MAKERERE UNIVERSITY

DEVELOPING A PEDESTRIAN ROUTE CHOICE MODEL FOR UNIVERSITY STUDENTS IN KAMPALA

Case Study: Makerere University

BY

BUKENYA ISMAIL

REG. No.: 2023/HD08/2653U

Signature: ……………………….

Date: 27-07-2026

Supervisors

Prof. Umaru Bagampadde
Signature: ……………………….
Date: 27-07-2026

Dr. Andrew Bwambale
Signature: ……………………….
Date: 27/07/26

A DISSERTATION SUBMITTED TO THE DIRECTORATE OF RESEARCH AND GRADUATE TRAINING IN PARTIAL FULFILMENT FOR THE AWARD OF THE DEGREE OF MASTER OF SCIENCE IN CIVIL ENGINEERING OF MAKERERE UNIVERSITY

OCTOBER 2025

KAMPALA, UGANDA

DECLARATION

I, **Bukenya Ismail**, hereby declare that this thesis is the result of my own original work carried out under the guidance and supervision of a formally approved university. It has not been submitted, either in whole or in part, for any degree or examination at any other university or institution of higher learning.

……………………………………………..

BUKENYA ISMAIL

REG. No.: 2023/HD08/2653U

APPROVAL

This research thesis has been submitted for examination with the approval of the following supervisors.

ENG. DR. BAGAMPADDE UMARU
UGANDA ENGINEERS REGISTRATION BOARD
ERB
Valid ONLY with a Practicing Licence
CIVIL ENGINEER ERB NO: 556

Prof. Umaru Bagampadde **Date:** 27 - 07 - 2026

College of Engineering, Design, Art, and Technology
Makerere University
Kampala – UGANDA

Dr. Andrew Bwambale **Date:** 27/07/26

College of Engineering, Design, Art, and Technology
Makerere University
Kampala – UGANDA

ABSTRACT

Universities in African cities are intensely walkable, yet evidence on how route features shape pedestrian choices is scarce. This study models student route choice at Makerere University in Kampala using both stated preferences from 106 respondents and revealed preferences for 309 trips from 105 participants captured with a mobile GPS application. Choice sets were generated with Yen's k-shortest paths on the OSM campus network, and the route attributes such as length, greenery, sidewalk availability, slope, and amenity presence were computed in ArcGIS. Multinomial and path-size logit models were estimated using Apollo package in R software. Whereas the SP model indicates disutility for time and slope and positive utilities for greenery and amenities, the RP model revealed that, students frequently choose longer, greener, off-street routes, avoiding busy routes with high amenity counts and sidewalks. Joint SP and RP estimation tightened model inference with five of six coefficients significant at 5% level of confidence and recovered a large SP scaling parameter, consistent with lower SP error variance and efficient experiment design. Out-of-sample validation with repeated 70/30 holdouts showed small estimation–validation gaps with the joint model exhibiting the best generalisation. Policy-wise, greening internal connectors and reducing pedestrian exposure to traffic appear more effective than adding sidewalks along busy corridors. The study's convenience sample and the controlled campus environment limit external validity, but results provide actionable evidence for institutional hubs and a foundation for larger, city-scale studies enriched with situational variables such as time of day, weather, lighting, safety and volume of traffic.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

AIC	Akaike Information Criterion
API	Application Programming Interface
BIC	Bayesian Information Criterion
CRD	Call Record Data
DEM	Digital Elevation Model
EPSG	EPSG code for coordinate reference systems (e.g., 32636)
GDP	Gross Domestic Product
GIS	Geographic Information System
GPS	Global Positioning System
HMM	Hidden Markov Model
IIA	Independence of Irrelevant Alternatives property
KCCA	Kampala Capital City Authority
LMICs	Low-middle income countries
MCA	Multi-Criteria Analysis
MNL	Multinomial Logit Model
NDP	National Development Plan
NMT	Non-Motorised Transport
OD	Origin–Destination
OSM	OpenStreetMap
OSRM	Open Source Routing Machine
PSL	Path-Size Logit
PST	Path-Size Term (used as ln(PST))
RP	Revealed Preference
SDGs	Sustainable Development Goals
SP	Stated Preference
UTM	Universal Transverse Mercator
VPS	Virtual Private Server
WGS 84	World Geodetic System 1984

1 INTRODUCTION

1.1 Background

Globally, walking, plays a central role in urban transportation systems, contributing strongly to urban liveability, public health, and sustainability (Baobeid et al., 2021; Westenhöfer et al., 2023). In African cities, walking is particularly prominent and in many urban areas, over 70% of daily trips are made on foot (Cobbinah & Finn, 2024; Okyere et al., 2024). Among low-income populations in African cities, more than 75% of daily trips are walked (Basil & Nyachieo, 2023). Despite its importance, pedestrian travel in many African cities occurs under conditions that pose significant safety and health risks. Pedestrian infrastructure is often inadequate, with limited sidewalks, poor road crossings, and insufficient separation from vehicular traffic (Chebe et al., 2025). As a result, pedestrians remain disproportionately affected by road traffic injuries and fatalities (Vissoci et al., 2017).

Within this context, a critical research gap remains in understanding pedestrian route choice behaviours in developing cities, where rapid urbanization and diverse socio-economic conditions create complex mobility patterns. University campuses offer practical pilot environments for studying these dynamics and generating insights that can be scaled to wider urban settings. Accordingly, using Makerere University as a living laboratory enables observation of concentrated pedestrian flows that can inform broader transport planning in comparable contexts. This approach helps fill the gap in route-choice modelling for universities in developing cities by basing model development on empirically observed walking behaviour (Jin, Luo, et al., 2024; Lieu & Guhathakurta, 2025; Tong & Bode, 2022a).

Previous research on pedestrian route choice has predominantly relied on stated preference (SP) methods, where respondents evaluate hypothetical route scenarios, and revealed preference (RP) methods, where actual pedestrian movements are observed. Each approach, however, has inherent drawbacks: SP designs are subject to hypothetical bias and oversimplification of real environments, while RP data, though behaviourally realistic may fail to capture unobserved or latent influences on route choice (Basu N. et al., 2022; Fulman et al., 2025).

This study seeks to bridge these methodological gaps and contribute to the existing literature by adopting a mixed-methods approach that integrates both SP and RP data, using Makerere University as a pilot site to capture the complexities of pedestrian behaviour within a developing urban context. By integrating both SP surveys and RP data collected through smartphone GPS tracking, this research aims to provide a more comprehensive and ecologically valid understanding of pedestrian decision-making. The outcomes of this model will offer critical insights into the factors influencing route choices among students, thus informing targeted improvements in pedestrian infrastructure and safety.

1.2 Problem Statement

Walking is a major mode of movement in Kampala and within university environments like Makerere University, yet pedestrian planning is often constrained by limited evidence on how people actually choose their walking routes. In many cases, infrastructure improvements are planned without sufficient understanding of how pedestrians trade off route attributes such as travel time, greenery, sidewalk availability, slope, amenities, safety and comfort. This creates a risk of investing in pedestrian facilities that may not correspond to actual user preferences or observed walking behaviour.

Existing pedestrian route-choice evidence is largely drawn from high-income and data-rich urban contexts, with limited application to African institutional settings such as university campuses. In addition, many studies rely on either stated preference data or revealed preference data, although each source explains only part of pedestrian decision-making. This study therefore addresses an evidence, context and methodological gap by developing a pedestrian route choice model for university students in Kampala, using Makerere University as a case study and combining survey-based stated preference data with GPS-based revealed preference data.

1.3 Main Objective

The main objective of this study is to develop a pedestrian route choice model for university students in Kampala. The case study is Makerere University.

1.3.1 Specific Objectives

The specific objectives of the proposed research are to:

i. Generate a pedestrian route choice set and quantify relevant attributes for each choice based on GPS trajectories on a Geographic Information System (GIS) platform.
ii. Investigate the factors affecting pedestrian route choice and quantify their effect.

iii. Develop pedestrian route choice models considering both stated and revealed preference data.

1.4 Research Questions

a. What are the primary factors (e.g., safety, convenience, aesthetics) influencing pedestrian route choices among university students in Kampala, as revealed through both stated and revealed preference data?
b. To what extent does a mixed approach, combining stated and revealed preference data, improve the accuracy of pedestrian route choice models compared to models based on a single data source?
c. What are the implications of the developed model for urban planning and transportation policy in Kampala, particularly concerning the needs and preferences of university students?

1.5 Expected Outputs

The expected outputs of this study include a pedestrian route choice model specifically calibrated for university students in Kampala, taking into consideration selected attributes. The model is developed based on both stated and revealed preference data elicited through student interviews and GPS tracking respectively.

1.6 Significance of the study

This research, incorporating the novel GPS tracking as a data collection method in Africa, is timely and essential for low-middle-income countries like Uganda. The study aims to create a more sustainable, safe, and inclusive urban environment by developing a contextually relevant pedestrian route choice model, aligning with the National Development Plan (NDP IV) goals, and contributing to the Sustainable Development Goals (SDGs).

1.7 Justification of the Study

Uganda has strengthened its commitment to walking and cycling through national Non-Motorised Transport (NMT) policies, road safety strategies and urban development frameworks that prioritise vulnerable road users. However, effective implementation of these policies requires behavioural evidence on how pedestrians choose routes under varying walking conditions. This study provides such evidence by quantifying how route attributes such as travel time, greenery, sidewalk availability, slope and amenities influence pedestrian path selection. The findings can therefore support more evidence-based prioritisation of pedestrian corridors, crossings, traffic calming, shaded walkways and other walking-environment improvements.

Makerere University was selected as a suitable living laboratory for this study because it represents a dense institutional pedestrian environment within Kampala. The campus is located approximately 5 km from the city centre, covers over 350 acres, hosts about 40,000 students, and has walking as the dominant mode of internal movement. Its semi-controlled circulation network, mix of formal roads, walkways and informal footpaths, varied topography, greenery, amenities and traffic exposure provide realistic route-choice conditions while still allowing detailed observation and modelling of pedestrian behaviour. The campus therefore provides a practical setting for developing and testing a pedestrian route choice model whose insights may inform planning in comparable universities, institutional hubs and selected urban pedestrian environments.

1.8 Scope of the study

The study was geographically limited to Makerere University, Kampala, Uganda, and focused on pedestrian route choices made within the campus boundary. The population scope was limited to university students because they form the dominant walking group within the campus. Technically, the study focused on developing a pedestrian route choice model using stated preference and revealed preference data, with travel time, greenery, sidewalk availability, slope and amenities as the selected route attributes. The analysis involved generating pedestrian route choice sets, quantifying route attributes and estimating discrete choice models to compare stated, revealed and joint stated–revealed preference results. The findings are therefore most applicable to Makerere University and comparable institutional pedestrian environments, while wider application to Kampala’s broader urban network would require further testing and recalibration. Figure 3-1 shows the location map of the study area.

1.9 Conceptual framework

The conceptual framework for this study is grounded in discrete choice theory, which assumes that pedestrians choose from a set of available route alternatives by evaluating the utility associated with each route. (Ben-Akiva & Lerman, 1985). In this study, route-specific attributes such as travel time, greenery, sidewalk availability, slope and amenities are treated as the main independent variables influencing perceived route utility.

The framework also recognises intervening variables that may shape how pedestrians perceive and weigh route attributes. These include gender, age, walking purpose, time of day, weather conditions and route familiarity. The dependent variable is the route chosen by the pedestrian, which is assumed to be the alternative with the highest perceived utility after accounting for observed route attributes, intervening factors and unobserved influences represented by the random error term.

Figure 1-1 below, shows the visual representation of the conceptual framework.

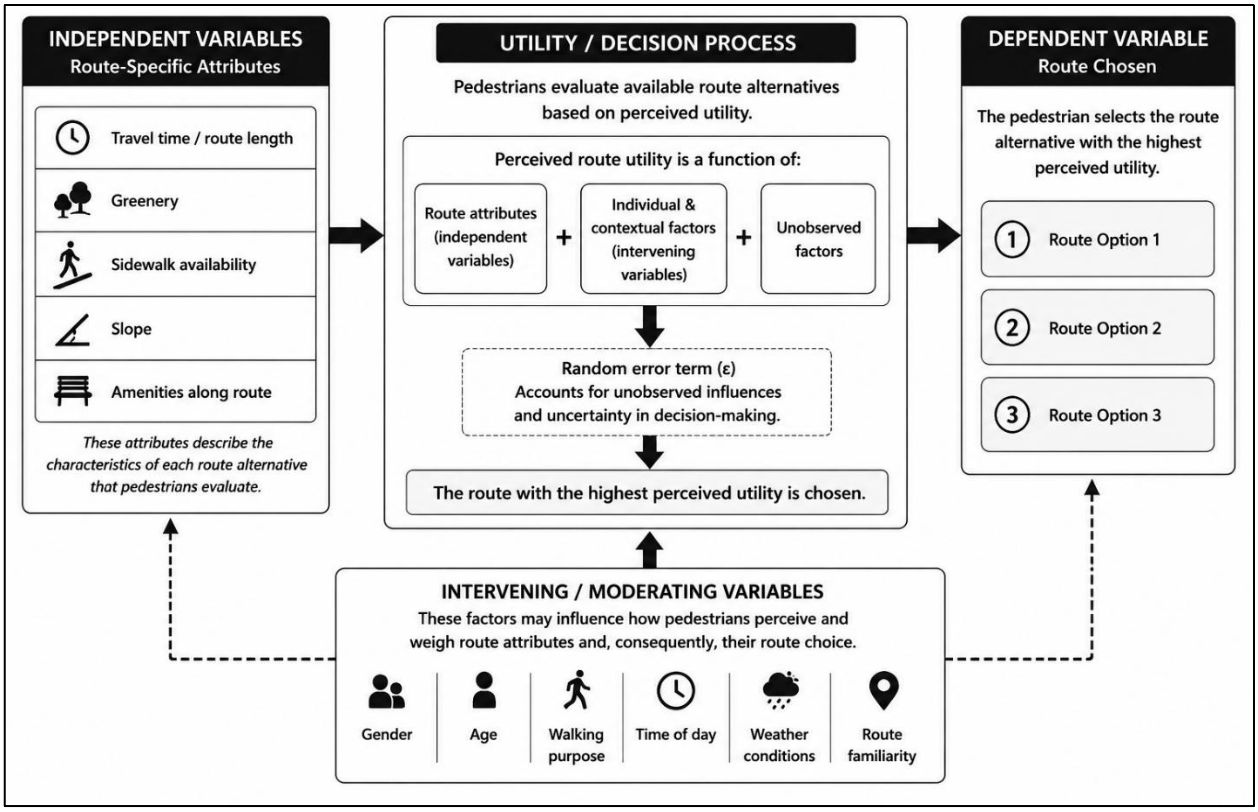


Figure 1-1: Conceptual Framework

Source: Developed by Author

2 LITERATURE REVIEW

2.1 Introduction

Understanding pedestrian route choice is an essential strand of urban transportation research that directly informs the design of walkable, safe, and inclusive infrastructure. Walking is the most basic form of human movement in cities, yet the decision-making process behind pedestrian route choice is complex and multi-faceted, involving perceptive, environmental, social, and technological factors (Tong & Bode, 2022). Understanding these factors is particularly crucial for institutional hubs located in rapidly urbanizing and yet resource-constrained contexts like Kampala, where fostering sustainable mobility solutions is an urgent development priority (Mogaji, 2022). This section synthesizes the evolution of pedestrian route choice modelling, discussing dominant modelling theories, empirical findings from recent studies, and implications for modelling in developing country contexts.

2.2 Theoretical Foundations of Pedestrian Route Choice

Pedestrian route choice is guided by random utility theory and discrete choice theory. Random utility theory assumes that when an individual is faced with a set of alternatives, each alternative provides a certain level of utility, and the individual is more likely to choose the alternative with the highest perceived utility (Marschak, 1960; Ben-Akiva & Lerman, 1985). In pedestrian route choice, the available alternatives are the possible routes between an origin and a destination, while utility is influenced by observed route attributes and unobserved factors. The utility that individual n obtains from choosing route r can be expressed as shown in Equation 1.

$$U_{nr} = V_{nr} + \varepsilon_{nr} \qquad \text{Equation 1}$$

where U_{nr} is the total utility of route r for individual n, V_{nr} is the systematic utility explained by observed route attributes, and ε_{nr} is the random component representing unobserved influences and uncertainty.

