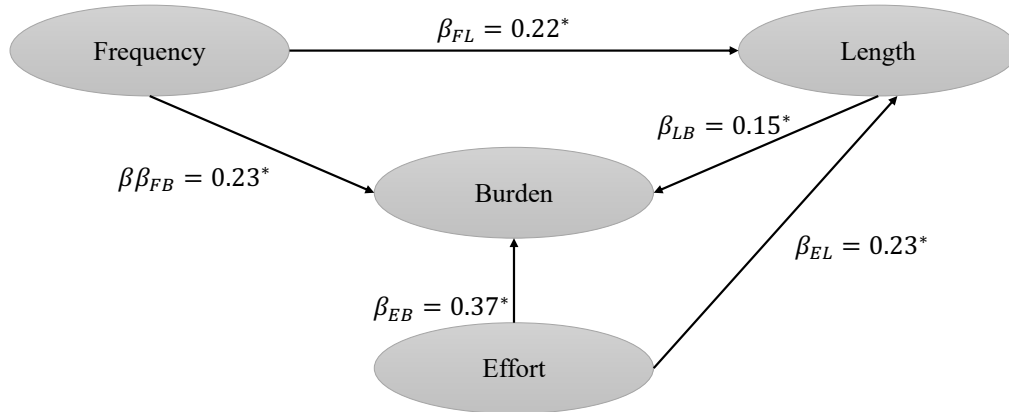


Figure 2.4 PLS path model of burden, replicated from (Fricker, Kreisler et al. 2012)

A later analysis by Yan, Fricker et al. (2014) on the Customer Expenditure Quarterly Interview Survey (face-to-face) found a significant relationship among subjective perceptions of the survey task and burden. In this work, Structural Equation Modelling (SEM) was developed to investigate how respondent motivation, task difficulty, recruitment effort, perception of survey content, and perception of survey task can lead to response burden as a subjective concept (Figure 2.5). Burden as the response variable was measured through a post survey question asking how burdensome the survey was for the respondent and predictors were all latent constructs. Details of variables associated with each latent construct are summarised in Table 2.4. This analysis revealed that survey features such as duration, which has been used a lot as a proxy measure of burden, cannot quantify burden directly. Those survey features were found to indirectly contribute to forming burden through respondent's subjective evaluation of the survey. Among the variables incorporated into the model, some indicators were measured by notes taken by interviewers, some by self-perception of respondents, and some were objective measures.

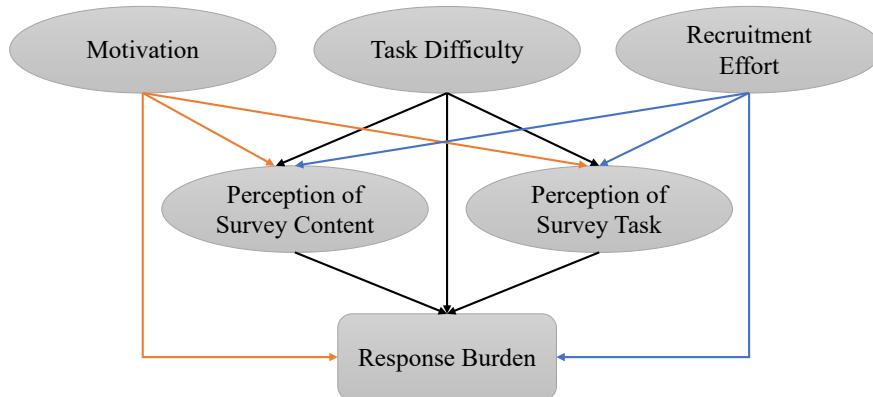


Figure 2.5 Structural model of respondent burden, replicated from (Fricker, Yan et al. 2014)

Table 2.4 List of latent constructs and associated indicators in the SEM model, replicated from (Fricker, Yan et al. 2014)

Factor	Indicator
Motivation	Cooperativeness
Respondent characteristics	Effort
	Low concern
	Trust
	Non-sensitivity
Recruitment effort	Frequent contact
Survey and task characteristics	Frequent personal visit
	Converted refusal
Task difficulty	Long interview
Survey and task characteristics	Multi-person household
Perception of survey content	Interesting survey
Respondents' perception	Easy survey
Perception of survey task	Long survey
Respondents' perception	Too many rounds

In the latest study by Yan, Fricker et al. (2020), SEM was used to investigate the direct and indirect effects of survey components, respondent characteristics, and respondents' perceptions of the survey task on burden in a set of five face-to-face and telephone interviews associated to consumer's expenditure and income. It was quite similar to the model used in their previous study (Fricker, Yan et al. 2014). Motivation, survey effort, task difficulty, and perception of the survey were considered as latent constructs leading to burden as a subjective phenomenon (Figure 2.6). An overview of variables associated to each latent construct is shown in Table 2.5. This study found that higher perceived burden is a consequence of lower motivation, more difficult task, more challenging survey effort, and negative impression of the survey task. All have significant direct effects on burden. However, the total effect of survey effort on burden formation was found to be non-significant. This study also conducted a Multiple Group Analysis (MGA) to examine the validity of the same burden model across different survey data collection modes, indicating no variance on the overall structural relationship across different data collection modes (Yan, Fricker et al. 2020).

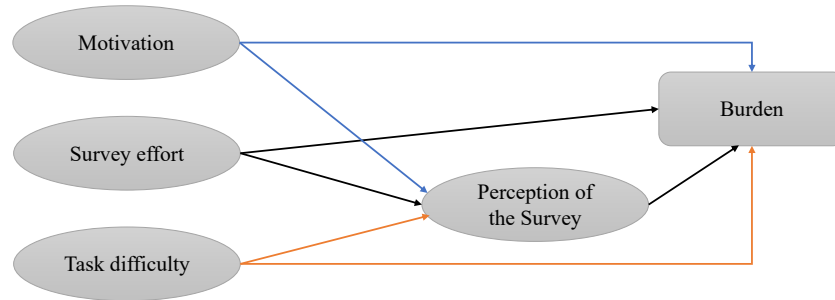


Figure 2.6 SEM formation of response burden, replicated from (Yan, Fricker et al. 2020)

Table 2.5 Indicators forming latent constructs and the source of data, replicated from (Yan, Fricker et al. 2020)

Latent construct	Indicator	Source of data
Burden		Self-report to survey item
Motivation	High concern Number of refusals Sensitivity of CE Low trust	Call records data Call records data Self-report to survey item Self-report to survey item
Task difficulty	Number of children in household Number of adults in the household Less than 65	Self-report to survey item Self-report to survey item Self-report to survey item
Survey effort	Duration of survey interview Number of interviews completed Using information book Using records	Call records data Call records data Paradata Paradata
Perception of survey	Not interesting survey Difficult survey Too many rounds of interview Too long interview	Self-report to survey item Self-report to survey item Self-report to survey item Self-report to survey item

A relatively recent study conducted by Schmid and Axhausen (2019) on a set of mail and mail-back surveys aimed to analyse the effects of response burden, monetary incentives, and respondents' recruitment levels on response rate. There were three levels of recruitment: “no prior recruitment, no incentive”, “prior recruitment, no incentive”, and “prior recruitment and incentive”. Burden scores were calculated only by considering objective features of survey questions (type and complexity of questions) without incorporating the perception dimension burden (Table 2.6). This

research considered burden as a predictor of response rate. Based on a logistic regression model, the response rate dropped by 6% for every 100 increases in response burden index. The results showed that response burden can be mediated by monetary incentives, however recruitment efforts were not found to significantly reduce burden.

Table 2.6 Burden point based on type and complexity of questions, replicated from (Schmid and Axhausen 2019)

Survey Item	Burden point
Question or transition (up to 3 lines)	2
Each additional line	1
Closed yes/no answer	1
Simple numerical answer (e.g., year of birth)	1
Rating with up to 5 possibilities	2
Rating with more than 5 possibilities	3
Left, middle, right rating	2
Scales with 3 and more grades	2
Best of ranking with cards	4
Second and each additional best ranking	3
Answer to sub-questions of up to 5 words	1
Answers to sub-question of up to 2 lines	2
a) Response to half-open question with ≤ 8 possibilities	2
Each additional one	2
b) Response to half-open question with ≥ 8 possibilities	4
Each additional one	3
Answer to “please specify”	2
First answer to an open question	6
Each additional answer to the open question	3
Mixing showcards	6
Giving/showing a card to the respondent	1
Per response category on a showcard	1
Filter	0.5
Branching	0.5
For each stated choice question with 2 alternatives	2
For each stated choice question with 3 alternatives	3
For each variable of the stated choice situation and each question	1

Bottone, Modugno et al. (2021) investigated how actual burden relates to perceived burden in a study based on two business surveys conducted with web and interactive PDFs which were conducted over several waves from 2004 to 2019. The study also aimed to investigate the direct effect of the perceived burden on data quality while controlling for other effective factors on the perceived burden. Actual burden was quantified by five indicators: the number of questionnaire pages (also includes survey duration), the number of questions, the number of people involved in

the survey, the use of external consultancies, and the average completion rate. In order to measure perceived burden, respondents were asked to rate their level of effort in completing the questionnaire. In the same regard, starting from 2017, respondents were asked to assess the degree to which the following factors made it difficult for them to complete the survey: too many questions, several people involved in answering the questions, the use of unclear terms, non-exhaustive response options, and difficulties in choosing the correct answer. Data quality was measured by unit non-response and item non-response. The analysis in this study was performed using logistic regression models. As shown in the results, there is a significant correlation between a higher actual burden and a higher measure of perceived burden. The findings indicate that, in addition to the actual burden, unobserved variables (e.g., interest in the survey) contribute to a high perception of burden. As a final point, the results showed that perceived burden directly impacts data quality while controlling for actual burden (Bottone, Modugno et al. 2021). Table 2.7 provides a summary of all models and variables used for analyzing burden which are discussed in this section.

Table 2.7 Summary of models and variables used for the analysis of burden in previous studies, replicated from (Keshavarz and Morency 2023)

Study	Study objective	Burden type	Variables used for the analysis	Used model	Survey method
(Bradburn 1978)	Identifying factors that contribute to perceived burden that are important in improving data quality	Perceived burden (dependent variable)	Survey duration Respondent effort Respondent stress Frequency of survey	No model	
(Fisher and Kydonieffs 2001)	Identifying sources of burden in surveys	Perceived burden (dependent variable)	Design burden Respondent burden Interaction burden	Theoretical model	
(Haraldsen 2004)	Finding the links between survey design, response burden and data quality	Perceived burden or gratification (dependent variable)	Survey properties Respondent characteristics	Conceptual model	Web survey
(Galesic 2006)	Understanding the reasons behind survey dropouts and the aftereffects	Perceived burden (independent variable)	Perceived burden Perceived interest Dropout rate Incentives	Cox proportional hazards regression model Hierarchical Multilevel Linear Model	Web survey
(Yan, Fricker et al. 2014)	To explore direct and indirect effect of survey characteristics, respondents' characteristics, and respondents' perception of the survey burden	Perceived burden (dependent variable)	Motivation Task difficulty Recruitment effort Perception of survey task Perception of survey content	Structural Equation Modelling	Face-to-face interview

Table 2.7 Summary of models and variables used for the analysis of burden in previous studies, replicated from (Keshavarz and Morency 2023) (cont'd)

Study	Study objective	Burden type	Variables used for the analysis	Used model	Survey method
(Fricker, Kreisler et al. 2012)	To develop burden index from indirect indicators of burden	Perceived burden (dependent variable)	Perceived duration Perceived interest Perceived frequency Perceived effort	Partial Least Square Modeling	Face-to-face interview
(Schmid and Axhausen 2019)	To study the effect of response burden, monetary incentives, and recruitment on response rate	Actual burden (independent variable)	Actual burden Recruitment type Monetary incentives Response rate	Logistic regression model	Mail and mail back survey
(Yan, Fricker et al. 2020)	To investigate direct and indirect features of survey, respondents' characteristics, and their perception of survey burden	Perceived burden (dependent variable)	Motivation Task difficulty Survey effort Perception of survey Burden	Structural Equation Modelling Multiple Group Analysis	Face-to-face and telephone interview
(Bottone, Modugno et al. 2021)	To find relationship between perceived and actual burden and then how perceived burden affects data quality	Actual burden (independent variable) perceived burden (dependent & independent variables)	Actual burden Perceived burden Data quality	Logit model Linear model	Web and Interactive pdfs

CHAPTER 3 DATA AND METHODOLOGY

There are four main parts to this study. The data used to conduct the first two analysis comes from a set of post-survey questions raised at the end of a web-based travel survey. This survey was conducted to study Quebecer's mobility behaviour before, during, and after the COVID crisis. The questions were designed to address respondents' self-assessment of survey task in terms of difficulty, interest, etc.

First, latent factor analysis is carried out to explore the effects of different items on subjective burden. The data collected from the survey serves as indicators of three latent constructs. The constructs are then used as predictor variables in a Structural Equation Modeling (SEM) to predict burden.

In the second phase, classification analysis is used to determine which factors lead people to perceive the survey as lengthy, brief, or of the right length. Machine learning techniques including Decision Tress and Random Forests are used to conduct the classification analysis. The two mentioned sections were considered more as reactive analysis meaning that it is associated with what happened earlier and is not forecasting a future event.

The third part of this study focuses on survey characteristics and respondents' demographics (e.g., question type, number of questions answered, respondent age and educational level) in order to determine the measured duration spent on answering questions and the perceived duration respondents. Although this is also considered a reactive analysis, the results can work as proactive guideline to future survey design. In simple words, this helps survey administration professionals to decide what to include in the survey and what to avoid. The data used to undertake this section also came from the same COVID survey and its administrative output.

Lastly, the fourth part of this study is related to design and administration of the first national survey on transportation justice and social exclusion by considering the things learned from the former parts.

3.1 Data

The data used to conduct this analysis comes from a web-based questionnaire which studies the mobility behaviour of Quebec residents pre-COVID, during the pandemic and after-COVID era.

The questionnaire was delivered in two languages: French and English and the invitations were sent to respondents via email followed by some announcement made on social media such as LinkedIn. The email list includes respondents to previous surveys who agreed to give their email address and expressed interest to participate in future surveys conducted by Chaire Mobilité. No surveying panel is involved and there is no incentive for respondents to participate. The sampling was opportunistic, and the target population were 17-87 years old.

The survey included different types of questions such as yes/no, open-ended, multiple-choice, Likert scale, and choose-on-map (i.e., for home address and frequently visited locations). The questions on Likert scale were all coded as sliders collecting continuous data on the scale of 0 to 100. Figure 3.1 shows an overview of questions layout and the graphical interface of the survey platform.

The figure is a montage of six screenshots from the 'mobilitystudy' survey platform. The top navigation bar across all screens reads: Home > Your household > Profile > Usual places > COVID Impacts > End.

- Top Left Screenshot:** The 'Home' screen. It features the 'mobilitystudy' logo and a welcome message: 'We are inviting you to participate in a study on the impacts of the COVID-19 pandemic on activity patterns and travel behaviours.' Below this is a 'Start' button and a note about data confidentiality. A language toggle for 'Français' is at the bottom.
- Top Middle Screenshot:** The 'Your household' section. It asks for household income range (optional), comments on the survey, and several Likert-scale questions about survey interest and time spent.
- Top Right Screenshot:** The 'COVID Impacts' section. It asks about plans to change the number of vehicles available in fall 2022 and travel modes usage (Walking, Bicycle) with sliders for frequency.
- Middle Left Screenshot:** The 'Usual places' section. It asks for the number of people living in the household, cars/vans, and scooters/motorcycles available, with options to include/exclude certain categories.
- Middle Right Screenshot:** The 'Usual places' section. It asks to specify locations visited regularly before the pandemic (before March 2020) with a list of examples and a 'Continue' button.
- Bottom Left Screenshot:** The 'Profile' section. It asks for personal information: Nickname, Age, Gender, Main occupation, and Driving license status.
- Bottom Middle Screenshot:** The 'Profile' section. It asks about main occupation before the pandemic, job loss, transit pass ownership, carsharing membership, and bicycle ownership.
- Bottom Right Screenshot:** The 'COVID Impacts' section. It contains three slider questions about the probability of being infected with COVID-19 in different scenarios (walking, personal bike, Mass Transit).

Figure 3.1 An overview of survey sections and questions layout (montage of different screens)

3.1.1 Sociodemographic and COVID related questions

The questionnaire includes six separate sections. The first three parts namely Home, Household, and Profile gather information on factual elements and sociodemographic characteristics of

respondents including: age, gender, home address, educational level, main occupation, household size, car ownership, driver's license ownership, transit pass, and shared services membership.

Section four includes questions on usual places people visited before COVID, their modes of transportation, and how their mobility behaviour changed during and after the COVID pandemic.

Section five includes two subsections. In the first subsection, the survey includes questions on residents' mobility behaviour changes during COVID. For example, if respondents are traveling more/less by a specific mode during or after COVID. In the second subsection, respondents are asked about their attitudes towards teleworking and online shopping and how these attitudes changed during and after the pandemic. After the pandemic behaviours are anticipated behaviours since the survey was administered before the official end of the pandemic.

3.1.2 Burden related questions

In the last section of the survey, a set of seven post survey questions are included. The questions were mandatory and they were asked to all participants. These questions are designed to gauge perceived difficulty, duration, burdensomeness, and interest in the survey. Respondents' assessment of survey duration (Fricker, Kreisler et al. 2012), willingness to participating in new surveys, and inclination towards abandoning the survey were also investigated throughout post-survey questions. Figure 3.2 shows the formulation of post-survey questions. These questions are considered as self-reported burdensome feeling. As seen in Figure 3.2, the questions concerning burden, interest, difficulty, and survey duration assessment are gathered by using sliders on the scale of 0 to 100. Based on the survey burden question, 0 indicates that the survey is not at all burdensome, and 100 implies that it is very burdensome. In terms of interest, 0 indicates that the survey is not at all interesting, while 100 indicates that it is a very interesting survey. Question regarding survey difficulty follows the same pattern, with 0 representing a very easy survey and 100 representing a very difficult one. Question associated with survey duration assessment has a little different formulation, such that 0 represents a survey that is too short, 40-60 (middle of the slider bar) represents the right duration, and 100 represents a survey that is too long. Willingness to participate in new surveys and inclination towards abandoning the survey were asked as yes/no questions. Perceived duration was gathered as a continuous variable indicating how many minutes respondents thought it took to complete the survey. Measured duration is gathered in minutes

through survey administration logs. To determine whether people underestimated or overestimated survey duration, the time ratio is calculated by dividing perceived duration by measured duration.

Would you be interested in participating to other studies conducted by the Chaire Mobilité at Polytechnique Montreal? ☐ Yes ☐ No

Do you feel that the length of the survey was too short, about right or too long?

Too short About right Too long

How much time do you think you spent answering this questionnaire (in minutes)?

How interesting was this survey to you?

Not at all Very interesting

How difficult or easy was it for you to answer the questions in this survey?

Very easy Very difficult

How burdensome was this survey to you?

Not at all Very burdensome

Have you been inclined, at any point, to abandon the survey? ☐ Yes ☐ No

Figure 3.2 Graphical interface of survey tool and design of survey questions

3.2 Latent Factor Analysis

One of the research objectives of this work was to identify the variables causing burden and their predictive power to explain variation in burden. Since post survey questions collected data on respondents' perception of survey feature, latent factors analysis is considered a suitable method in the context of this work. According to previous literature, Structural Equation Modeling (SEM) was found to be an appropriate method to explore the variables causing subjective burden in face-to-face and telephone interviews (Fricker, Kreisler et al. 2012). SEM is a known quantitative research technique in social science that is capable of analysing the causal relationship among multiple dependent and independent variables in a single model (Ullman and Bentler 2012). Hence, this model was used to investigate the influence of different variables on subjective burden along with their own relationship.

3.2.1 Data Preparation

The analysis relies on post survey questions data related to survey duration assessment, perceived difficulty, perceived interest, and perceived burden gathered on a scale of 0 to 100. Perceived duration collected in minutes from 1 to 60 minutes, willingness to participate in new surveys and inclination to abandon the survey were collected in binary format. Perceived burden was considered the observed dependent variable, and the other variables were regarded as predictive variables. In order to avoid negative variance estimates, using max-scaling, data collected in continuous form are re-scaled by dividing all instances of each variable by its maximum value.

3.2.2 Structural Equation Modeling

Structural equation modeling (SEM) is an analytical tool designed to model complex relationships among observed and unobserved variables as well as between multiple unobserved constructs. SEM is a technique that combines aspects of both factor analysis and multiple regressions. Observed variables, also known as manifests or indicators, are defined as variables that are measured directly (i.e., through surveys) and that actually exist in the data set. Unobserved variables also called latent variables/factors or constructs are the variables that are defined conceptually and measuring them is always accompanied by some level of error. The idea behind defining and measuring latent variables is that these concepts (i.e., perceived effort) have some levels of complexity and should be evaluated using multiple questions. Using SEM, the aim is to explain the covariance among the observed variables and to construct an observed sample covariance matrix. It will then be compared to the estimated covariance matrix to test how the theoretical model fits the data. Therefore, residuals are defined as the difference between estimated and observed covariance between any two pairs of measured variables.

SEM is composed of two primary pieces: the *measurement model* which specifies the relationship among measured and latent variables, and the *structural model* which represents the structural relationship among different latent constructs. SEM models alone are not capable of establishing causality and investigating the causal relationship among different variables. It requires undertaking either an experiment (by controlling for the causal variable) or a longitudinal analysis. Theoretical support is another means to make causality inferences. The hypothesis made in this study is based on theoretical support in order to construct a SEM model, which enables us to infer causal relationships between variables.

3.2.2.1 Structural model fit indices

When the structural model is developed, before investigating the relationship among different factors, goodness-of-fit measures are examined to ensure fit validity. Chi-square statistic with p-value smaller than 0.005 results in rejecting the null hypothesis. Since χ^2 statistic is a very sensitive global fit with higher sample sizes, it may cause biased results of p-value. Therefore, different metrics of approximate fit indexes that are not based on rejecting or accepting null hypothesis, such as absolute and incremental fit indices, were investigated. Incremental fit indices are different from absolute fit indices and tell us how well the estimated model (user model) fits the data in comparison to baseline models which are the worst fitted model by assuming no correlation among measured variables. The standard Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI) values above 0.95 indicate a very good fit (Hu and Bentler 1999, Sarmento and Costa 2019).

Among absolute fit indexes which compare the user model to the observed data, Root Mean Square Error of Approximation (RMSEA) was found appropriate in confirmatory contexts (Rigdon 1996). Hu and Bentler (1999) argued a good fit is indicated with RMSEA smaller than 0.06.

