Tambula nange

Relations between perceived walkability, built environment and pedestrian route choice in a developing country

S.C.A. Riemens

UNIVERSITY OF TWENTE.

TAMBULA NANGE

Modelling perceived walkability, built environment and pedestrian route choice in a developing nation context

S.C.A. Riemens

August 2018

Thesis for the purpose of completion of a
Master's degree in Civil Engineering and Management

Centre for Transportation Studies
University of Twente

Supervised by
Prof. E.C. van Berkum
Dr. T. Thomas
Prof. M.H.P. Zuidgeest (University of Cape Town)

SAMENVATTING

Het onderwerp van deze scriptie betreft het gebruik van een route-keuzemodel met smartphone GPS data verzameld in een ontwikkelingsland. Om precies te zijn: het Recursive Logit (RL) model, waarmee route-keuzes kunnen worden gemodelleerd zonder dat van te voren een deelverzameling van kiesbare routes aangegeven hoeft te worden. Het doel hiervan is om te ontdekken welke factoren nu eigenlijk belangrijk zijn bij het bepalen van de route die mensen gebruiken tijdens hun dagelijkse activiteiten. In eerste instantie wil men natuurlijk een zo kort mogelijke route lopen, maar bewust of onbewust wordt soms toch van de kortste route afgeweken en het is de vraag of hiervoor aanwijsbare oorzaken zijn. Dit onderzoek is vernieuwend omdat het RL model voor het eerst wordt toegepast voor routekeuze van voetgangers (in tegenstelling tot fietsers en automobilisten), en nog wel in een ontwikkelingsland wat niet bepaald bekend staat om zijn aantrekkelijke infrastructuur. Daarnaast wordt er voor het eerst een subjectieve component gebruikt in het RL model. De routes worden niet alleen aan externe, objectieve maatstaven gekoppeld, maar ook aan ratings die de voetgangers zelf gegeven hebben. Het blijkt dat deze ratings wel invloed hebben op het loopgedrag, maar dat de objectieve maatstaven ongeveer evenveel informatie geven. Interessant is dat een model waarin zowel objectieve als subjectieve factoren meegenomen werden, veruit de beste fit had. Dit suggereert dat het belangrijk is om zowel fysieke als belevingsfactoren in een routekeuzemodel mee te nemen.

Om het e.e.a. wat concreter te maken: de deelnemers aan dit onderzoek vermeden wegen met verkeersdrempels en met verkeerseilanden of andere oversteekhulpen. Ook wegen waar geparkeerde auto's de straatkant of stoep blokkeerden weren vermeden, evenals die waar zich veel (zwerf)afval ophoopte. Daarentegen werden wegen waar politie of beveiligers aanwezig waren juist vaker gebruikt. Van de zes rating-criteria hadden er slechts twee een significante invloed: "levendige" straten werden vaker gebruikt en "veilige" juist minder vaak. Niet van invloed waren de steilheid van wegen op de route noch de staat van de voetgangersinfrastructuur: hiervoor waren een heel aantal indicatoren opgenomen (waaronder kwaliteit van de bestrating en aaneengeslotenheid van het trottoir).

De data voor dit onderzoek is verzameld in Kampala (hoofdstad van Oeganda) gedurende een verblijf van vijf maanden. Het onderzoek is ondersteund door de lokale NGO *Advocates for Public Spaces*. Negenenvijftig deelnemers hebben een week lang een app gebruikt om hun bewegingspatronen op te slaan. In totaal zijn meer dan 500 looproutes verzameld en hebben de deelnemers op meer dan 1200 plekken ratings gegeven. De app is ook gedurende het verblijf in Oeganda ontwikkeld, op basis van een al bestaande app, en heet *Walkability*. De deelnemers gebruikten hun eigen telefoon of konden er een lenen.

Daarnaast is er ook data verzameld met de *GPStionnaire* app, die speciaal voor dit onderzoek is ontwikkeld. Hiermee kan eenvoudig een vragenlijst worden opgesteld en ingeprogrammeerd en de onderzoeker kan steeds langs de kant van de weg antwoorden geven. De app slaat dan automatisch op waar precies de vragenlijst ingevuld is. In dit geval waren er 36 vragen, welke op 137 plekken verspreid door Kampala zijn ingevuld.

Dit onderzoek past binnen een breder palet aan studies naar loopinfrastructuur en beleving daarvan. Op initiatief van de Universiteit Kaapstad zijn er inmiddels een aantal apps ontwikkeld met het doel om hiermee in ontwikkelingslanden aan de slag te gaan.

Titel van dit verslag is "tambula nange" en dat is Luganda voor "loop met me mee". Tijdens het verblijf in Oeganda werd er een zeer populair liedje uitgebracht met deze titel, waarin de zanger tot God bidt dat hij met hem meereist en het rechte pad aanwijst.

SUMMARY

The present thesis concerns the use of a route choice model with smartphone GPS data on walkability gathered in a developing country. Specifically, the Recursive Logit model is used which allows for choice modelling without choice sets of paths. The purpose of this is to uncover which factors determine the route people use when walking to their destination, even though such choices are often made subconsciously. Novel in this research is not only the application of the RL model to pedestrian route choice but also the inclusion of perceived walkability indicators: that is to say, the people who were doing the walking were also asked to give their opinion about various elements along their route. The influence of these subjective assessments is compared to that of more traditional objective indicators, measured using a Built Environment inventory form. We find that the subjective measures provide slightly more information about route choices as the objective indicators do, but that a model combining these factors produces a better fit by far. This suggests that it is important for future pedestrian route choice research to try and include subjective factors into the model if possible.

Concretely, the pedestrians in this research seemed to avoid roads with speed bumps and those with traffic islands or other crossing aids, as well as roads where the roadside was blocked by parked cars or those with heavy presence of litter and garbage. Paths with presence of police or private security guards were used comparatively more often. Of the subjective indicators, only two had a significant influence: roads which were deemed to be more lively were used more often, as well as those deemed to be unsafe because of traffic conditions. Interestingly, participants did not seem to avoid steep climbs and the state of the pedestrian infrastructure, including such indicators as pavement condition and path continuity, did not significantly influence route choices.

The data for this research was collected in Kampala, Uganda during a five-month stay assisted by local NGO *Advocates for Public Spaces*. Fifty-nine people agreed to be tracked via their cellphone for a one week period each. In total, over 500 walking routes were collected and the participants gave ratings to more than 1200 locations. This was done using a newly developed Android app, simply called *Walkability*, which collected and stored the routes and ratings. Participants used their own phone or borrowed one from a pool of phones which were made available for this project.

The second source of data was collected using another Android app developed for this research, called *GPStionnaire*. A Built Environment Inventory survey was conducted and the app was used to geo-reference the survey answers. In total, 137 locations were assessed using the survey tool, which included 36 questions.

This research project was proposed by the University of Cape Town and fits into a broader range of projects dealing with walking, walkability and the use of smartphone applications in developing countries.

The title of this research, "tambula nange", is a Luganda phrase meaning "walk with me" and is also the title of a very popular song that was released during the author's stay in Uganda. The singer beseeches the Lord to travel with him and to show him the way.

ACKNOWLEDGEMENTS

Vandaag ben ik gaan lopen
Ik heb de meningen geteld

Vandaag ben ik gaan lopen
en waar ik loop is van nu af aan een weg

CONTENTS

CHAPTER 1

INTRODUCTION

Walking is one of the main modes of transport in cities of the developing world, as many residents do not own private vehicles and public transport can be expensive or disorganised. At the same time, however, pedestrian infrastructure is often of low quality or wholly absent. This makes walking an unpleasant and often unsafe experience (it's said that "Africans walk despite the walking infrastructure, not because of it"[1]).

In developed countries, there has been a lot of interest in and research done on *walkability*, i.e. trying to measure the attractiveness and suitability of a region for pedestrians. There have also been some studies done on walkability in Africa (notably by Oyeyemi et al., 2016b). Links are primarily sought between walkability and physical activity, ostensibly to combat a rise in welfare diseases like obesity in emerging economies. However, research (Tainio et al., 2016) has recently shown that Kampala is one of the few cities in the world where walking can actually be detrimental to health - as a result of high levels of ambient air pollution. These health risks are accrued disproportionately by pedestrians who rely on walking for transport.

It is not clear how walking behaviour and walkability are related in the context of developing nations - in fact, there has been comparatively little research published on walkability in developing countries at large.

One of the goals of this research is to develop advanced forms of transportation research in Africa - specifically using smartphone apps. In Uganda it is said that there are "more mobile phones than lightbulbs" and indeed 17 million people have a phone capable of accessing the internet, nearly 41% of the population (GSM Association, 2017), fertile ground for these new research methods.

As part of this research, development was done on the Walkability smartphone app first created by the University of Cape Town, and it was tested in the field in a relatively large scale study. An important outcome of the study will also be to determine whether the app is suitable for use in walking behaviour studies. Traditionally, these studies are quite involved and there are many difficulties associated with self-reported walking behaviour and user-assessed walkability (Saelens & Handy, 2008); use of a a smartphone app could make data easier to collect and more accurate. This study shows that using such an app can provide reliable data and the kind of conclusions can be drawn about walking behaviour. The results of this study may help to facilitate further walkability surveys. Further, by identifying main drivers behind pedestrian behaviour, it is hoped that this research will help policy makers to identify efficient interventions to facilitate and promote walking.

The main research question addressed in this thesis is: *what does walkability mean for Kampala citizens and what factors influence their daily walking route choices?*. Five steps have been identified that will allow for answering this question and culminate in the definition of a Walkability index. Each of these steps will be considered as a separate research sub-question in the study and they are as follows:

1. *how do the users of the network perceive the walkability on its links and how do they navigate their way through the study area?*
2. *what are the physical characteristics of the street network in the study area and how conducive is the built environment to walking?*
3. *how representative are the user assessments and are the ratings correlated between individuals or between similar street types? (For example, do gender or presence of dedicated sidewalks influence ratings?)*
4. *can the user assessments or the built environment characteristics explain participants' travel behaviour and can this be modelled using a route choice model?*

This thesis document will first present findings from a literature review, in which some different interpretations of walkability are discussed and main methods for quantification are briefly presented. Then, each of the sub-questions is addressed in its own chapter of the report.

---

[1] https://thisbigcity.net/african-cities-are-walking-cities-but-are-they-walkable

LITERATURE REVIEW

Walkability is not a directly measurable quantity and it is often viewed in relationship to other, more quantitative factors such as physical activity. Also, walkability and walking behaviour have been theorised to be highly location-specific (Millington et al., 2009). It is important to distinguish between walkability at different spatial levels (at link or at network/neighbourhood level), for different purposes (transport or recreation) and using different indicators (objective/built environment or perceived/self-reported). Many walkability assessment tools have been developed, some more extensive than others. This section will review a number of these tools (listed in table 2.1); this will inform the factors included in the subjective (sub-question 1) and in the objective walkability surveys (sub-question 2). It should be noted that the difference between objective and perceived walkability can be relatively small here (Saelens & Handy, 2008); for every factor included in the objective analysis, the survey participants can be asked to give their opinion on that factor. As an example, consider a study wherein the researcher records the presence and quality of street lighting; this might include technical details about the lights and their placement. A user assessment on the same topic might ask the same questions or wholly different ones, depending on the research focus and on the amount of time assessments are expected to take.

One of the main types of walkability survey tools is known as *built environment assessment*, which involve *walking audits* performed by researchers (Saelens & Handy, 2008). A check-list or form is filled in per link segment, mainly concerning the presence or absence of certain built environment elements, such as a buffer between pavement and road surface . Further questions may involve rating factors on a Likert scale (e.g. 1-5); in that case some training may be necessary to ensure these ratings are sufficiently differentiable (Pikora et al., 2003). Much of the literature concerning these tools focusses on the reliability of these measures, i.e. the variation in ratings given by different researchers on the same road segments (Ma & Dill, 2015). To produce a walkability index directly from walking audit scores, the scores on all items can be weighed using Multi-Criteria Analysis (Pikora et al., 2003) - but this is of course a fairly crude method. A more common way of linking the built environment characteristics to walking behaviour is by correlation with physical activity, e.g. minutes walked per day, which is self-reported by study participants in the area (Saelens & Handy, 2008).

Many walking audit surveys are derived from the conceptual model developed in (Pikora et al., 2003), which divides the walkability factors into *functional*, *safety*, *aesthetic* and *destination* classes. In a companion paper, (Pikora et al., 2002) developed the seminal SPACES tool for built environment assessment which was then linked to physical activity. This has lead to a large growth in other, similar tools, of which the South African PEAT and the East Asian EAST are of the most direct interest to this study. The big advantage of walking audits is that all the data is readily available; it just requires researchers to go out and collect it. However, extensive audits can be somewhat time-consuming and methods have been sought to find more efficient ways of collecting built environment data (McMillan et al., 2010).

It was realised that physical characteristics of the built environment do not necessarily correspond to the general public's perception of walkability, causing "mixed findings from travel behaviour/built environment studies" (Ma & Dill, 2015). However, the response burden for a walking audit is generally far too high to be performed by study participants (see (Zandieh et al., 2017) for an exception). The solution is to present study participants with a survey that asks them to answer the same type of questions, but for their neighbourhood as a whole rather than per street. One of the most commonly used tools of this type is NEWS and its abbreviated version NEWS-A (Jack & McCormack, 2014). A version specific to African cities, NEWS-Africa, was recently developed and the subject of two papers (Oyeyemi et al., 2016a; Oyeyemi et al., 2016b) which also included data from Uganda.