The systematic utility is specified as a function of route attributes and their estimated parameters. The general utility specification is expressed as shown in Equation 2.

$$V_{nr} = \sum_{k=1}^{K} \beta_k X_{nrk} \qquad \text{Equation 2}$$

where X_{nrk} is the value of attribute *k* for route *r* and individual *n*, and β_k is the parameter representing the effect of that attribute on route utility. A positive β_k indicates that the attribute increases the likelihood of choosing a route, while a negative β_k indicates that the attribute reduces route attractiveness.

Discrete choice theory is appropriate for this study because pedestrians choose from a finite set of mutually exclusive route alternatives. A student walking from one point to another on campus may have several possible routes, but only one route is selected for a given trip. The selected route is therefore treated as the outcome of a choice process in which the route with the highest perceived utility is chosen. This can be represented as:

$$P_n(r) = P\left(U_{nr} > U_{nj}\right), \quad \forall j \neq r,\ j \in C_n \qquad \text{Equation 3}$$

where $P_n(r)$ is the probability that individual *n* chooses route *r*, and C_n is the choice set available to individual *n*.

The study also adopts a compensatory utility framework. This framework assumes that pedestrians may accept a disadvantage in one route attribute if it is compensated by advantages in other attributes. For example, a pedestrian may accept a longer walking time if the route is greener, less steep, more comfortable or has useful amenities. This is important because walking decisions are rarely based on travel time alone. Instead, pedestrians evaluate a combination of functional, environmental and comfort-related attributes when selecting a route.

The multinomial logit model provides the basic modelling structure for estimating route choice probabilities. The model assumes that the random error terms are independently and identically distributed according to a type I extreme value distribution, which leads to a closed-form expression for choice probabilities (McFadden, 1974). The probability that individual *n* chooses route *r* from choice set C_n is given by Equation 4.

$$P_n(r) = \frac{\exp(V_{nr})}{\sum_{j \in C_n} \exp\left(V_{nj}\right)} \qquad \text{Equation 4}$$

Model parameters are estimated by maximising the log-likelihood function in Equation 5.

$$LL = \sum_{n=1}^{N} \sum_{r \in C_n} y_{nr} \ln[P_n(r)] \qquad \text{Equation 5}$$

where $y_{nr} = 1$ if individual *n* chooses route *r*, and $y_{nr} = 0$ otherwise. The MNL model is useful because it is simple, interpretable and widely applied in travel behaviour analysis.

However, it relies on the independence of irrelevant alternatives assumption, which implies that the relative odds of choosing between two routes are unaffected by the presence or characteristics of other routes. This assumption may be restrictive in route choice analysis because alternative routes often overlap.

To address route overlap, correction models such as the C-Logit model and the Path Size Logit model were proposed (Cascetta et al., 1996; Ramming, 2002; Ben-Akiva & Ramming, 1998). These models modify the systematic utility function by introducing an overlap correction term shown in Equation 6.

$$V_{nr}^{*} = V_{nr} + \tau_{nr} \qquad \text{Equation 6}$$

where V_{nr}^{*} is the overlap-adjusted systematic utility and τ_{nr} is the correction factor for route r. In the Path Size Logit model, the correction term accounts for the degree to which a route shares links with other alternatives in the choice set. A route that shares many segments with other routes is treated as less distinct, while a route with less overlap is treated as more unique. This improves the behavioural realism of route choice estimation by reducing bias caused by similar alternatives.

Finally, the study is guided by the complementarity between stated preference and revealed preference data. Stated preference data capture how respondents say they would choose between hypothetical route alternatives with controlled attribute combinations, while revealed preference data capture actual route choices observed through GPS tracking (Lue & Miller, 2019; Hasnine et al., 2020a). Each data source has limitations when used alone: stated preference data may suffer from hypothetical bias, while revealed preference data may not fully explain the motivations behind observed choices. In joint SP/RP estimation, scale parameters are used to account for differences in error variance between the two data sources. A simplified joint utility specification can be expressed as shown in Equation 7 and 8.

$$U_{nr}^{SP} = \mu_{SP} V_{nr}^{SP} + \varepsilon_{nr}^{SP} \qquad \text{Equation 7}$$

$$U_{nr}^{RP} = \mu_{RP} V_{nr}^{RP} + \varepsilon_{nr}^{RP} \qquad \text{Equation 8}$$

where μ_{SP} and μ_{RP} are scale parameters for the SP and RP data sources, respectively. For identification, one scale is usually normalised, and the other is estimated. Combining SP and RP data allows the model to exploit the controlled trade-offs of SP data and the behavioural realism of RP data, thereby improving parameter stability, prediction and interpretation (Bradley & Daly, 1991; Hensher et al., 1999; Hasnine et al., 2020a; Qiao et al., 2016a). This theoretical foundation supports the study's use of separate and joint SP/RP models to understand pedestrian route choice among university students in Kampala.

2.3 Advances in modelling pedestrian route choice

The theoretical and empirical study of pedestrian route choice has matured rapidly in recent decades, largely due to advances in discrete choice modelling frameworks. Discrete choice models are predicated on the assumption that individuals select among discrete alternatives, in this case, routes that maximize their expected utility based on observable and unobservable factors (Ben-Akiva & Bierlaire, 1999; Sevtsuk et al., 2021; Sevtsuk & Kalvo, 2022; Tong & Bode, 2022).

The multinomial logit model (MNL), as formalized by McFadden (1974), emerged as the initial workhorse of route choice modelling owing to its simplicity. However, the MNL model is affected by the Independence of Irrelevant Alternatives (IIA) property. This can lead to biased or unrealistic results in contexts where alternative routes share common segments (Ramming, 2002). To address this limitation, more advanced models, such as the C-Logit (Cascetta et al., 1996) and Path Size Logit (Ben-Akiva & Bierlaire, 1999) were developed. These models introduce commonality or path size factors that account for correlations among overlapping routes, reducing bias in parameter estimation.

2.4 Attributes for pedestrian route choice

Attributes of alternative routes play a central role in pedestrian route choice modelling, as they directly influence individuals' decisions between competing paths (Basu N. et al., 2022; Basu R. & Sevtsuk, 2022). To better understand this, I reviewed previous studies to identify the key route attributes commonly used in pedestrian route choice models and assessed their relevance or availability for the context of Makerere university, Kampala, as summarized in Table 2-1.

Table 2-1: Attributes typically used in pedestrian route choice models

Attribute	Reference(s)	Location	Applicability & Availability	Remarks
Distance	Broach & Dill, 2016; Gregory Lue & Eric J. Miller, 2019; Sevtsuk et al., 2021	US (San Francisco), Canada (Toronto), US (Portland)	✓	Applicable; can be estimated from maps or GIS data

Attribute	Reference(s)	Location	Applicability & Availability	Remarks
Sidewalk Width	Gregory Lue & Eric J. Miller, 2019; Sevtsuk et al., 2021	US (San Francisco), Canada (Toronto)	✓	Possible consideration of sidewalk availability; Yes or No.
Elevation/Slope	Broach & Dill, 2016; Sevtsuk et al., 2021	US (San Francisco), US (Portland)	✓	Applicable based area topography; can be estimated from Google earth or digital elevation model (DEM)
Amenities	Erath et al., 2015; Guo & Loo, 2013; Sevtsuk et al., 2021	US (San Francisco), Singapore, US/NYC & HK	✓	Available; data on cafeterias, rest spots, printeries can be obtained
Traffic Speed/Volume	Basu N. et al., 2022; Broach & Dill, 2016; Erath et al., 2015; Lee & Lee, 2021	US (Portland), Singapore, Korea, Review	?	Possible consideration of route type; pedestrian only or mixed (Vehicles + pedestrians)
Turns	Basu N. et al., 2022; Broach & Dill, 2016	US (Portland), Review	✓	Applicable; can be counted from mapped routes
Safety & Secuirty	Basu N. et al., 2022; Broach & Dill, 2016; Lee & Lee, 2021	US (Portland), Korea, Review	✓	Safety perception can be merged with route type; pedestrian only or mixed (Vehicles + pedestrians)
Crossings	(Basu N. et al., 2022, 2023; Broach & Dill, 2016; Erath et al., 2015)	US (Portland), Singapore	*x*	No designated pedestrian crossings on campus.
Land Use/Mix	Clifton et al., 2016; Guo & Loo, 2013	US/NYC & HK, US (Portland region)	✓	Can be obtained from Land-use maps or planning data. Land-use aspect can be captured by amenities
Greenery	Andersson et al., 2022; Erath et al., 2015; Guo & Loo, 2013	Sweden, Singapore, US/NYC & HK	✓	Can be assessed from satellite imagery or field survey
Aesthetics	Andersson et al., 2022; Erath et al., 2015; Guo & Loo, 2013	Sweden, Singapore, US/NYC & HK	*x*	Largely subjective; perception-based data might be needed
Gender/Age	Basu N. et al., 2022	—	✓	Sociodemographic data can be collected in surveys
Crowding	Basu N. et al., 2022; Broach & Dill, 2016	—, US (Portland)	*x*	Not readily available; requires observation or crowd-sourced data

✓ = Available and applicable x = Not available or not applicable
Compiled from all cited sources
Source: As cited

Whereas the attributes influencing pedestrian route choice can be diverse and context-dependent, common trends emerge from prior studies. (1) Distance is almost always a primary determinant, with most individuals showing a preference for shorter routes (N. Basu et al., 2022; Sevtsuk et al., 2021). (2) Safety and comfort features such as sidewalk presence and width, crosswalk availability, and lower traffic volumes are crucial, especially in areas where streets are busy or not well maintained (Andersson et al., 2022; Basu R. & Sevtsuk, 2022; Broach & Dill, 2016; Gregory Lue & Eric J. Miller, 2019). (3) The presence of amenities like rest areas and cafes, as well as greenery and attractive aesthetics, can draw pedestrians to longer routes if these features are attractive (Andersson et al., 2022; N. Basu et al., 2023; Erath et al., 2015; Guo & Loo, 2013). (4) Elevation or slope is a relevant factor in hilly cities, like Kampala, with steeper paths generally being less attractive (Andersson et al., 2022; Basu N. et al., 2023; Lieu & Guhathakurta, 2025; Tong & Bode, 2022). (5) Finally, social factors like age, gender, and crowding shape which routes are perceived as most suitable, underscoring the importance of context-sensitive analysis (N. Basu et al., 2023; Lieu & Guhathakurta, 2025).

2.5 Emerging technologies in pedestrian route choice modelling

Early studies of pedestrian route choice modelling predominantly relied on stated preference (SP) surveys in which respondents were presented with hypothetical route scenarios aimed to discern preferences for various route attributes (N. Basu et al., 2023; Kožul et al., 2025; Sevtsuk et al., 2021). While SP data provides flexibility, it may be affected by hypothetical bias and scenario simplification resulting in not always accurately reflecting real behaviour (Hasnine et al., 2020; Jin et al., 2025; Kožul et al., 2025).

The advent and widespread usage of GPS-enabled devices and smartphones has revolutionized route choice data collection. Researchers can passively and accurately record real-world pedestrian trajectories. Utilization of this revealed preference (RP) data in recent years has accelerated the estimation of pedestrian discrete choice models, enabling analyses with unprecedented sample sizes and spatial completeness (Basu R. & Sevtsuk, 2022b; Bwambale et al., 2019; Gregory Lue & Eric J. Miller, 2019; Lieu & Guhathakurta, 2025).

Despite the powerful potential of GPS-based revealed preference (RP) studies to capture real pedestrian trajectories, there is a noticeable research bias: most such studies have been conducted in high-income or data-rich settings, with relatively few in developing or fast-urbanizing cities. This geographic bias constrains our understanding of pedestrian behaviour in regions undergoing rapid urbanization and mobility transition (Ankunda & Venter, 2025). Notably, a study by Lue and Miller (2019) showed that smartphone GPS data, even with only moderate spatial accuracy, can yield rich insights into the determinants of pedestrian route choice, including those less accessible by stated preference. With mobile technology penetration rapidly increasing in cities across the developing world and the growing availability of affordable, crowd-sourced data, there is now a compelling rationale for piloting GPS-based route choice studies in cities such as Kampala. Such efforts are critical to ensuring that urban transport planning and pedestrian infrastructure development in these settings are informed by empirically grounded and contextually relevant behavioural evidence (Stanitsa et al., 2023).

Another emerging trend in pedestrian route choice modelling is the blending of SP and RP data in discrete choice models to leverage the complementary strengths of each data type. This integration allows researchers to exploit the controlled experimental variation of SP data with the behavioural realism of RP data, ultimately enabling more stable identification of parameter estimates (Hasnine et al., 2020; Jin et al., 2025). Recent empirical work underscores the practical benefits of this integration. For instance, a study by Yanfu Qiao et al. (2016) shows that RP-SP joint models outperform single-source models, with composite utility functions yielding more accurate and policy-relevant forecasts of route choice.

2.6 Synthesis of Literature and Study motivation

The reviewed literature shows that pedestrian route choice is influenced by a combination of functional, environmental and contextual factors. Travel time and distance remain central determinants of route choice, but pedestrians may also consider greenery, sidewalk availability, slope, amenities, safety, comfort, crowding and the surrounding walking environment (N. Basu et al., 2023; Broach & Dill, 2016; Guo & Loo, 2013; Tong & Bode, 2022). This implies that pedestrian route choice is not simply a shortest-path problem, but a behavioural decision in which pedestrians trade off route efficiency against comfort, attractiveness, perceived safety and accessibility.

However, most empirical evidence on pedestrian route choice has been generated in high-income and data-rich urban contexts, where pedestrian infrastructure, data availability and travel conditions differ from those found in many African cities. This limits the direct transferability of existing models to rapidly urbanising, resource-constrained environments such as Kampala. Although university environments provide useful settings for studying concentrated pedestrian movement, African institutional hubs remain underrepresented in pedestrian route choice research. This creates a context gap in understanding how students and other pedestrians choose routes in settings where formal walkways, informal paths, greenery, amenities, slopes and traffic exposure interact within the same walking environment.

The review also reveals a methodological gap. Many pedestrian route-choice studies rely on either stated preference data or revealed preference data. Stated preference studies are useful for examining controlled trade-offs among route attributes, but they may be affected by hypothetical bias. Revealed preference studies capture actual walking behaviour, especially when GPS data are used, but they may not fully explain the preferences and motivations behind observed route choices. Combining both data sources is therefore important because stated preference data provide experimental control, while revealed preference data provide behavioural realism (Bradley & Daly, 1991; Hensher et al., 1999; Hasnine et al., 2020a; Qiao et al., 2016a).

In response to these gaps, the current study contributes to the literature in four ways. First, it provides empirical evidence on pedestrian route choice from an African institutional context using Makerere University as a living laboratory. Second, it uses GPS-based revealed preference data to observe actual student walking behaviour within a real campus network. Third, it integrates stated and revealed preference data to assess whether their complementarity improves model interpretation and predictive performance. Fourth, it quantifies key route attributes, including travel time, greenery, sidewalk availability, slope and amenities, using GIS-based methods in a resource-constrained setting. Through these contributions, the study develops context-specific evidence that can inform pedestrian planning within Makerere University and comparable institutional hubs, while providing a foundation for future city-scale studies in Kampala and similar African cities.

3 METHODOLOGY

3.1 Introduction

This section describes the methods employed in the current study with focus on the route attributes, design and implementation of both stated preference (SP) and revealed preference (RP) surveys.

3.2 Research Design

This study adopted a quantitative, cross-sectional, mixed-preference research design to analyse pedestrian route choice behaviour among university students. The design was quantitative because the study measured route attributes numerically and estimated discrete choice models to quantify how these attributes influenced route selection. It was cross-sectional because the stated preference survey and revealed preference tracking were conducted within a defined study period, providing a snapshot of pedestrian route choice behaviour within Makerere University.

The mixed-preference design combined stated preference and revealed preference data. The stated preference component used a controlled route-choice experiment in which respondents selected between hypothetical route alternatives with varying attributes such as travel time, greenery, sidewalk availability, slope and amenities. This enabled controlled assessment of trade-offs among route attributes. The revealed preference component used GPS tracking to observe actual walking routes made by students within the campus. Combining the two data sources allowed the study to compare intended route choices with observed walking behaviour and to estimate separate and joint SP/RP pedestrian route choice models. This design was appropriate because SP data provide experimental control, while RP data provide behavioural realism in the modelling of pedestrian route choice.

3.3 Study Site

Makerere University is situated on Makerere Hill in Kawempe Division, approximately 5 kilometres north of Kampala city centre, Uganda's capital. The campus spans over 350 acres and hosts a daytime population of about 40,000 students, approximately 90% of whom rely on walking as their primary mode of transport within the university. The internal road network is generally well maintained; however, some roads lack dedicated pedestrian walkways. The university is enclosed by a perimeter wall, and access to the campus roads is controlled through the issuance of parking tickets at designated entry points, thereby restricting public access. Figure 3-1 shows the location of Makerere University.