Prior to building up the structural model, measurement model also known as Confirmatory Factor Analysis (CFA) was undertaken to verify the factor structure and factor identification based on previous studies.

3.2.3 Confirmatory Factor Analysis (Measurement Model)

To confirm the previous theory and to investigate how new variables contribute to the model, CFA was carried out. CFA is a multivariate approach that tests how well a measurement theory describing relationships between observed variables and latent constructs fits the reality of data. The measurement model was built based on theoretical evidence and knowledge gained from previous literature. Followed by CFA, modification indices, derived from the chi-square goodness-of-fit differences between the current model and the modified one, has been used to improve the model by modifying the paths among measurement model variables. Modification index will represent how much the overall model χ^2 will decrease by adding new paths to be freely estimated by the model. During the course of model improvement, the research hypothesis and research questions were considered.

The final measurement model included three latent constructs each having two associated observed variables. In measurement model development, no cross-loading (one observed variable indicating on more than one construct), no within-construct error covariance, and no between-construct error covariance are hypothesized. Thus, the model can be called congeneric. However, it is worth noting that with limited variables in hand, the three-item rule was not accomplished. The relationship among latent variables and their indicators were considered reflective. According to reflective measurement theory, latent constructs cause the measured variables, and the error terms shows their inability to fully explain the measured variables.

The next step was to set the scale of latent variables. There are two ways to set the scale of latent factors in CFA. The first one is to fix one of the factor loadings on each construct to 1. The second is to fix the value of the variance of the latent constructs to 1 and letting the load factors to be freely estimated. This work was conducted in the latter manner (Hair and Babin 2018).

The analysis was carried out in R using the Lavaan (Latent Variable Analysis) package. Since two manifests were binary variables, instead of the Lavaan's default estimator which is Maximum Likelihood, another estimator know as Diagonally Weighted Least Squares (DWLS) that is based on polychronic correlation matrices was employed to carry out CFA (Xia and Yang 2019).

3.2.3.1 Measurement model assessment

3.2.3.1.1 Fit validity

Construct validity indicates how accurately the observed variables represent their intended theoretical constructs; hence, it is concerned with measurement accuracy. Prior to investigating measurement model validity, goodness-of-fit static measures are examined to ensure fit validity. Like structural model, fit validity, Chi-square statistics, CFI, TLI, and RMSEA metrics are required investigating to be within the determined thresholds.

3.2.3.1.2 Measurement Model Reliability and Validity Assessment

After confirming fit validity and prior to testing the relationships significance in the structural model, the measurement model should meet a certain level of validity and reliability. *Convergent validity* establishment indicates that manifests of a construct have a high proportion of variance in common. The first step in studying convergent validity is to determine the size and significance of path estimates. In this regard, previous researchers argued that loading values over 0.6 imply an

acceptable relation among manifests and the constructs they indicate (Malhotra, Kim et al. 2006, Hair 2010). Other works though used factor loadings greater than 0.4 as a criterion to keep items in the model (Nunnally 1978, Mehdizadeh, Shariat-Mohaymany et al. 2018).

In the second step, construct reliability (CR) value, another indicator of convergent validity, was studied to assess internal consistency termed *Reliability* (Raykov 1997). According to internal consistency, indicators of a scale should all measure the same construct and be highly intercorrelated. CR values between 0.6 and 0.7 are considered acceptable and represent the existence of internal consistency in exploratory research (Shrestha 2021). Cronbach Alpha is another indicator of reliability which assumes equally weighted indicator loadings. Nevertheless, given that the model is built based on reflectively measured constructs, only CR that is considered more reliable to evaluate construct reliability was used in this analysis (Hair and Babin 2018).

The amount of variation explained in an item by its associated construct is calculated by squaring the standardized factor loading. This latter value is also referred to as variance extracted or communality. As the third indicator of Convergent Validity, Average Variance Extracted (AVE) is calculated as the mean variance extracted from all items loading on a construct (squared standardized factor loadings of each manifest divided by the number of indicators associated to each factor). A value of AVE equal to or higher than 0.5 indicates adequate convergence. Thus, the construct is strongly correlated with the factors it indicates, defining its unidimensionality and reliability (Hair and Babin 2018). Thus, unidimensionality exists when a set of variables can only be explained by a single construct.

Having assessed the Convergent Validity and model reliability, the next step is to evaluate the Discriminant Validity (Divergent validity). *Discriminant validity* is the process of determining whether an indicator is measuring one construct, but not another. According to the Fornell-Larcker criterion, discriminant validity is fulfilled if the square root of the AVE for each latent variable is greater than the correlation value among the latent variables (Fornell and Larcker 1981).

3.3 Classification Analysis

The result of the SEM model suggests that survey duration assessment is strongly and positively associated with burden. Consequently, people are burdened out more during longer surveys. We aimed to study how respondents' perceptions of different aspects of the survey led them to evaluate

the survey duration as short, correct, or long by moving the slider. In this regard, we carried out binary classification analysis which are data mining methods assigning each instance in the data to a target group. The classification models are used to investigate how the combined effect of different variables and respondents' attitudes influence perceptions of how lengthy or brief the survey is.

3.3.1 Data Preparation

As mentioned earlier, the data associated to survey duration assessment was gathered as a continuous variable on a scale of 1 to 100 by asking “*Do you think the length of the survey was too short, about right, or too long?*”. This variable was found to have a great load factor associated to Perceived Length which was then used to predict burden in our latent factor analysis according to the CFA analysis. Hence, to carry out the classification tree analysis, we decided to discretize survey duration assessment variable and use it as a categorical response variable. We first ran a regression tree with burden as the response variable and found the cut-off point of 65 to split survey duration assessment variable. Thus, survey duration assessment is then categorised into two classes:

1. Perceived “Correct/Short”
2. Perceived “Long”

Since a decision tree can process multimodal training data (Du and Zhan 2002), the predictors (perceived burden, survey difficulty, interest, perceived time to complete the survey (min), and measured time (min)) are used as continuous variables, while willingness to participate in new surveys and inclination towards dropping out are implemented as binary data in the models. Respondents’ gender, age group, educational level and the platform used to complete the survey are included in the model as control variables.

3.3.2 Decision Tree Modeling

A decision tree (DT) algorithm divides the predictive space into more pure or homogeneous regions referred to as nodes. Decision trees are non-parametric supervised machine learning algorithms that are used for both classification and regression analysis. Although DTs are not prone to overfitting, they are famous for being easy to interpret, and being non-parametric meaning that there is no need to select unimodal predictive variables (Prajwala 2015). Since this work aims to

predict a categorical variable, decision tree classification is used. Starting from the topmost node referred to as the root node, decision trees split the data in a recursive manner to corresponding sub nodes and then to reach the most possible pure leaf nodes. DTs use different splitting criterion namely Gini Index and Entropy to learn the data and segment the features in the best possible way. The prior literature indicated that no criterion can be suggested as better than the other (Raileanu and Stoffel 2004). The Gini impurity is used as a splitting criterion in this study to partition the training dataset.

3.3.2.1 Classification and Regression tree (CART)

This study used Classification and Regression Tree (CART) technique, a tree based technique introduced by (Quinlan 1987, Breiman 2017). The goal was to create a binary tree that divides the data into two classes. There are some drawbacks associated with CART algorithm: first, it is biased in variable selection and tends to be influenced by predictor variables with larger number of outpoints and extreme values. This means that variables with outliers may be given a higher priority because the algorithm strives to maximize the separation between classes (Strobl 2014). Second, it is not prone to overfitting, and finally, no statistically significant relationships are included among variables and a significant or non-significant improvement in the tree is not differentiated (Hothorn, Hornik et al. 2015). Thus, unbiased Conditional Inference Trees (CTree) modeling has been further carried out in this work (Hothorn, Hornik et al. 2006).

Before building the trees, the data is split into training and testing datasets. The trees are trained using the training dataset and their performances are then evaluated on the testing dataset. The ratio of 0.8 is used to split the data.

Growing the tree, pruning the tree to overcome overfitting, and selecting the optimal tree are all steps taken in the CART process. The root node of the tree holds the whole learning sample. Next, the algorithm searches through all predictive features and splits the feature according to its popular default splitting criterion, Gini Index, in order to create the most homogeneous child-nodes. The Gini index is a measure of impurity which explains a certain class's purity after splitting along an attribute (Tangirala 2020). In a binary tree with two classes the Gini index is calculated by the following formula (equation 3.1):

$$Gini = 1 - \sum_{i=1}^c (P_i)^2 \quad 3.1$$

Where P_i is the relative frequency of observations belonging to a class c for a particular node. The Gini impurity ranges from 0 to 1, where 1 implies a pure node, in which all instances belong to one class, while 0 indicates an even distribution of instances among two classes, indicating maximum impurity or randomness. The splitting threshold is the value or condition that determines how instances are divided into different branches.

With each split the algorithm tries to minimize impurity, meaning that the algorithm tries to create each child node with less impurity compared to the parent node. This process will continue recursively until some pre-defined conditions are met or the moment that no remaining attributes exist for partitioning. As the tree is built, some branches may reflect noise or outliers on the training data, resulting in overfitting of the training dataset. To avoid the overfitting issue, pruning must then be performed to remove these branches and improve the accuracy of the tree's classification on testing data (Sheng and Gengxin 2010). To prune the tree, optimal values for hyperparameters should be selected, a process referred as hyperparameter tuning, which should be done before the model starts to learn using the training data. Since default hyperparameter values of the DT are not always promising in providing optimal model performance, complexity parameter (cp) is used to control the tree's growth. Model accuracy is set as the scoring metric for model performance, meaning that the cp value associated with the highest accuracy was used as the final value. A grid associated to cp with sequence of values from 0 to 0.1 with an increment of 0.001 is generated. Then 10-fold repeated cross validation with 3 repeats is used through cp grid and the optimal value is determined by the model achieving higher average accuracy across all folds. A K-fold cross validation is used to minimize the bias associated with random sampling. This technique splits the dataset into K subsets, in which K-1 folds will be used to train the model and 1-fold to test its prediction accuracy (Wong and Yeh 2019).

3.3.2.2 Conditional Inference trees (CTree)

Conditional Inference trees are constructed by recursively splitting the data according to statistically significant relationships between predictors and the target variable. In other words, a statistical approach which considers the distributional properties of variables is used to partition data (White and Liu 1994). Beginning at the root node, the algorithm seeks out the variable with

the highest significant association with the response variable using statistical tests usually permutation tests or chi-square test. Typically, p-values with a lower value indicate a more significant relationship between predictors and response variables. A split threshold is determined by testing different thresholds and conditions and identifying the threshold value that yields the greatest significance. As a result of this process, the most informative features and the best thresholds/conditions are determined to split the data. The splitting continues until a maximum depth of 5 is reached.

This work uses the caret R package and the party R package to implement the CART and Ctree algorithms, respectively. Both models were tested using the caret R package for their predictive performance.

3.3.3 Random Forest Modeling

Random Forest (RF) is a supervised ensemble learning algorithm that can be used for both classification and regression purposes (Breiman 2001). The advantages of RFs over other statistical classifiers include: high classification accuracy, able to report on variable importance, allows modeling complex interactions among predictor variables, and is robust with respect to noise and outlier by averaging the results (Breiman 2001, Cutler, Edwards Jr et al. 2007). RF uses the bagging technique that involves bootstrapping the data set to create a forest of several uncorrelated decision trees and then aggregate their results and averaging them to produce more accurate outcomes. By reducing correlation between trees, RF aims to increase the variance reduction of bagging. To achieve this, during tree development, split variables are chosen at random from among a subset of all variables a process which is called feature randomization (Kabir, Guikema et al. 2018). Random selection reduces error rates and increases the algorithm's predictive accuracy.

To create an ensemble of trees, the following steps are done by the RF algorithm. First, the algorithm uses 'bootstrapping' resampling technique to generate multiple random subsets with replacement (bootstrap sample) from the original data set, each having the same size as the original data set. Each bootstrap sample is further used to create a decision tree. In order to grow the tree, the RF algorithm uses the "random subspace" technique, where a random number of features are selected at each split point and the sample is partitioned in a recursive manner (Ho 1998). Typically, the number of features considered at each split is equal to the square root of the total number of features in the model to introduce variability and reduce correlation among the trees. The process

is repeated for each bootstrap subset to train multiple trees and to combine their prediction, an ensemble learning algorithm called bagging. When examining the model performance on the test data set, the final prediction on each test instance is determined by majority voting among predictions done by a unique decision tree. For each test instance, a class is predicted by each decision tree and the final predicted class is the class that receives the most votes.

To carry out RF analysis, it is not required to split the data into training and testing sets. Instead, a concept named Out-of-Bag (OOB) is used. OOB samples are the ones that are not selected in the bootstrapping process and that are used as test data set at the end to evaluate the model's performance. When deciding on the number of features used at each split, we choose the value associated with the lowest OOB error rate. OOB is also used to identify the total number of trees created by the algorithm. Lastly, the Gini Impurity indicator is used as the splitting criterion to assure the trees are developed in a way to create nodes with the highest purity (Equation 3.1).

Another advantage of RF is that it provides measures to evaluate the importance of each predictive variable. Mean Decrease in Gini and Mean Decrease in Accuracy are popular methods of estimating the importance of predictor variables in a classification task. As a measure of feature importance, Mean Decrease in Gini measures the average Gini impurity decrease across all decision trees within a Random Forest. The more important feature is the one with the larger reduction in Gini impurity. Using Mean Decrease in Accuracy, a feature's importance is determined by how much prediction accuracy decreases when one feature is randomly permuted or shuffled while other features are left untouched. Shuffling values of a more important variable yields to a greater decrease in prediction accuracy (Breiman 2001). This is mainly used in high-dimensional problems to screen predictive features and reduce dimensionality (Genuer, Poggi et al. 2010).

RF modeling and model's assessment in terms of Mean Decrease in Gini and Mean Decrease Accuracy were carried out using the randomForest R package.

3.3.4 Comparison Measures

To assess the prediction power of a model and to compare performance of different models, various measures are used. Classification accuracy is used commonly in previous studies to evaluate model score (Podgorelec, Kokol et al. 2002, Arentze and Timmermans 2004, Rachman and Ratnayake 2019). To calculate accuracy score, it is first required to understand the concept of Confusion

Matrix. A Confusion Matrix is a common measure in classification problems which shows the count of predicted and measured instances in each class in a matrix form (Figure 3.3) (Kulkarni, Chong et al. 2020). In this work, which aims to solve a binary classification issue, assessing the survey to be perceived “Correct/Short” is the positive class, and evaluating the survey as perceived “Long” is the negative class. The positive class is considered the class of interest in this study.

		Actual Values	
		Positive	Negative
Predicted Values	Positive Class	True positive (TP)	False negative (FN)
	Negative Class	False positive (FP)	True negative (TN)

Figure 3.3 Confusion matrix of a binary classification problem

Following is a breakdown of the components of a binary confusion matrix:

- True positive: Value count of observations that were correctly classified as perceived “Correct/Short”.
- False positive: Value count of observations that were wrongly classified as perceived “Correct/Short”.
- False negative: Value count of observations that were wrongly classified as perceived “Long”.
- True negative: Value count of observations that were correctly classified as perceived “Long”.

As the most common measure of model performance, model accuracy is defined as the ability of the model to make correct predictions for each class. In other words, accuracy is calculated by dividing the number of correct predictions by the total number of predictions made by the model (equation 3.2).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{Total\ correct\ predictions}{Total\ predictions} \quad (3.2)$$

Although accuracy represents how well the model performs in overall to classify the data, it does not do a good job in evaluating model performance in each class specially when the classes are imbalanced. In the case of imbalanced classes, the model tends to learn the class with the higher number of observations so the model accuracy will be biased in favor of the larger class. Thus,

even though the model achieves high overall accuracy, it has poor performance when it comes to classifying under-represented groups (Abd Elrahman and Abraham 2013).

Considering that this work investigates an event with imbalanced classes, it was necessary to use other measures namely sensitivity, specificity, precision, recall, and F1 score. These can be used to assess classifier performance on different classes (Nellore 2015).

Sensitivity (or recall) and specificity assess the predictive power of the model on each class. Sensitivity refers to the model's ability to correctly classify the actual positive class (respondents who perceived the survey length as (Correct/Short) and will be calculated using equation 3.3:

$$Sensitivity = \frac{TP}{TP + FN} = \frac{\text{Correct positive predictions}}{\text{Total actual positive instances}} \quad (3.3)$$

Specificity applies to the model's capability to correctly predict the instances belonging to the actual negative class (respondents who perceived the survey as long), which is calculated using equation 3.4:

$$Specificity = \frac{TN}{TN + FP} = \frac{\text{Correct negative predictions}}{\text{Total actual negative instances}} \quad (3.4)$$

Precision is another measure of model assessment that is commonly used in problems with focus on the model performance on a single class (Abd Elrahman and Abraham 2013). Precision and sensitivity complement each other. Equation 3.5 represents how precision is calculated.

$$Precision = \frac{TP}{TP + FP} = \frac{\text{Correct positive predictions}}{\text{Total positive predictions}} \quad (3.5)$$

To have a better classification result, the goal is to maximise both precision and sensitivity. Since sensitivity is not affected by the imbalanced class, another measure, the F1-score, is used which combines precision and recall into a single metric by taking their harmonic mean. F1- score is a good measure of model performance when comparing models, the higher the F1-score the better the classifier is. Equation 3.6 shows how to calculate F1-score.

$$F1 - score = \frac{2 * Precision * Recall}{Precision * Recall} = \frac{2TP}{2TP + FP + FN} \quad (3.6)$$

3.4 Regression Analysis

When designing a survey, one very important thing to consider is how question types and the number of questions can affect respondent experience in terms of survey duration and perceived burden. In order to reduce the effort required to complete the survey and maintain respondents' interest, it is extremely beneficial to identify the types of questions that should be avoided. Five different question types were included in the survey namely: yes/no, multiple-choice, choose-on-map, open-ended, and slider. Survey duration is a very important factor in the effort required to complete a survey followed by the burden imposed on respondents. Hence, we choose to study duration as the target variable. The objective is to understand how increasing the frequency of each question type changes perceived and actual duration of a survey.

3.4.1 Data Preparation

In the process of designing a survey, multiple scenarios with different question types and frequency may be designed which are randomly assigned to survey takers. Thus, there is always some variation in the number and types of questions answered by different respondents when completing a survey. In a survey related to transportation and mobility behaviour, frequency and type of questions may also vary depending on the scale of data collection. For example, in transportation surveys, a very common approach is to collect data at the household scale. In such surveys, duration will differ for each respondent when providing information about household members. The survey duration also varies when respondents provide details on the locations they visited and the modes they used (i.e., Origin-Destination surveys). Knowing this, the number of questions of each type that were answered by each respondent was collected. Next, a correlation analysis was conducted among the variables planned to be used as predictors in the final model. No other manipulation was done on the data since the objective was to see the variation in duration corresponding to one unit of change in each question type frequency. Data associated to perceived duration is gathered in minutes through a post-survey question asking respondents “*how much time do you think you spent answering this questionnaire?*”. Measured survey duration is gathered through survey logs and the values greater than 60 minutes are excluded from the analysis assuming that respondents either were multitasking or left the survey to do something else before completing it. A low frequency of instances with measured duration over 60 minutes supports the decision to exclude them (Figure 3.4). Unfortunately, there is no way of understanding this in detail through administrative data.

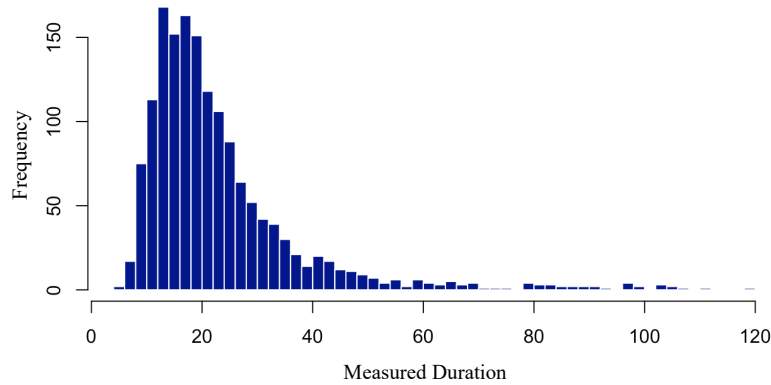


Figure 3.4 Distribution of measured duration in minutes

3.4.2 Multiple Regression Analysis

Multiple regression is a statistical technique and an extension of linear regression used to model the relationship among a single response variable and multiple predictors. Basically, it studies the joint influence of multiple factors on an outcome. The primary assumption in multiple regression analysis is the existence of a linear relationship between each explanatory variable and the outcome variable as expressed by equation 3.7:

$$Y_i = b_0 + b_1X_1 + b_2X_2 + \dots + b_nX_n + \varepsilon_i \quad i = 1, \dots, n \quad (3.7)$$

Where Y_i is the dependent variable, X_i s are explanatory variables, b_i s are the slope coefficients of independent variables, b_0 is the intercept term, n is the number of observations and ε_i is the error term of the i th observation. Only variables with appropriate variations and relatively strong correlations with the target variables are included in the final models.

The main idea behind a linear regression model is to fit a line or a plane or a higher dimensional object to the data by minimizing the sum of squared differences between observed data points and the predicted values by the model. The intercept coefficient represents the predicted value of the target variable while all independent variables are equal to zero. The estimated regression slope or unstandardized regression coefficients are measured in their original scale and used to examine the relationship among the predictors and the outcome variable. With an intuitive interpretation, unstandardized regression coefficients representing the changes in the outcome variable for an increase of one-unit in the predictor, holding all other variables constant. Standardized regression coefficients are normalised and unit-less. They are used to compare the effect of different predictors

and identify the most important ones. P-value is used as a standard practice to assess if there is a statistically significant relationship among predictors and outcome variables. A coefficient with a p-value smaller than 0.005 Shows the impact of the predictor on the outcome variable is significantly different than zero.

3.4.2.1 Regression assumptions

Besides linearity as the key assumption behind regressions, other assumptions in multiple regression includes independence of observations (absence of systematic relationship among observations), normality of errors, homoscedasticity (constant variance of residuals), and absence of multicollinearity among independent variables. The validity and interpretability of the regression model can be negatively affected if the forementioned assumptions are violated. For example, the sign of coefficients may change by implementing highly correlated explanatory variables into a model which prevents modelers from observing the true power of different explanatory variables in predicting the response variable. Therefore, residual analysis is done by plotting residuals against predicted values to identify the patterns and traces of violation of these assumptions. The pattern of residuals spreading around predicted values should be relatively constant and randomly scattered around the horizontal line ($y=0$) to suggest the assumptions of homoscedasticity is met.