A different approach to walkability was spearheaded by (Frank et al., 2005), who advocated a more macro-scale approach. In this type of study, which leans heavily on GIS data analysis, land use data and travel times are considered the main determinants for walkability. Frank's Walkability index is calculated for an area as follows (where z indicates
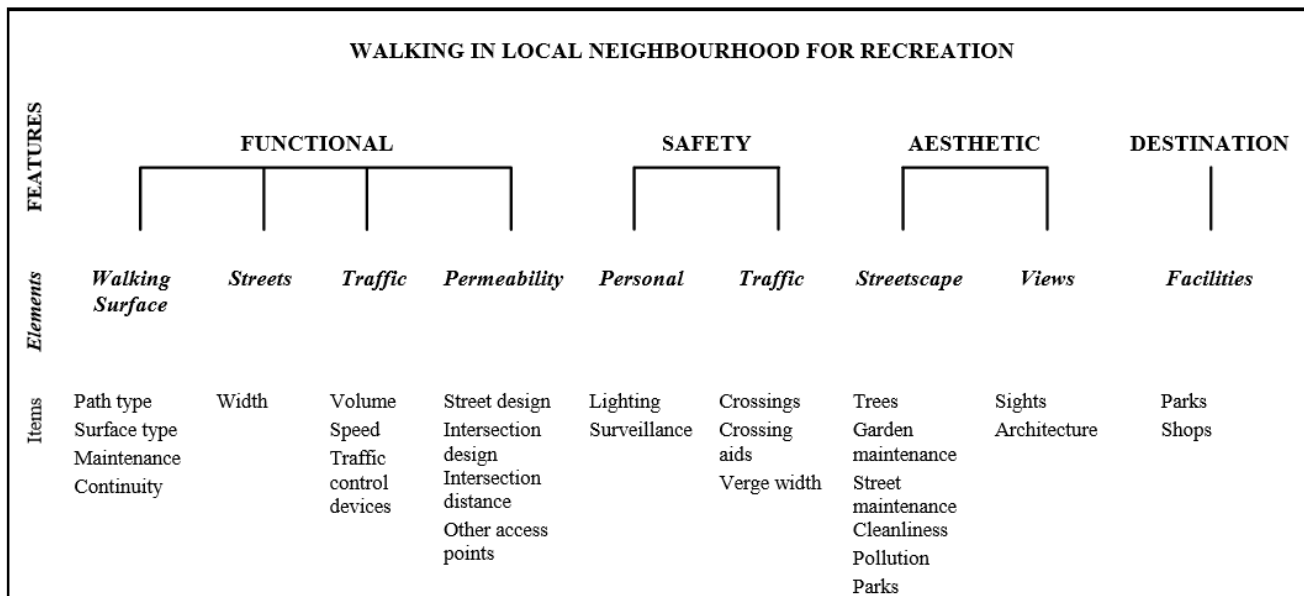
WALKING IN LOCAL NEIGHBOURHOOD FOR RECREATION

FEATURES									
	FUNCTIONAL				**SAFETY**		**AESTHETIC**		**DESTINATION**
Elements	*Walking Surface*	*Streets*	*Traffic*	*Permeability*	*Personal*	*Traffic*	*Streetscape*	*Views*	*Facilities*
Items	Path type Surface type Maintenance Continuity	Width	Volume Speed Traffic control devices	Street design Intersection design Intersection distance Other access points	Lighting Surveillance	Crossings Crossing aids Verge width	Trees Garden maintenance Street maintenance Cleanliness Pollution Parks	Sights Architecture	Parks Shops

Figure 2.1. Conceptual model of walking factors and the main elements. Source: Pikora et al. (2003).

z-scores within a study area):

$$Walkability = (z - \text{net residential density})$$
$$+ (z - \text{retail floor area})$$
$$+ 2(z - \text{intersection density})$$
$$+ (z - \text{land use mix})$$

Another similar tool is called *Pedshed* (see e.g. Manaugh & El-Geneidy, 2011); there walkability is defined from a single location, using a travel isochrone as follows:

$$Walkability = \frac{\text{Area reachable on foot within e.g. 5 min}}{\text{Area of a 5 min (as the crow flies) buffer}}$$

Pedshed is effectively a measure of the *porosity* of the neighbourhood, but includes no information on possible destinations within the area. The development of the (proprietary) Walk Score measure (also explained in Manaugh & El-Geneidy, 2011) filled this gap; it is basically an accessibility metric as it weighs destinations in an area around a location with the travel time on foot to those destinations using a (gravity) decay function. The advantage of macro-scale walkability studies is that they require relatively little time to complete if all the data is already available from GIS databases. However, this is likely not to be the case in developing countries, especially in East Africa where GIS data coverage is fairly poor. It should be noted that macro-scale walkability tools, especially those based on travel isochrones, have been criticised for failing to capture the pedestrian experience: "a mile in an American suburb is a lot longer than a mile in Rome."[1]

The final, most sophisticated type of walking behaviour investigation tool is *route choice modelling* using discrete choice models. While this type of choice models has existed for a long time, their application to pedestrian route choice has long been limited due to limited data availability and high cost of collection. Now, with the widespread availability of GPS tracking devices, these models can be applied fairly easily. Some recent studies of pedestrian and cyclist behaviour using route choice modelling include (Hintaran, 2016; Lue, 2017; Zimmermann, Mai & Frejinger, 2017). Route choice models allow for a much more fine-grained analysis of walkability than the previously mentioned survey types, which often rely on self-reported physical activity to show that their walkability metrics are actually related to walking behaviour. On the other hand, a choice model will show the trade-offs pedestrians make to get to their destination: do they choose the shortest route or the detour through a "more walkable" one?

A route choice model basically comes down to a function that assigns probabilities to different possible routes between the same origin and destination based on characteristics of that route and on a stochastic element of uncertainty. Since these probabilities are, in general, not observable, the choice probability function is fitted based on observed route choices taken by the study participants. Note that the set of all possible routes is exponential in the number of links and often intractably large; irrelevant alternatives such as improbably large detours may be pruned through the use of a route set generation algorithm. Assuming a certain form for the stochastic element of the choice function (usually

---

[1] https://www.citylab.com/transportation/2012/08/walk-score-great-it-still-doesnt-capture-walk-appeal/2858/

multinomial logit), the probability of taking a certain route is dependent on its **utility** (or, equivalently, its generalised cost) relative to that of the alternatives. The utility of a route is assumed to be linearly dependent on the route attributes through (unknown) parameters β. The values of these parameters can be estimated by fitting the choice model such that the total set of observed GPS traces has the maximum likelihood of occurring; i.o.w. the utility of observed trips should be high relative to the alternatives. The fitted values then indicate the contribution of the route characteristics of the used links to its utility. As an example, (Lue, 2017) found in a Toronto case study that every meter of length of a route contributed -0.0198 to its utility; this was offset, however, by sidewalk presence along the road ($+0.0073$) and signalised intersections ($+0.729$). This could be interpreted as follows: study participants preferred shorter routes, but perceive links with sidewalks are 33% shorter in length and they would on average walk 32 meters extra to get to a signalised intersection.

A major weakness in the route choice analysis used by the previously cited studies is the route set generation that needs to be used to generate plausible routes over the network in the study area. The observed GPS traces are often not generated (e.g. Hintaran reached 67% coverage, Lue only 55%) and need to be included manually, which raises questions what other plausible routes are not included in the choice model. A recent methodological advance in route choice modelling is the Recursive Logit model (Fosgerau et al., 2013), which allows for modelling route choices without pre-determining a restricted choice set of paths. Rather than assigning probabilities to entire routes at a time, routes are built up using sequential link choices. The probability of picking the next link is dependent on its (`instantaneous`) utility and on the `expected downstream` utility. Rewriting this relationship in the form of Bellman equations, the value function for each link choice can be calculated through dynamic programming, i.e. starting at the destination and working backwards (or equivalently by solving a set of linear equations). Then, the parameters should only be fitted such that the observed route choices have maximum likelihood compared to the alternative choices *at each link*, rather than in comparison to every other route on the network. This does mean that link properties should be link-additive for them to be able to be included in the fitting process; otherwise the value functions cannot be written as Bellman equations. Some non-additive properties can be made so by introducing dummy variables for them that interact with e.g. link length (Zimmermann et al., 2017). While restricting a route choice model to only using link-additive properties may lower its predictive power, it does make the results more easy to interpret; utilities can be expressed at the link level (rather than only for complete routes), which means that they can be mapped. Further, this leads to a direct formulation of accessibility between OD pairs and can be used to estimate traffic flows (Zimmermann et al., 2017).

Survey tool name	Reference	Approach	Nr. of factors	Location
SPACES	(Pikora et al., 2002)	Walking audit	71	Australia (Perth)
SWAT	(Millington et al., 2009)	Walking audit	113	Scotland (Glasgow)
PEAT	(Albers et al., 2010)	Walking audit	38	South Africa (Tswane)
EAST-HK	(Cerin et al., 2011)	Walking audit	91	China (Hong Kong)
Clean Air Asia Walkability	(Asia, 2011)	Walking audit	9	11 cities in East Asia
IMI	(Day et al., 2006)	Walking audit	162	United States (Minnesota)
BiWET	(Hoedl et al., 2010)	Cycling audit	15	Austria (Graz)
MAPS-mini	(Cain et al., 2015)	Walking audit	15	United States (San Diego)
MAPS Global	(Vanwolleghem et al., 2016)	Walking audit	119	Belgium
NEWS-Africa	(Oyeyemi et al., 2016a)	Questionnaire	76	6 cities in Africa
NEWS-A	(Jack & McCormack, 2014)	Questionnaire	54	
Perceived bikeability	(Ma & Dill, 2015)	Questionnaire	5	United States (Portland)
GWI	(Krambeck, 2006)	Questionnaire	45	India (Ahmadebad)
Frank's Walkability Index	(Frank et al., 2005)	GIS	4	
Walk Score	(Manaugh & El-Geneidy, 2011)	Travel distances	290	Canada (Montreal)
Pedshed	(Manaugh & El-Geneidy, 2011)	Travel distances		Canada (Montreal)
Lue's route choice model	(Lue, 2017)	GPS survey	18	Canada (Toronto)
Hintaran's route choice model	(Hintaran, 2016)	GPS survey	5	Switzerland (Zürich)
Bike route choice model	(Zimmermann et al., 2017)	GPS survey	16	United States (Eugene)

Table 2.1. Walkability assessment tools used to design the methodology of this study.

CHAPTER 3

PERCEIVED WALKABILITY ASSESSMENTS

As part of this research project, the Walkability smartphone app has been redeveloped for use in Kampala. The original app was developed by Banele Wasswa for his Master's Thesis at the University of Cape Town, together with Indian software development company KritiSolutions (Wasswa, 2016). The new version has been changed to be somewhat more suitable to Ugandan conditions as well as easier to use. A full guide to use, deployment and inner workings of the app can be found in appendix A. In this chapter, some aspects of the app (re)design and its use in procuring data for this study will be discussed.

Basic flow of the application

Screenshots of the app in use are shown in figures 3.1 through 3.3 in the rough order in that the user might encounter them. Briefly, the goal of the app is twofold: to collect GPS tracks of walking routes, and to collect information about the perceived walkability of links on the walking network. Users are asked to turn on the app whenever they go out walking. To ensure accurate GPS tracking, the phone must first acquire a lock on GPS satellites. As the user moves with their phone, GPS location updates will come in, together with an indication of their accuracy. Only points with a maximum deviation of 10 meters are accepted. The result is usually a very dense and straight path, without jumps or jitter - see e.g. figure 3.4. These routes will later be aligned to the street network.

On a regular interval, normally every 150 meters, the app user will be asked to give their opinions on their surroundings. This is done through a rating system, where 1 to 5 stars can be given on a total of six categories. The procedure for this is shown in figure 3.2. A question-mark button in the app shows the user more detailed explanations of each category, so they can know what to look for. The six questions and their corresponding explanations are shown in table 3.1 below. For a more detailed look at the application, please see Appendix A.

Changes made to the tool for this research

By the end of the redevelopment process, only the basic flow of the application remained the same; nearly every single line of Java code has been rewritten, as have the great majority of texts explaining the app to the user. The overriding motivation behind changes made was to reduce the complexity and thus the response burden on the users. There are now options for the user to configure the app and the various functionalities are easier to access. Further, the app has been made more robust: it is now usable without internet connection and in GPS-adverse conditions. Finally, the tool is now compatible with new versions of Android and a number of bugs and crashes have been resolved.

As a result of the rewriting effort, the Java code size has been reduced from nearly 6000 lines spread over 46 files to a more manageable 1300, in 13 files. Furthermore, there are now only 19 XML source files instead of 30, reducing the code from over 2600 lines to less than half of that. These reductions have been achieved primarily through rewriting the code to a more condensed style, taking advantage of inbuilt features and by removal of certain functionalities. Notably, the app no longer uses a database (rather, all data is written to CSV files) and a large number of internet-related

Category	Hint
Footpath	*The footpath is separated from the road, free of obstructions and of good quality.*
Walking comfort	*There are no parked cars, pools of water or other things blocking my way.*
Security	*I do not need to be afraid of other people or animals here and it is safe even at night.*
Accident safety	*I am not in danger from traffic here and can easily cross the street.*
Pleasantness	*I enjoy walking here, the environment is clean and pleasant.*
Liveliness	*There are many shops, facilities, places to go or things to do here.*