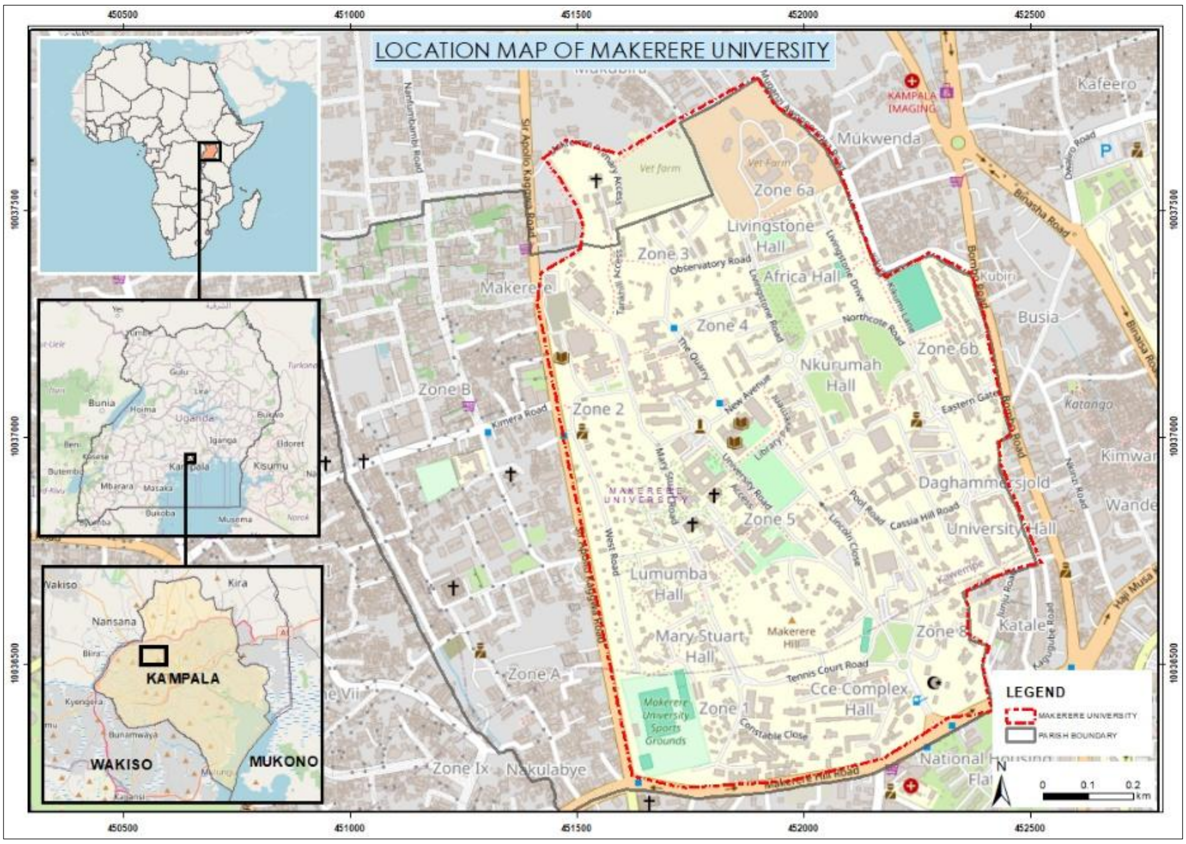


Figure 3-1: Location and layout of Makerere University
Source: Source: OpenStreetMap and Esri World Imagery basemap, 2025

3.4 Selection of Route Attributes

Route attributes were selected based on findings from the literature review (Chapter 2) and insights from a structured focus group session with Makerere University students. To systematically rank the route choice attributes obtained from previous research studies, each attribute was assessed using a Multi Criterial Analysis (MCA) with three yardsticks: Consensus & Influence, Measurability, and Literature Support. For each criterion, attributes were assigned a score reflecting the extent to which they met the criterion. The Consensus & Influence criterion assessed the level of agreement among students regarding the extent of each attribute’s influence during the focus group session. Measurability reflected the ease and reliability with which each attribute could be quantified in the context of the survey. Literature Support was established by reviewing previous research for empirical support of each attribute's importance. The individual scores across all three criteria were summed to calculate a total score for each attribute. Finally, attributes were ranked in descending order according to their total scores, with the highest-scoring attribute receiving the lowest rank (i.e. rank No. 1).

Table 3-1 presents the five highest-ranked attributes, selected from an initial list of twelve, that were considered for inclusion in the subsequent modelling and survey design.

Table 3-1: Summary of five highest-ranked attributes.

S/N	Attributes from Literature	Total Score	Rank
1	Travel Time	7	1
2	Presence of Trees / Greenery	3	5
3	Sidewalk availability	5	3
4	Slope	6	2
5	Landmarks / Amenities	4	4

Source: Multi-criteria analysis done by author

3.5 Determining Attribute Levels

A combination of the selected attributes and levels was used to enable respondents to make trade-offs. To determine appropriate levels for each route choice attribute, three Origin–Destination (OD) pairs (summarised in Table 3-2) were first analysed and possible route alternatives between each OD pair were identified. Figure 3-2 shows the location of the OD pairs and their alternative routes. For each alternative, the actual values of relevant attributes—namely, travel time, sidewalk availability, slope, and the presence of landmarks or amenities—were quantified. Attribute quantification was carried out using a combination of ArcGIS spatial analysis, data from OpenStreetMap, interpretation of satellite imagery, and physical surveys of the selected routes. Based on the range and distribution of observed values for each attribute, discrete attribute levels were established to accurately reflect real-world conditions within the study area (Table 3-3 and Table 3-4). Attribute levels were tailored to the characteristics of each OD pair; for example, if no route alternative within a given OD pair had a travel time greater than 10 minutes, this level was excluded from the set of attribute levels for that OD pair. This approach ensured that the attribute levels used in the experimental design were both realistic and representative of the actual variability present in the transport network.

Table 3-2: Summary of Origin Destination pairs

S/N	Origin	Destination	Data Code
OD Pair 1	University Hall (UH)	College of Agricultural and Environmental Sciences (CAES)	UH-CAES
OD Pair 2	Complex Hall (COMPLEX)	College of Engineering Design, Art and Technology (CEDAT)	COMPLEX- CEDAT
OD Pair 3	Livingstone Hall (STONE)	College of Business and Management Sciences (COBAMS)	STONE- COBAMS

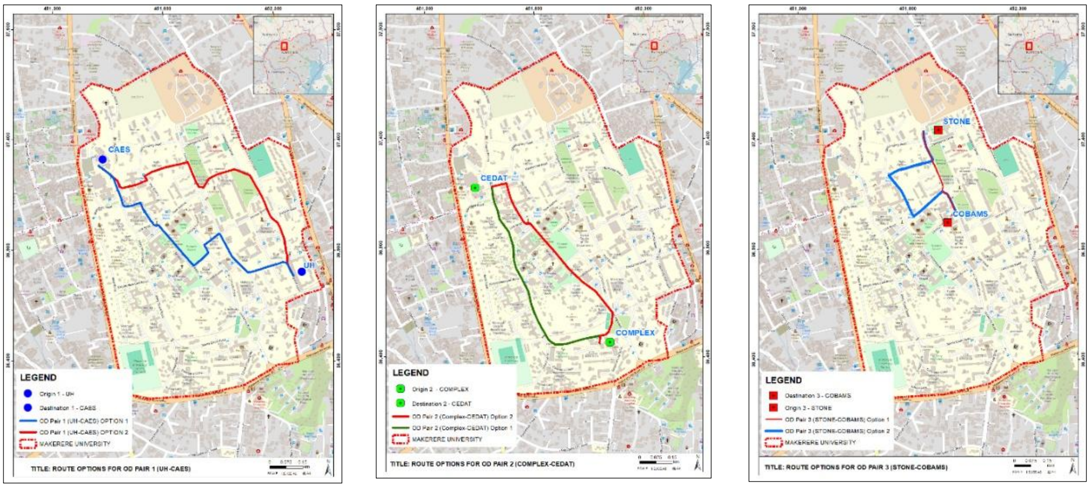


Figure 3-2: Map showing location of the OD pairs and alternative routes used in determining attribute levels.

Table 3-3: Attribute levels for OD 1 and OD 2

Attribute	Description	Levels
Travel Time	Time from Origin to Destination (TT)	2 (10,8) minutes
Presence of Trees	Presence of trees along route	2 (Present, Not present)
Sidewalk availability	Percentage presence of sidewalk along route	3 (100, 50, 0) %
Slope	General slope along the route	3 (10,5,0) %
Landmarks / Amenities	Presence of amenities (such as canteen, printery) along the route	2 (Present, Not present)

Table 3-4: Attribute levels for OD 3

Attribute	Description	Levels
Travel Time	Time from Origin to Destination (TT)	2 (4,3) minutes
Presence of Trees	Presence of trees along route	2 (Present, Not present)
Sidewalk availability	Percentage presence of sidewalk along route	3 (100, 50, 0) %
Slope	General slope along the route	3 (10,5,0) %
Amenities	Presence of amenities (such as canteen, printery) along the route	2 (Present, Not present)

Source: Origin – Destination analysis done by author

3.6 Stated Preference Survey

3.6.1 SP Experiment Design and Implementation

The stated preference survey was conducted in two phases: an initial pilot survey followed by the main survey. An efficient experimental design was employed for the pilot survey to incorporate prior information from the literature and preliminary analyses, thereby improving the precision of parameter estimates and reducing the required sample size. “Ngene”, a specialized software package was utilized for generating choice experiment designs. The minimum number of choice scenarios required in a discrete choice experiment is determined by K/(J-1), where K is the number of parameters to be estimated and J is the number of alternatives per choice set (Choice Metrics, 2024). For the present study, with K=5 and J=2, the minimum number of scenarios is 5. However, to enhance statistical efficiency and design coverage, a total of 18 hypothetical choice scenarios divided into 3 blocks of 6 choice sets each were specified in the Ngene syntax (see Appendix 1). Each respondent was presented with only one block with 6 choice scenarios.

The SP survey was implemented using Google Forms and comprised three sections: (1) an introduction outlining the study’s purpose and seeking informed consent, (2) a series of six choice tasks in which respondents selected their preferred pedestrian route from two tabulated alternatives, and (3) a section capturing personal and socio-demographic characteristics. The pilot was conducted with 32 respondents, with the primary objectives of assessing the clarity and practicality of the choice tasks and generating initial parameter estimates for optimizing the final experimental design. Appendix 3 shows the final SP survey tool utilised for this study.

3.6.2 SP Sample Size

The adequacy of the SP sample was assessed in relation to the requirements of discrete choice experiment design and model estimation. The experimental design involved five route attributes and two alternatives per choice set. Using the minimum choice-scenario rule (K/(J-1)), where (K) is the number of parameters and (J) is the number of alternatives, the minimum number of required scenarios was five. The final design exceeded this minimum by generating 18 choice scenarios, divided into three blocks of six choice tasks each. With 106 respondents each completing six choice tasks, the SP dataset produced 636 route choice observations. This provided substantially more observations than the minimum required for estimating the selected multinomial logit models and supported model stability, parameter identification and comparison of alternative model specifications.

3.7 Revealed Preference Survey

3.7.1 Data Collection

The open-source Traccar GPS tracking platform (https://www.traccar.org) was used for data collection and hosted on a virtual private server (VPS) to enable centralized device management and storage. One hundred-five (105) student volunteers were assisted in installing and configuring the Traccar mobile client. The application recorded timestamped GPS coordinates at 30-second intervals. To accommodate intermittent internet connectivity, offline buffering was enabled so that positions were stored locally and automatically uploaded once connectivity was restored. Participants' movements were monitored over a two-week period. At the end of the field phase, data were exported from the Traccar web interface as Excel files containing timestamped latitude–longitude coordinates for subsequent analysis. Figure 3-3 shows the process of data collection and processing.

3.7.2 GPS Data cleaning and Map matching

Raw GPS observations rarely coincide exactly with the walkable network because of signal noise, multipath, and duty-cycling of smartphone sensors. To obtain plausible pedestrian paths for route-choice analysis, the recorded GPS points were projected onto an Open Street Map Road network using the Open Source Routing Machine (OSRM) foot profile. OSRM implements a Hidden Markov Model (HMM) map-matching algorithm which has previously been applied to GPS data (Lieu & Guhathakurta, 2025). This algorithm balances per-point proximity to candidate edges with connectivity and travel feasibility between consecutive points.

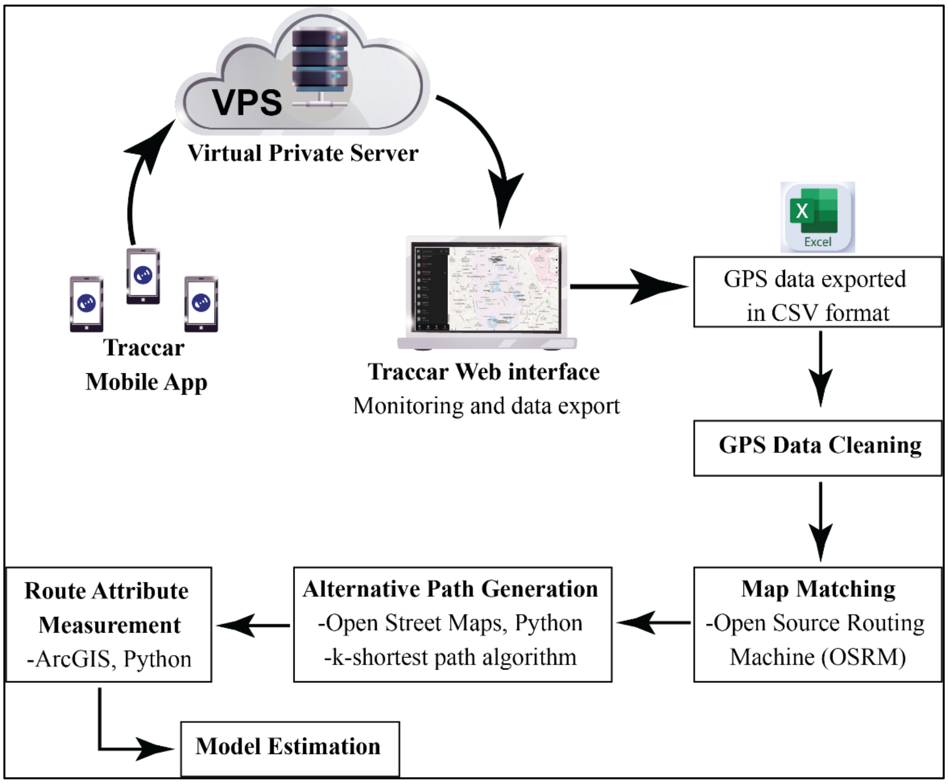


Figure 3-3: RP Data collection and processing
Source: Methodology design done by author

The raw data were imported from an excel file with fields for participant identifier, timestamp, latitude, and longitude. Records with missing identifiers, times, or coordinates were removed, and traces were segmented into trips on a per-person basis using the recorded start and end timestamps. Extremely short events were discarded by requiring a minimum trip duration of 1 minute as a conservative threshold to reduce false micro-trips created by app pauses or brief signal outages while preserving genuine walking episodes. To maintain relevance to the study area, trips outside the Makerere University boundary were excluded. In addition, traces exhibiting speeds greater than 2.0 m/s were also removed as likely non-walking movements. Finally, only trajectories that followed streets or mapped pedestrian paths were retained; segments located inside buildings or where no pedestrian facility exists were not considered. The output of the data cleaning and map matching process was a GIS shapefile with all map-matched trips and a tabular Excel summary of trip-level attributes both utilised in the subsequent GIS analysis.

3.7.3 Alternative path generation

A compact set of routes that are comparable in length yet sufficiently distinct in alignment to avoid excessive correlation among options is desirable for meaningful route choice analysis. To achieve this, alternative paths were generated using a deterministic constrained-enumeration procedure built on k-shortest loopless paths to mitigate violations of the independence from irrelevant alternatives (IIA) assumption while keeping the number of alternatives tractable for estimation. For each observed trip, a walking network was constructed from OpenStreetMap with a python package (OSMnx). Then the trip's origin–destination nodes were identified and enumerated the 12 shortest loopless routes by length using Yen's k-shortest paths algorithm (Yen, 1971).