3.4.2.2 Fit validity

Goodness-of-fit is evaluated using adjusted R-squared values, which indicate how much variation in outcome variable can be explained by explanatory variables. Higher adjusted R-squared values indicate a better fit on the data. F-statistic and its associated p-value were further used to evaluate the overall significance of the regression model and whether the model provides a better fit compared to a model with no predictors.

The analysis is carried out in Rstudio using `lm()` function to perform linear regression which uses Ordinary Least Square (OLS) technique to calculate the coefficients.

CHAPTER 4 RESULTS ON BURDEN AND SURVEY DURATION ANALYSIS

This chapter discusses the results of this research. The first section explores the data used in the analysis. Sections two, three, and four provide a detailed discussion on the results derived from various models and analysis carried out for this study. Finally, section five discusses how we based our findings from the rest of the analysis to design the first national survey on transport equity.

4.1 Descriptive Statistics

Table 4.1 and Table 4.2 present a descriptive statistic of sample sociodemographic and data associated with variables related to burden. In total 1704 questionnaires were completed, and 1507 responses were validated and used in this analysis.

Table 4.1 Descriptive statistic of the sample sociodemographic (n=1507)

Variable	Category	Nb. observations	Percentage
Gender	Male	683	45%
	Female	824	55%
Age category	Young adults 17-34	266	18%
	Middle-aged 35-54	731	48%
	Older adults ≥ 55	510	34%
Age category	17-24	39	2%
	25-34	227	15%
	35-44	404	27%
	45-54	327	22%
	55-64	269	18%
	65 and older	241	16%
Educational level	Non-university	365	24%
	University	1142	76%
Employment	Student	84	6%
	Full time worker	999	66%
	Part time worker	81	5%
	Worker and student	24	2%
	Retired	278	18%
	At home	25	2%
	Other	16	1%

Table 4.1 Descriptive statistic of the sample sociodemographic (n=1507) (cont'd)

Variable	Category	Nb. observations	Percentage
Income	<60K	281	19%
	60K-120K	475	31%
	>120K	493	33%
	Not stated	258	17%
Household size	Single household	414	27%
	2 persons	547	37%
	3 persons	241	16%
	4 persons & more	305	20%
Car ownership	No car	388	26%
	One car	738	49%
	2 cars & more	381	25%
Platform	Small screen (mobile)	439	29%
	Big screen (Tablet/Desktop)	1068	71%

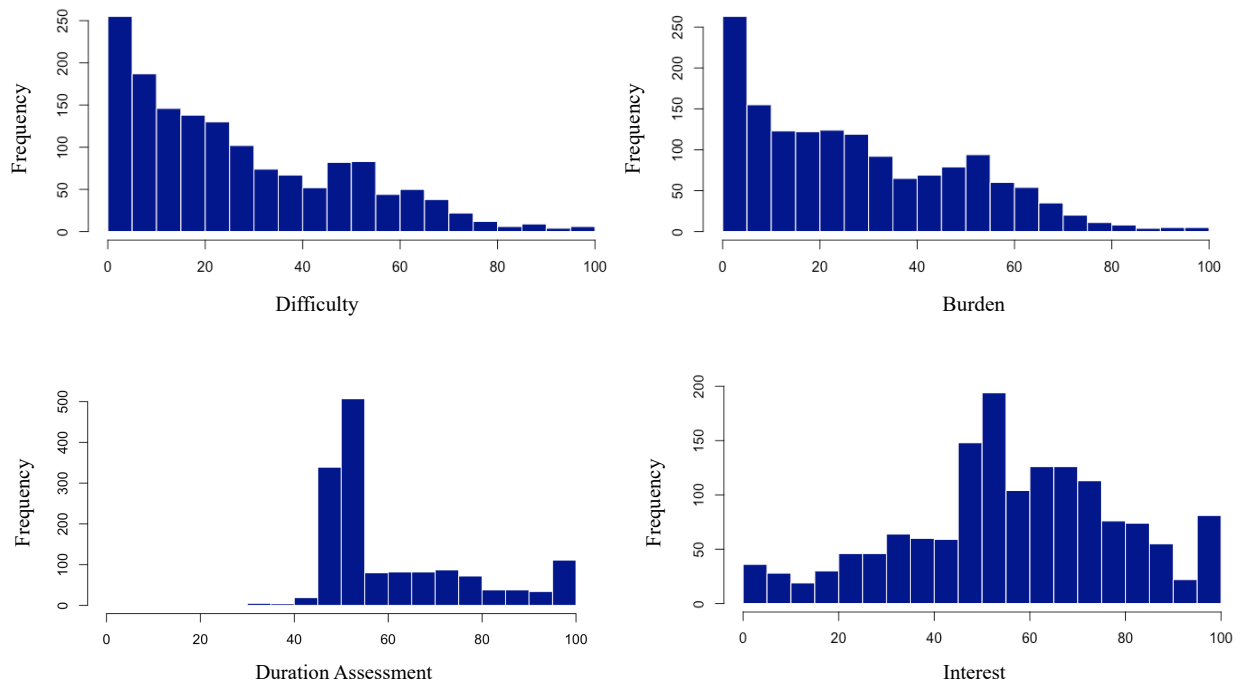
Table 4.2 Descriptive statistics of the observed variables

Variable	Description	Mean	Std	min	max
Burden	Continuous variable	28.34	21.85	0.00	100.00
Survey Duration Assessment	Continuous variable	61.28	16.36	2.00	100.00
Difficulty	Continuous variable	27.20	21.95	0.00	100.00
Interest	Continuous variable	56.78	22.93	0.00	100.00
Perceived time	Continuous variable	14.14	6.85	1.00	60.00
Measured time	Continuous variable	21.12	9.78	5.75	59.83
Time Ratio	Continuous variable	0.73	0.31	0.04	2.38
Inclined to abandon the survey	Binary	Yes: 501 No: 1006			
Willingness to participate in new surveys	Binary	Yes: 1070 No: 437			

Descriptive statistics of sociodemographic characteristics of respondents show that more respondents were middle-aged adults. More than 90% of respondents had university degree and about 66% percent were full time workers. More surveys were completed by women compared to men. Lastly, about 70% of surveys were filled out on desktop or tablet with a larger screen.

Figure 4.1 presents the distribution of observed and calculated variables. Burden, duration assessment, difficulty and interest are collected by slider questions in the scale of 0 to 100.

Perceived and measured duration are gathered in minutes. We can see the same trend of distribution for burden and difficulty, both graphs showing higher proportion of population belonging to the lower levels of burden and difficulty. This means that more respondents perceived the survey as not difficult and not burdensome. For interest, we can see the distribution is quite normal with the peaks representing the medium level of interest. Compared to lower levels of interest (0-40), more observations belong to medium and higher level of interest (40-60). In survey duration assessment, the graph shows more people believed the duration of the survey was about right (40-60), very few people believed it was too short (<40) and the rest thought the survey was too long (>60). Lastly in terms of perceived and measured duration, the same trend is seen in the distribution of data. However, the number of respondents who perceived the survey was shorter than 15 minutes is larger compared to the ones with actual measured duration of less than 15 minutes. This difference is better observed in the box plots presented in Figure 4.2.



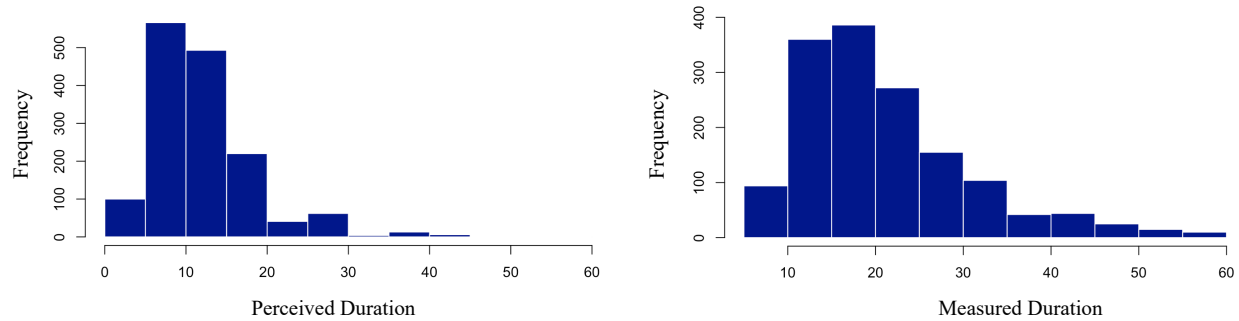


Figure 4.1 Distribution of measured variables

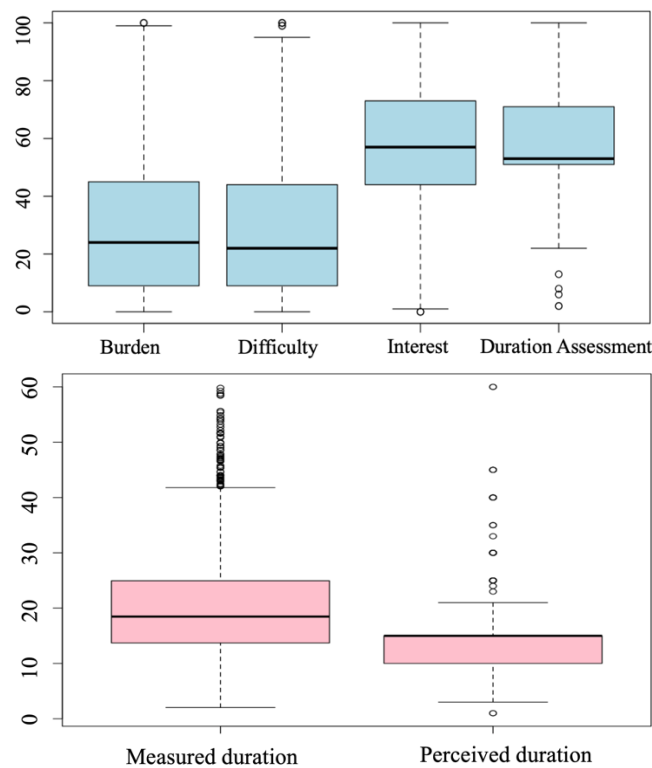


Figure 4.2 Box plot presenting the distribution of observed variables.

Looking into the descriptive statistics of perceived and measured duration, we can see that the mean value for perceived duration is 14.14 while this number is equal to 21.12 when it comes to actual measured duration. This is a very interesting observation. The ratio of duration was calculated by dividing perceived duration on measured duration and two categories were created. shows that 84% percent of respondents under-estimated the survey duration and reported perceiving the survey to be shorter than it actually was (Figure 4.3).

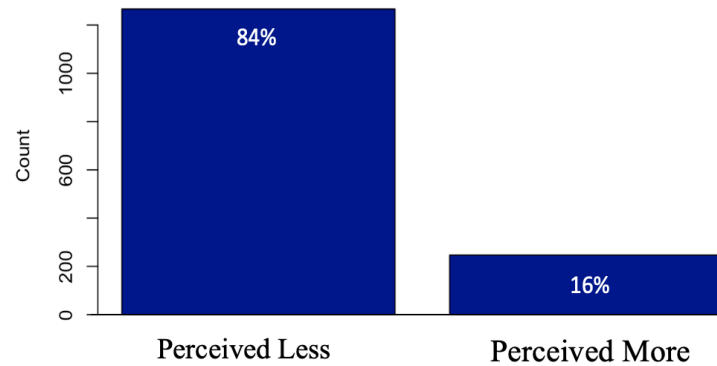


Figure 4.3 Perception of survey duration

Correlation among variables is presented in Figure 4.4. Pearson correlation is used among continuous variables and point biserial correlation is used to calculate correlation among continuous and binary variables. Looking into Figure 4.4, we can see that there is a relatively high positive correlation among burden, difficulty, inclination to abandon the survey, and survey duration assessment. The correlation among burden and interest is relatively high but negative. Among burden and willingness in new surveys there is a smaller negative correlation. The correlation among burden and willingness in new surveys there is a smaller negative correlation. The correlation among burden and perceived duration is smaller than the one among burden and survey duration assessment meaning that duration assessment has more to say about burden. Surprisingly, a very small correlation exists between burden and measured duration. This shows that measured duration does not have a lot to say about variation in burden.

	Burden	Interest	Duration Assessment	Difficulty	Perceived Duration	Measured Duration	Inclination to Abandon	Willing to New Survey
Burden	1							
Interest	-0.41	1						
Duration Assessment	0.5	-0.26	1					
Difficulty	0.42	-0.2	0.23	1				
Perceived Duration	0.18	-0.01	0.24	0.14	1			
Measured Duration	0.03	0.03	0.09	0.12	0.55	1		
Inclination to Abandon	0.43	-0.29	0.4	0.28	0.12	0.11	1	
Willing to New Survey	-0.25	0.34	-0.23	-0.1	0	0.03	-0.32	1

Figure 4.4 Correlation matrix among measured variables.

4.2 Latent Factor Analysis Results

Structured equation modeling (SEM) is an analytical technique for modeling complex relationships among observed and unobserved variables. SEM consists of two main components: a measurement model that specifies the relationship between measured and latent variables, and a structural model that depicts the structural relationships among different latent constructs.

4.2.1 Measurement Model

It was necessary to conduct a Confirmatory Factor Analysis (CFA), also referred as measurement model, before building the structural model to identify factors based on previous studies and to confirm the factor structure. The measurement model studies how observed variables are assigned to different latent constructs.

In the literature burden is measured as a factor with a single manifest (Fricker, Kreisler et al. 2012). Hence, the initial CFA model was built by considering burden as a single indicator latent factor. However, burden was explaining a big proportion of the variance among other observed variables and prevented the establishment of divergent validity among variables. Since burden formation is the primary objective of this study, we decided to consider burden as a single continuous observed outcome variable with no measurement error (Yan, Fricker et al. 2020). While building the measurement model, and in the course of hypothesis development, we accounted for what we learned from theory and previous works. Nevertheless, there were a few variables, namely willingness to participate in future surveys and inclination towards abandoning the survey, for which a rich literature was not available. Thus, we sought to investigate their contribution to burden formation. To accomplish this, new constructs are developed and validated. In this regard, perceived interest was investigated as a new construct to determine how it leads to burden. To ensure that the introduction of a new latent variable is statically representative of the data at hand, an Explanatory Factor Analysis (EFA) is correspondingly carried out. According to the results of EFA, the variables are behaving statistically logically according to our hypothesis and expectations. Finally, using learnings from EFA and accounting on theory, three latent constructs, each associated with two manifests, are identified to build up the measurement model. Three latent constructs each indicated on two observed variables are developed to conduct this analysis including:

- Perceived Length Measurement

Perceived length (PL) construct is indicated by two manifests namely “survey duration assessment” and “perceived duration” to complete the survey. Survey *duration assessment* is measured by a slider question on the scale of 0 to 100 where 0 represents the survey duration was too short, 40-60 represents the duration is about right and 100 indicates a too long survey. For *perceived duration*, respondents were asked to report how long they think it took them to complete the questionnaire (in minutes).

- Perceived Effort Measurement

Perceived effort (PE) factor is defined to be associated with “inclination to abandon the survey” and “perceived difficulty”. Respondent’s *inclination to abandon the survey* was collected as a binary question. Level of *difficulty* respondents felt to complete the survey was gathered with a slider question on the scale of 0 to 100 where 0 is very easy and 100 is very difficult.

- Perceived Interest Measurement

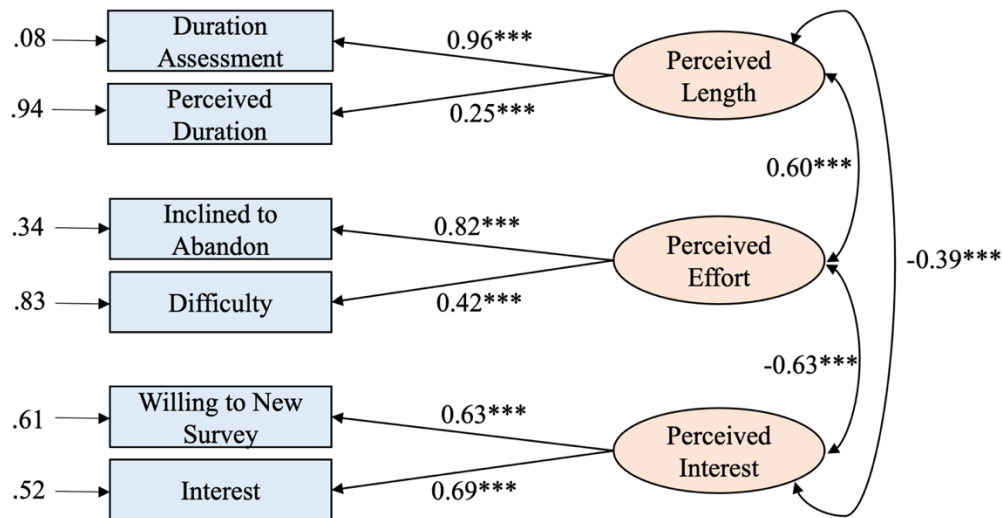
Perceived interest (PI) construct was defined to explain the variance in two observed variables, “Perceived interest” in survey and “willingness to participate in new surveys”. *perceived interest* was a slider question on the scale of 0 to 100 with 0 indicating the survey was not at all interesting and 100 indicating it was very interesting. *Willingness to participate in new surveys* was posed as a yes/no question.

Table 4.3 presents a summary of the latent constructs and the associated questions that were asked to serve as observed variables.

Table 4.3 Latent constructs and their indicators and survey questions

Latent variable	Indicator	Abbreviation
Perceived Length (PL)	Do you feel that the length of the survey was too short, about right or too long?	Duration Assessment
	How much time do you think you spent answering this questionnaire (in minutes)?	Perceived duration
Perceived Effort (PE)	How difficult or easy was it for you to answer the questions in this survey?	Difficulty
	Have you been inclined, at any point, to abandon the survey?	Abandon
Perceived Interest (PI)	How interesting was this survey to you?	Interest
	Would you be interested in participating to other studies conducted by the Chaire Mobilité at Polytechnique Montreal?	Willingness to participate in new surveys

To set the scale of latent variables, their variance was set to zero letting a free estimation of all load factors (Hair and Babin 2018). Figure 4.5 presents the path diagram of measurement model and the factor loadings of observed variables related to each latent factor. DWLS is used as the estimator to run the CFA model.



Standardized coefficients, *p-value < .05, **p-value < .01, ***p-value < .001
 $\chi^2=22.032$, df=6, P-value=0.001, RMSEA=0.042, CFI=0.987, TLI=0.968

Figure 4.5 Path diagram of measurement model specification, a three-factor solution (CFA model)

4.2.1.1 Measurement Model Fit, Validity, and Reliability Assessment

4.2.1.1.1 Fit validity

In order to ensure fit validity, several statistic measures of goodness-of-fit are examined. The Chi-square statistic ($=22.032$) has a p-value smaller than 0.005, resulting in rejecting the null hypothesis. The χ^2 static fit is very sensitive to sample size and leads to bias in the p-value. Therefore, other fit metrics including CFI, TLI, and RMSEA are investigated. The values of CFI and TLI equal 0.987 and 0.968, respectively. Both values are above 0.95 indicating a very good fit (Hu and Bentler 1999, Sarmento and Costa 2019). Lastly, RMSEA equals 0.042 which is smaller than 0.06 indicating a good fit (Hu and Bentler 1999).

4.2.1.1.2 Reliability and Validity Assessment

A measurement model should meet a certain level of validity and reliability before testing the results of the structural model. *Convergent validity* indicates that a high proportion of variance in common exists among indicators of each latent variables. To examine convergent validity, the size and significance of path estimates are studied. The estimations are all found to be statistically significant with p-values smaller than 0.001. In terms of size of load factors, among the six estimated loads, the load factor associated with difficulty equals 0.42, the load factor associated with perceived duration equals 0.25, and the other load factors all have values larger than 0.6. Previous studies have confirmed that loadings greater than 0.4 can be kept in the model. Following this threshold, all indicators except *perceived duration* are eligible to enter into the model. The items with poor loadings (<0.3) are considered inadequate and are removed from the model due insufficient relation with the construct. However, we keep *perceived duration* variable since we theorized that it was a contributing factor and it is among our research objectives.

Assessment of internal consistency *Reliability* is carried out using construct reliability (CR). The presence of internal consistency is considered acceptable in exploratory research if the CR value is between 0.6 and 0.7 (Shrestha 2021). Based on Table 4.4, only PI failed to satisfy the CR threshold.

Average Variance Extracted (AVE) is another metric of convergent validity which is calculated as the mean variance extracted from all items loading on a construct. In order to indicate adequate convergence, AVE should be greater or equal to 0.5. AVE values presented in Table 4.4 show that Convergent Validity is successfully established for PL and PE latent variables. However, full assurance of convergent validity is not proven for PI.

Discriminant validity refers to the extent to which it measures one construct and not another. As per Fornell-Larcker criterion, discriminant validity can be reached if the square root of the AVE for each latent variable exceeds the correlation value among them. Table 4.4 shows that in the case of PI-PE constructs, there is a little dispute, but since the difference is so small ($=0.004$), it can be ignored (Rahim and Magner 1995). We can therefore say that Discriminant Validity is met meaning that constructs are completely distinct from each other.

Table 4.4 Measurement model validity and reliability, square root of AVE in bold and correlation between constructs in off-diagonal

Construct	Convergent validity		Discriminant Validity		
	Composite reliability (CR)	Average Variance Extracted (AVE)	PL	PE	PI
PL	0.705	0.640	0.800		
PE	0.686	0.642	0.603	0.801	
PI	0.493	0.397	-0.395	-0.634	0.630

4.2.1.2 Measurement Model Findings and Discussions

The results of the CFA model show that a three-factor solution is promising according to the data in hand. For PL, we observe a high load factor for *survey duration assessment* indicating that PL is explaining about 92% of variation in this indicator (high communality). The second item, *perceived duration*, has a fairly small factor loading, meaning that less than 10% of variation in this manifest is explained by PL. Regardless of the small communality of the second indicator, AVE (>0.6) and CR are in the ideal range (>0.7) which confirms the factor structure.

PE factor is explaining about 70% of the variance of the first indicator (inclination to abandon) and about 20% of the second one (difficulty). Despite 20% communality not being ideal, AVE and CR are both within ideal limits. The factor structure of the PE is therefore confirmed.