Table 3.1. Categories on which the participants are asked to rate their environment, with further explanations.
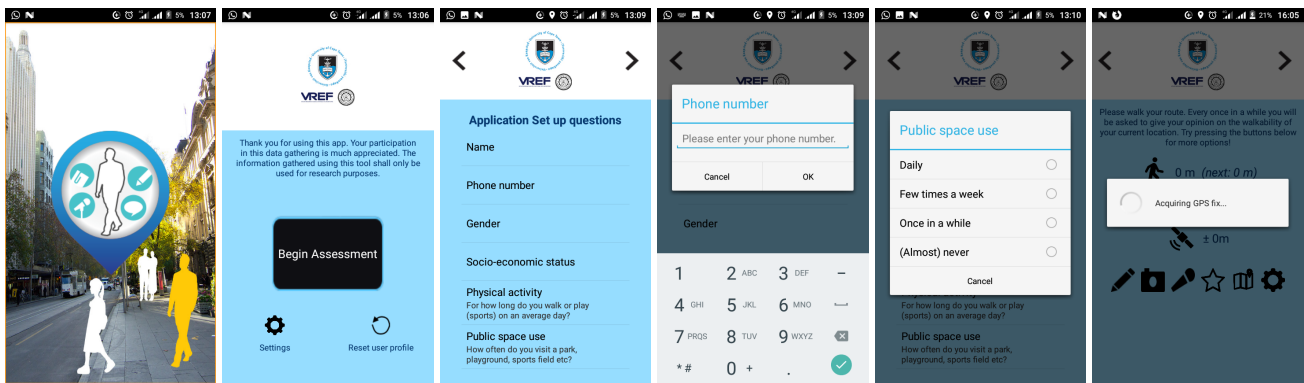
Figure 3.1. First screens of the app: welcome screen, initial questions and GPS fix dialog.
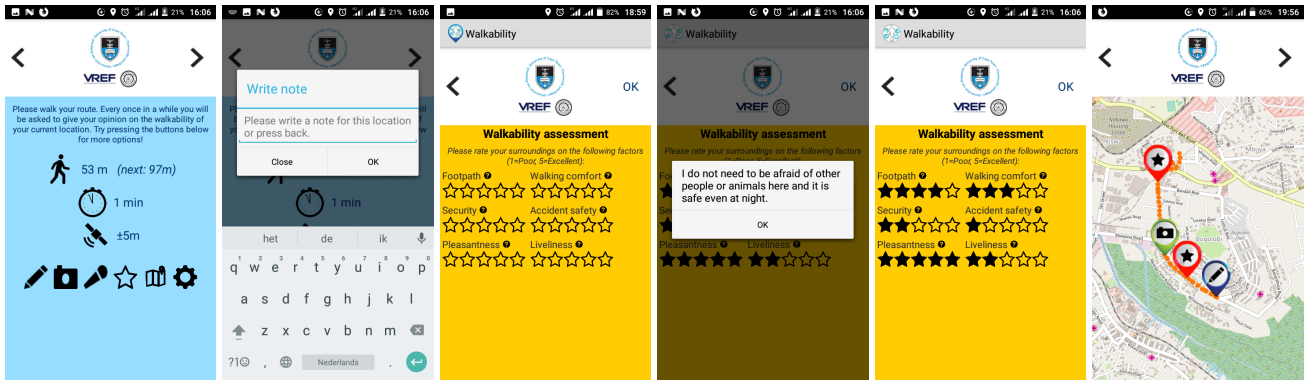

Figure 3.2. Screens shown during the assessment: the survey screen itself, the rating screen with explanations, the survey map and note taking facility (camera and voice note are similar)
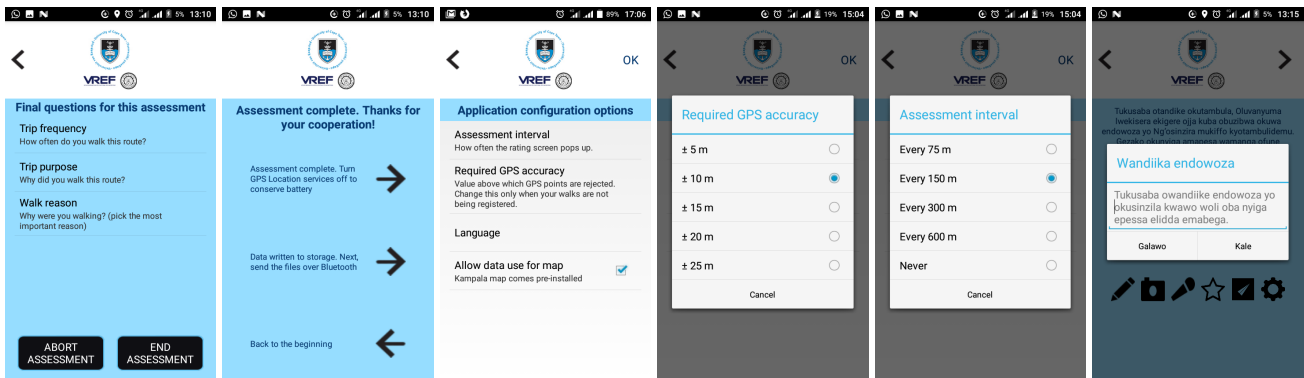

Figure 3.3. Other screens: first the survey ending screens and then the configuration options, with an example of a translated app screen (Luganda).
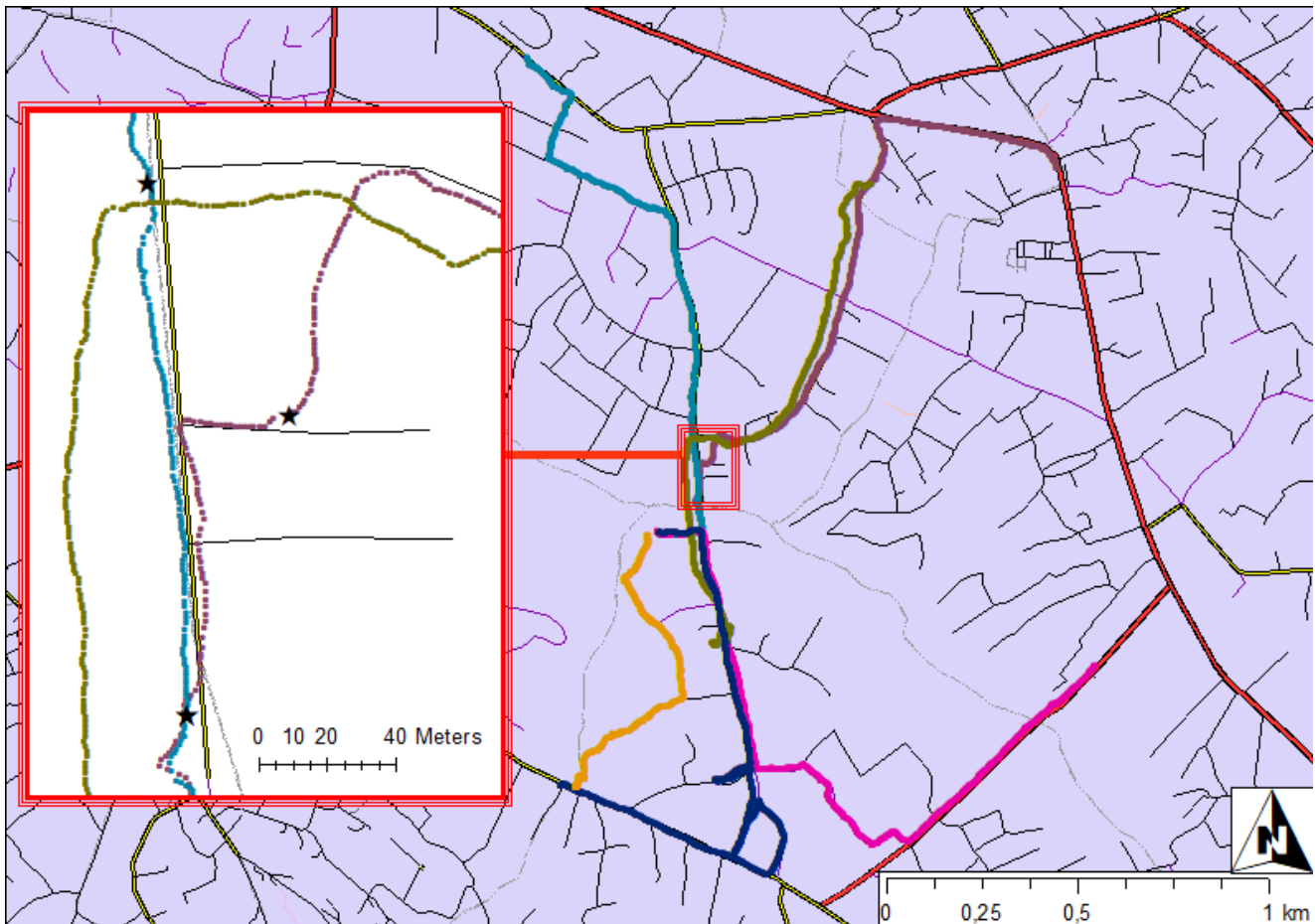
Figure 3.4. December 22 pilot survey in Lukuli-Nanganda with six participants. Routes in bold lines / small dots. In the inset, stars indicate locations where participants paused to give ratings and/or add media.

functionalities were removed. Previously, assessments were uploaded to a website run by the University of Cape Town but it was cumbersome to access, not secured and required making accounts for every participant. In future research, a new way to send assessments over the internet should be added in.

Pilot Surveys

A total of five pilot surveys were conducted to test the app in a real setting using untrained participants. Four of these were conducted in Kampala district Lukuli-Nanganda, where Advocates for Public Spaces is running a project on improvement of the local park (a popular sports and social gathering place). Teams of around 5 people were collected; in some cases they used their own smartphones and in others they borrowed phones from the researcher. A number of important conclusions were drawn from these pilots, which guided changes to the app. The wording of the questions was changed a number of times so as to align better with local terminology and the 'question mark' buttons with extra explanations were added. Also, the colour scheme of the app was changed to dark text on a light background, which is easier to read in the sun and on scratched phone screens. An interesting use pattern of the app was noticed: the participants are enthusiastic about the voice note and picture functionality, which they use to supplement the rating screen (rather than as an alternative). Contrary to expectations, the participants did give relatively varied ratings (rather than only one or only five) and were perfectly willing to give low scores to their own neighbourhood, even for safety indicators. They compared their experience in Lukuli to that in other neighbourhoods to back up these assessments; especially bemoaning the dusty nature of their (gravel) road, the frequent flooding and the lack of street-lights. On the other hand, traffic safety is seen as good and accidents are relatively rare, since speeds are low on the unpaved main road and most residential streets are not accessible by car.

Data collection & pre-processing

Data for the perceived walkability part of this study was collected in Kampala during the months of January, February and March using the previously described smartphone application. In total, 59 persons participated in the data collection, each of which used the app daily for a period of at least one week. The participants were mobilised by local

community leaders in different areas of the city. A training session would then be scheduled in which the participants met the researcher and the purpose and methods of the study were explained. All participants had to explicitly consent to being tracked with their phone. All those who successfully produced at least 10 routes were given, as a token of gratitude, a modest sum of 20,000 Uganda shillings (or about 5 Euro). The community leader who mobilised the app users was also given a sum of 50,000/= for their cooperation.

Seven separate groups participated, hailing from the following neighbourhoods of Kampala:*Lukuli*, *Nsambya*, *Kitintale*, *Namuwongo*, *Kamwokya*, *Namungoona* and *Kisenyi*. Groups from *Katwe* and *Wanukulukuku* were also invited but failed to produce satisfactory data sets. An overview of the study locations and the routes walked by participants is shown in figure 3.6.

The participants were instructed to use the app every day to record their regular walking patterns. This resulted in a total of 776 routes being recorded. Some pre-processing was done on these routes to exclude non-walking trips, merge incorrectly divided routes and split routes with (obvious) stopping points. This brought the total number of routes down to 517.
In addition to having their walking routes tracked, the participants were also asked to rate their surroundings on a regular basis. A total of 1652 locations were rated; in the vast majority of cases all six questions were answered.

The results of the perceived walkability assessments are shown in figure 3.7. For each of the six categories, a map is shown of ratings snapped to the closest links. In case of of multiple participants rating the same links, averages are shown. It is clear that the city centre (middle-north in the maps) consistently receives better ratings than the rest of the study area. This is hardly surprising given the great difference between infrastructure in that area compared to the rest of the city (an impression is shown in figure 3.5). It is also easy to see that ratings are lowest in the slum of Lukuli, especially along its western edge bordering Nsambya. It would seem that thoroughfares get higher ratings, with the exception of Lukuli road (west/east near the bottom right), which indeed is a dangerous and steep road, avoided by pedestrians.