Candidate routes with a detour ratio greater than 1.5 relative to the shortest route were discarded. From the remaining set, at most three alternatives per trip, were retained ensuring that geometric overlap of each alternative with the observed route was limited to less than 30%. This screening follows Basu and Sevtsuk (2022b), who reported that a 25% overlap between alternatives and the actual route optimizes attribute diversity in choice sets. Of the 309 observed trips, 91 yielded at least one plausible alternative; accordingly, the final choice set comprised 91 observed routes and 172 non-chosen alternatives.

3.7.4 Measurement of Route Variables

Five route attributes (length, slope, amenities, greenery, and sidewalk availability) were derived using ArcGIS's Spatial Analyst to characterize each observed and alternative path. All inputs with unknown coordinate reference systems were defined as WGS 84 and then projected to UTM Zone 36N (EPSG:32636) to ensure metric calculations. Route lengths were computed in meters on the projected geometry.

Travel Time

Travel Time is a fundamental descriptor in pedestrian route-choice analysis because people typically favour shorter paths, and it also provides a convenient common unit for interpreting other route attributes in "equivalent walking distance" terms. In this study, a walking speed of 80 metres per minute (Murtagh et al., 2021) was assumed and utilised to calculate travel time from length which was measured in ArcMap by projecting all route polylines to a local metric coordinate system (UTM Zone 36N).

Slope

A digital elevation model (DEM) of the campus was converted to a slope raster (percent rise) with the Slope tool. Each route was densified to 10 m vertex spacing, and slope values were sampled at these vertices. The mean of sampled values was taken as the route's slope indicator to provide a robust segment-weighted summary of grade along the path.

Amenities

Amenity locations were compiled from OpenStreetMap and augmented with local campus knowledge to reduce omissions. Amenities were measured with two variables; one with one as an indicator of presence or absence and the other as a count of amenities along the route. To measure exposure along each route, the route centreline was buffered by 20 m and performed a spatial join with the amenity points. The 20 m buffer was adopted as a practical exposure corridor around each walking route, intended to capture amenities that were close enough to be visible, accessible or likely to influence route choice, while avoiding over-counting facilities located farther away from the pedestrian path; the same buffer distance was applied consistently to both observed and alternative routes to ensure comparability. The resulting attribute records the count of amenity points intersecting the 20 m corridor around the route.

The amenities considered in this study were facilities located along or near pedestrian routes that support students' education-related walking trips by providing academic, food, social, financial or convenience services. These amenities are summarised in Appendix 3.

Greenery

Green space polygons were derived from Google Earth Engine land-cover processing exported as a vector layer. Greenery was measured with two variables; one as an indicator of presence or absence and the other as a greenery share ratio. For routes with any intersecting greenery polygon, the greenery share was determined by dividing the greenery length by the total route length. The resulting attribute is a ratio of route length traversing a green area and the total route length.

Sidewalk availability

A campus sidewalk inventory shapefile generated from a survey which mapped pedestrian facilities. Sidewalk availability was also measured with two variables; one as an indicator of presence or absence and the other as a sidewalk share ratio. Sidewalk features were buffered by 10 m and intersected with each route to compute the portion of the route co-located with sidewalks. Sidewalk share is determined by dividing length of route with sidewalks by total route length, a metric that captures how a route is served by pedestrian infrastructure.

3.7.5 RP Sample Size

The adequacy of the RP sample was assessed based on the number of usable observed walking trips and the resulting route choice sets rather than only the number of recruited participants. A total of 105 student volunteers participated in the GPS tracking exercise, producing 309 observed walking trips after data cleaning and map matching. For route choice modelling, the key requirement was that each observed trip could be linked to a feasible origin–destination pair, measurable route attributes and plausible non-chosen alternatives. After alternative route generation and screening, 91 observed trips yielded at least one plausible alternative, resulting in 91 chosen routes and 172 non-chosen alternatives for model estimation. This provided sufficient variation in travel time, greenery, sidewalk availability, slope, amenities and path overlap to support exploratory revealed preference route choice modelling within the Makerere University campus context.

The RP sample was considered adequate for exploratory modelling because it captured actual walking behaviour over a two-week monitoring period and produced multiple observed choices and alternatives suitable for estimating path-size logit models. However, the sample was based on voluntary participation and therefore does not constitute a statistically representative sample of the entire Makerere University student population or pedestrians in Kampala. The RP results are consequently interpreted as behaviourally grounded evidence from a campus-based living laboratory, rather than as population-wide estimates. Broader generalisation to Makerere University as a whole, to other universities, or to the wider Kampala urban network would require larger, more representative and longer-duration GPS datasets.

3.8 Modelling Framework

The present study uses discrete choice models since pedestrian route choices are inherently discrete and mutually exclusive. The modelling framework is grounded in random utility theory (Marschak, 1960), a widely recognized foundation for the analysis and estimation of discrete choice behaviour.

$$U_{nr} = V_{nr} + \varepsilon_{nr} \qquad \text{Equation 9}$$

In Equation 9, U_{nr} represents the utility that individual *n* associates with choosing route *r*. This utility is composed of a deterministic, or systematic, component V_{nr}, and a stochastic component ε_{nr} that captures uncertainty. The systematic part of the utility V_{nr}, can be modelled as a function of observed route attributes, as shown in Equation 10, where β_k denotes the parameter associated with attribute k which is assumed constant across individuals and is to be estimated, and X_{nrk} is the value of attribute *k* for individual *n* and route *r*

$$V_{nr} = \sum_k \beta_k X_{nrk} \qquad \text{Equation 10}$$

We assume that the stochastic term ε_{nr} is independently and identically distributed across alternatives according to a type I extreme value (Gumbel) distribution. We then use the Multinomial Logit (MNL) model to estimate the probability that individual *n* selects route *r* as follows (see McFadden, 1974 for details);

$$P_n(r) = \frac{\exp(V_{nr})}{\sum_{r^* \epsilon C_n} \exp(V_{nr*})} \qquad \text{Equation 11}$$

Where, $Pn(r)$ denotes the probability that individual *n* chooses route *r,* and Cn is the choice set. With these route choice probabilities, the model parameters can be estimated by maximizing the log-likelihood function (Equation 12) in which Z_{nr} is an indicator variable where Z_{nr} =1 if individual *n* actually chose alternative *r*; otherwise, Z_{nr} =0.

$$LL = \sum_n \sum_r [Z_{nr} * \ln(P_n(r))] \qquad \text{Equation 12}$$

Accounting for overlap

A major limitation of the MNL model is the independence of irrelevant alternatives (IIA) property, which can lead to illogical route choice probabilities for highly overlapping routes (Cascetta et al., 1996; Ramming, 2002). There are several models that account for overlap, however, in this study, we focus on the C-Logit model (Cascetta et al., 1996; Ramming, 2002) and the Path Size Logit model (Ben-Akiva & Ramming, 1998). Both models modify the standard MNL framework by incorporating correction factors into the systematic utility as follows;

$$P_n(r) = \frac{\exp(V_{nr} + \tau_{nr})}{\sum_{r^* \epsilon C_n} \exp(V_{nr*} + \tau_{nr})} \qquad \text{Equation 13}$$

Where τ_{nr} is the systematic utility correction factor for route r.

C-logit model

In the C-Logit model, the correction factor τ_{nr} is referred to as the commonality factor, which adjusts for the degree of overlap among routes (Cascetta et al., 1996). While several specification have been proposed in the literature, this study adopts the following common specification (Cascetta et al., 1996; Ramming, 2002):

$$\tau_{nr} = \beta_{CF}\ln\left[\sum_{r^{*}\epsilon C_n}\left(\frac{L_{rr*}}{\sqrt{L_r L_{r*}}}\right)^{\gamma_{CF}}\right] \quad \text{Equation 14}$$

Where L_{rr*} is the overlap length between routes r and $r*$, L_r and L_{r*} are the total lengths of routes r and $r*$ respectively, β_{CF} and γ_{CF} are the unknown parameters to be estimated.

Path size logit model

In the path size logit model, the correction factor τ_{nr} is known as the path size term, which accounts for the degree of overlap by representing the weighted average of the links that make up each route. The specification adopted in this study is as follows (Ben-Akiva & Ramming, 1998):

$$\tau_{nr} = \beta_{ps}\ln\left[\sum_{a\epsilon\Gamma_r}\left(\frac{l_a}{L_r} * \frac{1}{N_{ar}}\right)\right] \quad \text{Equation 15}$$

Where $\frac{1}{N_{ar}}$ is the inverse of the number of routes sharing the link a (the link size), and $\frac{l_a}{L_r}$ is a weight representing the proportion contributed by link a to the overall route size.

Models that account for route overlap are generally expected to provide a better fit than the standard MNL model. Model performance is assessed using the adjusted rho-squared statistic and likelihood ratio tests, following the recommended practices outlined by Ben-Akiva and Lerman (1985).

3.9 Model Validation

Model validation was undertaken to assess the stability and predictive performance of the estimated pedestrian route choice models. The validation focused on determining whether the models could produce consistent results when applied to data not used during model estimation. This was important for assessing whether the estimated parameters reflected meaningful route choice behaviour rather than overfitting to the estimation sample.

A repeated hold-out validation procedure was used. For each model category, namely the stated preference model, revealed preference model and joint stated–revealed preference model, the dataset was randomly divided into an estimation subset and a validation subset. The estimation subset comprised 70% of respondents, while the remaining 30% formed the hold-out validation subset. The models were first estimated using the 70% estimation sample, and their predictive performance was then evaluated on the 30% validation sample.

The validation procedure was repeated five times using different random splits of the data. For each split, the log-likelihood per observation was computed for both the estimation and validation samples. The percentage difference between the estimation and validation log-likelihood values was then calculated to assess the degree of predictive stability. Smaller differences between estimation and validation log-likelihoods were interpreted as evidence of better out-of-sample performance and lower risk of overfitting.

The validation results were compared across the SP, RP and joint SP/RP models. A model was considered more stable if its parameter signs remained consistent across repeated splits and if the validation log-likelihood was close to the estimation log-likelihood. The results of this validation procedure are presented in Chapter Four, where the relative predictive performance of the SP, RP and joint SP/RP models is discussed.

3.10 Parameter Estimation

Estimation of parameters was done using the Apollo package for R software (Hess & Palma, 2019), which is specifically designed for discrete choice modelling. The observed route choices and their associated attributes were input into the Apollo framework. Using the maximum likelihood estimation method, the software identified the parameter values that best explained the observed choices. This involved iteratively adjusting the parameters to maximize the likelihood function, ultimately producing the set of estimates that provided the best fit to the data. The results include both the estimated parameters and their standard errors, allowing for interpretation and assessment of model significance. Appendix 2 presents the code used in R software for parameter estimation.

3.11 Ethical considerations

All participants in this study were involved on a voluntary basis and provided informed consent prior to taking part in the structured interviews. The confidentiality and privacy of respondents were strictly maintained; all data collected was anonymised to remove any identifying information and securely stored to prevent unauthorized access. The data was used solely for research purposes and not shared for any other use, ensuring adherence to ethical standards throughout the study.

4 DISCUSSION OF RESULTS

4.1 Introduction

This chapter presents and discusses the results of the study in line with the three specific objectives. The discussion is organised according to the three modelling stages used in the study. First, the chapter presents the stated preference results by describing the SP route choice sets, the quantified route attributes and the factors influencing route choice based on the estimated SP model results. Second, it presents the revealed preference results by describing the observed GPS-based route choice sets, the measured attributes of observed and alternative routes, and the factors influencing actual walking route choices based on the estimated RP model results. Third, the chapter presents the joint stated–revealed preference model results and compares their performance with the single-source SP and RP models.

The chapter also presents the model validation results to assess whether the estimated models generalise to hold-out data. Particular attention is given to whether combining stated and revealed preference data improves model stability, interpretation and predictive performance.

4.2 Stated Preference Survey Results

4.2.1 Descriptive Statistics

The SP data sample is made up of 106 individuals and 636 route choice scenarios. The demographic composition is presented in Table 4-1.

Table 4-1: Characteristics of respondents

	N	%
Gender	106	100%
Female	58	55%
Male	48	45%
Age	106	100%
18-24	76	72%
25-34	27	26%
35-44	3	3%
45-54	0	0%
55-64	0	0%
65+	0	0%

	N	%
Employment Status	106	100%
Student	76	72%
Informal	1	1%
Formal in public sector	8	8%
Formal in private sector	13	12%
Not employed	8	8%
Monthly Income in UGX	106	100%
150,000	3	3%
150,001 - 300,000	2	2%
300,001 - 500,000	4	4%
500,001 - 1,000,000	7	7%
1,000,000 - 2,000,000	10	9%
2,000,000+	7	7%
No Income	73	69%

Source: Stated preference survey data

The demographic composition of the sample is characterized by a predominant representation of younger individuals, with the most common age group being 18 to 24 years (72%), followed by 25 to 34 years (26%). This pronounced peak in younger age categories reflects the sampling approach, as the survey was primarily distributed at Makerere University, where most respondents were students. Employment status data supports this, with 72% of the participants identifying as students. This substantial proportion of students explains the income distribution, where the majority (69%) reported having no income, and only a small fraction (7–9%) reported monthly earnings above 1,000,000 UGX. The gender distribution is relatively balanced, with 55% female and 45% male respondents. Notably, while there are small numbers of respondents in formal employment, either in the public (8%) or private sector (12%), these groups are considerably outnumbered by students. Compared to the broader population, this sample is younger and has lower household income, highlighting the influence of the university context on the sample's demographic profile.

4.2.2 Stated Preference Model

To investigate the determinants of pedestrian route choice, a series of multinomial logit (MNL) models were estimated using the stated-preference (SP) data. Table 4-2 summarizes the estimation results for eight alternative MNL specifications. Initially, individual route attributes; travel time, greenery, sidewalk availability, slope, and amenity presence; were assessed separately to understand their isolated impacts on pedestrian preferences. Subsequently, significant attributes were combined progressively to identify the model specification that yielded the highest adjusted ρ^2 value.

Each of the single-attribute models (M1–M5) isolates one route attribute at a time, estimating its effect on the utility of each alternative route independently. In all cases, the estimated coefficients for the focal attribute are statistically significant at 5% confidence level, indicating a consistent directional influence on route choice. However, the corresponding adjusted ρ^2 values remain low, suggesting that single-attribute specifications have limited explanatory power and that additional attributes are needed to better account for observed route choices.

Table 4-2: Stated Preference Model build-up.

MNL #		Travel Time β_{TT}	Greenery β_{GN}	Sidewalk Availability β_{SW}	Slope β_{SS}	Amenity Presence β_{AMN}	Final log-likelihood	Adjusted ρ^2
M1	Value	-0.2287*					-406.43	0.0844
	t-stat	-7.432						
M2	Value		0.5158*				-424.34	0.0442
	t-stat		5.573					
M3	Value			-0.02065*			-412.36	0.0711
	t-stat			-6.634				
M4	Value				-0.07331*		-429.06	0.0336
	t-stat				-5.548			
M5	Value					0.8003	-397.45	0.1046
	t-stat					7.427		
M6	Value	-0.13848*	0.32266*		-0.05509*	0.73869*	-360.26	0.1814
	t-stat	-3.906	3.084		-2.962	6.232		
M7	Value	-0.07099	0.24893	-0.02276	0.04427	0.72077*	-359.97	0.1798
	t-stat	-0.7029	1.5486	-0.7432	0.3373	5.9460		
M8	Value	-0.68652	0.25227	-0.01881	0.13219	0.72518*	-359.93	0.1799
	t-stat	-0.7595	1.5885	-0.7932	0.2737	5.9321		

Significance level: *p<0.05

Source: SP model estimation done by author

Interestingly, sidewalk availability has a negative low coefficient (-0.02065), contradicting prior expectations and previous studies (e.g., Sevtsuk & Kalvo, 2022), where sidewalk availability is expected to have a highly attractive influence. The signs of other attributes, however, aligned with theoretical expectations from literature.

In Model M6, attributes exhibiting theoretically consistent signs in the individual models were estimated jointly. It is observed that the parameter values do not differ very largely from the individually estimated attributes. The maximum deviation was 0.193 (for greenery), and the signs remained similar. Specifically, travel time and slope parameters were negative, suggesting repulsive effects, whereas greenery and amenity presence had positive parameters and attractive influences. Importantly, the adjusted ρ^2-value is remarkably higher than the previous individually estimated models.

In model M7, all attribute parameters were estimated jointly to test for possible improvements in model performance. However, this joint estimation resulted in a slight decrease in the adjusted ρ^2-value (0.1798), and only amenity presence was significant. It is seen that the travel time lost parameter significance and substantially reduced in magnitude. This reduction likely resulted from limited variability in the specified travel time levels, given the relatively narrow range (6–10 minutes) observed within the Makerere University context.