Lastly, PI explains less than 50% of the variation of the associated indicators (interest and willingness to participate in new surveys). Although the standardized factor loadings for both manifests fall in the acceptable range (>0.6), AVE and CR are not able to fully satisfy the ideal limit. In other words, there are aspects of PI that remain unexplained with the indicators available. To fulfill this, more questions should be included in the survey or questions should be asked in another way. This as a limitation of our study.

The direction of loadings for all indicators are in line with theory. Significant positive factor loadings are found for both items (perceived duration, and survey duration assessment) associated with PL. This means that longer duration perceived by respondents and indicating the survey as too long in question related to *duration assessment* is significantly and positively associated with PL. The factor loadings for the items under PE are both found to be positive and significant which shows that perceived survey *difficulty* and *inclination to abandon* are associated with greater effort to participate in surveys. And lastly, finding the survey to be interesting and be willing to participate in new surveys are found to positively and significantly related to PI.

In regard to relationship among constructs, PL and PE are positively correlated. It is consistent with the theory stating that greater effort required in survey task results in a longer survey perception. PI is found to be negatively correlated with PE and PL meaning that being interested in surveys will make the task less challenging and less time consuming which confirms the previous findings (Fricker, Yan et al. 2014).

4.2.2 Structural Model

SEM framework consists of two primary parts: measurement model and structural model. Through building the measurement model, three constructs namely *Perceived Length*, *Perceived Effort*, and *Perceived Interest* were identified. Having validate the measurement model, the next step is investigating the relationship among latent constructs by carrying out a structural model. As a basis for defining the structural model, we formulate the following hypotheses, some of which are theoretical and some of which are to be tested. Burden is considered as a single measured response variable.

H1: Longer surveys impose higher level of burden on respondents.

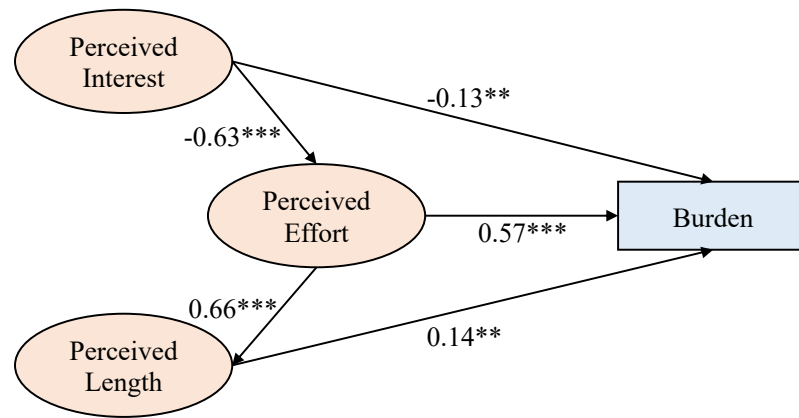
H2: Greater effort required to complete a survey imposes a higher level of burden on respondents.

H3: Higher levels of interest in survey leads to lower levels of imposed burden.

H4: Greater effort required to complete a survey increases participants' perception of time spent on the task.

H5: Higher levels of interest in survey leads to lower effort required to complete the survey.

The hypothesised structural model is then established by passing single headed arrows from PL, PE, and PI to burden, from PI to PE, and lastly from PE to PL (Figure 4.6).



Standardized coefficients, *p-value < .05, **p-value < .01, ***p-value < .001
 $\chi^2=37.386$, df=10, P-value<.001, RMSEA=0.043, CFI=0.988, TLI=0.975

Figure 4.6 Path diagram of structural and measurement models specifications

4.2.2.1 Structural Model Fit Assessment

4.2.2.1.1 Model Fit indices

To determine how well the model fits the data, different model fit indices are examined. Chi-square static is equal to 37.386 with a significant p-value resulting in null hypothesis rejection. Nevertheless, since χ^2 is too sensitive to sample size, other fit indices namely CFI, TLI, and RMSE are investigated. CFI and TLI are found to be equal to 0.988 and 0.975, respectively. Both being >0.95 represents a very good fit. RMSEA is equal to 0.043 which is under the 0.045 threshold referring as a very good fit. With all fit indices satisfying the determined thresholds, we can conclude that the data is well modelled.

4.2.2.2 Structural Model Findings and Discussions

The results of the SEM model show that all estimations are in the expected direction. All regressions are found to be statistically significant. First, the direct effect of each factor in burdensomeness of survey is investigated. There is a significant and negative relation between PI and burden meaning that respondents who expressed higher level of interest in the survey feel less burdened. Positive and statistically significant regression estimate among PE and burden reveal that greater effort required to complete the survey is significantly associated with higher

burdensomeness feeling. Furthermore, the positive and significant PL-burden regression estimates indicate that respondents feel more burdened over longer surveys. The effect size for PI-burden and PL-burden is quite small. However, a relatively larger effect (almost 4 time bigger than the other two) is estimated among PE and burden. Respondent's perception of how long it takes to complete the survey is shown to be significantly influenced by the effort required to complete the survey. The magnitude of effect is found to be relatively high. Furthermore, respondents' interest in participating in the survey significantly affects the effort required for the survey completion. This is in line with previous literature demonstrating that interest is an indicator of PE (Fricker, Kreisler et al. 2012).

The other aspect to be discussed with this empirical data is the mediation effect. Two mediation effects are found in this model:

1. PE acts as a mediator between PI and Burden ($PI \rightarrow PE \rightarrow Burden$)
2. PL plays the mediation role between PE and Burden ($PE \rightarrow PL \rightarrow Burden$)

Followed by direct effect, the indirect and total effect of relation among these variables is examined and presented in Table 4.5. The indirect effect is calculated by multiplying the path coefficients of predictive variable-mediator by the path coefficient of mediator-target variable. The statistical significance of both direct and indirect effects of two sets of factors suggests partial mediation effect (Cheung and Lau 2008, Shankar, Jebarajakirthy et al. 2020). If only indirect effect were significant, the interpretation would be a full mediation effect.

The total effects of the three factors on burden formation are calculated by summing up their direct and indirect effects and all were found to be statistically significant. The effect of PE is higher compared to the other two, followed by a relatively high regression estimation of PI. The smallest effect is associated to PL-burden.

Table 4.5 Relationship among constructs based on SEM estimates.

Hypothesis	Relationship	Direct effect	Indirect effect	Total effect
H1	Perceived Length → Burden	0.142**	-	0.142***
H2	Perceived Effort → Burden	0.568***	0.094**	0.662***
H3	Perceived Interest → Burden	-0.133**	-0.361***	-0.495***
H4	Perceived Effort → Perceived Length	0.663***	-	0.663***
H5	Perceived Interest → Perceived Effort	-0.636***	-	-0.636***

*p-value < .05, **p-value < .01, ***p-value < .001

4.3 Classification Analysis Results and Comparisons

In this work, three machine learning models are used for classification purposes, namely: Decision Tree using the CART algorithm, Ctree, and Random Forest. In this section, the results of each method are first presented. Then, a comparison is conducted to assess the models' performances. In all models, the response variable is a binary variable associated to survey length with two classes: "Long" survey and "Correct/Short" which came from discretizing the duration assessment variable. Correct-Short survey class is used as the positive class and is considered the class of interest for the analysis. The objective, nonetheless, is to create a model that can effectively classify data in both groups with a relatively high degree of accuracy.

4.3.1 Classification and regression tree (CART)

The predictive variables implemented in the model classification tree is burden, difficulty, interest, willingness to participate in new surveys, inclination to abandon the survey, perceived duration, and measured duration. Gender, age group, education level, and the platform used to complete the survey entered the model as control variables. In both decision tree models, 80% of the data is used to train the model and 20% of data is used for testing purposes.

For controlling tree growth, a grid of 101 different Complexity Parameter (CP) values is generated to find the optimal CP value associated with maximum accuracy. Then 10-fold repeated cross validation is carried out and a CP value equal to 0.011 is found to be associated with the highest average accuracy among the 10 folds (Figure 4.7).

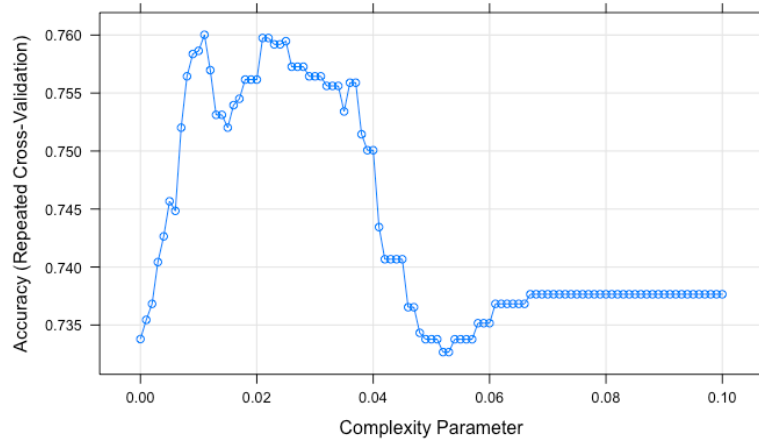


Figure 4.7 Model performance with varying complexity parameter (cp)

Figure 4.8 shows a graphical representation of the tree and how the training dataset set is split in accordance with different predictive features to undertake the classification task. The decision tree includes seven splits. “Burden” is used 3 times for splitting, followed by two splits for “perceived duration”, one split for “inclination to abandon the survey” and one for “difficulty”. *Burden* is selected as the root node to split the data, meaning that *burden* can provide the best distinction between classes of the target variable and results in the highest reduction in the impurity. The instances with burden smaller than 30 follow the left branch and receive the class label: “Correct/Short”. This is because 86% percent of instances in this leaf node belong to “Correct/Short” category and leaf nodes are labeled based on the class with most instances. Leaf nodes do not further split and represent the final classification made by the tree. Instances with burden greater than 30 traverse to the right node and then are split based on *perceived duration*. If perceived duration is smaller than 20 minutes, another split occurs based on *inclination to abandon the survey*. Respondents that were inclined to abandon the survey go to the right leaf node and are labeled as “Long”. In simple words, if burden is greater than 30, perceived duration is smaller than 20 minutes, and respondents were inclined to abandon the survey, the final leaf node corresponds to “Long” class. If instances were not inclined to abandon the survey and perceived the survey duration shorter than 13 minutes, they traverse to the final leaf classed as “Correct/Short”. Instances that are not inclined to abandon the survey and perceived the duration to be greater than 13 minutes will be split for the last time by burden. Higher values of burden (>61) will lead them to the right leaf labeled as “Long”, and relatively lower burden (<61) results in the left leaf with “Correct/Short” class label. Now going back to the second topmost split, for instances with *burden*

greater than 30, and *perceived duration* greater than 20 minutes, another split occurs based on *burden*. If *burden* is greater than 37, then the instances are classed as “Long” in the rightmost leaf. For instances with *burden* smaller than 37, another split occurs based on *difficulty*, sending all instances with *difficulty* greater than 47 to the right leaf classified as “Long” and the rest of instances with *difficulty* smaller than 47 to “Correct/Short” labeled left leaf.

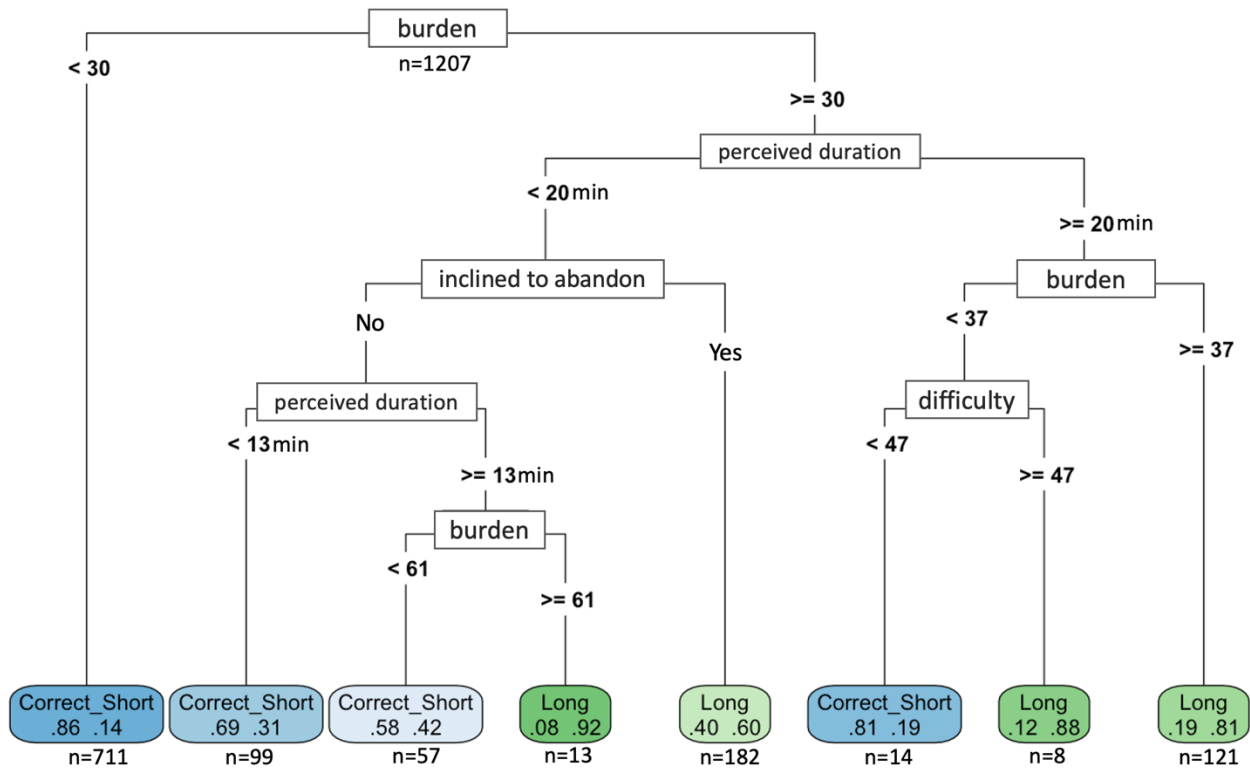


Figure 4.8 Results of classification and regression tree (CART)

In order to investigate the tree’s performance on testing dataset ($n=300$), a confusion matrix is plotted (Figure 4.9). There are various performance metrics to evaluate the classification model listed in Table 4.6. We can see that the tree does a better job in classifying the positive class (Correct/Short) instances compared to the negative class (Long) instances.

The figure shows a confusion matrix for the CARET model. It is a 2x2 table with 'Actual Values' as columns and 'Predicted Values' as rows. The columns are 'Correct/Short' and 'Long'. The rows are 'Perceived Correct/Short' and 'Perceived Long'. The values are: 179 (True Positive), 34 (False Positive), 25 (False Negative), and 62 (True Negative). The cells are color-coded: light blue for True Positives and True Negatives, and light orange for False Positives and False Negatives.

		Actual Values	
		Correct/Short	Long
Predicted Values	Perceived Correct/Short	179	34
	Perceived Long	25	62

Figure 4.9 Confusion matrix of classification and regression tree (CARET)

Table 4.6 Performance metrics of the decision tree according to the confusion matrix

Performance metric	Value	Interpretation
Accuracy	0.80	Proportion of correctly classified instances out of the total instances.
Precision	0.84	Proportion of correctly classified “Correct/Short” instances out of all “Correct/Short” predictions.
Recall or Sensitivity	0.88	Proportion of correctly classified “Correct/Short” instances out of all actual “Correct/Short” instances.
Specificity	0.64	Proportion of correctly classified “long” instances out of all actual “long” instances.
F1	0.86	Harmonic mean of precision and recall

4.3.2 Conditional Inference Tree (Ctree)

The CART algorithm tends to be biased when selecting variables by giving priority to variables with extreme values and not considering the statistical significance of relationships between variables when developing the tree. Consequently, a Ctree was developed to classify the training data. As the splitting criterion, Ctree uses statistical tests among the response and predictive variables to identify which variable has the most significant relationship with the target variable. To control the tree’s growth, a maximum depth of 5 is set.

Similar to what was found for the classification and regression tree, burden, difficulty, interest, willingness to participate in new surveys, inclination to abandon the survey, perceived duration,

and measured duration are considered as predictive variables for Ctree. Gender, age group, education level, and survey platform area all included as control variables in the model.

The result of the Ctree model is presented in Figure 4.10. A total of six splits are performed on the training data, with “burden”, “inclination to abandon the survey”, and “perceived duration” each being used twice as split nodes. *Burden* is selected as the root node indicating the highest significant association with *survey duration assessment*. After the initial split, the Ctree model identifies *inclination to abandon the survey* as the variable with the highest statistically significant relationship with the response variable. Thus, instances with burden higher than 34 go to the right leaf and are split again upon inclination to abandon the survey. The respondents who are inclined to abandon the survey will go to rightmost leaf and get labeled as “Long” class. Instances that were not inclined to abandon the survey will split upon *perceived duration* which is found by the algorithm as the third variable with the highest significant relationship with the target variable. If respondents perceived the survey duration is longer than 19 minutes, observations will be labeled as “Long” and the ones that perceived the survey to be shorter than 19 minutes will go to the left leaf which is labeled as “Correct/Short”. In the left side of the tree, instances with burden smaller or equal to 29, no inclination to abandon the survey, and perceived duration smaller than 25 minutes traverse to the leftmost leaf with “Correct/Short” label with about 90% homogeneity. Perceiving burden to be smaller than 34 and being inclined to abandon the survey will take instances to a leaf labelled as “Correct/Short” but this time with 70% purity. Leaves located in the left side of the tree are all labelled as “Correct/Short”, but their purity varies. The leftmost leaf node has the highest purity of Correct/Short instances. The outcomes of the decision tree analysis provided novel perspectives on survey duration, including both perceived and measured duration, as well as thresholds for perceived burden. These insights guide survey designers in creating surveys that aren't perceived as excessively long by participants.

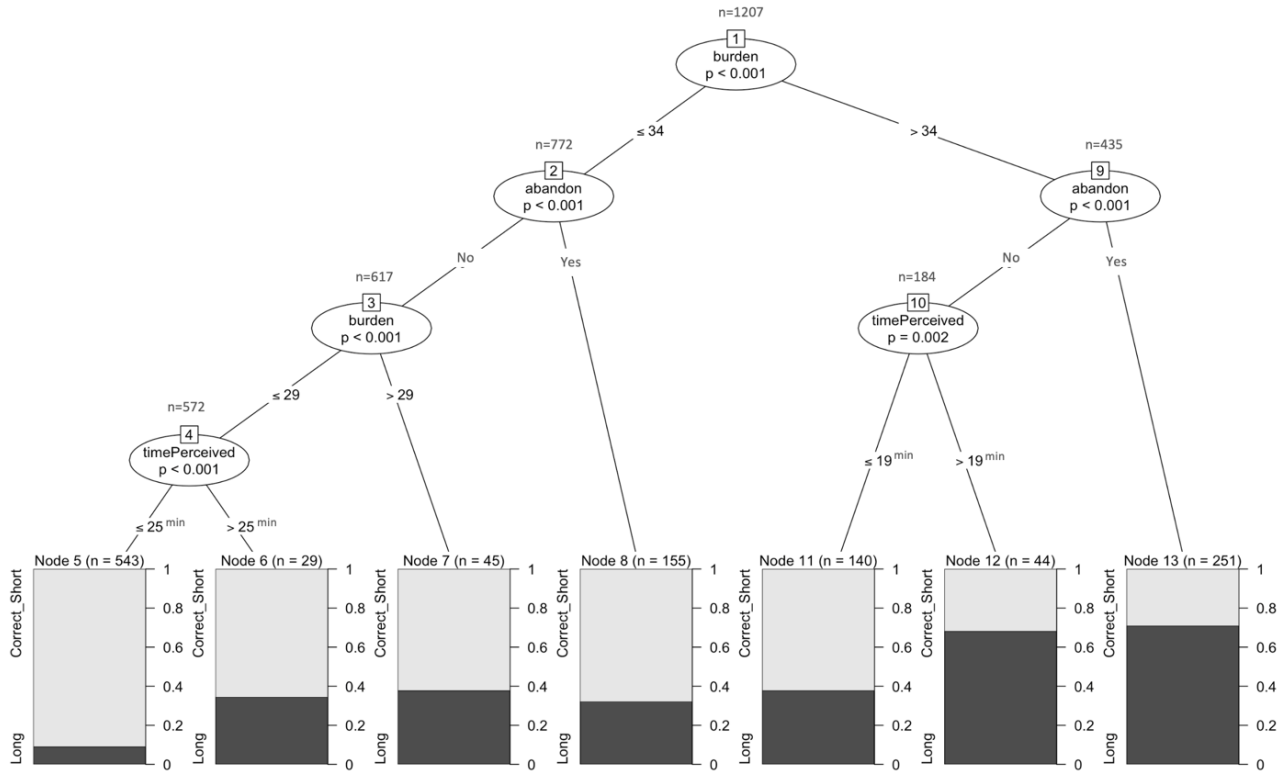


Figure 4.10 Results of conditional inference tree (CTree)

The confusion matrix presents how the Ctree performs in predicting the test data (Figure 4.11). Table 4.7 shows a summary of the different performance metrics illustrating how the tree performs in classifying unseen instances. Studying these metrics shows that Ctree performs better at classifying “Correct/Short” instances than “Long” instances.

		Actual Values	
Predicted Values		Correct/Short	Long
	Perceived Correct/Short	190	41
	Perceived Long	14	55

Figure 4.11 Confusion matrix of conditional inference tree (CTree)

Table 4.7 Performance metrics of the CTree according to confusion matrix.

Performance metric	Value	Interpretation
Accuracy	0.82	Proportion of correctly classified instances out of the total instances.
Precision	0.82	Proportion of correctly classified “Correct/Short” instances out of all “Correct/Short” predictions.
Recall or Sensitivity	0.93	Proportion of correctly classified “Correct/Short” instances out of all actual “Correct/Short” instances.
Specificity	0.57	Proportion of correctly classified “long” instances out of all actual “long” instances.
F1	0.87	Harmonic mean of precision and recall

4.3.3 Random Forest results

Random forest (RF) classifier is used for its high prediction accuracy and robustness against overfitting as a result of its ensemble nature and built-in bagging techniques. RFs provide insight on variable importance and their contribution to a model’s predictive power. Same as CART and CTree, burden, difficulty, interest, willingness to participate in new surveys, inclination to abandon the survey, perceived duration, and measured duration were used as predictive variables. The control variables included in this model were gender, age group, education level, and survey platform.