Figure 3.5. Stark inequality: rural-like conditions with skyscrapers in the background. Picture taken looking from Nsambya (middle of the study area) to the city centre (top of the study area).
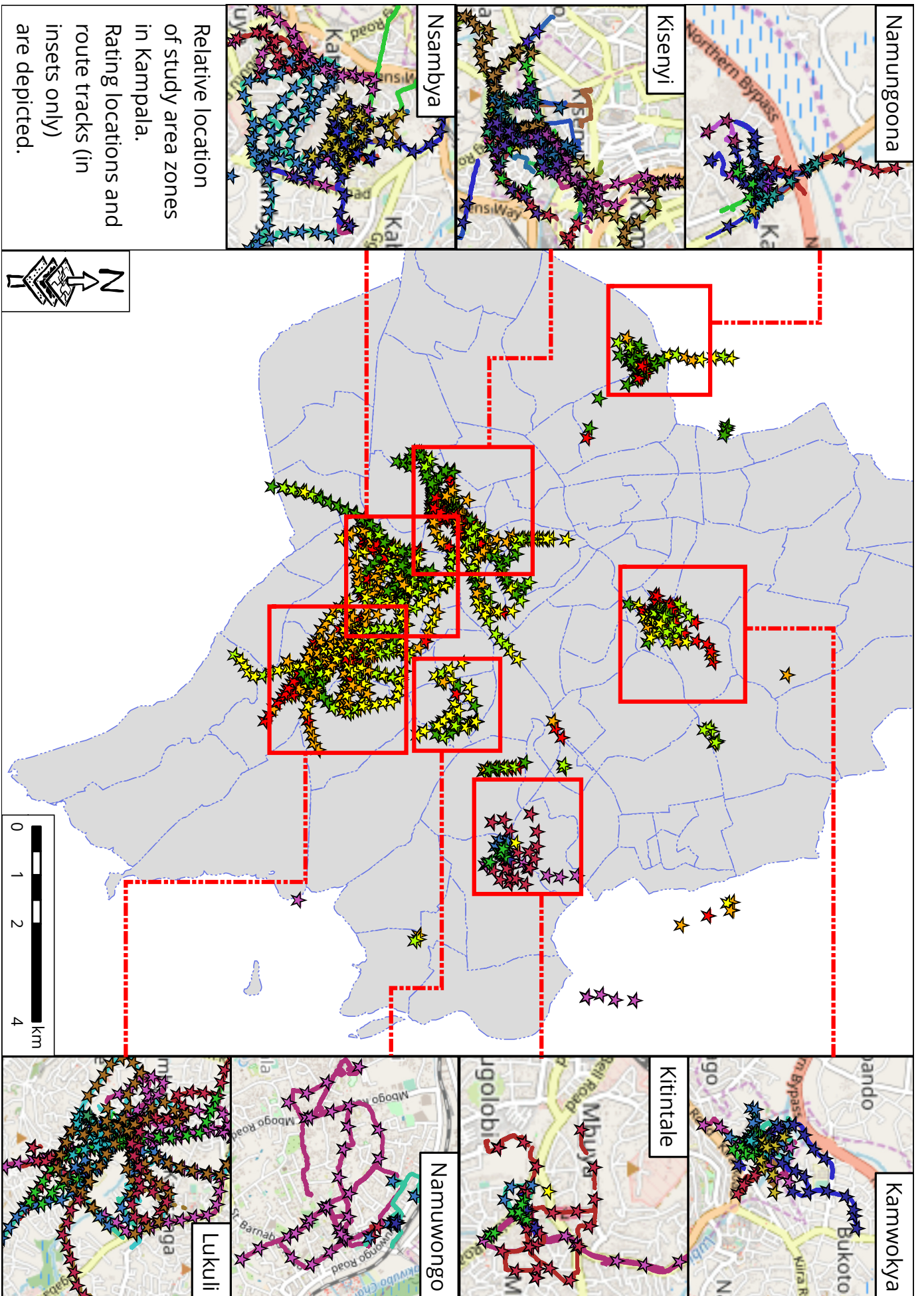
Figure 3.6. Overview of data collected. Stars indicate locations where ratings were given while solid lines show tracked walking routes. Note: rating stars have been given different colours to make them easier to distinguish, not to convey any meaning.
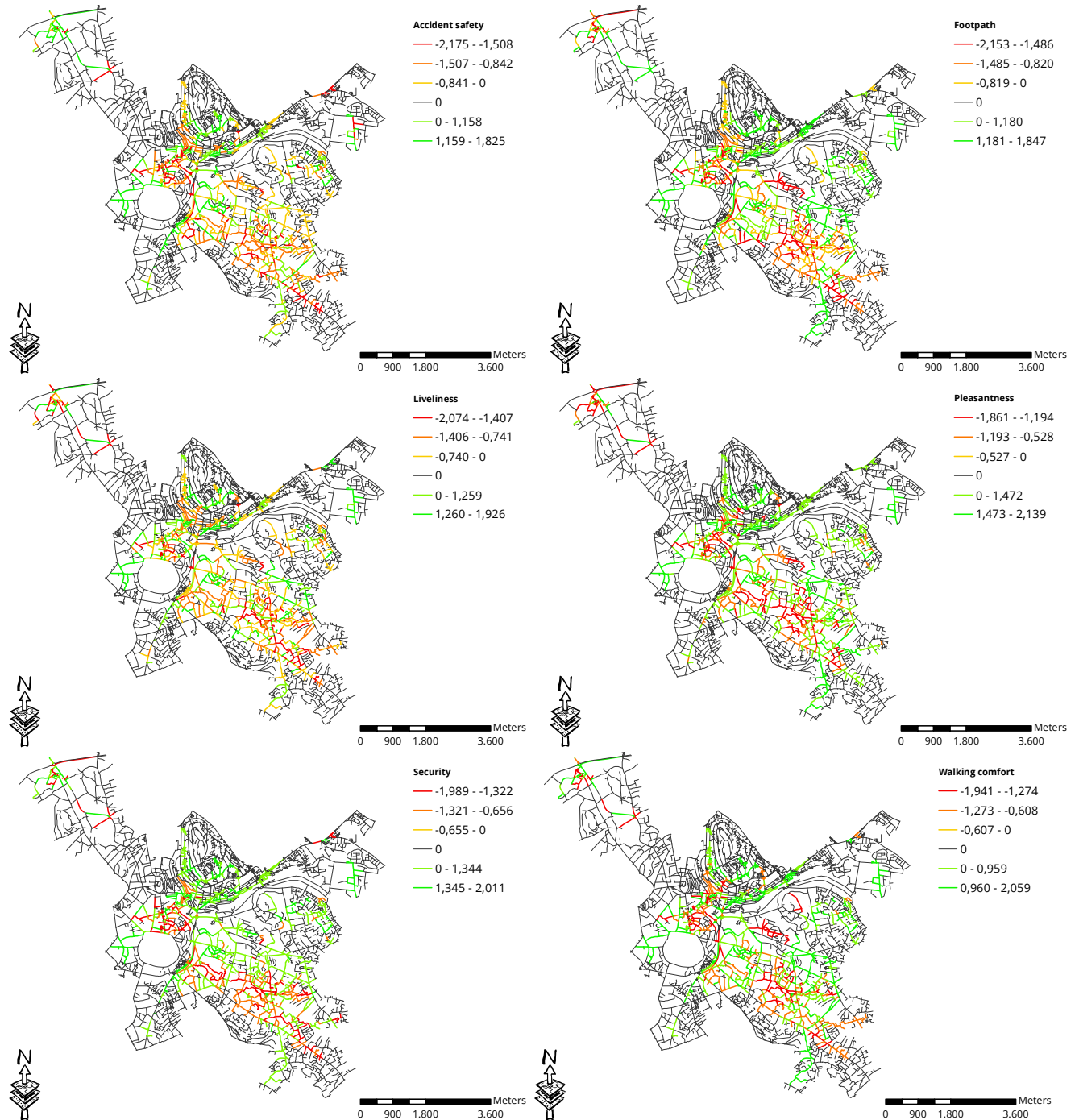
Figure 3.7. User ratings as edge attributes. Links were assigned the closest rating value (mean-shifted), from any user.

CHAPTER 4

BUILT ENVIRONMENT INVENTORY

In addition to the assessments made by users of the Walkability app, a second more 'objective' data collection was carried out to create a more complete picture of the situation on Kampala streets. This was done in the form of a built environment inventory, or in other words a street survey in which a questionnaire was filled in on different locations in the city. We write 'objective' here in quotes as there is certainly still a degree of subjectivity involved in a built environment inventory; the term objective is used to contrast with the user assessment survey which is explicitly about user perceptions. In any case, the inventory consisted of a total of 36 questions in 8 categories, whereby some questions allowed multiple answers. The questions were answered in 137 locations spread around Kampala, in an attempt to cover as many as possible of the streets where participants walked, or could have walked - see the map in figure 4.1.

The full questionnaire that was used for the inventory is shown in table 4.1. Most of the questions are answered on a yes/no basis (absence/presence), while others are measured on a continuous scale. There are also questions that have a scale from, e.g., 0-4; this should not be understood as a Likert scale but rather as a factor with different levels: for example, the *Dedicated side-walk* factor was a choice between *Walk on road*, *Walk in berm*, *Intermittent side-walks*, *Full-length side-walks* and *Full-length side-walks with crossing aids*. The full set of questions with meanings for all the levels is reproduced in appendix C in its original format. Some examples of the difference between different levels on a single item are shown in figures 4.2 through 4.4 on the items *Litter*, *Dedicated side-walk* and *Parked vehicle obstructions*.

Literature on built environment studies is strongly focused on developed nation contexts and there was thus no questionnaire for link-level built environment surveys available. Therefore, a bespoke survey tool was created. The tool was based strongly on the Environment in Asia Scan Tool–Hong Kong (or EAST-HK; see Cerin et al., 2011), which itself mostly follows the methodology outlined in the foundational SPACES tool (Pikora et al., 2002). In SPACES and tools inspired by it, nine key aspects for a pleasant walking environment are identified: *Walking surface*, *Street*, *Traffic*, *Permeability*, *Personal safety*, *Traffic safety*, *Streetscape*, *Views* and *Facilities (Destinations)*. The tools include several indicators to try and measure the built environment on these categories. As an example, refer back to figure 2.1 which shows the questionnaire items used in SPACES.

Although EAST-HK can be used in a link-level built environment inventory, not all of the questions are relevant to the Kampala context and some aspects were missing. Therefore, the EAST-HK questionnaire was compared to another tool, *NEWS-Africa* (Oyeyemi et al., 2016b) - which is a neighbourhood level built environment survey - to determine key aspects and for the addition of more relevant questions.

The built environment inventory was conducted using a second smartphone application, which was also developed during the study period. This app, called *GPStionnaire*, allows the user to conduct geo-referenced surveys and the results can then easily be loaded into GIS software. An introductory guide for use and deployment of this app is located in appendix B.

Item	Range	Item	Range
Walking surface		**Streetscape**	
Dedicated side-walk	0-4	Litter	0-3
Side-walk pavement	0-4	Presence of trees or greenery	0-1
Sidewalk continuity	0-4	Air pollution; smells	0-1
Path width	meter	Noises	0-1
Drainage channel presence	0-1	Overhead electric wiring	0-1
Drainage channel covering	0-2	**Permeability**	
Slope	0-3	Dead end on either side of segment	0-2
Permanent path obstructions	0-2	Three-way intersection on either side	0-2
Street shop obstructions	0-2	Four-way intersection on either side	0-2
Parked vehicle obstructions	0-2	Roundabout on either side	0-2
Crowdedness	0-2	**Traffic safety**	
Street width	meter	Traffic speed	km/h
Road surface condition	0-4	Traffic island or other crossing aid	0-1
Personal safety		Matatu route	0-1
Liveliness	0-3	Boda boda route	0-1
Presence of children	0-1	Speed bumps	0-1
Beggars or homeless people	0-1	**Views**	
Police or armed guards	0-1	Abandoned buildings/construction works	0-1
Street lighting	0-1	Advertisements/signage	0-1
Stray animals	0-1	Building height	stories
Destinations		**Buildings**	
Boda boda stage	0-1	Office buildings	0-1
Matatu stops	0-1	Detached dwellings (compounds)	0-1
Street market or kiosks	0-1	Single-story dwellings	0-1
Strip mall	0-1	Multi-story dwellings mixed with shops	0-1
Other shop buildings	0-1	Multi-story dwellings without commercial	0-1
School or church	0-1	Apartment block	0-1
Field, park or court	0-1		
Water tap point	0-1		

Table 4.1. Built environment assessment tool, containing 60 items.
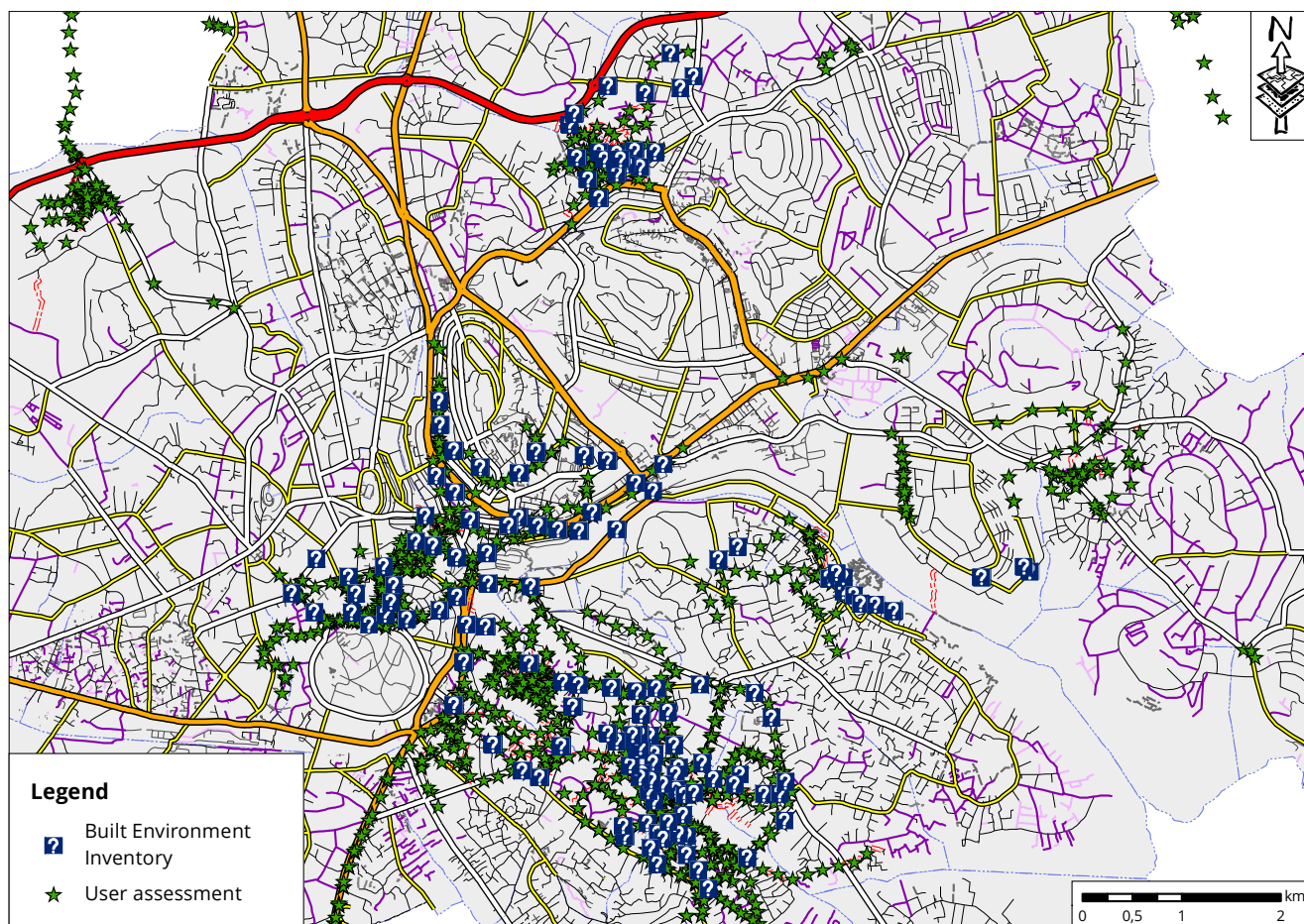
Figure 4.1. Comparison of locations where the Built Environment questionnaire was filled in and where Walkability app users rated street segments.

Figure 4.2. Examples of built environment item levels. Here, the four different levels of **Litter**: *None*, *Low (scattered and avoidable)*, *Medium (piles of garbage)* and *High (unavoidable heaps of garbage)*.