Finally, Model M8 explored log-transformations of travel time and slope to accommodate potential non-linearities and diminishing sensitivities. Attributes such as greenery, sidewalk availability, and amenity presence, represented as binary indicators, were retained in their original form. Nonetheless, this specification produced negligible improvements in adjusted ρ^2-value.

Overall, the results suggest that while separately estimated attribute parameters significantly influence pedestrian choices, a minimalist model capturing multiple relevant attributes yields superior explanatory power.

Table 4-3 presents the results for Model M1, which estimates the effect of travel time alone, and Model M6, which, after excluding sidewalk availability, is selected as the final SP model because all estimated parameters exhibit the expected signs.

Table 4-3: Parameter estimates for the selected SP Model

Variable	M1		M6	
	Estimated Coef.	t-stat	Estimated Coef.	t-stat
Travel Time	-0.2287*	-7.4316	-0.13848*	-3.906
Greenery			0.32266*	3.084
Slope			-0.05509*	-2.962
Amenity Presence			0.73869*	6.232
No. of decision makers	106		106	
No. of observations	636		636	
Number of parameters	1		4	
Final Log-likelihood	-406.43		-360.26	
Adj. Rho-squared	0.0844		0.1814	
AIC	814.86		728.53	

Significance level: *p<0.05

Source: SP model estimation done by author

4.2.3 Trade-off Analysis

To further interpret the estimated coefficients, trade-offs between route attributes were computed by calculating the ratio of coefficients for the model parameters relative to travel time, which serves as the reference variable in this study. This approach is consistent with prior work that applies the concept of equivalent walking distance to quantify the perceived effort or disutility associated with different route features (R. Basu & Sevtsuk, 2022). Assuming an average walking speed of 80 meters per minute (Murtagh et al., 2021), the coefficients can be translated into distance-based measures that enable a more intuitive understanding of the relative importance of route characteristics.

For instance, in Equation 10, the ratio of the coefficient for greenery to that of travel time indicates that students are willing to extend their walk by approximately 186.4 meters in order to select a route with greenery, all else held constant. Put differently, the presence of greenery provides the same utility gain as a route without greenery that is 186.4 meters shorter. Similarly, the coefficient for slope suggests that a 5% increase in route gradient is associated with a perceived disutility equivalent to an additional 31.8 meters of walking. By contrast, amenities emerge as the most influential attribute: the presence of a single amenity along a route is valued equivalently to a reduction of 426.7 meters in walking distance.

$$\frac{\boldsymbol{\beta_{GN}}}{\boldsymbol{\beta_{TT}}} = \frac{0.32266}{-0.13848} \approx \boldsymbol{-2.33\,min \times 80\,m/min \approx -186.4\,meters} \qquad \text{Equation 10}$$

Table 4-4: Equivalent walking distance

Attribute	Equivalent walking distance interpretation	Value (meters)
Greenery	The presence of greenery along a route is associated with a utility increase equivalent to a reduction of:	186.40
Slope	An average increase in slope by 5% is associated with a utility decrease equivalent to an increase of:	31.83
Amenity Presence	A route with one amenity along a route is associated with a utility increase equivalent to a reduction of:	426.74

Source: Trade-off analysis done by author

Table 4-4 summarizes the equivalent walking distances associated with the three key attributes. These findings underscore the substantial role of amenities in shaping pedestrian route choice, while also highlighting gendered sensitivities observed in the segmented analysis. Whereas greenery and slope influence route decisions to a lesser degree, their effects remain significant, reinforcing the importance of designing pedestrian environments that balance aesthetic, functional, and comfort-enhancing features.

4.2.4 Segmented model results based on Gender

Table 4-5 presents the results of a route choice analysis by gender, highlighting differences in preferences for various route variables between males and females. The same attributes as of Model 6 (M6) are adopted.

Table 4-5: Summary for SP Models for Female Vs Male

Variable	Female		Male	
	Estimated Coef.	t-stat	Estimated Coef.	t-stat
Travel Time	-0.17323	-3.44	-0.16322	-3.2027
Greenery	0.17808	1.175	0.54692	3.5888
Slope	-0.09664	-3.727	-0.01658	-0.661
Amenity Presence	1.20996	6.518	0.29796	2.1836
No. of decision makers	58		48	
No. of observations	348		288	
Number of parameters	4		4	
Final Log-likelihood	-169.56		-171.61	
Adj. Rho-squared	0.2805		0.1203	
AIC	347.12		351.22	

Significance level: *$p<0.05$

Source: SP model estimation done by author

Consistent with expectations, both genders demonstrate a negative sensitivity to travel time, with females exhibiting slightly higher disutility (–0.173) compared to males (–0.163). This indicates that women are marginally more time-sensitive in their walking choices. Regarding route slope, a strong negative effect is observed for females (–0.097), suggesting pronounced aversion to steep routes, whereas the effect is statistically insignificant among males. This finding aligns with broader evidence that women are more constrained by route gradients, likely reflecting differences in physical comfort and perceived walking effort (Lieu & Guhathakurta, 2025).

Interestingly, greenery exhibits contrasting patterns across genders. While the coefficient is insignificant for females, males display a strong and positive preference (0.547), indicating that the presence of green features substantially enhances route attractiveness for them. This divergence may reflect gendered perceptions of greenery, where men prioritize aesthetic or environmental quality more strongly in their route evaluations.

Amenities emerge as the most influential attribute for both groups, though the magnitude of the effect differs considerably. Females demonstrate a markedly stronger preference (1.210) compared to males (0.298), underscoring the centrality of amenities in shaping women's route choices. This heightened sensitivity echoes established literature, which links women's walking preferences to features that enhance comfort, safety, and convenience in urban environments.

4.3 Revealed Preference Model Results

This section describes the pedestrian route choice model results derived from revealed preference data.

4.3.1 Sample Size

The dataset was derived from a survey conducted as described in section 3.6. The recruitment was voluntary, and only willing participants were enrolled on the tracking system. The sample size for the RP survey is 105, representing approximately 0.25% of Makerere University student population. Therefore, this convenience sample is utilized for exploratory analysis of walking educational trips. Thus, there is no question of extrapolating the results to the entire Makerere Population nor even to the population of pedestrians in Kampala, but the analysis can still provide insight into route attributes considered of importance particularly for institutional hubs.

4.3.2 Descriptive Statistics

After cleaning the GPS data as highlighted in section 3.6, the analysis utilized 309 trips made by 105 individuals. The trips were made within the boundary of Makerere University, as shown in Figure 4-1, with user demographics shown in Table 4-6. The RP sample was male-dominated, with males constituting 79% and females 21% of the participants. This imbalance arose because participation in the GPS tracking exercise was voluntary, and more male students were willing to install and use the tracking application, while many female students opted out after the GPS tracking requirements were explained.

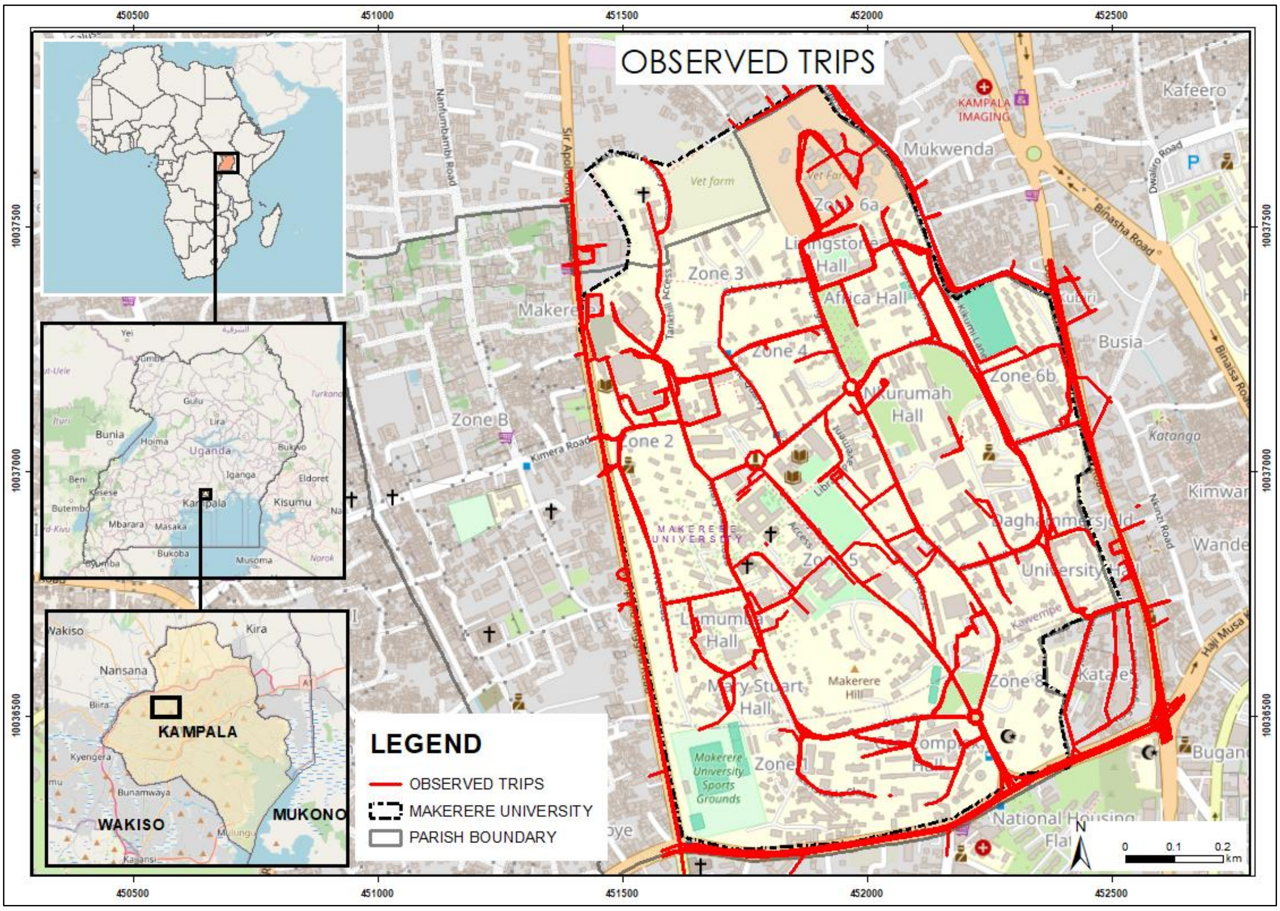


Figure 4-1: Observed trips
Source: GPS tracking data processed in ArcGIS, 2025

Table 4-6: Demographic Characteristics of RP Survey respondents

	N	%
Gender	105	100%
Female	22	21%
Male	83	79%
Age	105	100%

	N	%
18-24	92	88%
25-34	13	12%
35-44	0	0%
45-54	0	0%
55-64	0	0%
65+	0	0%
Employment Status	105	100%
Student	105	100%
Informal	0	0%
Formal in public sector	0	0%
Formal in private sector	0	0%
Not employed	0	0%
Monthly Income in UGX	105	100%
150,000	0	0%
150,001 - 300,000	2	2%
300,001 - 500,000	1	1%
500,001 - 1,000,000	0	0%
1,000,000 - 2,000,000	0	0%
2,000,000+	0	0%
No Income	102	97%

Source: Revealed preference survey data

4.3.3 Route Variables

Table 4-7 describes the mean value and descriptions of the route variables for the actual chosen route and three alternatives used in the model. The alternatives for each observation are classified by the value of PS, from low to high.

The chosen trips have an average travel time of 16.36 minutes, which is noticeably longer than the alternatives (about 8–10 minutes). The greenery share on the chosen routes (0.58) is higher than Alt 1 and close to Alt 2, though it remains below Alt 3 (0.62). By contrast, sidewalk availability is lower for the chosen routes (0.42) than for any alternative (0.48–0.51). The average slope on the chosen routes (14.26) exceeds that of Alt 1 and Alt 2 but is lower than Alt 3. Amenities are much fewer along the chosen routes (0.29 on average) than along the alternatives (0.55–1.38).

Table 4-7: Mean Values of Attributes for the chosen route and alternatives

Variable	Description	Mean			
		Chosen	Alt 1	Alt 2	Alt 3
Time (m)	Travel time in minutes	16.36	8.79	11.16	9.99
Greenery	Share of route length passing green areas.	0.5782	0.5167	0.5654	0.6152
Sidewalk availability	Share of route length with sidewalks.	0.4219	0.4767	0.5079	0.4932
Slope	Average slope along the route	14.26	12.51	13.41	15.20
Amenities	Number of amenities along a route	0.2890	0.8772	1.3784	0.5524
PS	Overlap indicator for alternative paths for the same trip	0.865031	0.404598	0.411932	0.494083

Source: ArcGIS analysis of observed GPS routes and generated alternative routes.

Figure 4-2 presents the correlations among route variables pooled across all alternatives. Overall, the coefficients are small, suggesting limited multicollinearity. Green share shows modest positive associations with slope (0.20) and with amenities (0.18), consistent with greener routes tending to traverse slightly hillier segments and areas with somewhat more nearby destinations. Time exhibits a weak negative correlation with greenness (-0.22) and near-zero relationships with sidewalk share (-0.06) and slope (-0.07), indicating that longer routes are only marginally less green and not systematically steeper or more sidewalk-rich. Sidewalk share is essentially uncorrelated with amenities (0.01) and slightly negative with slope (-0.04). Given that all non-diagonal coefficients remain well below |0.25|, the variables capture largely distinct aspects of the pedestrian environment and are retained together in subsequent models.

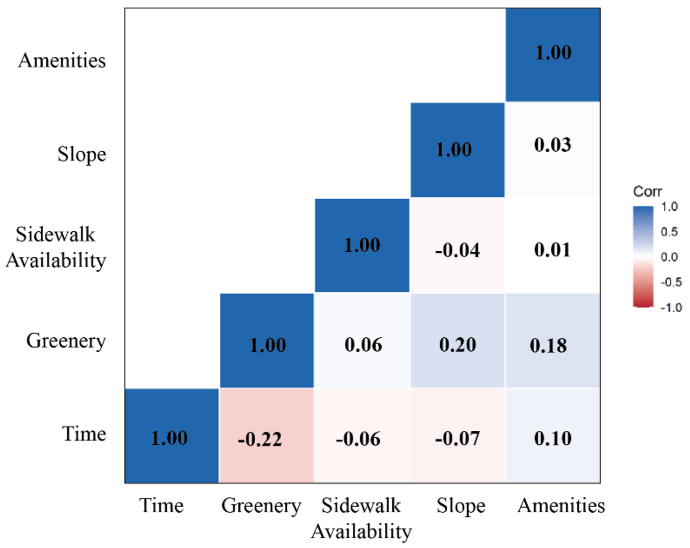


Figure 4-2: Route attributes correlation Matrix

Source: Correlation analysis of observed and alternative route datasets, 2025.

4.3.4 Revealed Preference Model Results

Table 4-8 summarizes estimates for two specifications based on 105 decision makers and 309 observations. Goodness of fit is compared using adjusted rho-squared and discussion the sign and significance of key attributes ensues.

Model RP 1 is estimated with greenery, amenities and sidewalk availability measured as indicators of presence and absence whereas model RP 2 is estimated with greenery and sidewalk availability measured as ratios of availability and amenities as count along the route.