OOB error is first used to find the optimal number of variables tried at each split node (m_{try}) and a value of 2 is found to be associated with the lowest OOB error Figure 4.12.

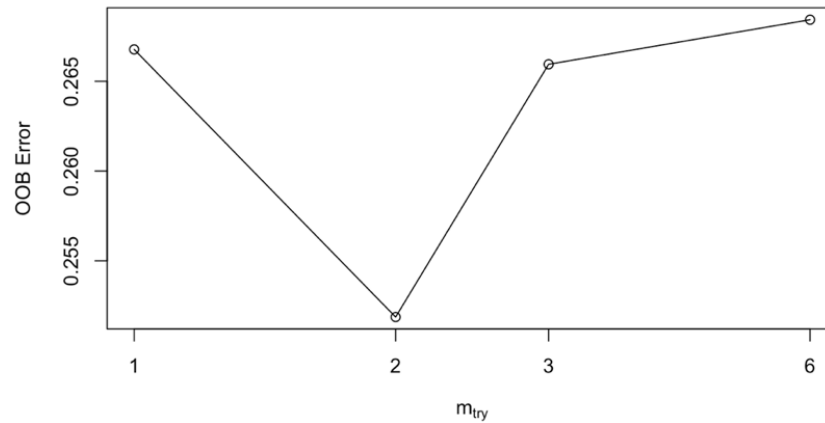


Figure 4.12 Exploring model tuning: OOB Error vs. m_{try}

To identify the optimal number of trees in the forest, we plot the OOB error rate, the error rate associated with “Correct/Short” class (frequency of misclassifying the Correct/Short class instances), and error rate associated with “Long” class against the number of trees. The data comes from `err.rate` matrix that R `randomForest` function creates by default. The results show that with the increase in the number of trees, the error rate goes lower and stabilizes when 500 trees are created (Figure 4.13). 500 trees are hence found to be an optimal value for RF.

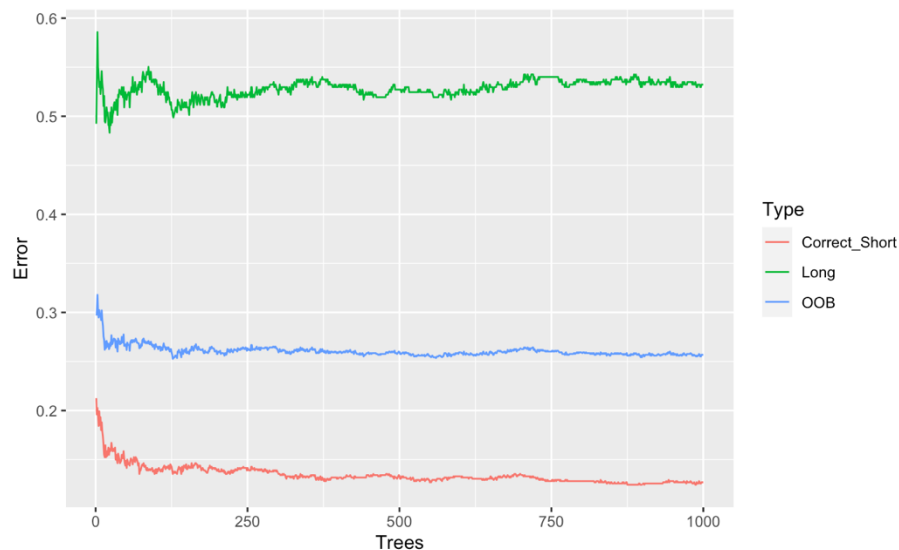
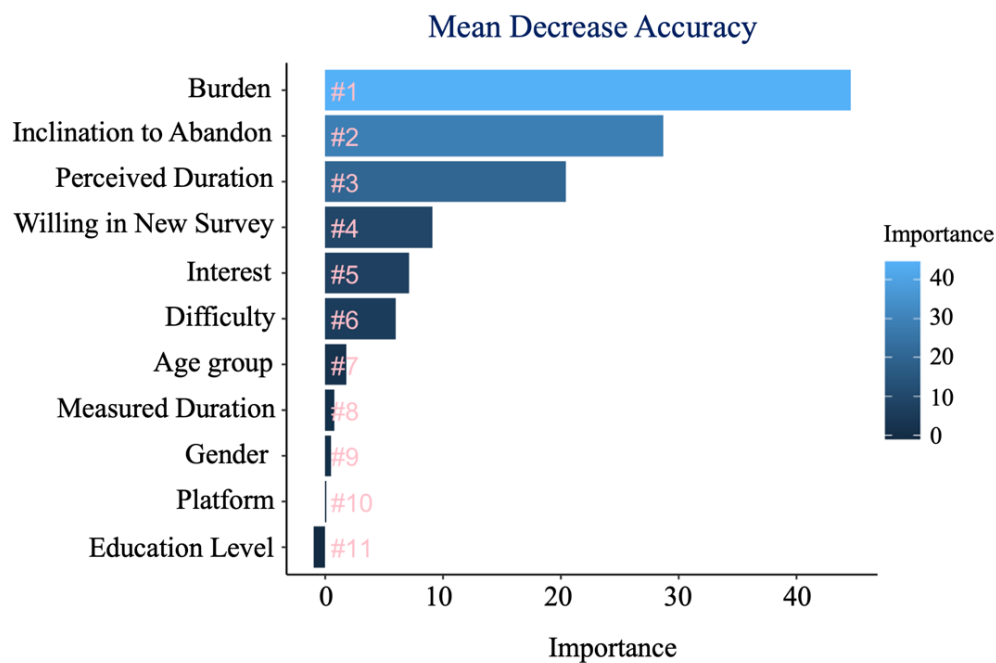


Figure 4.13 Exploring model tuning: OOB Error vs. number of trees in the forest.

One of the key features of running RF is to provide feature importance, a way to quantify the relative importance of each predictor variable in the model’s performance. The graphs associated with the Mean Decreases Gini (MDG) and the Mean Decrease Accuracy (MDA) ranks *burden* as the

most important factor. In other words, self-reported burden is the most important factor related to perceiving a survey as long or short. Burden is associated with larger reduction in Gini impurity across all trees and permuting its values will lead to the larger reduction in the model's accuracy. MDA shows that *inclination to abandon the survey* and *perceived duration* as second and third important factors. The first three variables associated with Mean Decrease Accuracy are exactly the ones included in CART and CTree. MDG graph shows *burden*, *measured duration*, and *difficulty* as the three most important factors in reducing the impurity of nodes. Gender, platform, and education level are found the least important variables in the models' predictive power (Figure 4.14). The classification methods are not used to explore the causal relationship among variables and their contribution is to capture relationships among different features and the target variables. Therefore, merely being an important variable in a classification analysis does not make it a causal one.



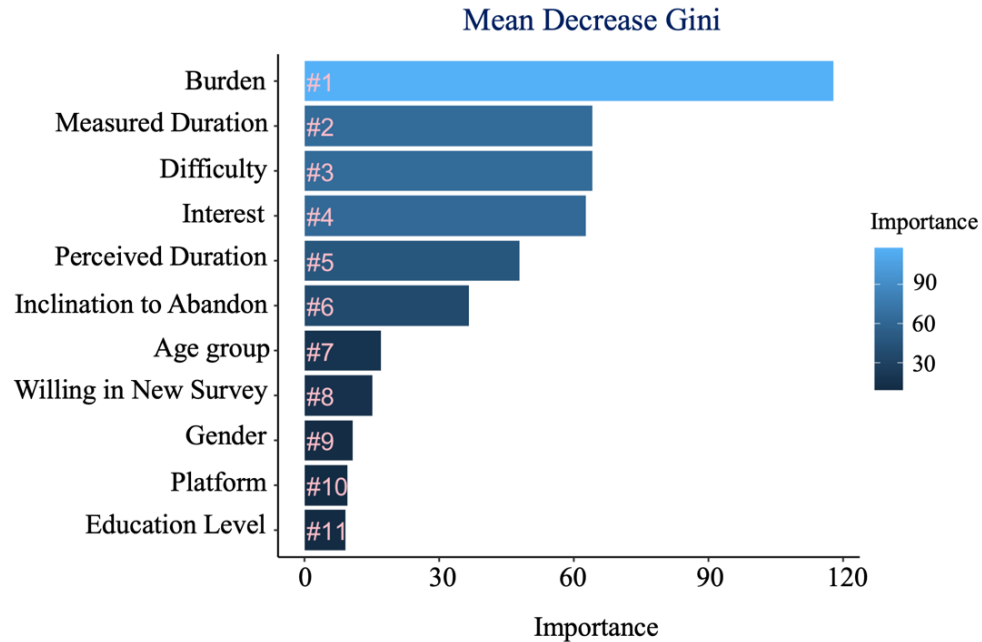


Figure 4.14 Variable importance based on Mean Decrease Gini and Mean Decrease Accuracy metrics in Random Forest

Lastly the confusion matrix shows the performance of the trained model on testing dataset (Figure 4.15). The performance metrics presented in Table 4.8 shows that random forest is doing a better job in classifying the correct/short class compared to long class.

		Actual Values	
		Correct/Short	Long
Predicted Values	Perceived Correct/Short	190	39
	Perceived Long	14	57

Figure 4.15 Confusion matrix of Random Forest

Table 4.8 Performance metrics of the Random Forest according to confusion matrix

Performance metric	Value	Interpretation
Accuracy	0.82	Proportion of correctly classified instances out of the total instances.
Precision	0.83	Proportion of correctly classified “Correct/Short” instances out of all “Correct/Short” predictions.
Recall or Sensitivity	0.93	Proportion of correctly classified “Correct/Short” instances out of all actual “Correct/Short” instances.
Specificity	0.59	Proportion of correctly classified “long” instances out of all actual “long” instances.
F1	0.88	Harmonic mean of precision and recall

4.3.4 Comparison among classification methods

To compare performance of different models and assess their prediction power, various measurement metrics namely: accuracy, precision, recall, specificity, and F1 score are used. A summary of the measurement metrics is provided in Table 4.9.

Table 4.9 Model performance scores and comparison

Metric	Classifier Model		
	CART	CTree	Random Forest
Accuracy	0.80	0.82	0.82
Precision	0.84	0.82	0.83
Recall or Sensitivity	0.88	0.93	0.93
Specificity	0.64	0.57	0.59
F1	0.86	0.87	0.88

According to F1 score, which is the harmonic mean of precision and recall, Random Forest is doing a better job in classifying the instances. Considering Recall, “Correct/Short” class prediction rate, CTree and Random Forest are performing better than CART by correctly classifying 93% of positive class instances out of all positive instances. According to Specificity, the negative class prediction rate, the CART algorithm has a better performance compared to the other two models with 64% correct prediction of negative class instances out of all negative instances. Since in this study the research interest is the “Correct/Short” class, we can conclude that Random Forest is the

best classifier to use. If our focus is rather on the prediction of the long class, the CART algorithm is the one to opt for.

4.4 Regression Analysis

To investigate the impact of different question types and the number of questions on perceived and observed duration of the survey, we performed multiple regression analysis. The survey questions were a mixture of five question types including yes/no, multiple-choice, choose-on-map, slider, and open-ended questions. An overview of the average number of different question types and observed survey duration in each survey section is shown in Figure 4.16 and Figure 4.17. Globally, more multiple-choice questions are answered in the visited places and COVID impacts sections. On average the majority of slider questions are included in the “COVID impacts” section. A higher number of yes/no questions are responded in the profile section. Lastly, in the “visited places” section, most responses are choose-on-map. The graph associated with survey duration illustrates that on average all sections are completed in less than 5 minutes and the longest duration is associated with the visited places section. This may be due to the time-consuming nature of choose-on-map questions. A relatively short duration is found in the COVID impacts section where on average 20 slider questions were responded, revealing how quickly these questions can be answered.

Caveat 1: As a result of respondents going backwards in survey sections (e.g., from visited places to profile), 59 observations were dropped out for this part of analysis. Survey logs provide a negative value for these instances in different sections which made the duration analysis for separate survey sections meaningless. 1446 observations were used to conduct multiple regression analysis.

Caveat 2: A large difference is observed when comparing the average and the maximum time spent to fill out each survey section. During those times, some respondents may have been multitasking or may have been away from their device. Unfortunately, there is no way to detect precisely how long respondents were doing other activities and not filling out the questionnaire. This is considered a limitation for this research.

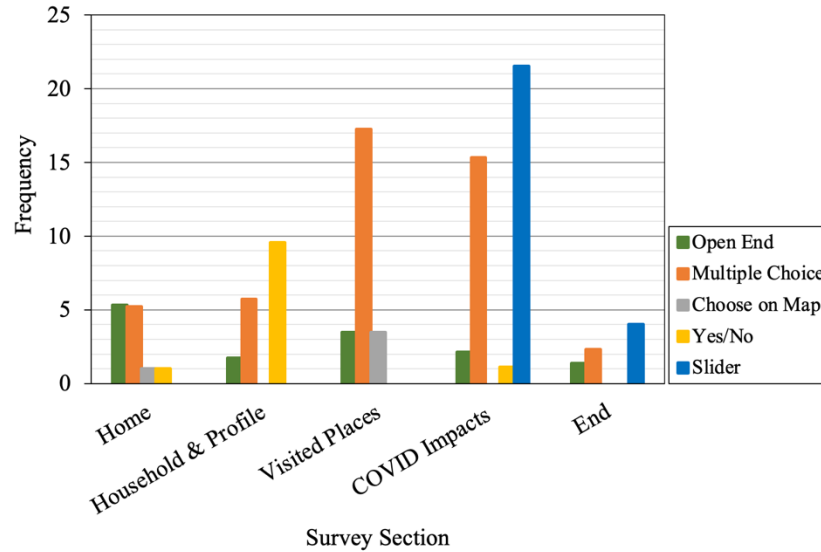


Figure 4.16 Average number of different question types in each survey section

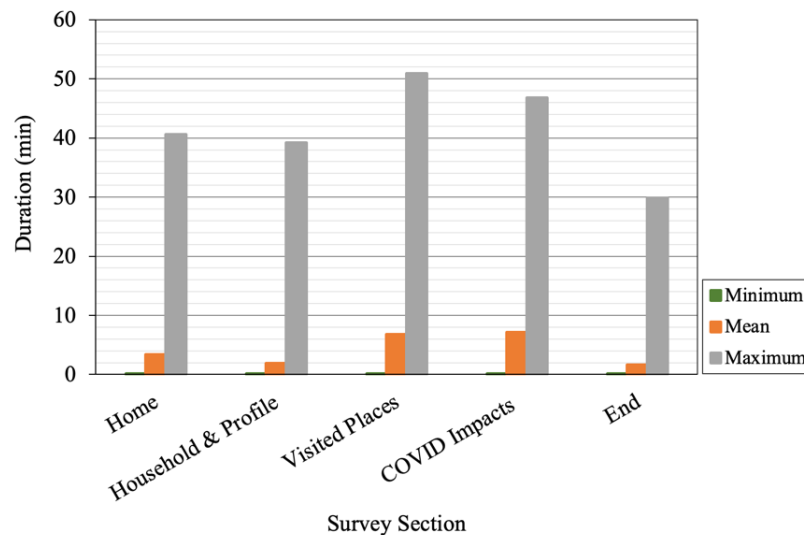


Figure 4.17 Minimum, mean, and maximum survey measured duration in each survey section

4.4.1 Multiple Regressions

To undertake this analysis, two multiple regression models are estimated once considering *measured duration* as the dependent variable and once considering *perceived duration* as the target variable. The objective of this analysis is to investigate how survey duration varies by changes in frequency of different question types. Figure 4.18 shows the correlation matrix among different question types and perceived and measured duration. A box plot presenting the distribution of different question types responded by respondents is shown in Figure 4.19. According to these two

figures, since sliders and yes/no questions have a very small variation and pretty weak correlation with target variables, they are excluded from the regression analysis.

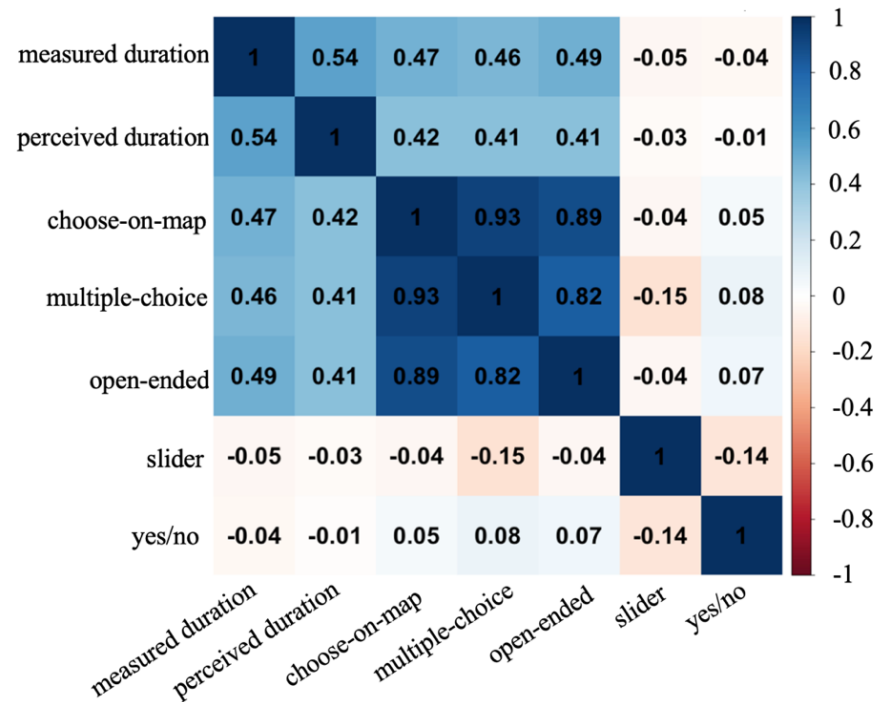


Figure 4.18 Correlation matrix between predictors and target variables

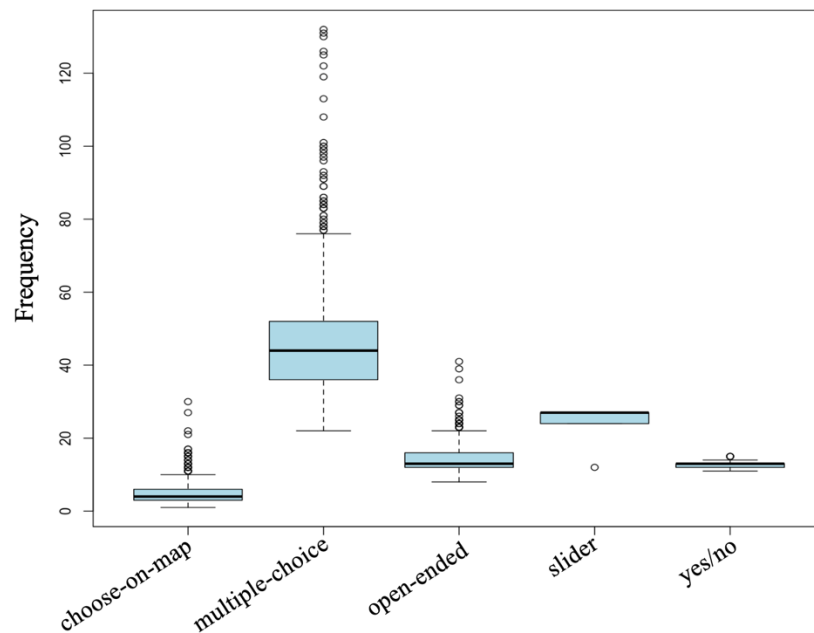


Figure 4.19 Box plot presenting the distribution of various question types

Number of choose-on-map, multiple-choice, and open-ended questions are potential predictor variables. Due to the existence of relatively strong multicollinearity among these variables, we exclude the variables that make the models unstable. Following this, the model predicting measured duration as the target variable only includes the number of multiple-choice and the number of open-ended questions as independent variables. The number of choose-on-map question can also be used as a proxy of number of places visited by respondents. Gender, age group, education level, and survey platform are included as control variables in the models. No previous study has examined the relationship between sliders and choose-on-map questions and survey duration.

The following hypothesis was made prior to performing regression models.

H1: Increasing the number of questions will increase the survey duration.

H2: Among different question types yes/no questions are less time consuming while open-ended and multiple-choice questions take more time.

H3: Choose-on-map questions are time consuming and slider questions will not take long to answer.

H4: Completing a survey on a small-screen will take longer compared to completing it on a big-screen.

H5: Women perceive surveys to be shorter compared to men.

H6: Respondents with university level education complete surveys in a shorter time compared the ones without university degree.

H7: Older people will complete the survey in a longer time compared to young adults.

4.4.1.1 Evaluation of regression models goodness of fit

The adjusted R-squared shows that 31% of the variability of measured duration is explained by predictors. The F-statistics equals to 94.35 with a p-value pretty smaller than 0.05 which illustrates that the regression model is statistically significant.

The results of the second multiple regression model performed to predict perceived survey duration, shows that about 22% of the target variable variation is explained by predictors. The model

is found to be statistically significant and useful as a whole by the F-statistic which equals 51 with a p-value smaller than 0.05.

4.4.1.2 Multiple Regression Results

The results of the two multiple regression models are presented in Table 4.10. According to the results, the number of multiple-choice and open-ended questions contributes to the model in a statistically significant manner with pretty small p-values. The unstandardized regression coefficients show that every multiple-choice question will increase the measured survey duration by 7 seconds on average while holding other independent variables constant. The similar situation applies to perceived duration, showing 7 seconds increase in perceived duration by a unit increase in the number of multiple-choice questions. Adding an open-ended question is associated with 66 seconds increase in measured survey duration and 31 seconds increase in perceived duration. Standardized regression coefficients reveal that the frequency of open-ended questions is almost twice as important as the frequency of multiple-choice questions in predicting measured duration by having higher relative weight of coefficients. When predicting perceived duration, the frequency of multiple-choice and the one of open-ended questions are relatively equally important.