Figure 4.3. Examples of built environment item levels. Here, the five different levels of **Sidewalk presence**: from *No sidewalks*, *Berm only*, *Sidewalk one side*, *Sidewalk both sides* to *Not car accessible*.

Figure 4.4. Examples of built environment item levels. Here, the three levels of **Parked vehicles blocking the path**: *No obstacles*, *Some obstacles (avoidable)* and *Unavoidable obstacles*.

RATING ANALYSIS

Over the course of this research, data was collected from a number of different sources. The different datasets have varying degrees of completeness and reliability and it is thus considered important to do some checking on whether there are underlying correlation structures that can help validate the results of the different studies. The datasets are as follows: from the *Walkability app* survey, there are **ratings** given by users of the app to their environment, **routes** they took (with some characteristics) and **personal information**. From the *built environment inventory* there are survey questions filled in at specific locations, mostly close to locations where Walkability ratings were done. Lastly, from OpenStreetMap there is information about the road features; of particular interest are the *road classifications*.

The choice was made to do regression on the user ratings with the other datasets as explanatory variables. The idea behind this is to assess whether differences in ratings are mostly brought about by differences in individuals doing the scoring, or whether differences in the environment are the most important. It is also of interest to see whether there are strong linkages between some of the subjective (user) assessment categories and the objective built environment scores, which might reveal something about what users find important. Lastly, we wish to see how strong the OSM classification correlates with the user ratings. If it has strong relations, then that may mean that this readily available - and quite comprehensive - dataset can also be of use.

5.1   Factor / component analysis

To get an initial idea of the relationships between user assessment scores, a simple principal component and factor analysis has been carried out using the standard tools in R. We should be careful not to read too much into this, as this type of analysis is dependent on assumptions of normality which may not hold on the 1-5 Likert scale on which the variables are ranked. In any case, principal component analysis (PCA) suggests that two factors may be enough to explain most of the variance in the dataset. A factor analysis (FA) aligning the six rating categories to two factors reveal that Accident safety is strongly viewed as its own category, whereas the rest of the walkability factors are often given closely related scores. These two factors explain a total of $62,7\%$ of the variance in the dataset; again under (possibly flawed) normality assumptions.

PCA	Standard deviation	Proportion of variance	Cumulative proportion	FA	Loading 1	Loading 2
Comp 1	1,855	0,574	0,574	Accident safety		1,019
Comp 2	0,862	0,124	0,698	Footpath	0,779	
Comp 3	0,774	0,100	0,797	Liveliness	0,577	
Comp 4	0,709	0,084	0,881	Pleasantness	0,853	-0,130
Comp 5	0,627	0,066	0,947	Security	0,634	0,109
Comp 6	0,565	0,053	1,000	Walking comfort	0,786	

Table 5.1. Results of Principal Component Analysis (left) and Factor Analysis (2 factors, right) on the user assessment scores.

5.2   Ordinal regression

Unfortunately, the variables from the other datasets cannot simply be used in a PCA analysis. Instead an ordinal regression model should be used; in a choice has been made to use the proportional-odds logit model (McCullagh, 1980). Standard software packages such as R have routines for use of this model and it is one of the most common ways of dealing with levelled data.

We shall briefly go into the specifics of the proportional odds logit model using McCullagh's notation. If Y is our response variable, in this case ratings on one of the six categories, and we denote $\pi_1, \ldots, \pi_5$ the probabilities of each of the five levels being chosen, then of course $P(Y \leq j) = \pi_1 + \cdots + \pi_j$. Now we define the cumulative logits: $\lambda_j = \log(\frac{P(Y \leq j)}{P(Y > j)}) = \log(\frac{\pi_1 + \cdots + \pi_j}{\pi_{j+1} + \cdots + \pi_5})$. Now we perform regression on these logits by adding our explanatory variables. We assume that the coefficients are the same in each of the equations - this is the proportional odds assumption that gives the model its name:

$$\lambda_1 = \theta_1 + \beta_1 X_1 + \cdots + \beta_n X_n = \log(\frac{\pi_1}{\pi_2 + \pi_3 + \pi_4 + \pi_5})$$

$$\lambda_2 = \theta_2 + \beta_1 X_1 + \cdots + \beta_n X_n = \log(\frac{\pi_1 + \pi_2}{\pi_3 + \pi_4 + \pi_5})$$

$$\lambda_3 = \theta_3 + \beta_1 X_1 + \cdots + \beta_n X_n = \log(\frac{\pi_1 + \pi_2 + \pi_3}{\pi_4 + \pi_5})$$

$$\lambda_4 = \theta_4 + \beta_1 X_1 + \cdots + \beta_n X_n = \log(\frac{\pi_1 + \pi_2 + \pi_3 + \pi_4}{\pi_5})$$

Maximum likelihood then gives values for the parameters θ and β. Now, most of the explanatory variables provided as an input are themselves also categorical in nature. R will automatically define indicator variables for contrasts, such that the effect of a certain factor level on the expected score can be compared to a baseline - that means to say, for a factor B with e.g. levels $no\ data$, low, $medium$ and $good$, R will create indicator variables $X_{B=low}$, $X_{B=medium}$, $X_{B=high}$ and estimate three related coefficients β. These coefficients for the effect of the contrast do not relate directly to the assessment scores but rather to odds ratios: a coefficient of x indicates that presence of this factor level will increase the odds of a higher score by e^x. We are interested mostly in sign and relative magnitude of the coefficients.

What we can see from the results (table 5.2) is that the ratings seem to be strongly correlated to certain values of the personal characteristics, e.g. people with Other occupation give quite low scores in a lot of categories. People from Kamwokya are, on the other hand, quite positive and give high scores in most categories; people from Nsambya and Katwe are positive on some categories but not on others. On security, Namuwongo scores highly whereas Namungoona receives poor ratings.
The liveliness category seems not to relate to many aspects, not even to the liveliness category from the Built environment inventory. This may indicate that Kampala residents have a different view on liveliness; or they might have understood the question in many different ways.
An interesting correlation is that between security and litter. Whereas street light presence had no significant impact, cleanliness is apparently associated with more secure streets. Walking comfort is higher in places that are not too crowded.

Accident safety

Factor	Level	Coef.	p-value	
Walk reason	No/Other reason	-1.341	0.048	*
Trip purpose	Education	1.674	0.007	**
Occupation	Caretaker	1.865	0.003	**
Occupation	Employed	-1.628	0.001	***
Occupation	Other	-6.565	9e-16	***
Occupation	Student	-2.364	5e-07	***
Physical activity	30 min - 1h	3.095	2e-11	***
Physical activity	Over 2 hours	1.397	6e-05	***
Residence	Katwe	1.318	0.007	**
Residence	Kamwokya	1.623	0.0002	***
Residence	Nsambya	1.385	0.013	*
Roadclass	Footway	3.192	8e-05	***
Roadclass	Service	-1.045	0.035	*
Roadclass	Tertiary	1.105	0.019	*
Speed bumps	Yes	-1.486	5e-06	***

Liveliness

Factor	Level	Coef.	p-value	
Public space use	(Almost) never	1.978	3e-12	***
Public space use	Few times a week	-0.666	0.039	*
Residence	Kamwokya	1.216	2e-08	***
Residence	Nsambya	0.805	6e-04	***
Pres. beggars/homeless	true	0.361	0.007	**
Presence of police/guards	true	0.275	0.040	*

Security

Factor	Level	Coef.	p-value	
Walk reason	Leisure	1.203	0.036	*
Walk reason	Too much traffic	2.563	0.001	**
Walk reason	Transport affordability	1.698	0.020	*
Trip purpose	Education	-1.478	0.020	*
Trip purpose	Personal business	-1.976	0.003	**
Occupation	Caretaker	-2.882	0.002	**
Occupation	Employed	-2.073	5e-05	***
Occupation	Other	-7.372	1e-08	***
Occupation	Student	-1.545	0.002	**
Public space use	Once in a while	-1.136	0.010	**
Public space use	Few times a week	-1.653	0.000	***
Public space use	Daily	-1.157	0.002	**
Gender	Male	0.858	0.010	**
Residence	Kamwokya	1.603	4e-06	***
Residence	Namuwongo	16.128	2e-16	***
Residence	Nsambya	2.154	5e-10	***
Residence	Namungoona	-2.350	0.017	*
Roadclass	Footway	3.623	7e-06	***
Roadclass	Primary	1.470	0.015	*
Roadclass	Tertiary	1.597	0.000	***
Litter	Low	0.684	0.019	*
Litter	None	3.622	4e-10	***

Footpath

Factor	Level	Coef.	p-value	
Trip purpose	Personal business	-1.390	6e-04	***
Occupation	Other	-3.288	2e-06	***
Occupation	Student	-0.803	0.030	*
Physical activity	Over 2 hours	1.101	4e-06	***
Residence	Kamwokya	2.323	4e-11	***
Residence	Namuwongo	2.578	0.001	**
Roadclass	Footway	3.823	4e-05	***
Roadclass	Primary	-1.201	0.018	*
Roadclass	Service	-1.463	0.001	**
Sidewalk	Wide berm	0.575	0.004	**
Sidewalk	Not car-accessible	-1.642	6e-04	***
Path continuity	Full-length	1.982	2e-06	***
Path continuity	Intermittent	2.067	7e-04	***
Traffic islands	yes	-0.935	0.014	*

Pleasantness

Factor	Level	Coef.	p-value	
Drainage channels	Not covered	-0.490	0.047	*
Road surface	Decent condition	0.821	0.029	*
Children presence	yes	-0.395	0.021	*
Pres. police/guards	yes	0.534	0.004	**
Litter	Low	0.601	0.027	*
Litter	None	1.381	0.002	**
Physical activity	30 min-1 hour	1.245	0.000	***
Physical activity	Less than 30 min	1.347	0.031	*
Public space use	Once in a while	-2.169	2e-06	***
Public space use	Few times a week	-1.545	0.002	**
Public space use	Daily	-1.604	1e-04	***
Occupation	Other	-4.388	3e-06	***
Trip purpose	Personal business	1.803	0.004	**
Trip purpose	Work	1.754	0.001	**
Walk reason	Distance is short	-1.366	0.006	**
Walk reason	No/Other reason	-1.429	0.009	**
Walk reason	Public transport not available	-1.773	0.001	**

Walking comfort

Factor	Level	Coef.	p-value	
Occupation	Employed	-0.941	0.012	*
Occupation	Other	-4.570	3e-10	***
Occupation	Student	-1.495	3e-05	***
Physical acitivity	1-2 hours	-1.133	3e-05	***
Physical activity	30 min-1 hour	1.639	5e-11	***
Physical activity	Over 2 hours	0.733	3e-04	***
Roadclass	Primary	-1.621	0.005	**
Road surface	Decent	0.999	0.005	**
Road surface	Good	0.958	0.005	**
Road surface	Excellent	1.443	0.003	**
Crowdedness	Low	1.465	3e-05	***
Crowdedness	Medium	1.124	7e-04	***
Pavement condition	Bad	1.529	0.028	*
Sidewalk	None	-0.683	0.017	*
Sidewalk	Not car-accessible	-1.116	0.015	*

Table 5.2. Results of Ordinal regression between the user assessments and the supporting datasets.

ROUTE CHOICE MODELLING

6.1 RL model specification

The route choice model to be used is the Recursive Logit model (developed by Fosgerau, Frejinger & Karlstrom, 2013). The choice model can be estimated using reasonably standard methods for log-likelihood optimisation. A MATLAB-implementation of a procedure (Mai, Fosgerau & Frejinger, 2015a) for model estimation is publicly available and has been used for this study.

6.1.1 Traditional route choice models

Before explaining the RL model, let us briefly recall how traditional route choice models function. Those models operate on the assumption that there is a finite number of possible routes from an origin O to a destination D. Each of these routes is assigned a value U, which may depend on such link-level characteristics as well as attributes of the entire route (e.g. maximum elevation change) and, in a mixed model, characteristics of the individual who is walking the route. These attributes are often called X and U is usually a linear combination of the attributes, such that $U = \beta X$ and we are interested in the parameter values of β, as these show the relative importance of the different attributes. It is then assumed that the individual will choose one of the routes to walk according to a multinomial probability distribution, often a logit distribution: $P_j = \frac{e^{\frac{1}{\mu}U_j}}{\sum_i e^{\frac{1}{\mu}U_i}}$. The value of the parameters β can then be estimated through e.g. maximum likelihood methods. The underlying assumption behind such models is that route choice can be viewed as utility-maximizing behaviour; albeit with some uncertainty which is controlled by the scale parameter μ.

Two particular issues with traditional route choice models have been noted. The first is that of *overlap*: commonly, routes that follow mostly the same path, thus overlapping in many links, may have very similar attributes. Intuitively, paths that differ only on a small number of links should "share" some of the probability assigned - for example, in the network in the left of figure 6.1 one would assume that the top route have probability $1/2$ while the bottom two routes have probability $1/4$. However, a naive model would give all three routes equal probability ($1/3$). Therefore, a correlation term between the utilities of routes sharing links should be introduced.

The second issue is that of route sampling. A finite number of choices is needed to estimate a multinomial logit distribution, whereas the number of possible routes in a network is exponential in the number of links - thus such problems become intractable. Traditionally this would be solved with a sampling algorithm, which generates a choice set of routes. Ideally this choice set includes the actual chosen routes in the dataset, plus some other plausible options which are sufficiently different. There are a number of methods for sampling, which we shall not go into further; noting only that these sampling techniques often fail to generate the actually chosen routes, which casts doubts on the validity of the generated choice set (what other plausible routes were missed?).
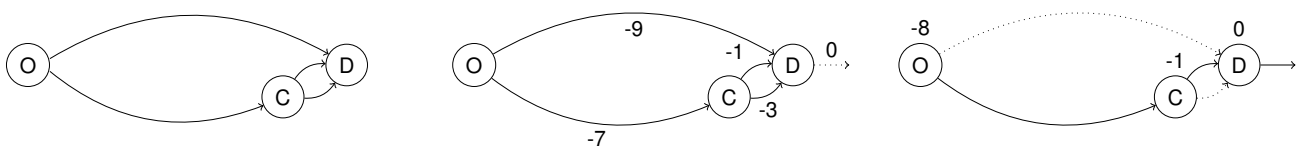

Figure 6.1. Classic three-path example network, adapted from (Mai et al., 2015a). Left: incidence matrix. Middle: instantaneous (link) utilities. Right: expected downstream utilities.