Table 4-8: Parameter estimates for PSL models derived from Revealed Preference data

Variable	**RP1**		**RP2**	
	Estimated Coef.	**t-stat**	**Estimated Coef.**	**t-stat**
Travel Time (min)	**0.3233****	2.8538	**0.5502****	3.1945
Greenery	0.8417	0.8791	**7.0683****	3.9576
Sidewalk Availability	-0.0104	-0.8575	-0.0454	-1.0971
Slope	-0.0290	-0.7866	**-0.8236****	-2.6233
Amenity Presence	-0.8751	-1.4808	**-2.8026***	-2.0026
ln(Path Size)	**2.5781***	2.2636	1.0400	1.0126
No. of decision makers	105		105	
No. of observations	309		309	
Number of parameters	6		6	
Final Log-likelihood	-52.91		-42.54	
Adj. Rho-squared	0.320		0.4395	
AIC	117.81		97.08	

Significance level: *$p<0.05$, ** $p<0.01$

Source: RP model estimation done by author

Model RP 1 is a PSL specification that includes all six attributes described in Table 4-7 but with greenery, amenities and sidewalk availability measured as indicators of presence and absence. Only travel time and path size factor are significant at 5% confidence level. The positive travel-time coefficient in the RP model is unexpected under the conventional assumption that pedestrians minimise travel time. However, in this study it should be interpreted in relation to the campus walking context and the compensatory nature of pedestrian route choice. The observed GPS routes show that students sometimes accepted longer walking routes where these routes provided better environmental quality, lower traffic exposure, more greenery, or access to useful amenities (Andersson et al., 2022; Jin, Wu, et al., 2024). This suggests that actual walking behaviour was not governed by travel-time minimisation alone, but by a broader utility framework in which environmental and functional benefits could compensate for additional walking time. The positive coefficient therefore indicates that, within the observed Makerere University campus network, longer routes were sometimes preferred because they offered route qualities that were more attractive than the shortest or most direct alternatives. The parameter estimate for greenery is positive indicating that respondents prefer greener routes which aligns with recent studies showing that greener corridors enhance aesthetics and comfort thereby increasing walking utility (Andersson et al., 2022). Sidewalk availability is negative which is unexpected since pedestrians are expected to prefer routes with available sidewalks. However, on Makerere university campus, like many other campuses, routes with sidewalks often run along vehicular roads which pedestrians tend to avoid in preference to internal off-street paths. Noise and traffic exposure are known deterrents, which make road-edge sidewalks less attractive even if they are available (Andersson et al., 2022). Slope is negative as expected but not significant. The negative sign matches evidence that steeper grades hinder walking, but the weak effect is plausible on generally short trips (Xie et al., 2024). Amenity presence is negative and not significant. The negative sign contradicts expectation. In institutional settings, routes with high amenity counts can coincide with vehicular streets and localized crowded spots. Recent work shows that crowding and flow constraints can shift route choices away from otherwise convenient segments (Jin, Wu, et al., 2024) and that walking inconvenience can raise detour ratios (Loo et al., 2024). ln (Path size) is positive as expected but and significant at 5% confidence level, which is reasonable given that the study's alternative-generation method limits severe overlap.

Model RP 2 is a PSL specification that includes all six attributes but with greenery and sidewalk availability measured as ratios of availability and amenities measured as count along the route. Only travel time, greenery, slope and amenities presence are significant at 5% confidence level. The signs of the parameters are similar to those of Model RP 1. Model RP 2 has a higher adjusted rho-squared indicating that measuring greenery and sidewalk availability as share ratios and amenities as counts yields models with better goodness of fit compared to indicators of presence or absence.

4.3.5 Model results for Joint estimation on RP and SP data

Joint estimation of the SP and RP data employed a specification with source-specific scale parameters that rescale the SP data relative to the RP data to equalize error variances (Bradley & Daly, 1991). For identification, the RP scale was fixed at 1, and an SP scale parameter (μ_SP) was estimated to accommodate cross-source differences in variance (Hensher et al., 1999). Estimation was implemented in R using the Apollo package (Hess & Palma, 2019). The RP data used for the joint estimation included greenery, sidewalk availability and amenities measured as indicators of presence or absence along a route for compatibility issues with the SP data. Table 4-9 shows the joint-estimation results for two specifications: a full model including all attributes and a restricted model excluding the slope variable.

The travel time coefficient is positive and statistically significant at the 5% level, which is an unexpected sign. A plausible interpretation is that pedestrians deliberately detour from the shortest path to select greener, lower-traffic routes, so longer travel time is correlated with preferred environmental attributes.

Table 4-9: Parameter estimates for Joint estimation for SP and RP data

Variable	**SP + RP1**		**SP + RP1 (without slope)**	
	Estimated Coef.	**t-stat**	**Estimated Coef.**	**t-stat**
Travel Time (min)	**0.0264***	2.0415	**0.0615***	2.1462
Greenery	0.0273	0.7533	**0.3235***	2.3218
Sidewalk Availability	**-0.0170****	-2.9386	**-0.0219****	-3.3031
Slope	**0.0585****	2.8760		
Amenity Presence	**0.1670***	2.3013	**0.7132****	3.5204
μ SP	**4.2295***	2.5389	**1.0434****	3.2164
No. of decision makers	105		105	
No. of observations	309		309	
Number of parameters	6		5	
Final Log-likelihood	-435.38		-445.45	
Adj. Rho-squared	0.1632		0.1592	
AIC	882.77		900.9	

Significance level: *$p<0.05$, ** $p<0.01$

Source: SP + RP model estimation done by author

As anticipated, greenery has a positive estimate, indicating a preference for vegetated corridors. Sidewalk availability is negative, also unexpected, but likely reflects that sidewalks in Makerere University are concentrated along high-traffic routes which many pedestrians avoid in favour of off-street links without designated sidewalks. Finally, amenity presence is positive, consistent with a preference for routes that offer accessible services and facilities.

4.3.6 Comparison of SP, RP and Joint Model results

Table 4-10 compares results from the SP, RP, and joint models. The travel time parameter is negative in the SP model but positive in the RP and joint models, indicating a divergence between stated and revealed preferences. Whereas respondents stated that they prefer shorter routes, their observed behaviour reflects choices of longer routes, likely due to detours toward greener or less congested paths. Greenery is positive across all three models, consistently indicating a preference for greener corridors. Sidewalk availability is negative in all models, suggesting a revealed and stated preference for low-traffic, off-street links over high-traffic routes where sidewalks are prevalent. Slope is positive in SP but negative in RP, implying that respondents under-value slope in stated choices relative to the actual effort required to overcome slope during walking. Amenity presence is positive in SP but negative in RP, suggesting that while respondents state a preference for routes with amenities, in practice they avoid routes with amenities likely due to congestion and activity clusters around those locations.

4.4 Validation Results

The stability and predictive performance of the SP, RP and joint SP+RP models was assessed using a hold-out validation. The dataset was randomly split into five subsets comprising 70% of respondents as the estimation subset and 30% as the hold-out subset. Models were re-estimated on the estimation data with predictions evaluated on the hold-out users. Per-observation log-likelihood on the hold-out sample was compared with the in-sample log-likelihood, following Apollo's out-of-sample validation routine.

Across the repeated splits, the sign and relative magnitude of the key parameters remained stable between estimation and validation, indicating that the results are not driven by overfitting but by explanatory power in the covariates. Consistent with the tables, the differences between estimation and hold-out log-likelihood are modest, with the joint SP+RP specification showing the smallest average percentage change among the three models. This is evidence that joint estimation with SP and RP data improves generalisation to unseen individuals.

In simple terms, the validation results show how well each model performs when tested on respondents who were not used to estimate the model. A small difference between the estimation and hold-out log-likelihood means that the model is not only fitting the original data well, but is also able to predict choices in new data with reasonable consistency. The joint SP+RP model had the smallest percentage difference, meaning that it generalised better to unseen respondents than the SP-only and RP-only models. This suggests that combining stated and revealed preference data reduced the risk of overfitting and produced a more robust representation of pedestrian route choice behaviour.

Table 4-10: Comparison of SP, RP and Joint Model results

Variable	SP		RP1		SP + RP1	
	Estimated Coef.	t-stat	Estimated Coef.	t-stat	Estimated Coef.	t-stat
Travel Time (min)	-0.07099	-0.7029	**0.3233**	2.8538	**0.0264**	2.0415
Greenery	0.24893	1.5486	0.8417	0.8791	0.0273	0.7533
Sidewalk Availability	-0.02276	-0.7432	-0.0104	-0.8575	**-0.0170**	-2.9386
Slope	0.04427	0.3373	-0.0290	-0.7866	**0.0585**	2.8760
Amenity Presence	**0.72077**	5.946	-0.8751	-1.4808	**0.1670**	2.3013
ln(Path Size)			**2.5781**	2.2636		
μ SP					**4.2295**	2.5389
No. of decision makers	106		105		105	
No. of observations	636		309		309	
Number of parameters	5		6		6	
Final Log-likelihood	-359.97		-52.91		-435.38	
Adj. Rho-squared	0.1798		0.320		0.1632	
AIC	729.94		117.81		882.77	

Source: SP, RP and joint SP/RP model estimation results.

Table 4-11: Model Validation results

	LL per obs in estimation sample (70%)	LL per obs in validation sample (30%)	Percentage Difference (%)
Model		**SP Model**	
Subset 1	-0.5577	-0.5521	-0.9942
Subset 2	-0.5624	-0.5136	-8.6694
Subset 3	-0.5596	-0.5367	-4.0932
Subset 4	-0.5621	-0.5219	-7.159
Subset 5	-0.5596	-0.5404	-3.2454
Average	**-0.56028**	**-0.53294**	**-4.83224**
Model		**RP Model**	
Subset 1	-0.6369	-0.5702	-10.478
Subset 2	-0.6371	-0.616	-3.303
Subset 3	-0.6369	-0.5702	-10.478
Subset 4	-0.6453	-0.523	-18.953
Subset 5	-0.6305	-0.6408	1.642
Average	**-0.63734**	**-0.58404**	**-8.314**
Model		**SP + RP Model**	
Subset 1	-0.6067	-0.5867	-3.2952
Subset 2	-0.6124	-0.5586	-8.7733
Subset 3	-0.6059	-0.5967	-1.5253
Subset 4	-0.6063	-0.5968	-1.556
Subset 5	-0.6065	-0.5863	-3.3186
Average	**-0.60756**	**-0.58502**	**-3.69368**

Source: Model validation results

5 CONCLUSIONS AND RECOMMENDATIONS

5.1 Introduction

This chapter presents the conclusions, recommendations, study limitations and transferability considerations arising from the study. The conclusions are organised according to the three specific objectives: generating pedestrian route choice sets and quantifying route attributes from GPS trajectories; analysing the factors affecting pedestrian route choice by comparing stated preference and revealed preference models; and developing a pedestrian route choice model that combines stated and revealed preference data. The chapter also highlights practical planning implications, policy implications, future research needs and the conditions under which the model may be transferred to other contexts.

5.2 Conclusions

The study successfully generated pedestrian route choice sets and quantified route attributes from GPS trajectories within Makerere University. Actual student walking movements were captured using the Traccar GPS tracking application and map-matched to the campus pedestrian network using OSRM. Plausible alternative routes were generated using the OpenStreetMap-based campus network and Yen's k-shortest path algorithm, resulting in observed routes and non-chosen alternatives suitable for route choice modelling. The study further demonstrated that travel time, greenery, sidewalk availability, slope and amenities can be measured using GIS-based methods in a resource-constrained institutional setting. Continuous and ratio-based measures, such as greenery share, sidewalk coverage and amenity counts, provided richer explanatory value than simple binary indicators, confirming the importance of detailed route-attribute quantification in pedestrian route choice analysis.

The SP and RP models showed that pedestrian route choice is influenced by a combination of efficiency, environmental quality, physical effort and functional accessibility. In the SP model, students generally stated a preference for shorter, greener, less steep and amenity-supported routes, with travel time and slope reducing route utility while greenery and amenities increased route attractiveness. The RP model, however, revealed that actual walking behaviour differed from stated choices, particularly through the positive travel-time coefficient, which indicated that students sometimes accepted longer routes where these routes offered better walking conditions, such as greener surroundings, lower traffic exposure, or more comfortable off-street corridors. The negative sidewalk effect in some specifications should therefore be interpreted in context: it does not imply that sidewalks are undesirable, but rather that some formal sidewalks along busy road-edge corridors may be less attractive than quieter internal pedestrian paths. Overall, the results confirm that pedestrian route choice is not governed by travel-time minimisation alone, but by compensatory trade-offs among route efficiency, comfort, environmental quality and accessibility.

The joint SP/RP model provided stronger and more stable evidence than the single-source SP and RP models. By combining stated preference data with revealed preference data, the model captured both intended choices under controlled experimental conditions and actual walking behaviour observed through GPS tracking. The statistically significant SP scale parameter confirmed that the two data sources had different error structures and that scale adjustment was necessary. The joint model produced more statistically significant coefficients than the single-source models and achieved the smallest difference between estimation and validation log-likelihoods during hold-out validation. This indicates that the joint model generalised better to unseen respondents and was less likely to be overfitted to the estimation data. The study therefore concludes that SP and RP data are complementary, and their integration provides a more behaviourally realistic and robust approach for modelling pedestrian route choice in data-constrained African institutional settings.

5.3 Recommendations

For practical pedestrian planning within Makerere University, priority should be given to developing continuous, shaded, safe and comfortable pedestrian corridors linking lecture rooms, residences, libraries, administrative buildings, transport stops, student service areas and amenities. The findings show that students may accept longer walking distances when routes provide better environmental quality, comfort and lower traffic exposure. Therefore, campus pedestrian planning should go beyond providing sidewalks along busy vehicular corridors and should instead improve the quality of walking routes through tree planting, lighting, traffic calming, safe crossings, physical separation from vehicles, management of pedestrian obstructions, improved route surfaces and better connectivity between formal walkways and internal off-street paths. Amenities that support education-related trips, such as printeries, canteens, restaurants, reading areas, stationery points, ATMs and rest areas, should also be considered in pedestrian network planning, while ensuring that amenity-rich corridors are managed to reduce congestion and conflict with vehicles.

At policy level, the study demonstrates the need for pedestrian planning to be guided by behavioural evidence rather than infrastructure provision alone. Transport and campus planners should recognise that pedestrians value comfort, greenery, safety, accessibility and route quality in addition to travel time. Policies and design guidelines for universities, institutional hubs and urban centres should therefore promote walkability audits, pedestrian route inventories, GIS-based route-attribute mapping and routine collection of walking behaviour data. The results also support the use of low-cost GPS-based data collection and SP/RP modelling as practical tools for evidence-based non-motorised transport planning in resource-constrained urban settings.

Future research should use larger and more representative samples, longer GPS tracking periods and targeted recruitment strategies to improve female participation and reduce gender imbalance in revealed preference data. Future studies should also derive local walking speeds from GPS trajectories instead of relying only on literature-based walking speeds. Additional route variables should be incorporated, including perceived safety, security, crowding, lighting, weather, time of day, route condition, number of turns, traffic volume and pavement quality. Further modelling work should test advanced approaches such as mixed logit and latent variable models to capture unobserved heterogeneity, psychological factors and differences among pedestrian groups. Future studies should also test the model in other campuses such as Uganda Christian University or Kyambogo University, and in more complex urban settings such as the Kampala Central Business District.

5.4 Study Limitations

The study was limited to Makerere University, a semi-controlled campus environment with restricted access and lower traffic complexity than many parts of Kampala. The findings are therefore most applicable to university campuses and similar institutional hubs and should not be directly generalised to the wider Kampala urban network without further testing. The RP sample was based on voluntary participation and convenience recruitment. Although 105 participants and 309 observed trips provided useful evidence for exploratory route choice modelling, the sample was not statistically representative of the entire Makerere University student population. The RP sample was also gender-imbalanced, with males representing a much larger share than females, which limits the strength of gender-based inference from the observed GPS data.

The GPS tracking period was relatively short, limiting the ability to capture seasonal variation, day-to-day changes and long-term walking behaviour. The GPS data were also subject to possible accuracy errors arising from smartphone positioning limitations, signal interference, missing points and map-matching uncertainty. Although data cleaning and map matching were undertaken, some uncertainty in the exact walking paths may remain. The model included five main route attributes: travel time, greenery, sidewalk availability, slope and amenities, but did not fully incorporate other potentially important variables such as perceived safety, security, crowding, lighting, weather, time of day, route condition, traffic exposure and number of turns. In addition, although participation was voluntary and informed consent was obtained, formal Research Ethics Committee approval was not obtained, which is acknowledged as a limitation because GPS tracking involves sensitive personal mobility data.

5.5 Transferability

The model is transferable in principle to other university campuses, institutional hubs and selected urban pedestrian environments, but it should not be applied without local recalibration. Its findings are most relevant to settings with high pedestrian activity, mixed formal and informal walking paths, internal road networks, varied topography, greenery and student-oriented amenities. If applied to another campus such as Uganda Christian University or Kyambogo University, the local pedestrian network should be reconstructed, route attributes should be re-measured, and new SP and RP data should be collected to recalibrate the model. If applied to a more complex urban environment such as the Kampala Central Business District, additional variables such as traffic volume, roadside trading, pedestrian delay at crossings, crowding, public transport interaction, security, air quality and motorcycle or vehicle conflicts would be necessary. With more time and resources, the model could be improved through larger samples, longer GPS tracking, locally derived walking speeds, richer contextual variables and more advanced discrete choice models.

6 REFERENCES

Andersson, D., Wahlgren, L., & Schantz, P. (2022). Pedestrians' perceptions of route environments in relation to deterring or facilitating walking. *Frontiers in Public Health*, *10*. https://doi.org/10.3389/fpubh.2022.1012222

Ankunda, G., & Venter, C. (2025). Studying transfers in informal transport networks using volunteered GPS data. *Travel Behaviour and Society*, *39*. https://doi.org/10.1016/j.tbs.2024.100936

Baobeid, A., Koç, M., & Al-Ghamdi, S. G. (2021). Walkability and Its Relationships With Health, Sustainability, and Livability: Elements of Physical Environment and Evaluation Frameworks. In *Frontiers in Built Environment* (Vol. 7). Frontiers Media S.A. https://doi.org/10.3389/fbuil.2021.721218

Basil, P., & Nyachieo, G. (2023). *Exploring barriers and perceptions to walking and cycling in Nairobi metropolitan area*.