As for control variables, the relationship between age group, platform, educational level, and measured duration as response variable were found statistically significant with very small p-values. This means the mentioned variables are significant predictors of outcome variables. Control variables: gender, age group, and educational level contributed significantly to the model with perceived duration as the target variable. According to the results, men perceived survey duration about 60 seconds longer compared to women. For older adults (≥ 55) the measured survey duration was found to be about 4 minutes longer compared to middle-aged respondents. Among older respondents, perceived survey length was about 2 minutes longer than among middle-aged respondents. For young survey takers, measured survey duration was found to be about 78 seconds shorter compared to middle-aged respondents. Concerning the platform used to take the survey, whether on mobile with small screen, or on desktop/tablet with bigger screen, no statistically significant association was found when predicting survey perceived duration. This can be an indication of an appropriate graphical design and survey platform compatible with small screens. However, completing the survey on a small screen was found to take on average about 47 seconds longer than filling it out on a big screen when predicting measured survey duration. Lastly,

respondents with university degrees completed the survey about 2 minutes faster than respondents without university degrees. These respondents perceived the survey duration about 1 minute and 20 seconds shorter compared with survey takers with non-university education.

Table 4.10 The results of multiple regression models

Explanatory variable	Unit of measurement	Response variable = Measured duration (min)					Response variable = Perceived duration (min)				
		Unstandardized coefficient		Standardized coefficient	t	P value	Unstandardized coefficient		Standardized coefficient	t	P value
		B	std.error	Beta			B	std.error	Beta		
(Intercept)		-0.320	1.014	—	-0.316	7.52E-01	1.375	1.300	—	1.058	2.90E-01
No. choose-on-map	Number	—	—	—	—	—	0.001	0.194	0.000	0.003	9.98E-01
No. multiple-choice	Number	0.119	0.026	0.173	4.513	6.92E-06 ***	0.113	0.031	0.230	3.644	2.78E-04 ***
No. open-ended	Number	1.100	0.106	0.402	10.332	3.47E-24 ***	0.526	0.101	0.269	5.195	2.34E-07 ***
Gender <i>women: ref</i> men	Dummy	0.097	0.428	0.005	0.227	8.21E-01	1.009	0.327	0.073	3.083	2.09E-03 **
Age group <i>middle-aged: ref</i> older Adults (≥ 55)	Categorical	4.505	0.483	0.221	9.322	4.16E-20 ***	2.106	0.372	0.145	5.668	1.74E-08 ***
young Adults (<35)		-1.305	0.592	-0.051	-2.204	2.77E-02 *	-0.086	0.453	-0.005	-0.189	8.50E-01
Platform <i>big screen: ref</i> small screen	Dummy	0.786	0.474	0.037	1.658	9.76E-02 .	0.287	0.363	0.019	0.791	4.29E-01
Education level <i>non_university: ref</i> university	Dummy	-1.950	0.501	-0.087	-3.889	1.05E-04 ***	-1.328	0.383	-0.083	-3.472	5.31E-04 ***
Adjusted R_squared: 0.3114							Adjusted R_squared: 0.2168				
F-statistic: 94.35 on 7 and 1438DF, p-value<2.2e-16							F-statistic: 51 on 8 and 1437DF, p-value<2.2e-16				

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Number of observations: 1446

4.4.1.3 Testing regression models assumptions

Lastly the residual plots of both models are investigated to assure linear regression model's primary assumption including linearity, normality, and lack of heteroscedasticity are not violated. This confirms that the linear regression is offering a good fit to the data. Residual analysis is done by plotting real values against predicted values for both models to examine trends or deviations from linearity between predicted and actual values. Based on the linear patterns for both models, it is clear that the models are quite capable of capturing the relationships among variables (Figure 4.20 and Figure 4.21).

To check for homoscedasticity, standardised residuals are plotted against predicted values. The random distribution of residuals around horizontal line ($y=0$) supports homoscedasticity assumptions being met. In other words, the random dispersion of residuals around the horizontal line shows the error terms do not have higher importance compared to others. The residual plots for both models suggest the lack of heteroscedasticity meaning that all linear regression assumptions are met, and the models are reliable. This indicates that linear regression is a good choice and provides a good fit to the data (Figure 4.20 and Figure 4.21).

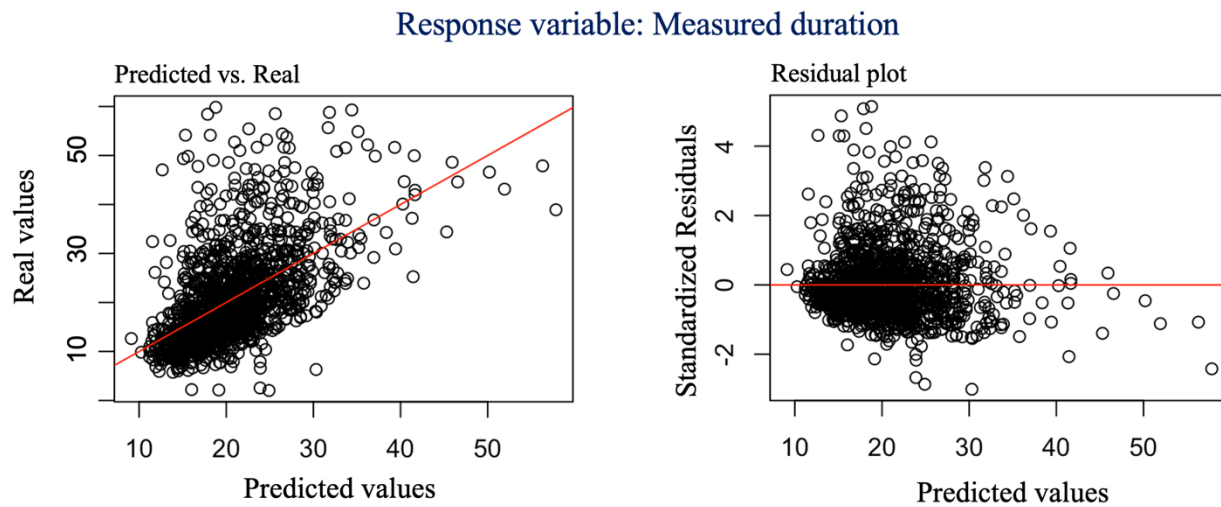


Figure 4.20 Residual analysis of measured duration model, predicted values vs. Real values (left), and residual plot (right)

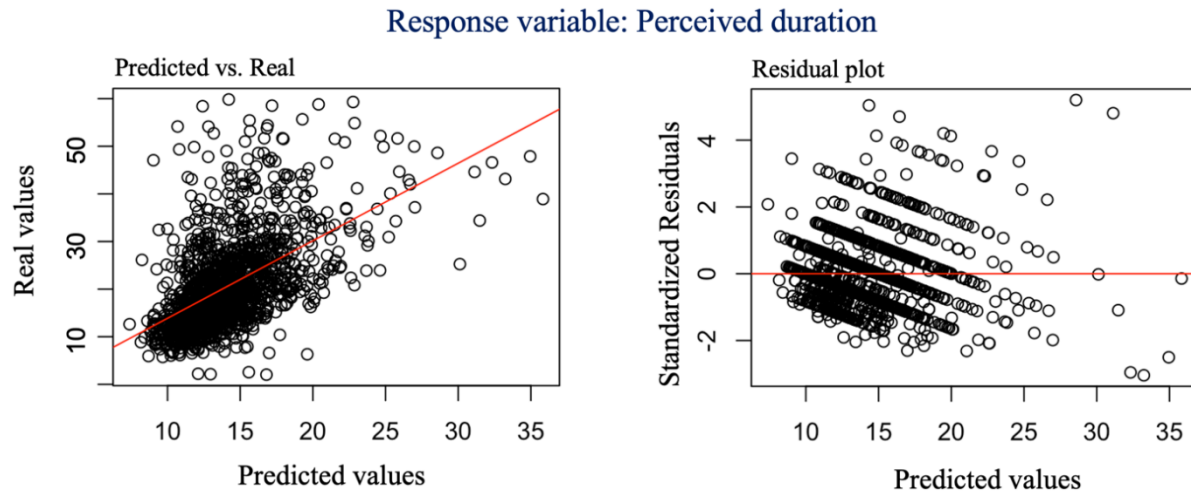


Figure 4.21 Residual analysis of perceived duration model, predicted values vs. Real values (left), and residual plot (right)

CHAPTER 5 DESIGN OF THE FIRST NATIONAL SURVEY ON TRANSPORT EQUITY

5.1 Mobilizing Justice Project and its Goals

Transportation systems are an inseparable part of people's lives. There are multiple limitations and barriers that people face in their daily travel which negatively affects individuals' health, well-being and life outcomes which leads to social inequality among people. This is where the government and policy makers should intervene to provide a more equitable system and environment to citizens.

Mobilizing Justice is a project aiming to address transportation inequities in Canadian cities through a five-year multi-sector and multi-disciplinary partnership. This project is funded by the Social Science and Humanities Research Council (SSHRC) and is a collaborative project that involves municipalities, governments, industries, and non-profit organizations and researchers. To benefit from the expertise and insights of experts in the field of transportation and mobility justice, the team is divided into six different working groups:

1. National Survey
2. Pilots and Policies
3. Transportation Modes
4. Data Driven Standards
5. Prioritizing Populations
6. Equitable Planning

This is to assure different aspects of an equitable transportation system will be reflected throughout this project ([Mobilizing Justice.ca](http://MobilizingJustice.ca)). Each working team is guided by an academic lead and a partner lead to achieve its goals. National survey working group is led by Professor Catherine Morency and Howaida Hassan and one of the contributions of this research is to discuss the steps and procedures involved in designing and administrating the first Canadian national survey on transportation poverty and social exclusion.

5.2 Contribution to the Survey Design

Learnings from the analysis in the previous sections of this work were intended to be inform into the design process of the First National Survey of Transport Equity. The aim was to pay special attention to burden formation while proceeding through various steps of the survey design along

with satisfying Mobilizing Justice project's primary objectives. The design process is planned to involve multiple iterations. During each iteration, the insights gained from the previous rounds are employed to improve and adjust various aspects of the survey. Three following resources are used to guide the decisions at each step:

1. Literature insights: Existing literature provides a strong theoretical foundation for the survey design. Whether it is related to specific aspects of the survey design, like question type and formatting, or key concepts and theories associated with different elements to address in order to satisfy the project objectives and answer the research question. Incorporating fundamental elements and theoretical framework ensures that the survey is rooted in previous findings and contributes to the advancement of knowledge in the field.
2. Research results: The results obtained and discussed in previous sections of this research study are used to guide the formulation of survey questions, response options, question types in different survey sections, and overall structure of the survey. This provides assurance to project stakeholders that respondent burden is considered as the key factor throughout the design steps and a great effort is made to reduce burden in the final version.
3. Consultation and input from stakeholders and Community Equity Advisory Table: Getting input from experts, stakeholders, and also potential respondents can provide valuable insights in the survey design. This feedback can help identify gaps, ambiguities, or areas that require improvement or modification in the survey design. Stakeholders and different research teams were involved in the design process to assure the survey is capable to address the primary research questions and the project objectives. Furthermore, the Community Equity Advisory Table (will be discussed in further detail) provided valuable guidance by reviewing the design process and by providing feedback on elements that were vague or ones that were not thoroughly examined.

An overview of the survey and the design process are discussed thoroughly in the following sections.

5.3 Introduction to the Survey

Traditional transportation surveys collect data on current mobility behaviour of residents, including the choice of mode of transportation and their travel time to different destinations. By using the

data collected from these surveys, planners decide on future infrastructure investments such as new bridges, roads, walking and cycling infrastructures. The investments aim to facilitate traveling for residents and will affect mobility choices and the time individuals spend traveling. However, traditional surveys are not capable of identifying the barriers which may prevent people from reaching their desired destinations, the trade-offs people make to afford transportation in terms of cost and time, the challenges they face in their daily travel, the amount of suppressed or unwanted travel people experience. Overall, the current surveys cannot capture the way people feel about the transportation options available to them and the choices they make in this regard. For example, someone may spend 1.5 hours in public transit to commute to work because of his dwelling location, or they may be forced to buy a car that they cannot afford due to the lack of transportation services or employment opportunities close by. A person may need/be forced to decline a job offer or limit their choices because of insufficient availability of transportation services. This means that when planners decide on new investments and projects planning, they only rely on the data which barely measures the real-life situation and take actions without considering the full impact of transportation systems on people's lives. The Mobilizing Justice project aims to conduct the *First national survey on transportation poverty and transportation-related social exclusion in Canada*, so that constraints people face and the trade-offs they make can be considered in the design of future mobility and urban plans.

Transportation poverty is the product of socio-economic marginalization combined with limited transportation options which induce accessibility issues. For example, lack of adequate transport services may challenge someone to get required healthcare, may force a person to take longer trips to meet his needs which is undesirable on a daily basis, or may cost more than one can afford to make the trip. Constraints such as experiencing racism, homophobia, fear of COVID-19 transmission, lack of elevators in metro stations, etc., can discourage people from using certain modes. People with low incomes are amongst the most vulnerable population groups when it comes to transport poverty, which makes them a key target for the first National Survey on Transportation Justice. This survey aims to gather data on people's aspiration with respect to home location and transportation choices, the barriers, and constraints they face, the level of satisfaction they experience throughout their daily routines, and the social and economic outcomes they gain. The data will allow these people's voices to be heard and considered in future infrastructure

investments, strategies, plans, and non-infrastructure interventions to insure a more equitable urban future (Mobilizing Justice.com).

5.4 Survey Design

To make sure the survey satisfies its objectives various steps are taken as following:

1. Workshop on national survey
2. Forming Community Equity Advisory Table (CEAT)
3. Collaboration with CEAT
4. Create a pool of all potential questions
5. Survey MJ leads with the pool and gather feedbacks
6. Meet MJ leads and exchange ideas
7. Develop 4 different scenarios of questions for pre-pilot phase

5.4.1 Workshop on National Survey

In the first step, in order to brainstorm and to identify the main elements to be included in the survey, an online workshop of 70 stakeholder from federal and regional agencies, municipalities, academia, non-profit, industry and community advocacy was held in November 2021. During the workshop, attendees discussed the following questions in separate breakout room sessions:

- How should we ask about mobility aspirations in the survey?
- How should we ask about mobility barriers and constraints in the survey?
- How should we ask about social, economic, and wellbeing outcomes in the survey?
- How should we ask respondents about their satisfaction with mobility, access, and the built environment?
- How should we ask about respondents suppressed and excess travel in the survey?

Table 5.1 summarizes all elements discussed in the workshop that could be included in the survey.

Table 5.1 Elements discussed in the workshop (replicated from brainstorm session report)

Theme	Potential elements to be asked
Mobility aspiration	Residential aspiration Access to destination aspiration Public transit aspiration Active travel aspiration Automobility aspiration Person specific aspiration Quality of life aspiration
Barriers and constraints	Social and institutional barriers Digital and transactional barriers Safety barriers Design and active travel barriers Public transit barriers General and multimodal barriers
Social, economic, and well-being outcomes	Employment and educational outcomes Well-being outcomes Participation outcomes Social connections Health outcomes Financial outcomes General well-being outcomes
Mobility satisfaction	Satisfaction with functional ability of travel to meet needs Satisfaction with experiential components of travel Mode-specific satisfaction Satisfaction with built environment and access Emotional and psychological satisfaction
Suppressed and excess travel	Suppressed essential activities Missed or late to work or school Suppressed physical/outdoor activities Missed social/cultural/civic activities Missed children's activities General characteristics and causes Public transit excess travel by traveling further distances Public transit excess travel by spending more time traveling compared to other modes Waiting time for buses Delays induced by waiting for a less crowded bus. Destination induced excess travel (driving to amenities) Costs and trade-offs (avoiding the costs of faster modes) Forced car ownership Living further from work to afford housing Household induced (social norms to chaperon dependents)

The main outcomes of the National Survey Workshop include data requirements and points of interest from stakeholders and project partners followed by recommendations on national survey. First and foremost, an important recommendation related to survey development was the equitable co-creation process and engagement plan, which led the National survey team to collaborate with the Community Equity Advisory Table (CEAT).

5.4.2 CEAT Formation

An advisory committee of 20 people has been formed under the name Community Equity Advisory Table (CEAT) to provide advice on the scope of work, the selection of criteria, and recruitment process. The committee included members from across Canada with different mobility needs and expertise in different aspects affecting equity deserving communities who provided insights into survey design to working team and guided and reviewed Mobilizing Justice research and engagement.

5.4.2.1 Collaboration with CEAT

5.4.2.1.1 Pre-meeting survey

The survey design team decided to meet CEAT members to benefit from their knowledge and expertise. CEAT members were sent a draft questionnaire of suggested survey questions and elements before the meeting so they can provide feedback and ideas on these potential elements. This pre-workshop questionnaire was crafted based on five themes discussed at the National Survey brainstorming session in November 2021 with the entire partnership. As part of the original brainstorming process, the National Survey Working Group came up with more than 100 possible survey questions that would be representative of the topics discussed at the November 2021 workshop. A representative subset of these possible questions was sent to CEAT members prior to the meeting, encompassing the following elements presented in Table 5.2.

Table 5.2 Content of pre-meeting survey

Survey section	Details
Open-ended box	Topics that must absolutely be included in the survey
Evaluation of proposed questions in 4 categories	Home location Access to destination Access to mobility services and infrastructures Mode choice and mobility behaviour
Survey design components	Optimal survey length Question types Language

Written feedback was provided by CEAT members on these draft elements through a questionnaire ('a survey on survey questions'). A dozen CEAT members contributed their input through this pre-workshop questionnaire regarding what topics should remain in the survey and what topics needed to be added, removed, or reworded.

5.4.2.1.2 Feedback received from CEAT

The feedback received from CEAT members on the second component of pre-meeting survey can be summarised in two main sections. What is asked in a vague way and what is missed (Table 5.3).

Table 5.3 Feedbacks from CEAT on second component of pre-meeting survey

Vague elements	Missed elements
• Satisfaction with home location • Safety barriers • Suppressed and excess travel • Trade-offs • The share of income spent on transportation	• Availability/cost of paratransit services • Rural transportation poverty • Concerns about carbon footprint and climate change • Appreciation of multimodality

Several CEAT members suggested that satisfaction with the location of the home should be addressed in a more focused manner to determine if satisfaction is influenced by the location of the house itself or by its neighborhood. In addition, respondents believed that people would interpret the concept of safety differently if they were asked how safe or unsafe they feel in their daily commutes or other situations. In this regard, it is necessary to provide different definitions of safety (e.g., safety from harassment, safety from crime, safety from Covid transmission, etc.). CEAT also

suggested to define suppress and excess travel for respondents before asking the associated questions. CEAT members also shared their perspectives on the trade-offs people make based on where they live and what they have access to. They argued that asking what people have sacrificed because of their home location or what they wish to have access to can be beneficial.

The third component of pre-meeting survey is associated to survey design. Figure 5.1 illustrates feedbacks from CEAT members on complexity of different question types and optimal survey duration. Regarding question type, the highest share of respondents believed that multiple response questions are the easiest type of questions to answer, followed by yes/no questions. This was against the findings of our regression analysis where multiple-choice questions were found to be quite time consuming. There may be a connection between the perceived value of such questions to CEAT and their expertise in the field, which makes responding to these questions easier than it would be for a typical respondent. Questions including maps were ranked as the most difficult ones to answer. After that, open-ended questions got the highest rate of complexity among CEAT members. About optimal survey duration, higher share of respondents (69%) mentioned that 6-10 minutes is the optimal survey length. 23% percent believed that they prefer to fill-out a survey that takes less than 5 minutes to complete (Figure 5.2). And lastly only 8% percent were of the opinion that 11 to 15 minutes survey is optimal. This finding is in line with previous studies indicating that longer questionnaire imposes more burden on respondents. However, the results of our current analysis and some other research revealed that duration is not the sole indicator of burdensomeness and other factors such as survey design have impact on burden and response rate.

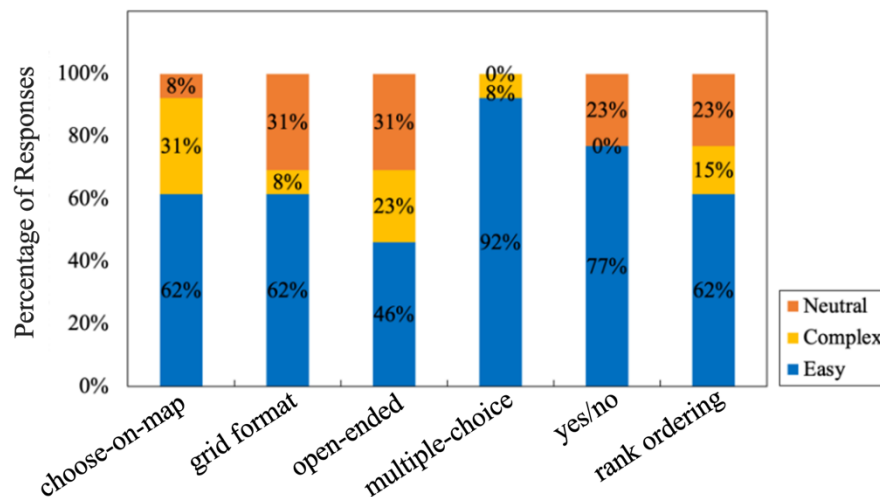


Figure 5.1 CEAT member feedback question types and perceived complexity

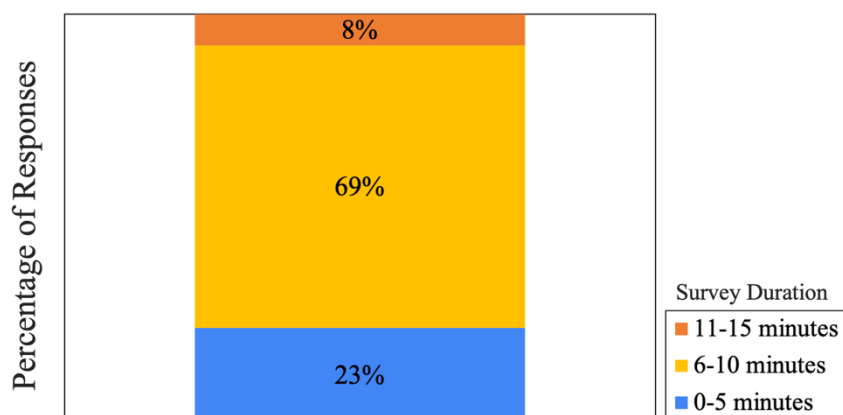


Figure 5.2 CEAT members feedback on optimal survey duration

5.4.2.1.3 Meeting with CEAT

As a result of the feedback received on the pre-meeting questionnaire, the meeting mainly focused on evaluating proposed questions, including four main elements:

- Satisfaction with home location
- Suppressed and excess travel
- Safety barriers
- Trade-offs made on the basis of home location and availability of transportation options

The survey design team met the CEAT in July 2022 to discuss these elements and evaluate new approaches to address these questions. The discussions were carried out in multiple break-out room sessions so that participants could discuss their perceptions on structure and comprehensibility of the questions, as well as their relevance and importance. Through these discussions, the goal was to ensure that the survey can respectfully collect data associated to the needs of equity deserving populations. Bellow, a summary of what was discussed in the meeting with CEAT is presented.