6.1.2   Recursive Logit model

The RL model was developed in (2013) by Fosgerau, Frejinger & Karlstrom specifically to address the second problem explained above; that is, the need for sampling of a choice set of paths. In order to explain the RL model, we first need to introduce some notation; for clarity the notation of Fosgerau et al. will be used where possible.

We consider a **directed** and **connected** graph G with a set of links A (the nodes play no role in the model). Each path is a sequence of consecutive links, ending in a dummy link - for each destination, a dummy link is added to the network. An individual choosing a route through the network is now considered to be making a series of choices: from each link k, they can proceed to any of the neighbouring links $a \in A(k)$. Each option has an *instantaneous* utility

$$u_n(a|k) = v(x_{n,a|k}; \beta) + \mu \epsilon_n(a)$$

which consists of a deterministic and a random component. The instantaneous utility of a choice may depend on the individual (n), on the originating link (k) and the outgoing link (a), but not on the origin. This allows for a wide range of link attributes to be included in the model, even those that are not link-additive or directional (e.g. turn angles, slope). It is important to note that utilities should be strictly negative, except for the dummy links which have utility 0 from all ingoing links.

When the traveller arrives at the node, the random components of all instantaneous utilities are sampled. The utility of a link choice is then the sum of the instantaneous utility and the **expected downstream utility** of that link. The latter can be written in a recursive form (Bellman equation):

$$V_n^d(k) = E \left[\max_{a \in A(k)} \left(u_n(a|k) + V_n^d(a) \right) \right] \forall k \in A$$

These equations can be solved (written in a direct form) through a dynamic programming approach: starting at the destination, which has a known utility of 0, then working backwards to cover the whole network (a very simple example, with only deterministic utility components, is shown in figure 6.1). If we assume, again, a multinomial logit distribution, then the expectation reduces to a logsum and the value function becomes (equation 4 of Fosgerau et al., 2013):

$$V_n^d(k) = \mu \ln \sum_{a \in A(k)} e^{\frac{1}{\mu} v_n^d(a|k) + V_n^d(a)} \forall k \in A$$

Although these value functions are non-linear, they can be transformed and then written as a system of linear equations of the form $(I - M)z = b$ (equations 5-7 of Fosgerau et al., 2013) which can be solved to obtain the utility values $V(k)$. The likelihood of a path $(k_0, k_1, .., k_N)$ being chosen can derived and is equal to

$$P = e^{-\frac{1}{\mu} V(k_0)} \prod_{i=0}^{N} e^{\frac{1}{\mu} v(k_{i+1}|k_i)}$$

In the preceding we have assumed the instantaneous utilities to be known, but this is of course not the case. It is the intent to estimate the utility values *given* the witnessed route choices; we can do this using maximum likelihood estimation. The likelihood function for each path is calculated as above. If we assume the deterministic part of the instantaneous utility to be a linear combination of the link attributes as follows: $v_n(a|k) = \sum_i \beta^i x_{n,a|k}^i$, then the gradient of the log-likelihood function with respect to β can even be derived. This means that the estimation can be done with standard minimisation algorithms, e.g. BFGS. Note that computing the log-likelihood function does require solving the full linear system once for each route. If there are no individual attributes, then it is easy to see that the utilities are the same per destination and thus the system only needs to be solved once for routes sharing destinations. Even better, if the utilities are also destination-independent then a clever decomposition method invented by Mai, Bastin & Frejinger (2016) allows us to calculate the log-likelihood value by solving only a single linear system, for all destinations at once. This greatly reduces computation times, at the cost of reducing the possible breadth of the model, as the instantaneous utility v is no longer allowed to depend on destination or individual.

We remain with the issue of overlap, as discussed in the beginning of section 6.1.1. Fosgerau et al. (2013) introduced a special parameter that can be included in the Recursive Logit model: the *Link size attribute*, which is somewhat similar to Path Size Logit. Link size (LS) can be calculated as an attribute for each edge after the RL model has been run for the first time, in the following fashion: each link is assigned an expected link flow to each of the destinations according to its probability of being used in a path, according to the non-LS model estimation. This is then the LS attribute; links with high expected flow (many overlapping routes) receive a higher utility value when LS has a positive parameter value. It should be noted that use of the LS attribute does greatly increase computation times and it was unfortunately not possible to include in this model.

Calculation example Recursive Logit model

To develop a better intuition for the properties of the R.L model, we now work through a small calculation example. The same three-path network from figure 6.1 is used, but now stochastic components are added to the instantaneous path utilities. See figure 6.2. These error terms, ϵ_k, are i.i.d. Gumbell distributed and with zero mean; the scale parameter μ has been normalised to 1. The network can be characterised by its incidence matrix, E, and the matrix with deterministic instantaneous utilities, v, as shown below:

$$E = \begin{array}{c} \\ e_0 \\ e_1 \\ e_2 \\ e_3 \\ e_4 \end{array} \begin{array}{c} \begin{array}{cccccc} e_0 & e_1 & e_2 & e_3 & e_4 & d \end{array} \\ \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array} \quad v = \begin{pmatrix} 0 & -9 & -7 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -3 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} M = \begin{pmatrix} 0 & e^{-9} & e^{-7} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & e^{-3} & e^{-1} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

From E and v we calculate the matrix M, shown above, as the exponentiation of all elements of v, then element-wise multiplied with E. We then use the equation $(I - M)z = b$, with b all zeros except the last element, to find the elements of z. Finally, $V_k = \ln(z_k)$:

$$z = \begin{pmatrix} 0.0005 & 1 & 0.4177 & 1 & 1 & 1 \end{pmatrix} \qquad\qquad V = \begin{pmatrix} -7.5924 & 0 & -0.8731 & 0 & 0 \end{pmatrix}$$

Remember that V represents the expected downstream utility values. It is immediately clear why $V_1 = V_3 = V_4 = 0$: every action taken after edge 1, 3 or 4 must be to use edge d which is the destination and thus has utility 0. From $V_2 = -0.87$ we can see that edge 3, with utility -3+ϵ_3, reduces the expected cost of moving from C to D by 0.13 (as its utility might be higher than that of edge 4). Similarly, having edge 1 and edge 3 as options increases the expected utility of traversing the network from -8 to -7.59. Travellers seeking to traverse this network would choose to use edge 1, with its utility of $-9 + \epsilon_1$, with probability: $\frac{e^{-9}}{e^{-9}+e^{-3-0.8731}} = 5.9\%$.
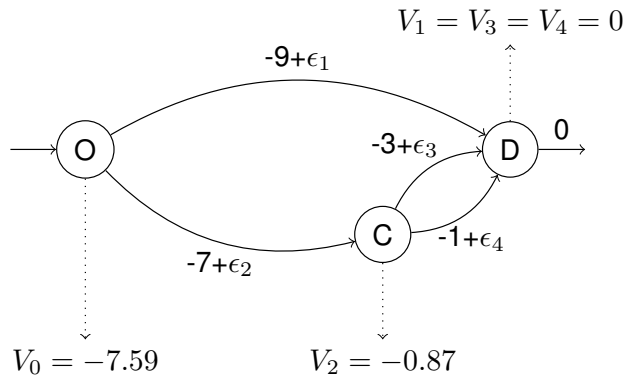

Figure 6.2. Network used in the calculation examples on this page, with utilities and expected downstream utilities shown at edges.

We now turn our attention to the calculation of log-likelihood values. From (Fosgerau et al., 2013), we have: $LL(\beta) = \frac{1}{\mu} \sum_{paths} \left[\sum_{i=0}^{I_n-1} v(k_{i+1}|k_i) - V(k_0) \right]$, or in words: for each path terminating in d, the log-likelihood of it being chosen is minus the difference between: the summed deterministic utility of links in that path, and: the expected value of traversing the network from the origin. Thus, the log-likelihood of the top path is equal to $-9 - -7.524 = -1.4076$, for the middle path it is $-10 - -7.524 = -2.4076$ and the bottom path has likelihood $-8 - -7.5924 = -0.4076$. For the sake of easier comparisons, the final LL will be divided by the number of paths in the dataset.

We can imagine that the utility values in u are not constant, but depend on the attributes of the links via some parameters β. We would then try to find the maximum likelihood of the paths in our network occurring by changing the utility values. A trivial way of doing this would of course be to set these values so that each link has exactly the utility corresponding to the frequency in the dataset; in that case the log-likelihood would be equal to 0. However, when using link attributes, the LL will be negative; to make this a minimisation problem the negative log-likelihood is used.

The above example included only a single destination. For datasets that have paths with different destinations, the same procedure must be followed for each destination separately; thus a linear system must be solved for each destination in every iteration of the algorithm.

6.2   Data pre-processing

Some processing is required before the RL model can be applied to data from the Walkability app. First of all, the route tracks have been cleaned up (see section 3) and the street network from OpenStreetMap was improved and made navigable (see chapter 4). In particular, the RL model requires the given link network to be completely connected to avoid matrix inversion issues. A procedure was developed in ArcGIS-Python to identify disconnected sections of the map, which could otherwise easily escape the eye.

The next major step is to turn the GPS tracks produced by the Walkability app into series of consecutive visited links on the network. This is known as *Map Matching*. There are a number of different ways to accomplish this - the most advanced of which is based on a stochastic approach developed by (Newson & Krumm, 2009). An implementation in ArcGIS-Python is freely available, but the calculation times on the Kampala network proved to be excessively long. Therefore, a buffer-based approach was used; this method was first proposed by (Dalumpines & Scott 2011), and also used by (Lue 2017). A script written in ArcPy performs the map matching in a relatively quick manner. The procedure works as follows: for each track, all links within 15 meters of any GPS point are extracted from the network and loaded into a NetworkX graph. The start and endpoint of the track are snapped to the closest node of this graph (where nodes represent intersections). Then, the NetworkX library is used to determine a shortest path between the start and end nodes; these are then considered to be the walked route and the IDs of links used are written to file. The endpoint of each route is also recorded for later use in the pre-processing procedures. An example of the map matching procedure is shown in figure 6.3. Figure 6.4 shows which links on the network participate in any of the matched routes.

Two more components are required for use of the RL model: an edge incidence matrix and one or more edge attribute matrices. These have mostly the same structure and thus similar procedures are used to generate the matrices. First the network is loaded from ArcGIS into a NetworkX graph data structure. NetworkX can generate the required matrices fairly easily by looping through the edge sets (unfortunately, the library itself can only generate a node adjacency matrix, not edge incidence). For our incidence matrix it is important to add dummy links, which represent route destinations; these were collected in the Map Matching procedure. For the attribute matrices, it is usually easiest to prepare the values as a field in ArcGIS and just read this out in the script. In any case, the matrices are quite sparse and it is thus advantageous to use the SciPy library to write them to files.

The map matched routes and edge matrices are written to file for use by the MATLAB model estimation software. It can then estimate parameter values for the vector β, which represents how strongly participants value different aspects of links on their routes. Remember that the values in β are optimised such that the observed routes have maximum (log-)likelihood to occur on the network for the given origin-destination pairs.

Since the algorithm uses an iterative method to establish optimal parameter values, appropriate starting values should be chosen to ensure quick convergence. Crucially, the RL likelihood function is not defined for all parameter values of β due to matrix inversion issues (which are discussed extensively on page 9 of (Fosgerau, Frejinger & Karlstrom, 2013)). This means that some trial-and-error is required to find a feasible starting point.
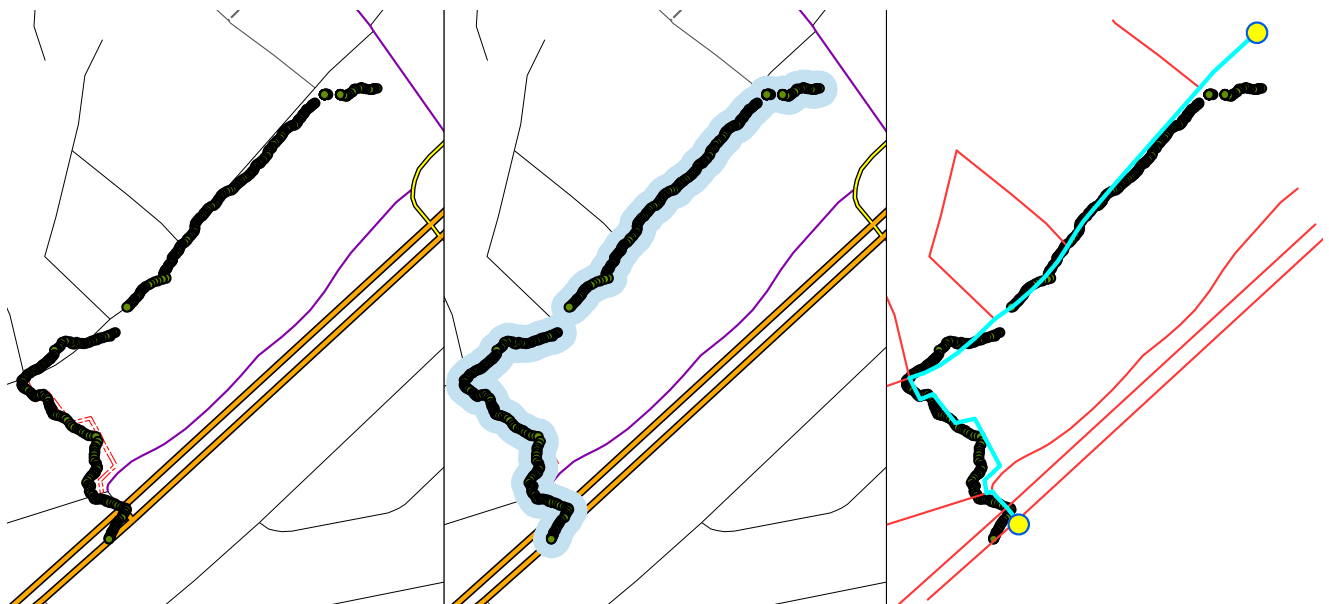