Basu, N., Haque, M. M., King, M., Kamruzzaman, M., & Oviedo-Trespalacios, O. (2022). A systematic review of the factors associated with pedestrian route choice. *Transport Reviews*, *42*(5), 672–694. https://doi.org/10.1080/01441647.2021.2000064

Basu, N., Oviedo-Trespalacios, O., King, M., Kamruzzaman, M., & Haque, M. M. (2023). What do pedestrians consider when choosing a route? The role of safety, security, and attractiveness perceptions and the built environment during day and night walking. *Cities*, *143*. https://doi.org/10.1016/j.cities.2023.104551

Basu, R., & Sevtsuk, A. (2022a). How do street attributes affect willingness-to-walk? City-wide pedestrian route choice analysis using big data from Boston and San Francisco. *Transportation Research Part A: Policy and Practice*, *163*, 1–19. https://doi.org/10.1016/j.tra.2022.06.007

Basu, R., & Sevtsuk, A. (2022b). How do street attributes affect willingness-to-walk? City-wide pedestrian route choice analysis using big data from Boston and San Francisco. *Transportation Research Part A: Policy and Practice*, *163*, 1–19. https://doi.org/10.1016/j.tra.2022.06.007

Ben-Akiva, M., & Bierlaire, M. (1999). Discrete Choice Methods and their Applications to Short Term Travel Decisions. In *Handbook of Transportation Science*.

Ben-Akiva, M., & Lerman, S. R. (1985). *Discrete Choice Analysis - Theory and Application to Travel Demand* (1st ed.). The MIT Press.

Ben-Akiva, M., & Ramming, S. (1998). *Lecture notes: Discrete choice models of traveler behavior in networks. Prepared for Advanced Methods for Planning and Management of Transportation Networks. Capri, Italy, 25*.

Bradley, M., & Daly, A. (1991). Estimation of Logit Choice Models Using Mixed Stated Preference and Revealed Preference Information. *6th International Conference on Travel Behaviour*.

Broach, J., & Dill, J. (2016). Using predicted bicyclist and pedestrian route choice to enhance mode choice models. *Transportation Research Record*, *2564*, 52–59. https://doi.org/10.3141/2564-06

Bwambale, A., Choudhury, C., & Hess, S. (2019). Modelling long-distance route choice using mobile phone call detail record data: a case study of Senegal. *Transportmetrica A: Transport Science*, *15*(2), 1543–1568. https://doi.org/10.1080/23249935.2019.1611970

Cascetta, E., Nuzzolo, A., Russo, F., & Vitetta, A. (1996). A modified logit route choice model overcoming path overlapping problems: Specification and some calibration results for interurban networks. *Transportation and Traffic Theory*, 697–711.

Chebe, P. C., Nkurunziza, A., & Karel Martens. (2025). Urban roads: Enablers or barriers to walking? insights from Kigali, Rwanda. *Travel Behaviour and Society*, *41*.

ChoiceMetrics. (2024). *Ngene 1.4 USER MANUAL & REFERENCE GUIDE The Cutting Edge in Experimental Design*. www.choice-metrics.com

Clifton, K. J., Singleton, P. A., Muhs, C. D., & Schneider, R. J. (2016). Development of destination choice models for pedestrian travel. *Transportation Research Part A: Policy and Practice*, *94*, 255–265. https://doi.org/10.1016/j.tra.2016.09.017

Cobbinah, P. B., & Finn, B. M. (2024). On Pedestrian Accessibility: Spatial Justice and Progressive Planning in African Cities. *Journal of Planning Literature*, *40*(2).

Dzisi, E. K. J., & Lugada, T. (2021). Modeling the potential shift from motorcycles (boda bodas) to bicycles among young people on a Ugandan university campus. *Scientific African*, *12*. https://doi.org/10.1016/j.sciaf.2021.e00741

Erath, A., Van Eggermond, M., Ordonez, S., & Axhausen, K. (2015). *Modelling for walkability; Understanding pedestrians' preferences in Singapore*. https://doi.org/10.3929/ethz-b-000106077

Fulman, N., Huber, J., Lloyd, A. T., Foshag, K., Grinblat, Y., Türk, U., Lautenbach, S., Amcoff, J., Toger, M., Jokinen, J., & Zipf, A. (2025). Route choice to inform navigation system design and accessibility analysis for older pedestrians: a scoping review. *Journal of Transport & Health*, *44*, 102151. https://doi.org/10.1016/J.JTH.2025.102151

Gregory Lue, & Eric J. Miller. (2019). Estimating a Toronto pedestrian route choice model using smartphone GPS data - ScienceDirect. *Travel Behaviour and Society*, *14*, 34–42. https://doi.org/https://doi.org/10.1016/j.tbs.2018.09.008

Guo, Z., & Loo, B. P. Y. (2013). Pedestrian environment and route choice: evidence from New York City and Hong Kong. *Journal of Transport Geography*, *28*, 124–136.

Hamouni, P. (2018). *A Pedestrian Route Choice Model Concerning Quantified Built Environment Factors*. CONCORDIA UNIVERSITY.

Hasnine, M. S., Rashedi, Z., & Nurul Habib, K. (2020a). Making use of elicited confidence ratings of the SP survey in a joint RP-SP model: a case study of commuting mode choices for TDM evaluations. *Transportation Letters*, *12*(4), 274–287. https://doi.org/10.1080/19427867.2019.1585687

Hasnine, M. S., Rashedi, Z., & Nurul Habib, K. (2020b). Making use of elicited confidence ratings of the SP survey in a joint RP-SP model: a case study of commuting mode choices for TDM evaluations. *Transportation Letters*, *12*(4), 274–287. https://doi.org/10.1080/19427867.2019.1585687

Hensher, D., Louviere, J., & Swait, J. (1999). Combining sources of preference data. In *Journal of Econometrics* (Vol. 89).

Hess, S., & Palma, D. (2019). Apollo: A flexible, powerful and customisable freeware package for choice model estimation and application. *Journal of Choice Modelling*, *32*. https://doi.org/10.1016/j.jocm.2019.100170

Jin, C. J., Li, J., Wu, C., Xiu, X., & Li, D. (2025). Findings in pedestrian route choice experiments: Tradeoff and hypothetical bias. *Travel Behaviour and Society*, *40*. https://doi.org/10.1016/j.tbs.2025.101000

Jin, C. J., Luo, Y., Wu, C., Song, Y., & Li, D. (2024). Exploring the Pedestrian Route Choice Behaviors by Machine Learning Models. *ISPRS International Journal of Geo-Information*, *13*(5). https://doi.org/10.3390/ijgi13050146

Jin, C. J., Wu, C., Song, Y., Liu, T., Li, D., Jiang, R., & Fang, S. (2024). The route choices of pedestrians under crowded and non-emergency conditions: Two-route experiments and modeling. *Journal of Choice Modelling*, *50*. https://doi.org/10.1016/j.jocm.2023.100463

Kožul, N., Novačko, L., Babojelić, K., & Brlek, P. (2025). Stated Preference Approach in Shaping Urban Sustainable Multimodal Transport—A Literature Review. *Systems*, *13*(10), 853. https://doi.org/10.3390/systems13100853

Lam, S.-H., & Xie, F. (2002). Transit Path-Choice Models That Use Revealed Preference and Stated Preference Data. In *Transportation Research Record*.

Lee, S., & Lee, M. H. (2021). Impact of neighborhood environment on pedestrian route selection among elementary schoolchildren in Korea. *International Journal of Environmental Research and Public Health*, *18*(13). https://doi.org/10.3390/ijerph18137049

Lieu, S. J., & Guhathakurta, S. (2025). Exploring pedestrian route choice preferences by demographic groups: Analysis of street attributes in Chicago. *Transportation*

Research Part A: Policy and Practice, *195*. https://doi.org/10.1016/j.tra.2025.104437

Loo, B. P. Y., Lian, T., & Frank, L. D. (2024). Walking (In)Convenience: An In-Depth Study of Pedestrian Detours to Daily Facilities. *Journal of the American Planning Association*, *90*(4), 742–757. https://doi.org/10.1080/01944363.2024.2310613

Marschak, J. (1960). Binary Choice Constraints and Random Utility Indicators. *Economic Information, Decision, and Prediction*, *7*(1), 218–239.

Mcfadden, D. (1974). The Measurement of Urban Travel Demand. *Journal of Public Economics*, *3*, 303–328.

Mogaji, E. (2022). Cycling in Lagos: The challenges, opportunities, and prospects. *Transportation Research Interdisciplinary Perspectives*, *14*. https://doi.org/10.1016/j.trip.2022.100608

Murtagh, E. M., Mair, J. L., Aguiar, E., Tudor-Locke, C., & Murphy, M. H. (2021). Outdoor Walking Speeds of Apparently Healthy Adults: A Systematic Review and Meta-analysis. In *Sports Medicine* (Vol. 51, Issue 1, pp. 125–141). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/s40279-020-01351-3

National Planning Authority. (2024). *Fourth National Development Plan*. *1*.

Okyere, S. A., Frimpong, L. K., Oviedo, D., Mensah, S. L., Fianoo, I. N., Nieto-Combariza, M. J., Abunyewah, M., Adkins, A., & Kita, M. (2024). Policy-reality gaps in Africa's walking cities: Contextualizing institutional perspectives and residents' lived experiences in Accra. *Journal of Urban Affairs*. https://doi.org/10.1080/07352166.2023.2296105

Qiao, Y., Huang, Y., Yang, F., Zhang, M., & Chen, L. (2016a). Empirical study of travel mode forecasting improvement for the combined revealed preference/stated preference data-based discrete choice model. *Advances in Mechanical Engineering*, *8*(1). https://doi.org/10.1177/1687814015624836

Qiao, Y., Huang, Y., Yang, F., Zhang, M., & Chen, L. (2016b). Empirical study of travel mode forecasting improvement for the combined revealed preference/stated preference data-based discrete choice model. *Advances in Mechanical Engineering*, *8*(1). https://doi.org/10.1177/1687814015624836

Ramming, M. S. (2002). *Network Knowledge and Route Choice*.

Sevtsuk, A., Basu, R., Li, X., & Kalvo, R. (2021). A big data approach to understanding pedestrian route choice preferences: Evidence from San Francisco. *Travel Behaviour and Society*, *25*, 41–51. https://doi.org/10.1016/j.tbs.2021.05.010

Sevtsuk, A., & Kalvo, R. (2022). Predicting pedestrian flow along city streets: A comparison of route choice estimation approaches in downtown San Francisco.

International Journal of Sustainable Transportation, *16*(3), 222–236. https://doi.org/10.1080/15568318.2020.1858377

Sietchiping, R., Permezel, M. J., & Ngomsi, C. (2012). Transport and mobility in sub-Saharan African cities: An overview of practices, lessons and options for improvements. *Cities*, *29*(3), 183–189. https://doi.org/10.1016/j.cities.2011.11.005

Stanitsa, A., Hallett, S. H., & Jude, S. (2023). Investigating pedestrian behaviour in urban environments: A Wi-Fi tracking and machine learning approach. *Multimodal Transportation*, *2*(1). https://doi.org/10.1016/j.multra.2022.100049

Tong, Y., & Bode, N. W. F. (2022a). The principles of pedestrian route choice. In *Journal of the Royal Society Interface* (Vol. 19, Issue 189). Royal Society Publishing. https://doi.org/10.1098/rsif.2022.0061

Tong, Y., & Bode, N. W. F. (2022b). The principles of pedestrian route choice. In *Journal of the Royal Society Interface* (Vol. 19, Issue 189). Royal Society Publishing. https://doi.org/10.1098/rsif.2022.0061

Uganda Police Force. (2023). *Annual Crime Report*.

UN Habitat. (2023). *Annual Report - UN Habitat*.

Vissoci, J. R. N., Shogilev, D. J., Krebs, E., Andrade, L. de, Vieira, I. F., Toomey, N., & Staton, C. A. (2017). Road traffic injury in sub-Saharan African countries: A systematic review and summary of observational studies. *Traffic Injury Prevention*, *18*(7), 767–773.

Westenhöfer, J., Nouri, E., Reschke, M. L., Seebach, F., & Buchcik, J. (2023). Walkability and urban built environments—a systematic review of health impact assessments (HIA). *BMC Public Health*, *23*(1). https://doi.org/10.1186/s12889-023-15394-4

World Health Organisation (WHO). (2023). *Global status report on road safety*.

Xie, W., Wai Ming Lee, E., Li, T., Jiang, N., & Ma, Y. (2024). Pedestrian dynamics on slopes: Empirical analysis of level, uphill, and downhill walking. *Safety Science*, *172*. https://doi.org/10.1016/j.ssci.2024.106429

Yen, J. Y. (1971). Finding the K Shortest Loopless Paths in a Network. *Management Science*, *17*(11), 712–716.

7 APPENDIX

7.1 Appendix 1: Ngene syntax

```
design
;alts = opt1*, opt2*
;rows = 18
;block = 3
;eff = (mnl, d)
;alg = swap

;model:
U(opt1) = Btt[-0.955] * TT[6,8,10]
        + Bgn[0.787] * GN[0,1]
        + Bsw[0.791] * SW[0,50,100]
        + Bss[-3.98] * SS[1,5,10]
        + Bamn[0.201] * AMN[0,1]
/
U(opt2) = Btt[-0.955] * TT
        + Bgn[0.787] * GN
        + Bsw[0.791] * SW
        + Bss[-3.98] * SS
        + Bamn[0.201] * AMN
;formatTitle = 'Scenario <scenarionumber>'
;formatTableDimensions = 3, 7
;formatTable:
1,1 = '' /
1,2 = 'Travel Time (minutes)' /
1,3 = 'Greenery' /
1,4 = 'Sidewalk availability (%)' /
1,5 = 'Slope (%)' /
1,6 = 'Amenities' /
1,7 = '' /
2,1 = 'Option 1' /
2,2 = '<opt1.tt>' /
2,3 = '<opt1.gn>' /
2,4 = '<opt1.sw>' /
2,5 = '<opt1.ss>' /
2,6 = '<opt1.amn>' /
2,7 = '' /
3,1 = 'Option 2' /
3,2 = '<opt2.tt>' /
3,3 = '<opt2.gn>' /
3,4 = '<opt2.sw>' /
3,5 = '<opt2.ss>' /
3,6 = '<opt2.amn>' /
3,7 = ''
;formatTableStyle:
1,1 = 'default' /
1,2 = 'headingattribute' /
1,3 = 'headingattribute' /
1,4 = 'headingattribute' /
1,5 = 'headingattribute' /
1,6 = 'headingattribute' /
1,7 = 'headingattribute' /
2,1 = 'heading1' /
2,2 = 'body1' /
2,3 = 'body1' /
```

```
2,4 = 'body1' /
2,5 = 'body1' /
2,6 = 'body1' /
2,7 = 'choice1' /
3,1 = 'heading2' /
3,2 = 'body2' /
3,3 = 'body2' /
3,4 = 'body2' /
3,5 = 'body2' /
3,6 = 'body2' /
3,7 = 'choice2'
;formatStyleSheet = Blue buttons.css
;formatAttributes:
opt1.tt(6=# minutes, 8=# minutes, 10=# minutes) /
opt1.gn(0=NOT Present, 1=Present) /
opt1.sw(0=#%, 50=#%, 100=#%) /
opt1.ss(1=Gentle Slope (1%), 5=Medium Slope (5%), 10=Steep Slope (10%)) /
opt1.amn(0=NOT Present, 1=Present) /
opt2.tt(6=# minutes, 8=# minutes, 10=# minutes) /
opt2.gn(0=NOT Present, 1=Present) /
opt2.sw(0=#%, 50=#%, 100=#%) /
opt2.ss(1=Gentle Slope (1%), 5=Medium Slope (5%), 10=Steep Slope (10%)) /
opt2.amn(0=NOT Present, 1=Present)
$
```

7.2 Appendix 2: Code for parameter estimation in R software an output

1. Install Apollo

```
# Install Apollo (only once per computer)
install.packages("apollo")   # You can skip this if already installed

# Load the Apollo library (do this every R session)
library(apollo)
```

2. Set Your Working Directory (Optional but recommended)

```
setwd("C:/Path/To/Your/Folder")

getwd()
apollo_initialise()
```