- Satisfaction with home location

The survey design team emphasized the importance of asking about satisfaction with respect to home location, which can indicate if there is a mismatch between where people live and where they would prefer to live. Also, asking this question helps to determine what trade-offs people make based on their location, and evaluating people's sense of belonging and social cohesion with the

place they live. There was a consensus that “home location” is a vague concept, so asking questions about “current residence” or “neighborhood” is a clearer approach. This allows the respondent to distinguish if the question is about the dwelling itself or what is available in its surrounding. According to CEAT, it is useful to ask what is available around the respondent’s place of residence, and what is missing. It was believed that this would aid in understanding what people would like access to and what are their aspirations. It was also discussed that the majority of people do not have the option to choose where they live, and when they do have a choice, it is more affected by their families (e.g., being close to families, neighbourhood with the same culture) and not amenities nearby, so the best approach would be to not ask about the choice of home location. Finally, a member suggested to use a less formal term, such as “happy with”, rather than “satisfied with” through the survey questionnaire.

- Trade-offs based on home location and available transport options

The survey team elaborated the importance of asking questions associated with trade-offs people make because of their home location to understand what people are giving up (e.g., time, money, safety, leisure activity) and to uncover the reason for dissatisfaction. The first comment was to re-word the term “trade-off” and instead use a less academic terms such as the things that are “left-behind” or “given-up” or even the sacrifices people made. CEAT also brought up the comment of including broader categories of sacrifice i.e., social and cultural. The possibility of questioning the concept of trade-off using a completely different approach was also suggested. As such, asking how much of a transportation situation is being caused by the place people live, or what are the things people are trying to accomplish and whether they are happy with their accomplishments. Asking more specific questions such as what they are losing because of long travel time, or what they could do if they did not have to spend so much time commuting was also a suggested approach. One mentioned that the question can be asked from the perspective of not having access to personal transport option such as how happy/unhappy people may be because of choosing to live close to metro since they do not own or cannot afford a personal vehicle.

- Suppressed and excess travel

The survey team also aimed to understand why people make longer or more trips than they would like to. In addition to learn what factors contribute to people missing or delaying important appointments and how much of it can be attributed to travel or home locations. To address

suppressed travel, the survey team proposed to ask respondents whether they have missed or delayed important opportunities because of available transportation options and if yes how often they experienced such situation. Excess travel has been addressed by asking respondents whether they took longer trips to be safe or to save money, or if they had to leave very early to work because of the un/availability of transit options. CEAT argued that a broader range of options e.g., job opportunities, community engagement or volunteer work, social opportunities such as youth club, and recovery meetings and support groups should be listed in potential suppressed trips. They also suggested exploring the reasons behind suppressed and excess travel, such as affordability, safety, accessibility to safe routes for people with mobility restrictions or avoiding multiple transfers. Members also suggested setting a timeline for assessing the frequency of excess and suppressed activities.

- Barriers and constraints: Safety

Lastly, the survey team has the objective to explore the barriers people face in their daily activities, how these barriers affect the way they travel and how much these barriers may impact people's satisfaction with their home location and mobility options. The survey team suggested asking how safe people feel from crime, harassment, profiling, and traffic when traveling. There were multiple comments indicating that the concept of "safety" must be clearly defined. Commenters suggested that it should be clearly indicated whether safety is associated with the mode of transport or the built environment. One suggested asking respondents if there was any part of their trip where they felt unsafe. This is believed to be helpful in distinguishing the time and situation people feel unsafe and the probable reasons. CEAT suggested to consider the flexibility around the personal context, i.e., one may feel unsafe travelling alone vs traveling with children or friends. And lastly, distinguishing between daily/non-regular trips and local/non-local trips seemed important to avoid confusion.

The National Survey working team incorporated many suggestions to create a pool of potential questions.

5.4.3 Question pool

5.4.3.1 Survey sections

Using feedback from CEAT members a pool of possible questions to include in the final survey was created. It was decided to test different scenarios (combination of questions) in a pilot survey to assess the best approach(es) for the National Survey. The assessment will be based on two criteria:

1. Relevance for MJ objectives

The relevance of various questions will be assessed by MJ working groups and MJ partners to see which one(s) allows gathering the most relevant data.

2. Burden/duration/response rate

The post-survey questions will be asked at the end of the pilot survey to understand how burdensome the questionnaire is and how the survey is perceived in terms of task difficulty. This is important since higher burden is associated with higher drop-outs and careless responding which may bias the results. Future works may benefit from the knowledge gained from the burden analysis of the previous sections in order to select scenarios and modify them to reduce the burden level.

The pilot survey will include 4 main sections of questions as presented in Figure 5.3.

Descriptive & factual elements				Perception / aspiration				Burden, further control variables
Home	Household members	Respondent profile	Trip & Activity	TP & TRSE				END
Location Dwelling features Car ownership	Features of all members (age, gender, driving license, etc.)	Mobility tools Type of occupation Education Cell phone Limitations	Main activity locations Transportation modes	4 scenarios (random)				Complexity Perceived duration Missed points Income
				A	B	C	D	

Figure 5.3 Schematic overview of pilot survey design

The break ups of different sections are as follow.

1. Descriptive and factual elements

This section gathers data associated to sociodemographic characteristics of respondents, i.e., information on respondent's home location, age, gender, household size, car ownership as well as information associated with main activity locations and transport modes. It is essential to gather this information so they can be further used as control variables.

2. Questions on perception and aspiration

These questions can be considered as the survey core since they are designed to capture the equity and justice elements of the survey. Questions regarding the satisfaction with travel behaviour and residential location, the mismatch between the aspired behaviour and the real one, self-perception of people's mobility behaviour and respondent life outcomes will be addressed in this part. Different scenarios will be built by changing the questions in this section and the rest will remain untouched. The proposed scenarios will address MJ objectives through different lenses including aspiration, satisfaction, barriers, trade-offs, suppressed and excess travel/activities.

3. Questions regarding burden and further control variables

In the last section of the questionnaire, post-survey questions associated to survey perceived burdensomeness, interestingness, difficulty, value, and perceived duration will be posed. Additional socio-economic questions will also be added such as a question capturing respondents income level.

5.4.3.2 Developing the question on perception

To make sure the National Survey captures all MJ objectives and that questions will address all aspects of individual's daily mobility activities, we thought of 4 anchor points.

1. Home location (the dwelling and the neighbourhood)
2. Mobility (associated to the trips made by individuals)
3. Destination
4. Well-being and outcomes

An individual's home location is chosen as the first anchor point because it is where his first and last daily trips begin and end. The second anchor point includes all respondent's mobility activities (e.g., walk from home to transit station, the station itself, transit service) starting from the home location to destination. The third anchor point deals with all elements associated to destination alone. The factors associated to *access* to destination will be addressed in the mobility anchor point. Lastly, well-being and life outcomes constitute a separate anchor point since it addresses respondent's self-perception of what they achieve in life and whether they are satisfied with what they achieve considering where they live and what they have access to.

To make sure the survey is easily understood by the target population and to avoid confusion, we decided to develop a set of simple generic question structures for each type of question. Although it is not applicable to all questions, the goal is to use questions as simple as possible. Below is the list of generic question structures:

To ask about satisfaction: Agree/disagree question.

“To what extent do you agree with the following.”

- I am happy/unhappy with _____. (dwelling, neighbourhood/means of transport)
- I am able to/can _____.
- I can participate/accomplish/access _____ as much as I would like.

To ask about aspiration/mismatch: agree/disagree question.

Following the descriptive question,

- I would rather _____.
- I wish I could _____.
- I wish I had access to _____.

To ask about trade-off: Agree/disagree question.

- I am losing _____ because of _____.
- I am giving up _____ because of _____.
- I am sacrificing _____ because of _____.
- I do _____ because of _____.

To ask about safety: Agree/disagree question.

- I feel safe from _____.
- I am concerned about _____.

5.4.3.2.1 Questions associated with the first anchor point: Home location

When thinking about home location, *dwelling* and *neighbourhood* attributes should be distinguished. In this anchor point, the very first element to consider is the dwelling itself and the second is what is around (neighbourhood). This helps respondent to know what aspect of their experience and their perception is under study. All potential variables associated with dwelling and neighbourhood variables are then recognized and listed (Figure 5.4). Each of these factors can be looked at from different lenses (e.g., satisfaction, aspiration, trade-off) but this does not mean that all these variables will be asked in the questionnaire. The next step is to recognize the best approaches of asking questions related to each of the listed variables to reach higher participation rate and to gather high quality of data.

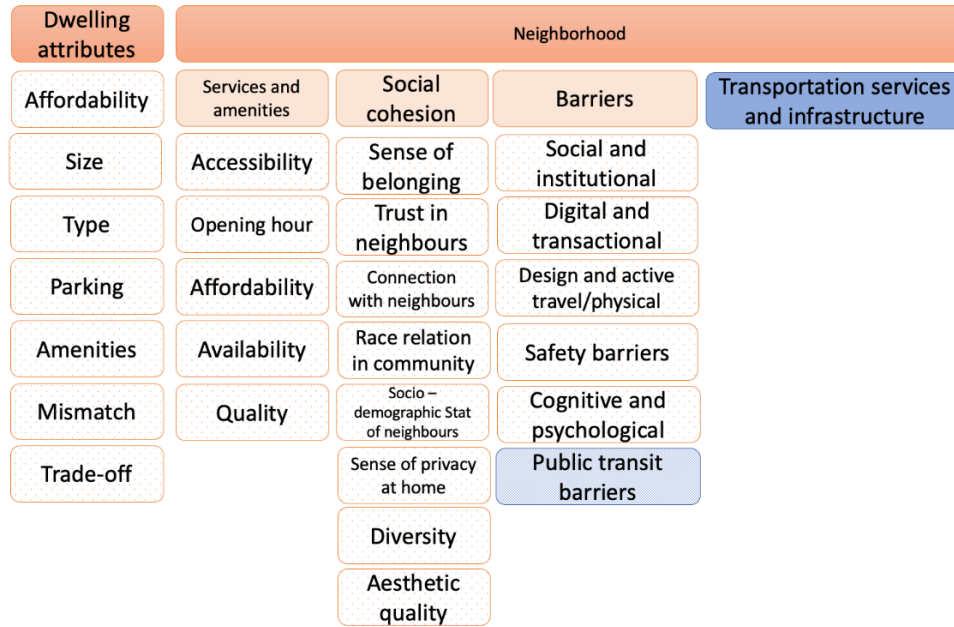


Figure 5.4 Elements associated with home location

The elements came in blue boxes will be addressed in the second anchor point since they target transportation services and infrastructures.

5.4.3.2.2 Questions associated with the second anchor point: Mobility

The second anchor point is identified as mobility anchor. In this section all the elements associated to one's mobility activities from the doorstep of his home location to his destinations are considered. This includes the transportation mode one uses, car ownership, membership of shared services, infrastructure, and available services (Figure 5.5).

Mobility						
Ownership	Shared-services	Infrastructure	Public Transit		Mode choice	
Aspiration	Aspiration	Quality	Transit station and stops	Service		Mode to access primary destination
satisfaction	satisfaction	Universal design	Accessibility (walking distance)	Affordable	Comfort	Cost
Mismatch	Mismatch	Safety from crime	Information board	Availability	Reliability	Preferred mode
	Distance	Safety traffic accident	Purchase station	Accessibility	Perceived value	Mismatch
	Affordability	Social Safety	Shelter design	Access to destination	Perceived Duration	Alternative for primary mode
	Multimodality	Greenery	Availability of seat	Universal design	Payment method	Captive: Only option to Access to destination
		Noise level	Universal design	On-board Safety	Barriers	Flexibility and freedom
		pollution	Signage(lang uage)	Frequency of use	Off-peak options	To keep fit
		Active travel design	barriers	Number of transfers	Service coverage	Environmental concern
		Road design		Route finding		Parking availability

Figure 5.5 Variables associated to mobility anchor point

5.4.3.2.3 Questions associated with the third anchor point: Destination.

The third anchor point is the destination which includes questions about access to transit services at the destination, availability of parking and its effect on mode choice, safety elements, and quality of infrastructure (Figure 5.6).

5.4.3.2.4 Questions associated with the fourth anchor point: Well-being.

Lastly the final anchor point investigates different aspects of life outcomes one gains based on his residential location, places to go frequently, and with available transportation options. The questions in this section address subjective well-being, employment/educational, social, participation, health, and financial outcomes. In addition, questions regarding suppressed and excess travel one may experience are addressed in this section. This includes a variety of elements including missed or delayed opportunities, forced car ownership, excess travel times, and cutting leisure time (Figure 5.6).

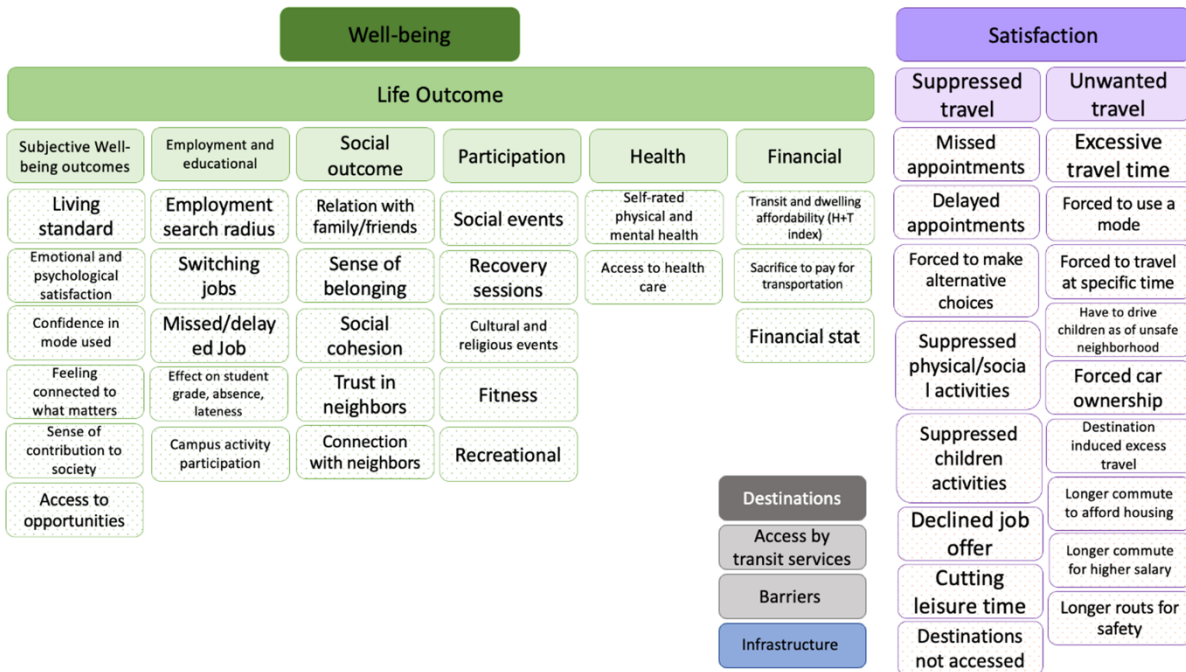


Figure 5.6 Variables associated to well-being and destination anchor points

Considering the elements associated to each anchor point and using various perspectives, an excessive list of questions was designed and coded in the survey platform.

5.4.4 Mobilizing Justice Survey Test

5.4.4.1 Testing the pool of questions by MJ leads

After implementing the pool of questions into the survey platform, the following objective was to ensure that the questions are designed in accordance with the research objectives of the various MJ groups. To accomplish this, the questionnaire link, alongside with a feedback document including all questions was sent to MJ project leads. The feedback document allowed respondents to: 1) indicate whether a question should be mandatory included in the final survey, 2) evaluate the importance of each question, and 3) provide comments, suggestions, or proposals related to each question or additional questions. This is helpful to identify the most important questions, relevant ways to ask those questions, and missing elements. The final goal was to generate different scenarios including only a subset of questions in the perception section.

The written feedback was collected from MJ leads. Leadership feedbacks on objective elements were associated to unclarity of the need of asking about respondents' precise address, occupation, type of car and employment/educational location. According to the survey design team, the precise

address is required for data fusion purposes with various sources using various zoning systems, and to generate precise features of the neighbourhood namely walking access distance to transit stops and local amenities. The information of respondent's occupation is considered important for segmentation of people. Information regarding respondent's type of car was necessary to estimate costs and to assess access to electric cars. Lastly, questions on employment/educational locations are referred as vital to estimate transportation options for constrained trips.

According to the overall feedback from the MJ leads, questions concerning "sense of belonging" are regarded as of "low importance" were excluded in the final scenarios. The expected uses of survey results are diverse and not all anticipated and this further confirmed the interest of testing multiple pools of questions.

In addition to written feedback, two online interviews were conducted with the two MJ working groups namely: Data Driven Equity Standards, and Transportation Modes. The objective was to gain a clear and precise understanding of their needs and expectations regarding the survey as the teams planning to use MJ survey data directly.

According to discussion with leads of Transportation Modes working group the following recommendations were collected:

- **Frequency of usage** of usage of different modes accompanied by asking about the travel purpose.
- **Focus on non-educational and non-work trips** to study the big picture of transport complexity and challenges experienced by vulnerable groups.
- **Meet consistency on different modes** and include questions addressing cycling and walking and be less focused on public transit.
- **Examine respondents' perception of cycling, walking, driving, and transit**, regardless of whether they use these modes frequently or not. Capture if respondents find these modes convenient, affordable, reliable, safe, practical, and enjoyable.
- **Study enjoyment of travel** experienced by respondents using different modes.
- **Not concentrate on driving experience** and car ownership and parking facilities.

The following recommendations were made by Data Driven Equity Standards working group to be included in the survey.

- **Affordability** of housing and the link to transportation
- **Suppressed travel** and trips people cannot make because of lack of transportation.
- **Language and digital barriers** to using transit, bikeshare and ride hailing, etc.
- **Perceived safety** as the significant barrier to travel for equity deserving groups.

5.4.4.2 Strategies for Scenario development

Relying on the learnings from the MJ leadership team feedback, the survey design group articulated questionnaire scenarios. Each scenario was developed by focusing on only one challenge (i.e., suppressed/excess travel) and only one component (i.e., dwelling and home location, accessibility to services and opportunities, and availability of transit options).

Consequently, the pre-pilot survey includes 4 scenarios:

1. First focusing on equity of access to opportunities
2. Second focusing on excess and suppressed travel + car ownership.
3. Third focusing on safety and trade-offs.
4. Fourth focusing on home locations and access to transit.

These scenarios are planned to be launched in social media and in a pre-pilot phase. The results then will be examined by different MJ groups to assure their objectives are met. Different survey scenarios will also be assessed in terms of burden and required effort. The results then will be discussed with CEAT to make decisions on the required modifications before launching the real pilot survey. An overview of survey platform visual interface is presented in Figure 5.7.

Home > Household > Profile > Usual places > **Perceptions** > End

MOBILIZING JUSTICE

We invite you to participate in this pilot study, carried out as part of a pan-Canadian research project on transportation justice, which aims to document the barriers and constraints encountered by people in their daily travels as well as their aspirations and preferences about transportation options in their neighborhoods. This pilot version will help determine the final questionnaire for the national survey scheduled for 2023.

[Start](#)
Français

Home > Household > Profile > Usual places > **Perceptions** > End

What was your household income range before taxes (gross income), in 2022? (optional)

Your comments and suggestions on the survey

Would you be interested in participating to other studies conducted by the Chaire Mobile at Polytechnique Montreal?

Do you feel that the length of the survey was too short, about right or too long?

How much time do you think you spent answering this questionnaire (in minutes)?

How interesting was this survey to you?

How difficult or easy was it for you to answer the questions in this survey?

How burdensome was this survey to you?

Have you been inclined, at any point, to abandon the survey?

Home > Household > Profile > Usual places > **Perceptions** > End

Are you going to an educational institution (school, college, CEGEP or university) to study or attend courses/lectures there? If you are registered there, but never go, choose "No"

For which reason do you never go to an educational institution to take courses?

[Save and continue](#)

Home > Household > Profile > Usual places > **Perceptions** > End

To what extent do you agree with the following?

I can easily afford my dwelling.

I feel I should pay less for my housing (based on the quality of the home and access to transportation options).

I would rather live in a bigger place/different type of dwelling.

I would rather live in another neighbourhood.

I would rather own my dwelling compared to renting.

I would rather have affordable indoor parking.

I have chosen my home location to have better access to public transit.

I chose my home to have better access to personal needs e.g. being close to family, children's school

Home > Household > Profile > Usual places > **Perceptions** > End

Place a friend

Main activity at this location

Location name or description

Please locate friends on the map

Home > Household > Profile > Usual places > **Perceptions** > End

Perceptions to other activities

Home > Household > Profile > Usual places > **Perceptions** > End

Home > Household > Profile > Usual places > **Perceptions** > End

Home > Household > Profile > Usual places > **Perceptions** > End

Figure 5.7 An overview of Mobilizing Justice survey sections and questions layout (montage of different screens)

CHAPTER 6 CONCLUSION

In this chapter a summary of this research project is be presented. The focus of this work was to understand, define, and quantify respondent burden as the main driver of low participation rate in surveys. The findings are aimed to provide insights into the development of the first national survey on transportation justice. Typically, the definition of burden is mostly concerned with objective survey characteristics; only recently few studies considered subjective features of the questionnaire. The research however was mostly focused on face-to-face and telephone interviews and not much work was undertaken on web-based surveys. In addition, to the authors knowledge, no study examined respondent burden in transportation related surveys.

6.1 Summary and Contribution

In this work, first a thorough overview of previous studies done in this topic is provided. The review includes research on the concept of respondent burden as a multidimensional phenomenon. Previous research classified respondent burden into two types: objective burden and subjective burden, with subjective burden having greater influence on response rates and data quality than objective burden. Therefore, the best way to quantify survey burden is to consider objective features of the survey, respondent's characteristics and the way survey tasks are perceived by respondents. The first objective of this work was to explore previous works to identify factors affecting respondent burden and to examine the magnitude of their effect in burden formation.