Figure 6.3. Map matching procedure: for each GPS track, select all links within a buffer of the track, then snap first and last GPS points to nodes on this network and calculate the shortest path between them.
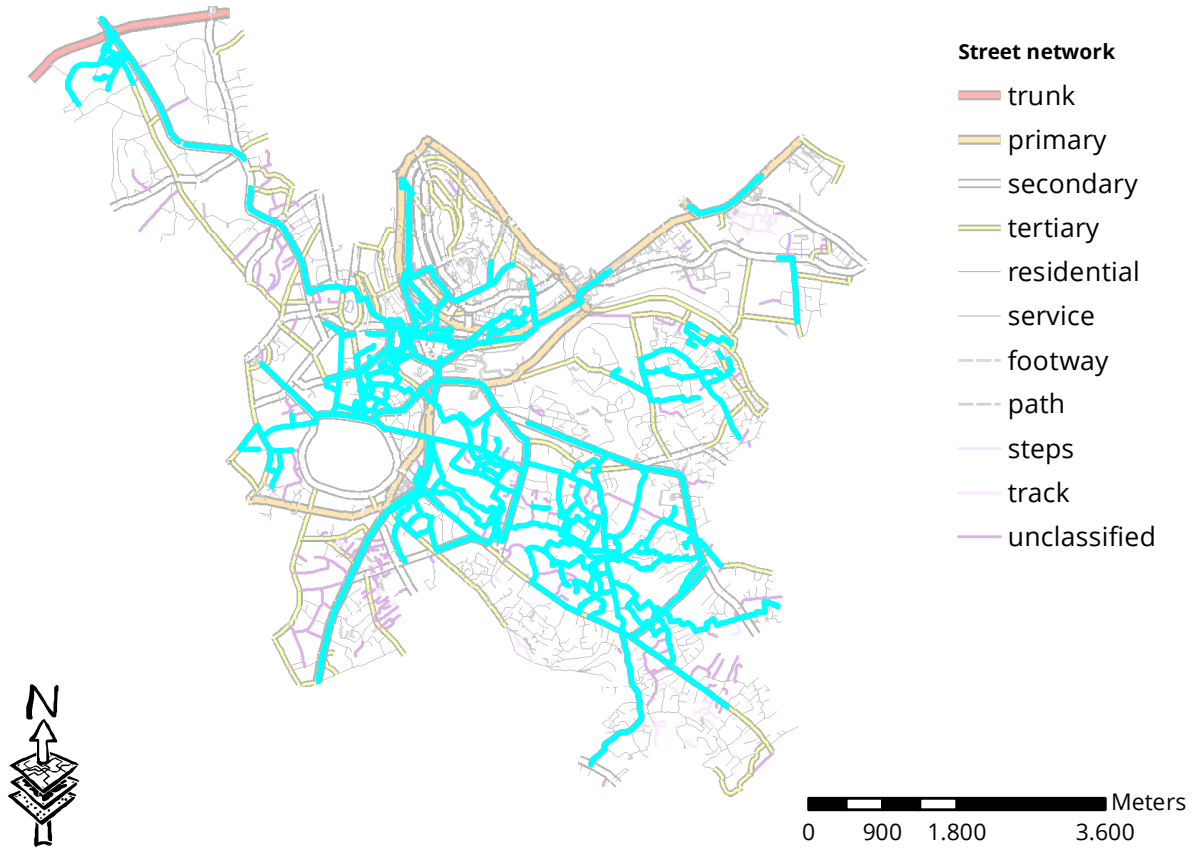
Figure 6.4. Highlighted in blue: links that participate in one or more walked routes.

6.3 Model estimation

Four model specifications have been successfully fitted: one with data from GIS sources, one with data only from averaged user ratings, one with data only from the Built Environment Inventory and finally one that incorporates all data sources. The first three models have somewhat comparable Log-Likelihood values, although the subjective model scores slightly better. The combined model, however, performs quite a bit better than the others. Before explaining the models in full, we quickly recall the value function against which optimisation is performed: this is the logsum $V_n^d(k) = \mu \sum_{a \in A(k)} e^{\frac{1}{\mu} v_n^d(a|k) + V_n^d(a)}$ and the model specification arises from the choice of instantaneous utility function v_n^d. For now, v will not depend on the individual performing route choice (n) nor on the destination (d) so as to allow use of the Decomposition method.

The GIS only model uses only attributes which are already available on the entire network: **link length**, **link type** (as given by OpenStreetMap) and **link rise** (altitude meters). The method of (Zimmermann, Mai & Frejinger, 2017) was used to construct the value function. Indicator variables are use to specify whether an attribute applies or not; the value is then multiplied by the link length. The corresponding parameter is then basically an interaction term between the variable and the link length. The addition of a further attribute, the *link constant*, was inspired by the same source. This attribute simply takes the value 1 for each link used, thereby penalising paths made up of more links irrespective of their length. The instantaneous utility function for the objective model then becomes:

$$v(a|k) = \beta_{LC} \cdot 1 + L \cdot (\beta_L + \beta_{L,primary} \cdot \delta_{a,primary} + \cdots + \beta_{L,walkway} \cdot \delta_{a,walkway} + \beta_s \cdot S_a)$$

The δ functions indicate whether a link corresponds to that type; omitted are a number of other road types - we will see that some of them are significant, others are not. The variable S corresponds to the slope on a link; multiplied by its length L this becomes the height difference between start and end. This variable turns out to not be significant, which is an interesting result in itself.

The second model is an (averaged) subjective one. We assume that all users would rate all roads the same; if a road has not been rated, the closest one is used or the average rating is assigned. Further, the scores are shifted so that their mean becomes 0; i.o.w. the scale becomes -2 to +2 instead of 1-5. This produces attributes AS, FP, LV, PL, SC and WC which are also allowed to interact with the link length (respectively: Accident Safety, Footpath, Liveliness, Pleasantness, Security and Walking Comfort). Recall that these attributes were shown on a map in figure 3.7. The value function is as

follows:

$$v(a|k) = \beta_{LC} \cdot 1 + L \cdot (\beta_L + \beta_{L,AS} \cdot AS + \cdots + \beta_{L,WC} \cdot WC)$$

The third model contains variables from the built environment inventory (B.E.I.). Recall that these are all given in table 4.1. Again, the closest survey point is used to assign a score to the link, within a radius of 40 meters. A number of variables from the B.E.I. were scored on a 0-1 scale (presence or absence / yes or no), while others were scored on numeric scales with explicitly defined levels. Always, the variables were shifted in such a way that unscored links had value 0. The value function becomes:

$$v(a|k) = \beta_{LC} \cdot 1 + L \cdot (\beta_L + \beta_{L,dedicatedsidewalk} \cdot Presence_{dedicatedsidewalk}$$
$$+ \cdots + \beta_{L,apartmentblock} \cdot Presence_{apartmentblocks})$$

The final model simply combines all the previous ones.

All models were run through the log-likelihood optimising procedure. All attributes with parameter values close to zero, relative to their standard deviation, where successively eliminated, producing the results shown in the table below. The length parameter in all models is around -0.02, which means that a pedestrian would be half as likely to choose a route that is 35 meters longer. This is consistent with results of Lue (2017), whose Path Size Logit model assigned a weight of -0.0198 to the length parameter. It is also interesting to compare this outcome to that of Zimmerman et al. (2016), where the value was -0.0074 - apparently cyclists are three times as tolerant to longer routes than pedestrians are. The link constant parameter is also similar to the latter study in all four models (they arrive at -1.6); its precise value will depend mostly on the density of the network.

We can see that only the Accident Safety (AS), Liveliness (LV) and Pleasantness (PL) attributes have a significant effect in the subjective model. Recalling the results from the Factor Analysis, it may not be strange that some of the attributes (other than AS) would be eliminated. What is, however, unexpected is the sign of AS: it is negative. Apparently, unsafe roads are preferred over safer ones.

The results from the Built Environment Inventory are less strong than those from the GIS or the Rating models (lower LL). From a wide range of 36 indicators, only five were actually close to significance. Many indicators which did have some significant effect on ratings in the Ordinal Regression model failed to produce the same result in the route choice model. Of those that did, presence of speed bumps and traffic islands (crossing aids) did not have the expected sign; apparently these features correlate with less desirable roads. The other three indicators did have the expected sign: participants avoided links where many cars were blocking the road and also those with high garbage presence, while those with police or guards were considered more attractive.

Attr.	Coef.	S.D.	Attr.	Coef.	S.D.
L.C.	-1.4283	0.0466	L.C.	-1.364	0.0562
Length	-0.02135	0.0034	Length	-0.02109	0.0039
Secondary	0.00689	0.0014	AS	-0.00154	0.0008
Residential	0.00231	0.0013	LV	0.00204	0.0009
Service	-0.00491	0.0041	PL	0.00195	0.0007
GIS only model			Ratings only model		
$LL = -3.8691$			$LL = -3.8281$		
Attr.	Coef.	S.D.	Attr.	Coef.	S.D.
L.C.	-1.4324	0.0566	L.C.	-1.364	0.0802
Length	-0.01941	0.0033	Length	-0.02176	0.0032
SpeedBumps	-0.00264	0.0023	Secondary	0.00604	0.0013
TrafficIslands	-0.00577	0.0039	AS	-0.00147	0.0005
ParkedCars	-0.00352	0.0015	LV	0.00346	0.0008
Litter	-0.00321	0.0010	SpeedBumps	-0.00524	0.0028
Guards	0.00446	0.0024	TrafficIslands	-0.00548	0.0030
			ParkedCars	-0.00382	0.0012
			Litter	-0.00259	0.0007
			Guards	0.00376	0.0014
B.E.I. only model			Combined model		
$LL = -3.8750$			$LL = -3.6720$		

Table 6.1. Results of log likelihood estimation: for each attribute the coefficient and its standard deviation is shown.

From these parameter values we may calculate walkability scores; this will become our walkability index. The link constant is removed and then the score is adjusted for link length. Parameter values are also multiplied by 1000 for ease of reading. For example, in the case of the subjective model, this means that the link score will be: $score = -1.54AS + 2.04LV + 1.95PL$. For the combined model, it would be $score = 6.04 \cdot is\_secondary + \cdots + 3.76 \cdot presence\_guards$. These scores are shown on a map in figure 6.5.
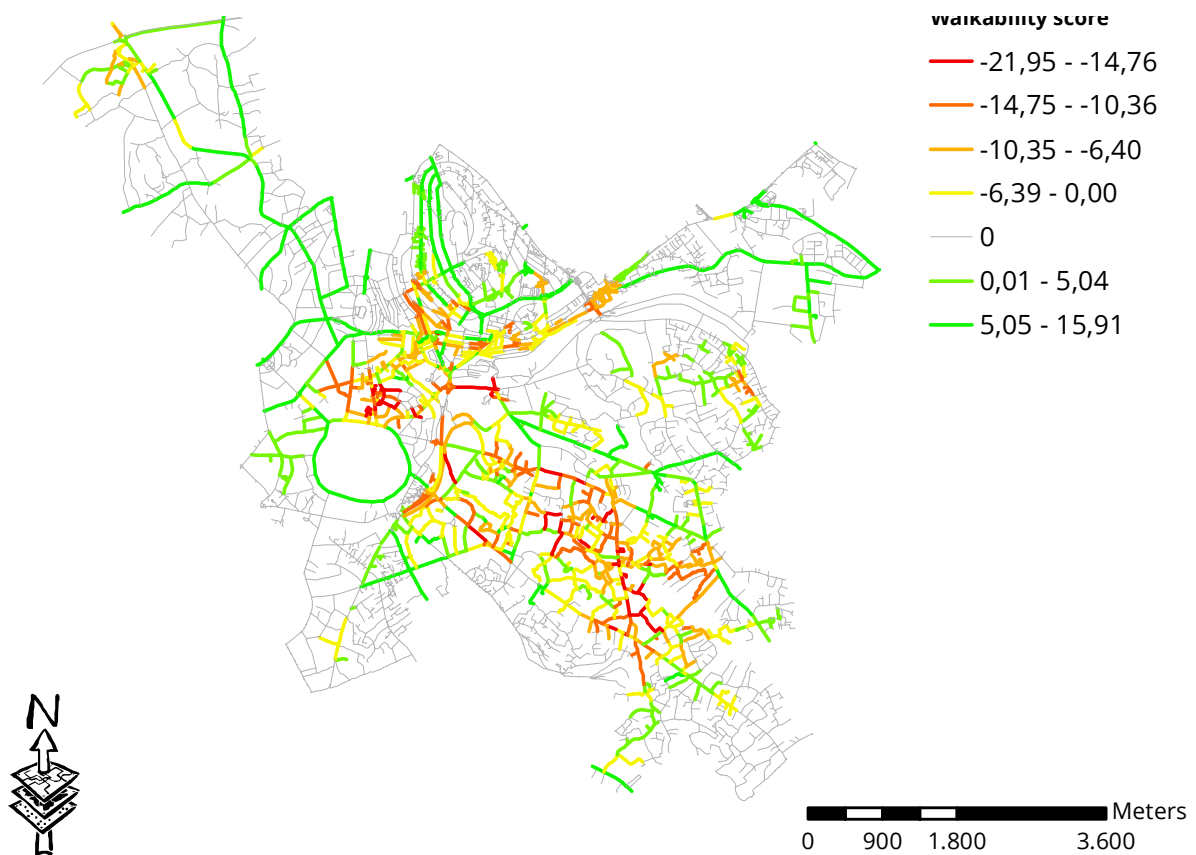
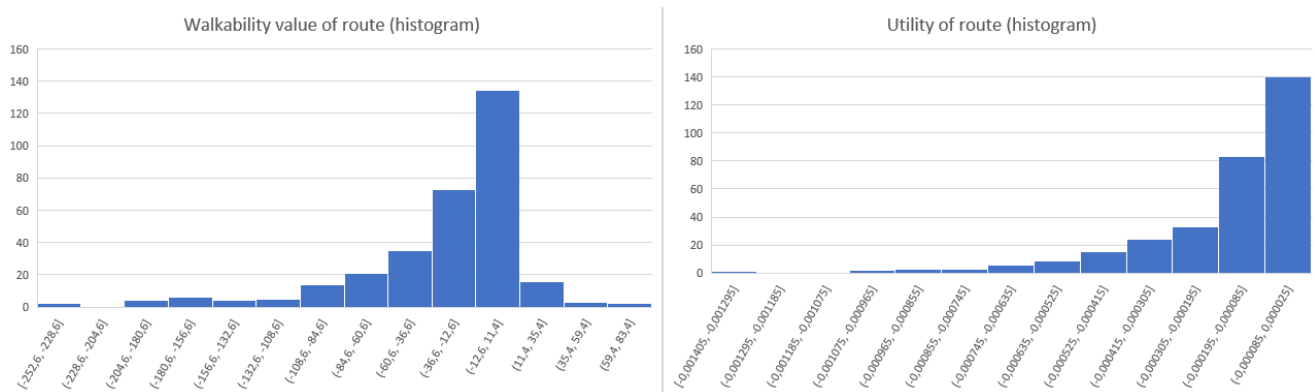**Figure 6.5.** Walkability scores based on the combined model.