3. Load Your Data

```
database <- read.csv("Main_Study_results.csv", header=TRUE)
```

4. Basic Apollo Setup

```
apollo_control = list(
  modelName = "PedestrianRouteChoiceMNLMainStudy",
  modelDescr = "Main Study MNL model for pedestrian route choice (Kampala 
students)",
  indivID = "ID",             # Must match your ID column in data
  panelData = TRUE,           # TRUE: multiple choices per respondent
  outputDirectory = "output" # Results will be saved here
)
```

5. Define Model Parameters

```
apollo_beta = c(
  B_tt  = -0.955,
  B_gn  =  0.787,
  B_sw  =  0.791,
  B_ss  = -3.98,
  B_amn =  0.201
)

apollo_fixed = c()   # Leave empty to estimate all parameters
```

6. Validate Inputs

```
apollo_inputs = apollo_validateInputs()
```

7. Define the Utility and Probabilities Function

```
apollo_probabilities = function(apollo_beta, apollo_inputs, functionality 
= "estimate") {
  apollo_attach(apollo_beta, apollo_inputs)
  on.exit(apollo_detach(apollo_beta, apollo_inputs))
  P = list()
```

```
  V = list()
  V[["opt1"]] =
B_tt * opt1.tt +
B_gn * opt1.gn +
B_sw * opt1.sw +
B_ss * opt1.ss +
B_amn * opt1.amn

  V[["opt2"]] =
B_tt * opt2.tt +
B_gn * opt2.gn +
B_sw * opt2.sw +
B_ss * opt2.ss +
B_amn * opt2.amn

  mnl_settings = list(
    alternatives = c(opt1 = 1, opt2 = 2),
    avail = list(opt1 = 1, opt2 = 1),
    choiceVar = choice,
    utilities = V
  )
  P[["model"]] = apollo_mnl(mnl_settings, functionality)
  P = apollo_panelProd(P, apollo_inputs, functionality)
  P = apollo_prepareProb(P, apollo_inputs, functionality)
  return(P)
}
```

8. Estimate the Model

```
model = apollo_estimate(apollo_beta, apollo_fixed, apollo_probabilities,
apollo_inputs)
```

9. View and Save Results

```
apollo_modelOutput(model)
apollo_saveOutput(model)
```

Output

```
Model run by ismab using Apollo 0.3.5 on R 4.5.0 for Windows.
Please acknowledge the use of Apollo by citing Hess & Palma (2019)
  DOI 10.1016/j.jocm.2019.100170
  www.ApolloChoiceModelling.com

Model name                                  :
NoSSNoSWPedestrianRouteChoiceMNLMainStudy
Model description                           : NoSSNoSW Main Study MNL
model for pedestrian route choice (Kampala students)
Model run at                                : 2025-07-09 12:56:14.744379
Estimation method                           : bgw
Model diagnosis                             : Relative function
convergence
Optimisation diagnosis                      : Maximum found
     hessian properties                     : Negative definite
     maximum eigenvalue                     : -113.419186
     reciprocal of condition number         : 0.101773
Number of individuals                       : 106
Number of rows in database                  : 636
Number of modelled outcomes                 : 636

Number of cores used                        :  1
Model without mixing

LL(start)                                   : -607.05
LL at equal shares, LL(0)                   : -445
LL at observed shares, LL(C)                : -444.55
LL(final)                                   : -365.24
Rho-squared vs equal shares                  :  0.1792
Adj.Rho-squared vs equal shares              :  0.1725
Rho-squared vs observed shares               :  0.1784
Adj.Rho-squared vs observed shares           :  0.1739
AIC                                         :  736.47
BIC                                         :  749.87

Estimated parameters                        : 3
Time taken (hh:mm:ss)                       :  00:00:0.24
     pre-estimation                         :  00:00:0.13
     estimation                             :  00:00:0.06
     post-estimation                        :  00:00:0.05
Iterations                                  :  8

Unconstrained optimisation.

Estimates:
         Estimate        s.e.   t.rat.(0)    Rob.s.e.
B_tt      -0.1922     0.03014      -6.375     0.03506
B_gn       0.3066     0.09061       3.384     0.10257
B_amn      0.6938     0.09014       7.697     0.11115
      Rob.t.rat.(0)
B_tt         -5.481
B_gn          2.989
B_amn         6.242


Overview of choices for MNL model component :
                                 opt1   opt2
Times available                642.00 642.00
Times chosen                   309.00 333.00
```

```
Percentage chosen overall         48.13  51.87
Percentage chosen when available  48.13  51.87



Classical covariance matrix:
             B_tt        B_gn       B_amn
B_tt   9.0850e-04  2.8369e-04  -6.032e-05
B_gn   2.8369e-04    0.008210 -6.4018e-04
B_amn  -6.032e-05 -6.4018e-04    0.008125

Robust covariance matrix:
             B_tt        B_gn       B_amn
B_tt     0.001229 -3.0719e-04 -4.6256e-04
B_gn  -3.0719e-04     0.01052 -9.7842e-04
B_amn -4.6256e-04 -9.7842e-04     0.01235

Classical correlation matrix:
             B_tt        B_gn       B_amn
B_tt      1.00000     0.10388    -0.02220
B_gn      0.10388     1.00000    -0.07838
B_amn    -0.02220    -0.07838     1.00000

Robust correlation matrix:
             B_tt        B_gn       B_amn
B_tt      1.00000    -0.08543    -0.11871
B_gn     -0.08543     1.00000    -0.08582
B_amn    -0.11871    -0.08582     1.00000

 20 most extreme outliers in terms of lowest average per choice
prediction:
  ID Avg prob per choice
  61           0.2576807
  79           0.2930379
  86           0.3247260
  76           0.3460415
  94           0.3460415
  98           0.3460415
 107           0.3834611
  78           0.3935229
  39           0.3936421
  19           0.4093744
  14           0.4199164
   4           0.4203160
   5           0.4311397
  63           0.4360770
  80           0.4360770
  55           0.4368318
   1           0.4481192
  11           0.4596589
  37           0.4655061
   2           0.4655460

Changes in parameter estimates from starting values:
          Initial    Estimate  Difference
B_tt      -0.9550     -0.1922      0.7628
B_gn       0.7800      0.3066     -0.4734
B_amn      0.2010      0.6938      0.4928

Settings and functions used in model definition:
```

```
apollo_control
--------------
                       Value
modelName              "NoSSNoSWPedestrianRouteChoiceMNLMainStudy"
modelDescr             "NoSSNoSW Main Study MNL model for pedestrian
route choice (Kampala students)"
indivID                "ID"
panelData              "TRUE"
outputDirectory        "output/"
debug                  "FALSE"
nCores                 "1"
workInLogs             "FALSE"
seed                   "13"
mixing                 "FALSE"
HB                     "FALSE"
noValidation           "FALSE"
noDiagnostics          "FALSE"
calculateLLC           "TRUE"
analyticHessian        "FALSE"
memorySaver            "FALSE"
analyticGrad           "TRUE"
analyticGrad_manualSet "FALSE"
overridePanel          "FALSE"
preventOverridePanel   "FALSE"
noModification         "FALSE"

Hessian routines attempted
--------------------------
numerical jacobian of LL analytical gradient

Scaling used in computing Hessian
---------------------------------
          Value
B_tt  0.1921576
B_gn  0.3066123
B_amn 0.6937807


apollo_probabilities
----------------------
function(apollo_beta, apollo_inputs, functionality = "estimate") {
    apollo_attach(apollo_beta, apollo_inputs)
    on.exit(apollo_detach(apollo_beta, apollo_inputs))
    P = list()
    V = list()
    V[["opt1"]] = B_tt * opt1.tt + B_gn * opt1.gn + B_amn * opt1.amn
    V[["opt2"]] = B_tt * opt2.tt + B_gn * opt2.gn + B_amn * opt2.amn

    mnl_settings = list(
        alternatives = c(opt1 = 1, opt2 = 2),
        avail = list(opt1 = 1, opt2 = 1),
        choiceVar = choice,
        utilities = V
    )
    P[["model"]] = apollo_mnl(mnl_settings, functionality)
    P = apollo_panelProd(P, apollo_inputs, functionality)
    P = apollo_prepareProb(P, apollo_inputs, functionality)
    return(P)
}
```

7.3 Appendix 3: Amenities Considered in the study

Amenity category	**Examples considered**	**Relevance to education-related walking trips**
Printing and photocopying services	Printeries, photocopying points, stationery shops	Support printing, copying, scanning and preparation of academic materials
Food and refreshment facilities	Canteens, restaurants, cafés, snack kiosks	Provide meals, drinks and rest opportunities between classes or study sessions
Academic service points	Libraries, reading rooms, computer laboratories, bookshops	Support study, access to learning resources and academic work
Financial service points	ATMs, mobile money points, banking agents	Support access to cash and payments for student needs
Religious and community facilities	Chapel, mosque	Support routine student activities and route attraction for non-class trips

7.4 Appendix 4: Stated Preference Survey tool

Pedestrian Route Choice Survey (Block1)

Dear Participant,

Thank you for taking part in our survey.

This survey is part of a research study on pedestrian route choices conducted by the Department of Civil and Environmental Engineering at Makerere University. The survey will take approximately 10 minutes.

Your responses will be treated confidentially and used only for this research. Data will not be shared with third parties. There are no right or wrong answers.

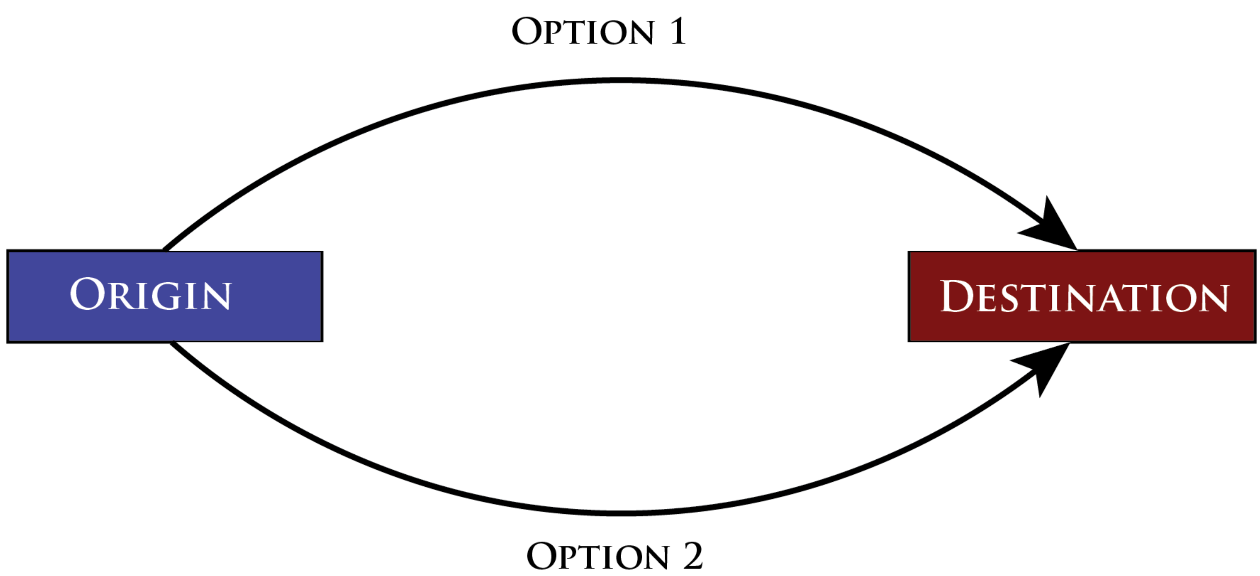
By proceeding, you consent to the use of the information you provide for research purposes only.

CHOICE MOMENTS

Imagine you are on the university campus about to begin your usual activities. Based on the two route options presented, each with their respective characteristics, which route would you prefer to take?

* Indicates required question

Choice Moments



Choice Moments 1 and 2

Please proceed and answer the choice questions below.

1. **Choice Moment 1:** Imagine you are on the university campus about to start your usual activities. Considering the characteristics of the two route options shown, which route would you prefer? *

NOTE:
Amenities: Facilities such as printers and restaurants along the route.
Slope: Indicates how steep the route is.
Greenery: Presence of trees or green areas for shade and resting.

	Option 1	Option 2
Travel Time (minutes)	8 minutes	10 minutes
Greenery	NOT Present	Present
Sidewalk availability (%)	0%	0%
Slope (%)	Medium Slope (5%)	Medium Slope (5%)
Amenities	NOT Present	Present

Mark only one oval.

- Option 1
- Option 2

2. **Choice Moment 2:** Imagine you are on the university campus about to start your usual activities. Considering the characteristics of the two route options shown, which route would you prefer? *

NOTE:

Amenities: Facilities such as printers and restaurants along the route.
Slope: Indicates how steep the route is.
Greenery: Presence of trees or green areas for shade and resting.

	Option 1	Option 2
Travel Time (minutes)	10 minutes	6 minutes
Greenery	Present	NOT Present
Sidewalk availability (%)	100%	50%
Slope (%)	Steep Slope (10%)	Gentle Slope (1%)
Amenities	NOT Present	Present

Mark only one oval.

- Option 1
- Option 2

Choice Moments 3 and 4

Please proceed and answer the choice questions below.

3. **Choice Moment 3:** Imagine you are on the university campus about to start your usual activities. Considering the characteristics of the two route options shown, which route would you prefer? *

NOTE:

Amenities: Facilities such as printers and restaurants along the route.
Slope: Indicates how steep the route is.
Greenery: Presence of trees or green areas for shade and resting.

	Option 1	Option 2
Travel Time (minutes)	8 minutes	6 minutes
Greenery	NOT Present	Present
Sidewalk availability (%)	100%	50%
Slope (%)	Steep Slope (10%)	Gentle Slope (1%)
Amenities	Present	NOT Present

Mark only one oval.

- Option 1
- Option 2

4. **Choice Moment 4:** Imagine you are on the university campus about to start your usual activities. Considering the characteristics of the two route options shown, which route would you prefer? *

NOTE:

Amenities: Facilities such as printers and restaurants along the route.
Slope: Indicates how steep the route is.
Greenery: Presence of trees or green areas for shade and resting.

	Option 1	Option 2
Travel Time (minutes)	8 minutes	8 minutes
Greenery	Present	NOT Present
Sidewalk availability (%)	100%	100%
Slope (%)	Medium Slope (5%)	Medium Slope (5%)
Amenities	NOT Present	Present

Mark only one oval.

- Option 1
- Option 2

Choice Moments 5 and 6

Please proceed and answer the choice questions below.

5. **Choice Moment 5:** Imagine you are on the university campus about to start your usual activities. Considering the characteristics of the two route options shown, which route would you prefer? *

NOTE:
Amenities: Facilities such as printers and restaurants along the route.
Slope: Indicates how steep the route is.
Greenery: Presence of trees or green areas for shade and resting.

	Option 1	Option 2
Travel Time (minutes)	6 minutes	10 minutes
Greenery	NOT Present	Present
Sidewalk availability (%)	50%	100%
Slope (%)	Gentle Slope (1%)	Steep Slope (10%)
Amenities	NOT Present	Present

Mark only one oval.

- Option 1
- Option 2

6. **Choice Moment 6:** Imagine you are on the university campus about to start your usual activities. Considering the characteristics of the two route options shown, which route would you prefer? *

NOTE:
Amenities: Facilities such as printers and restaurants along the route.
Slope: Indicates how steep the route is.
Greenery: Presence of trees or green areas for shade and resting.

	Option 1	Option 2
Travel Time (minutes)	8 minutes	8 minutes
Greenery	Present	NOT Present
Sidewalk availability (%)	0%	50%
Slope (%)	Gentle Slope (1%)	Steep Slope (10%)
Amenities	Present	NOT Present

Mark only one oval.

- Option 1
- Option 2

PERSONAL DETAILS

In this section, you will provide us with some of your personal details. All information provided will be used solely for research purposes.

Choose the applicable option for each of the next questions.

7. **Gender** *

Mark only one oval.

- Male
- Female

8. **Age** *

Mark only one oval.

- 18-24
- 25-34
- 35-44
- 45-54
- 55-64
- 65+

9. **Employment Status** *

Mark only one oval.

- Student
- Informal
- Formal in public sector
- Formal in private sector
- Not employed

10. **Monthly Income in UGX** *

Mark only one oval.

- 150,000
- 150,001 - 300,000
- 300,001 - 500,000
- 500,001 - 1,000,000
- 1,000,000 - 2,000,000
- 2,000,000+
- No Income

Survey Complete

Thank you very much for taking off sometime to participate in this survey.