Among elements associated with respondents characteristics, age, gender, race, household size, and educational level were found to be effective in burden level. Survey duration, announced duration, question type, number of questions, graphical interface, and survey sponsorship were regarded as important factors influencing burden level. Lastly, among elements related to interaction between survey and respondents, perceive length, interest in the survey, contact mode, invitation letter, incentives, perceived complexity of the task, survey frequency, and perceived value were noted to influence burden formation.

Following identification and distinction of factors affecting respondent burden, different methods used in previous works to quantify burden were examined. Different models including theoretical models, Structural Equation Model, Partial Least Square Model, Logistic Regression, linear model,

Cox proportional hazards regression models and hierarchical multilevel linear models were applied in previous works to measure burden as the response variable or as a predictor. Some studies implemented objective and some used subjective burden in their analysis.

According to what we learned from previous works; the second objective was to measure subjective burden in a web-based travel survey through asking a set of post-survey questions in which respondents declared their perception of survey. The questions covered respondents' perception of survey length, interest, difficulty, burdensomeness, willingness to participate in new surveys, and their tendency to drop-out. Some of these variables were examined in previous works as drivers of respondent burden and some were aimed to use for the first time through our research. A SEM model with reflective approach is used to predict subjective burden. Three latent constructs namely perceived effort, perceived interest, and perceived length each indicating on two measured variables were used to build the measurement model. The contribution of these latent constructs was then examined in measuring burden. Our study confirms that burden is a multidimensional concept that is influenced by various factors which was also found by Fricker, Yan et al. (2014). The results show that longest survey duration and higher effort required to complete a survey will increase the level of perceived burden. The relationship among interest and perceived burden is found to be significant showing that higher levels of interest mediate burden feeling. The contribution of perceived interest on burden was investigated for the first time in this analysis.

Survey duration was often used as a proxy for burden assessment. However, the question of why respondents perceive a survey as lengthy, or brief has never been addressed in previous literature. As the third objective, classification analysis was carried out to determine the important variables in survey duration assessment. Two decision trees and a random forest model were developed to fulfill this objective. Subjective burden is used as a predictive variable in these models. The models aimed to classify the data into two classes: perceived "Long", and perceived "Correct/Short". The results found perceived burden as the most important variable classifiers revealing its ability to provide the best distinction among two classes. Inclination to abandon the survey, and perceived duration are regarded as the second and third important variables in classifying instances into pure leaves. According to the results, lower level of perceived burden (smaller than 30~37 gathered on the scale of 0 to 100) is significantly associated to perceiving a survey as "Correct/Short". In addition, respondents were more likely to label a survey "Correct/Short" if they perceived it took shorter than 20 minutes to complete. A very interesting finding was that perceived duration is a

more effective factor in assessing whether the survey is “Long” or “Correct/Short”. Thus, special attention should be made to perceived duration rather than real measured duration in the process of survey design. Decreasing perceived duration can be fulfilled by paying more attention to question formulation and the survey interface to keep respondents’ engaged in the survey and increase the probability of full completion.

As the fourth objective, we investigated how subjective characteristics of the questionnaire were affecting perceived and actual measured duration spent completing it. This analysis is noted very beneficial for providing insights to survey designers and to understand what type of questions to include and what types to avoid in their surveys. Exploratory data analysis showed that respondents spent a relatively short time to answer slider questions. Moreover, two multiple regression models were developed with measured and perceived duration as target variables. According to the results, the relationship among multiple-choice and open-ended questions with both target variables were statistically significant. Every multiple-choice question was found to increase perceived and measured duration by about 10 seconds. For open-ended questions, the increase in measured duration was found to be 66. This is estimated to about 30 seconds for perceived duration.

Lastly, learning gained from the beforementioned analysis was used in the process of designing the first national survey related to transportation equity and social exclusion in Canadian cities. This work was accompanied by collaboration with different partners from academia and industry. To assure the survey covers different aspects of respondent’s mobility behaviour, different anchor points namely: home location, mobility, destination, and life outcomes with multiple components were considered in developing questions. Considering the objectives of the MJ team as a priority, we tried to design the questions in a way to minimise subjective effort required for its completion. Multiple slider questions were included in different scenarios of the pre-pilot surveys and open-ended questions were avoided.

In conclusion, this research sought to develop approaches to quantify respondent burden taking to account both subjective and objective aspects of the burden concept. The objective was to then establish some guidelines for future survey designs to keep respondents engaged in surveys while collecting data of high quality.

6.2 Limitations

The very first limitation of this work is concerned with the sampling technique. Opportunistic sample includes respondents with certain characteristics which makes it difficult to make inference to a broader population. In this work, survey respondents were the ones who already expressed their interest in participation in future surveys and this sampling approach introduces some level of bias into the results.

Secondly, this analysis is mostly based on post-survey questions addressing burden. Consequently, it was not possible to ask many questions from respondents who may already feel burdened. A limited number of questions prevented us from measuring a large number of indicators, which were then used in the latent factor analysis. For this reason, during latent model analysis, we set aside the rule of using three indicators for each construct and defined each latent construct with only two manifests. This means that each of these unidimensional two-factor construct are solely under identified. Although integrating these constructs into a model with other constructs lead the final model to be in overall identified, the best practice would have been being able to build identified latent variables. Another prerequisite of measurement model is to establish convergent and divergent validity and to fulfil internal consistency. As mentioned earlier, despite satisfying discriminant validity for all three factors implemented into the measurement model, internal consistency and convergent validity was only fully established for PL and PE and not for PI. This means that there are some shortcomings in the conduct of this study that should be addressed for future works. For PI factor, the results of CFA show there are other hidden aspects to be explained by this construct. In this regard, more questions are required to be asked or the existing ones need to be re-worded to address different aspects of PI. By doing these, in addition to increasing the chance of fulfilling construct validity, more indicators will be available and the ≥ 3 indicators rule can be satisfied.

Another limitation related to the use of post-survey questions to capture respondent's self-assessment is that these questions are asked to individuals who were motivated and encouraged enough to complete the survey and proceed to the final section. This demonstrates their appropriate level of interest and value towards the different components of the survey. This may lead to a selection bias in the sample since their attitudes towards the survey will likely be relatively positive. To overcome this, questions regarding self-perceptions can be included at the end of each section

of the survey with the risk though to further increase the survey duration and burden. As a result, questions will be asked at a time when respondents are not completely exhausted or burdened. On the other hand, the analysis of burden can be undertaken in a relative micro level. This helps survey designers to better examine drivers of subjective burden and to understand whether burden feeling comes from some specific questions or from the overall cumulative effect of all questions. Another practice would be to define target groups and monitor their perception of surveying task while completing a survey. Some approaches such as tracking mouse movements could be developed to see whether they are interested or bored in the course of survey.

In the section regarding the regression analysis, a limitation was associated with the lack of variation in the frequency of yes/no and slider questions, which resulted in the exclusion of these questions from the regression models. Therefore, their contribution to measured and perceived duration variation must be investigated in other surveys on transportation and mobility behaviour. This can be helpful in developing guidelines for future survey designs.

6.3 Research perspectives

Considering what was learned from CFA and SEM, the concept of respondent burden still lacks an adequate conceptualization. To more accurately develop latent constructs more questions are required to be asked or the current ones need to be reformulated. This will provide more variables to conduct burden related analysis and to measure different aspects of the concept. More specific attention should be given to explore the positive dimensions of burden feeling called gratification. In other words, being a survey respondent is not solely about being burdening out and there are always some positive aspects involved. Some people value the effort they make to share their experience and to contribute to the future developments. That being said, it is quite interesting to include questions capturing positive feeling of respondents such as perceived value at the end of the survey. Since there is not adequate literature to rely on for the conduct of future studies addressing burden, looking into survey-based studies conducted in other fields can shed some light in developing pre-tested reliable questions.

A second point to be discussed is that the current study only considered respondents feelings and self-perception of the survey, in which burden was viewed solely as an individual's subjective task. Although subjective burden is regarded as more important in data quality and participation rate, fully understanding and quantifying burden necessitates investigating burden in both objective and

subjective manners. The number of questions, different question types, word count on each question or each choice, and graphical interface play important roles on burden development. Hence, in future studies, indicators should be developed to quantify the objective burden and attempts should be made to find links between the subjective and objective aspects of the burden concept.

Lastly, this research only focused on studying burden and how different elements lead to burden formation without defining measures to examine the outcomes. It is suggested to develop indicators of data quality (i.e., unit or item nonresponse, response quality) to look at burden and what it causes from a wider perspective.

CHAPTER 7 REFERENCES

- Abd Elrahman, S. M. and A. Abraham (2013). "A review of class imbalance problem." Journal of Network and Innovative Computing **1**(2013): 332-340.
- Arentze, T. A. and H. J. P. Timmermans (2004). "A learning-based transportation oriented simulation system." Transportation Research Part B: Methodological **38**(7): 613-633.
- Backor, K., et al. (2007). Estimating survey fatigue in time use study. international association for time use research conference. Washington, DC, CiteSeer.
- Bergman, L. R. and R. Brage (2008). "Survey experiences and later survey attitudes, intentions and behaviour." Journal of Official Statistics **24**(1): 99.
- Bottone, M., et al. (2021). "Response Burden and Data Quality in Business Surveys." Journal of Official Statistics **37**(4): 811-836.
- Bourbonnais, P.-L. and C. Morency (2018). "Factors affecting interview duration in web-based travel surveys." Transportation Research Record **2672**(42): 33-44.
- Bradburn, N. (1978). Respondent burden. Proceedings of the Survey Research Methods Section of the American Statistical Association, American Statistical Association Alexandria, VA, USA.
- Breiman, L. (2001). "Random forests." Machine learning **45**: 5-32.
- Breiman, L. (2017). Classification and regression trees, Routledge.
- Briz-Redón, Á. (2021). "Respondent Burden Effects on Item Non-Response and Careless Response Rates: An Analysis of Two Types of Surveys." Mathematics **9**(17): 2035.
- Busby, D. M. and K. Yoshida (2015). "Challenges with Online Research for Couples and Families: Evaluating Nonrespondents and the Differential Impact of Incentives." Journal of Child and Family Studies **24**(2): 505-513.
- Cheung, G. W. and R. S. Lau (2008). "Testing mediation and suppression effects of latent variables: Bootstrapping with structural equation models." Organizational research methods **11**(2): 296-325.
- Chung, B., et al. (2021). "An inductive experimental approach to developing a web-based travel survey builder: developing guidelines to design an efficient web-survey platform." Transportation Planning and Technology **44**(5): 487-502.
- Cobanoglu, C., et al. (2022). Embedded Questions in Online Survey Email Invitations: The Impact on Response Rate and Quality. Contemporary Research Methods in Hospitality and Tourism, Emerald Publishing Limited.

Crawford, S. D., et al. (2001). "Web Surveys: Perceptions of Burden." Social Science Computer Review **19**(2): 146-162.

Cutler, D. R., et al. (2007). "Random forests for classification in ecology." Ecology **88**(11): 2783-2792.

Du, W. and Z. Zhan (2002). "Building decision tree classifier on private data."

Dykema, J., et al. (2013). "Surveying Clinicians by Web: Current Issues in Design and Administration." Evaluation & the Health Professions **36**(3): 352-381.

Fang, J., et al. (2012). "Participation willingness in web surveys: exploring effect of sponsoring corporation's and survey provider's reputation." Cyberpsychology, Behavior, and Social Networking **15**(4): 195-199.

Fisher, S. and L. Kydoniefs (2001). Using a theoretical model of response burden (RB) to identify sources of burden in surveys. 12th International Workshop on Household Survey Nonresponse, Oslo, Norway, September.

Fornell, C. and D. F. Larcker (1981). "Evaluating Structural Equation Models with Unobservable Variables and Measurement Error." Journal of marketing research **18**(1): 39-50.

Frankel, J. (1980). "Measurement of Respondent Burden: Study Design and Early Findings."

Fricker, S., et al. (2012). An exploration of the application of PLS path modeling approach to creating a summary index of respondent burden. JSM Proceedings.

Fricker, S., et al. (2014). Response burden: What predicts it and who is burdened out. JSM proceedings.

Galesic, M. (2006). "Dropouts on the web: Effects of interest and burden experienced during an online survey." Journal of Official Statistics **22**(2): 313.

Galesic, M. and M. Bosnjak (2009). "Effects of questionnaire length on participation and indicators of response quality in a web survey." Public opinion quarterly **73**(2): 349-360.

Genuer, R., et al. (2010). "Variable selection using random forests." Pattern recognition letters **31**(14): 2225-2236.

Giesen, D. (2012). Exploring causes and effects of perceived response burden. Proceedings of the Fourth International Conference on Establishment Surveys (ICES IV), Montreal, Canada.

Görizt, A. S. (2014). "Determinants of the starting rate and the completion rate in online panel studies1." Online Panel Research: Data Quality Perspective, A: 154-170.

Grandmont, J., et al. (2010). Grappling with grids: How does question format affect data quality and respondent engagement. annual meeting of the American Association for Public Opinion Research.

Groves, R. M., et al. (2004). "The role of topic interest in survey participation decisions." Public opinion quarterly **68**(1): 2-31.

Hair, J. F. (2010). Multivariate data analysis : a global perspective. Upper Saddle River, N.J., Pearson Education.

Hair, J. F. and B. J. Babin (2018). Multivariate Data Analysis, Cengage.

Haraldsen, G. (2004). "Identifying and reducing response burdens in internet business surveys."

Haraldsen, G., et al. (2013). "Understanding and coping with response burden." Designing and conducting business surveys: 219-252.

Ho, T. K. (1998). "The random subspace method for constructing decision forests." IEEE transactions on pattern analysis and machine intelligence **20**(8): 832-844.

Hoogendoorn, A. W. (2004). "A questionnaire design for dependent interviewing that addresses the problem of cognitive satisficing." Journal of Official Statistics **20**(2): 219.

Hothorn, T., et al. (2006). "Unbiased recursive partitioning: A conditional inference framework." Journal of Computational and Graphical statistics **15**(3): 651-674.

Hothorn, T., et al. (2015). "ctree: Conditional inference trees." The comprehensive R archive network **8**.

Hu, L. t. and P. M. Bentler (1999). "Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives." Structural Equation Modeling: A Multidisciplinary Journal **6**(1): 1-55.

Kabir, E., et al. (2018). "Statistical modeling of tree failures during storms." Reliability Engineering & System Safety **177**: 68-79.

Keusch, F. (2015). "Why do people participate in Web surveys? Applying survey participation theory to Internet survey data collection." Management review quarterly **65**(3): 183-216.

Krosnick, J. A. and D. F. Alwin (1987). "An evaluation of a cognitive theory of response-order effects in survey measurement." Public opinion quarterly **51**(2): 201-219.

Krosnick, J. A. and M. K. Berent (1993). "Comparisons of party identification and policy preferences: The impact of survey question format." American Journal of Political Science: 941-964.

Kulkarni, A., et al. (2020). 5 - Foundations of data imbalance and solutions for a data democracy. Data Democracy. F. A. Batarseh and R. Yang, Academic Press: 83-106.

Lavrakas, P. J. (2008). Encyclopedia of survey research methods, Sage publications.

Malhotra, N. K., et al. (2006). "Common method variance in IS research: A comparison of alternative approaches and a reanalysis of past research." Management science **52**(12): 1865-1883.

Mehdizadeh, M., et al. (2018). "Accident involvement among Iranian lorry drivers: Direct and indirect effects of background variables and aberrant driving behaviour." Transportation research part F: traffic psychology and behaviour **58**: 39-55.

Moore, R. and L. Varghese (2021). "An examination of the role of funny reminders and embedded survey items on response rate." Social Science Computer Review **39**(1): 162-172.

Nellore, S. B. (2015). "Various performance measures in Binary classification-An Overview of ROC study." IJISSET-International Journal of Innovative Science, Engineering & Technology **2**(9): 596-605.

Nunnally, J. C. (1978). Psychometric Theory 2nd ed, Mcgraw hill book company.

Patrick, M. E., et al. (2013). "Incentives for college student participation in web-based substance use surveys." Addictive behaviors **38**(3): 1710-1714.

Podgorelec, V., et al. (2002). "Decision Trees: An Overview and Their Use in Medicine." Journal of Medical Systems **26**(5): 445-463.

Porter, S. R., et al. (2004). "Multiple surveys of students and survey fatigue." New directions for institutional research **2004**(121): 63-73.

Prajwala, T. (2015). "A comparative study on decision tree and random forest using R tool." International journal of advanced research in computer and communication engineering **4**(1): 196-199.

Puleston, J. and D. Sleep (2008). "A look at the impact of boredom on the respondent experience." Retrieved June **28**: 2013.

Quinlan, J. R. (1987). Decision trees as probabilistic classifiers. Proceedings of the Fourth International Workshop on Machine Learning, Elsevier.

Qureshi, K. N. and A. H. Abdullah (2013). "A survey on intelligent transportation systems." Middle-East Journal of Scientific Research **15**(5): 629-642.

Rachman, A. and R. M. C. Ratnayake (2019). "Machine learning approach for risk-based inspection screening assessment." Reliability Engineering & System Safety **185**: 518-532.

- Rahim, M. A. and N. R. Magner (1995). "Confirmatory factor analysis of the styles of handling interpersonal conflict: First-order factor model and its invariance across groups." Journal of Applied Psychology **80**(1): 122.
- Raileanu, L. E. and K. Stoffel (2004). "Theoretical comparison between the gini index and information gain criteria." Annals of Mathematics and Artificial Intelligence **41**: 77-93.
- Raykov, T. (1997). "Estimation of Composite Reliability for Congeneric Measures." Applied Psychological Measurement **21**(2): 173-184.
- Revilla, M. (2017). "Analyzing survey characteristics, participation, and evaluation across 186 surveys in an online opt-in panel in Spain." methods, data, analyses **11**(2): 28.
- Revilla, M. and C. Ochoa (2017). "Ideal and maximum length for a web survey." International Journal of Market Research **59**(5): 557-565.
- Rigdon, E. E. (1996). "CFI versus RMSEA: A comparison of two fit indexes for structural equation modeling." Structural Equation Modeling **3**: 369-379.
- Rogelberg, S. G., et al. (2000). "Employee attitude surveys: Examining the attitudes of noncompliant employees." Journal of Applied Psychology **85**(2): 284.
- Rolstad, S., et al. (2011). "Response burden and questionnaire length: is shorter better? A review and meta-analysis." Value in Health **14**(8): 1101-1108.
- Rudžionienė, J., et al. (2018). "Drop out factors in data literacy and research data management survey: experiences from Lithuania and Finland." Informacijos mokslai **82**: 115-130.
- Sakshaug, J. W., et al. (2019). "Paper, e-mail, or both? Effects of contact mode on participation in a web survey of establishments." Social Science Computer Review **37**(6): 750-765.
- Saleh, A. and K. Bista (2017). "Examining factors impacting online survey response rates in educational research: Perceptions of graduate students." Online Submission **13**(2): 63-74.
- Sarmiento, R. P. and V. Costa (2019). "Confirmatory factor analysis--a case study." arXiv preprint arXiv:1905.05598.
- Schmid, B. and K. W. Axhausen (2019). "Predicting response rates of all and recruited respondents: A first attempt." Transport Findings.
- Shankar, A., et al. (2020). "How do electronic word of mouth practices contribute to mobile banking adoption?" Journal of Retailing and Consumer Services **52**: 101920.
- Sharp, L. M. and J. Frankel (1983). "Respondent burden: A test of some common assumptions." Public opinion quarterly **47**(1): 36-53.

Sheng, B. and S. Gengxin (2010). Data Mining in census data with CART. 2010 3rd International Conference on Advanced Computer Theory and Engineering (ICACTE), IEEE.

Shrestha, N. (2021). "Factor analysis as a tool for survey analysis." American Journal of Applied Mathematics and Statistics **9**(1): 4-11.

Singer, E., et al. (1999). "The effect of incentives on response rates in interviewer-mediated surveys." Journal of Official Statistics **15**(2): 217.

Smyth, J. D., et al. (2006). "Comparing check-all and forced-choice question formats in web surveys." Public opinion quarterly **70**(1): 66-77.

Stocké, V. and B. Langfeldt (2004). "Effects of survey experience on respondents' attitudes towards surveys." Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique **81**(1): 5-32.

Strobl, C. (2014). "Discussion on fifty years of classification and regression trees." International Statistical Review **82**(3): 349-352.

Tangirala, S. (2020). "Evaluating the impact of GINI index and information gain on classification using decision tree classifier algorithm." International Journal of Advanced Computer Science and Applications **11**(2): 612-619.

Ullman, J. B. and P. M. Bentler (2012). "Structural equation modeling." Handbook of Psychology, Second Edition **2**.

Verzosa, N. M. (2020). Let's Play for the Future of Transport: The Gamification Effect on Travel Survey Participation, Acceptance, and Data Quality, University of Sydney.

Walston, J. T., et al. (2006). "The influence of web-based questionnaire presentation variations on survey cooperation and perceptions of survey quality." Journal of Official Statistics **22**(2): 271.

White, A. P. and W. Z. Liu (1994). "Bias in information-based measures in decision tree induction." Machine learning **15**: 321-329.

Wright, K. B. (2017). "Researching Internet-Based Populations: Advantages and Disadvantages of Online Survey Research, Online Questionnaire Authoring Software Packages, and Web Survey Services." Journal of Computer-Mediated Communication **10**(3).

Xia, Y. and Y. Yang (2019). "RMSEA, CFI, and TLI in structural equation modeling with ordered categorical data: The story they tell depends on the estimation methods." Behavior Research Methods **51**(1): 409-428.

Yan, T., et al. (2010). "Should I Stay or Should I go: The Effects of Progress Feedback, Promised Task Duration, and Length of Questionnaire on Completing Web Surveys." International Journal of Public Opinion Research **23**(2): 131-147.

Yan, T., et al. (2014). The impact of response burden on data quality in a longitudinal survey. Proceedings of the International Total Survey Error Workshop, October.

Yan, T., et al. (2020). "Response burden: What is it and what predicts it?" Advances in Questionnaire Design, Development, Evaluation and Testing: 193-212.

Zhang, J., et al. (2011). "Data-driven intelligent transportation systems: A survey." IEEE Transactions on Intelligent Transportation Systems **12**(4): 1624-1639.

Zillmann, D., et al. (2014). "Survey topic and unit nonresponse." Quality & quantity **48**(4): 2069-2088.

Keshavarz, H., and Morency, C. "An overview of methods for measuring respondent burden in surveys", World Conference on Transport Research Society (WCTR2023)

Retrieved from <https://mobilizingjustice.ca>, on Aug 12th, 2022.

Retrieved from <https://mobilizingjustice.ca>, the brainstorm session report, on Aug 12th, 2022.