Figure 6.6. Histogram of total utility derived from walking a route (right) and accrued walkability score (left)

The central question addressed in this thesis was: *what does walkability mean for Kampala citizens and what factors influence their daily walking route choices?* To answer this question, a route choice model was estimated which allows for modelling the relative importance of link-level attributes of the walking network. As an added benefit, it was also possible to create a link-level walkability index which can be used for accessibility studies. Before this could be done, however, four research sub-questions had to be answered.

The first sub-question related to the perceived walkability of links in the network. This question was answered through the collection of data using the Walkability app. The methodology for data collection and the main findings have been presented in chapter 3. Links in a reasonably wide area of Kampala were rated by one or more participants; also their routes through the city were recorded. We have noted that walkability was perceived as being much better in the city centre than in the surrounding sub-urban and slum areas, which is not surprising given the great difference in infrastructure quality.

Secondly, more objective measures of the walkability of links were sought. A survey with more detailed assessments of the infrastructure was developed and carried out in the study area. Though there were too many indicators to give a concise overview of the collected data, a clear pattern arises which points to decades of neglect of the built environment, especially outside of the city centre and main thoroughfares. Furthermore, motor vehicles are given much more than their fair share of the road space. In many streets, pedestrians are sandwiched between open sewers and speeding vehicles. The alternative is walking through slums, which often have to deal with flooding and which have many security issues.

Some analysis was then done to investigate how representative the user assessments were and whether they were at all correlated with the objective measures. The first of these, actually, two questions, was attempted to be answered using a Factor and Principal Component Analysis, although some remarks can be made about the validity of these methods. The result was that the answers of five of the six categories were fairly strongly correlated, indicating that the differentiation between them is somewhat weak. We did see, however, that all ratings 1 through 5 are given without odd skews or an extreme split between very high and very low ratings.

To dive further into this topic and to find correlations with other factors a more appropriate model had to be used; in this case the choice was made for a Proportional-Odds Logit model for ordinal regression. Unfortunately, the results of this model are not so easy to interpret, although it can be seen in table 5.2 that both characteristics of the individual and the objective walkability measures of the link being rated have a relationship with the score that user gives to that link. In some cases, the neighbourhood the link is in also has an influence, indicating that there is still some variance between built environment in different areas which is not captured by the survey questions. Further, the fact that individual characteristics of the participants had an influence on their ratings means that researchers should be careful to take a representative sample of the population, not just geographically but also socio-economically. In this study the majority of participants were male students, as smartphone possession is highest in this demographic.

On the other hand, the fact that objective characteristics were in some cases significantly related to the user assessments shows that there is some agreement between these measures. Speed bumps were associated with less safe roads (-1.486 to odds of higher Accident Safety). Presence of beggars contributed to lively streets (+0.361) as did presence of police or private security guards (+0.275). Security is positively impacted by low (+0.684) or no presence of litter (+3.622). It was expected that presence of police, street lights and stray animals would have a significant relation, but this was not the case.

The score for footpath was associated with sidewalk presence, but not as expected: +0.575 when there was only a berm to walk in, -1.642 on not-car-accessible links; actual pavement had no significant influence. However, uninterrupted paths did have the expected positive influence: +1.982 for full-length paths and +2.067 for paths with some interruptions for e.g. access roads. Curiously, traffic islands were negatively correlated (-0.935), indicating that users would rather

cross roads fully all at once.

For pleasantness then, open drainage channels were a negative influence (-0.490) as was presence of children (-0.395), the latter of which was unexpected. Again, low or no litter (+0.601 and +1.381, respectively) were a positive influence, as was presence of police or guards (+0.534) and the road being in decent condition (+0.821).

Finally, walking comfort was positively impacted by the state of the road surface (+0.999 for decent road, +0.958 for a good road and +1.443 for an excellent road). Apparently, this is considered more important than the condition of the pavement, which only had a -0.683 for absent sidewalk and +1.529 for one that is separated but not paved. As expected, though, walking comfort was positively associated with low (+1.465) or medium (+1.124) crowdedness.

All of the above relationships are interesting in themselves, but even more important is whether these subjective and objective factors influenced the travel behaviour of participants (sub-question 4). After all, giving an opinion is one thing, but actually acting on it is another and this is what the walkability index will be based on. Of course, length of the path is by far the most important factor and receives parameter weight -0.02. Apparently, our participants were half as likely to use a path 35 meters longer that the shortest one. Given the prevalence of mostly shorter routes in the dataset, this is not too surprising; in fact the value is very close to some other studies as discussed in the previous section. To improve model fit a Link Constant was added as in Zimmerman et al. (2016). This parameter gives a flat penalty, around -1.4, to each link used in a path, which serves primarily to control some statistical properties of the model relating to network density. In very dense networks, many different paths can be used and this increases utility of moving from A to B, even though the existence of many not-very-different alternatives can be quite irrelevant.

The interesting bit comes when we start adding more factors corresponding to the links as explanatory variables. This was done first for each data source separately, then in a combined model. The combined model had, of course, the best fit (LL=-3.672). Second was a model using only the user assessments (LL=-3.828) and only then came models using OpenStreetMap link types (LL=-3.869) and indicators from the built environment dataset (LL=-3.875). This validates the use of user assessments, a welcome result.

In all cases, the factors were added as an interaction term with the length of the link, for easier comparison. Three of the six categories have a significant influence: a point in Accident Safety decreases utility by -0.00154 times -0.02 times the link length. This can be interpreted as follows: the link is perceived to be 7.1% longer for each point of Accident Safety (which, again, was rated -2 to +2). A point in liveliness similarly decreases the perceived link length by 9.7% and pleasantness decreases it by 9.2%. In the combined model the results are similar. The following factors are important: the link being a secondary thoroughfare (27.8% shorter), presence of speed bumps (24.1% longer), presence of traffic islands (25.2% longer), presence of parked cars blocking the path (17.6% longer), presence of litter (11.9% longer) and presence of police or security guards (17.2% shorter). A point of Accident safety makes the link 6.8% longer and one of Liveliness reduces length by 15.9%.

The above analysis answered the main question of this research, namely which factors are important in walking route choice and thus which are important for walkability. It is possible to go one step further with the model estimation results; namely by constructing a walkability index. Each link is assigned a score that relates to its (instantaneous) utility in the choice model. Note that, here, the scores are the same for everybody and also do not depend on the origin or destination where the user is travelling. The score may depend on the direction in which someone is walking over the link; it was expected that slope would be significant, but this was not the case.

In any case, the index relates to the link utilities as follows: both the link constant and the link length are removed from the score, and what remains is a sum-product of the model coefficients and the attributes. For illustration: a link in the ratings only model is scored as follows: $score = -1.54 \cdot Accident\ safety + 2.04 \cdot Liveliness + 1.95 \cdot Pleasantness$. The scores for the full model are shown on a map in figure 6.5 and this is an important output that helps show which areas can be considered well-walkable and which have low walkability.

CHAPTER 8

DISCUSSION

The outcomes of this research are reasonably in line with other recent route choice modelling studies, with respect to the obtained parameter values. It should be no surprise that the lack of comprehensive datasets describing Kampala meant that the results may not be as complete as those of comparable studies. A further shortcoming is that the ratings given by individual participants could not be matched to their own routes, so only aggregate scores could be used. There are two reasons for this: first, computation times of the RL model are prohibitively long when such individual components are included. Second, only a selection of links in the study area was rated during Built Environment Inventory, which meant that missing data imputation on links not visited by the participants ("predicting the scores they would have given") would not have been able to score many of the non-used routes.

The results of the factor analysis showed that there was not a lot of differentiation between ratings on a number of the categories. Only ratings of Accident Safety were consistently very different from the other categories. This shows that the questions asked to the participants could be improved. Also, explanation of the research goals and the rating criteria should be better.

It can be questioned whether it was worth it to ask so many participants to use the app rather than having a few surveyors do the work by themselves. However, the results of the different route choice models clearly show a better fit when perceived walkability indicators are added, even better than the built environment indicators. This may show that it is indeed worthwhile to collect data from members of the public.

For policy makers, the results of this research show that "soft" factors such as cleanliness and security on the streets is as important as safety of the infrastructure. It can be worthwhile to take a more proactive approach on parking, litter and police presence.

The results of this thesis are however mostly of importance to the academic literature. We have shown that the Recursive Logit model can be successfully applied to pedestrian route choice and can be used with GPS data collected from ordinary smart-phones. Further, we have shown that perceived walkability factors should have a more prominent role in route choice models. It is hoped that more research will be done on route choice and that the methods used in this project will be used more often. All in all, processing the gathered data and preparing the route choice model was much more cumbersome than initially expected and a wide variety of technologies had to be applied: Java programming for the two smartphone apps, ArcGIS for visualisation and importing of spatial datasets, Python for map matching and importing the smartphone data into ArcGIS, R and Excel for statistics and finally Matlab to estimate the model. Hopefully, future researchers can make use of the programs and methods developed over the course of this research to skip many of these steps, which are in themselves not of any academic value.

CHAPTER 9

BIBLIOGRAPHY

Albers, P. N., Wright, C. & Olwoch, J. (2010).
Developing a South African pedestrian environment assessment tool: Tshwane case study.
South African Journal of Science, *106(9-10)*, 1–8.

Asia, C. A. (2011).
Walkability survey tool.
`http://cleanairasia.org/wp-content/uploads/portal/files/documents/18_Walkability_Survey_Tool_2011.pdf`.

Bailey, C. (2016).
The spaces in-between: Learning from Kampala's slums.
Master's thesis, University of Washington.

Bigazzi, A. Y., Broach, J. & Dill, J. (2016).
Bicycle route preference and pollution inhalation dose: Comparing exposure and distance trade-offs.
Journal of Transport & Health, *3(1)*, 107–113.

Cain, K. L., Millstein, R. A., Sallis, J. F., Conway, T. L., Gavand, K. A., Frank, L. D., Saelens, B. E., Geremia, C. M., Chapman, J., Adams, M. A., Glanz, K. & King, A. C. (2015).
Microscale Audit of Pedestrian Streetscapes: MAPS.
University of California, San Diego.

Cerin, E., Chan, K., MacFarlane, D. J., Lee, K. & Lai, P. (2011).
Objective assessment of walking environments in ultra-dense cities: Development and reliability of the Environment in Asia Scan Tool—Hong Kong version (EAST-HK).
Health & place, *17(4)*, 937–945.

Dalumpines, R. & Scott, D. M. (2011).
Gis-based map-matching: Development and demonstration of a postprocessing map-matching algorithm for transportation research.
In *Advancing geoinformation science for a changing world* (pp. 101–120). Springer.

Day, K., Boarnet, M., Alfonzo, M. & Forsyth, A. (2006).
The Irvine-Minnesota Inventory to measure built environments: Development.
American Journal of Preventive Medicine, *30(2)*, 144 – 152.

Fosgerau, M., Frejinger, E. & Karlstrom, A. (2013).
A link based network route choice model with unrestricted choice set.

Transportation Research Part B: Methodological, *56*, 70–80.

Frank, L. D., Schmid, T. L., Sallis, J. F., Chapman, J. & Saelens, B. E. (2005).
Linking objectively measured physical activity with objectively measured urban form: findings from smartraq.
American journal of preventive medicine, *28(2)*, 117–125.

GSM Association (2017).
The mobile economy: Sub-saharan africa 2017.

HealthBridge (2015).
Informal places for active recreation and children's play in Kampala slums, Uganda.
`https://healthbridge.ca/library/entry/informal-places-for-active-recreation-in-kampala-slums`.

Hintaran, E. (2016).
Unravelling urban pedestrian trips: Developing a new pedestrian route choice model estimated from revealed preference GPS data.
Master's thesis, TU Delft.

Hoedl, S., Titze, S. & Oja, P. (2010).
The bikeability and walkability evaluation table: Reliability and application.
American Journal of Preventive Medicine, *39(5)*, 457–459.

Jack, E. & McCormack, G. R. (2014).
The associations between objectively-determined and self-reported urban form characteristics and neighborhood-based walking in adults.
International Journal of behavioral nutrition and physical activity, *11(1)*, 71.

Jun, H.-J. & Hur, M. (2015).
The relationship between walkability and neighborhood social environment: The importance of physical and perceived walkability.
Applied Geography, *62*, 115–124.

Kiggundu, A. & Mukiibi, S. (2012).
Land use and transportation planning in the greater Kampala, Uganda.
44(1), 1–11.

Krambeck, H. V. (2006).
The global walkability index.
Master's thesis, Massachusetts Institute of Technology.

Lee, R. E., Mama, S. K., Medina, A. V., Ho, A. & Adamus, H. J. (2012).
Neighborhood factors influence physical activity among African American and Hispanic or Latina women.
Health & place, *18(1)*, 63–70.

Lue, G. (2017).
Estimating a Toronto pedestrian route choice model using smartphone GPS data: It's not the destination, but the journey, that matters.
Master's thesis, Univsersity of Toronto.

Ma, L. & Dill, J. (2015).
Do people's perceptions of neighborhood bikeability match "reality"?

In *Transportation Research Board 94th Annual Meeting*, number 15-5739.

Mai, T., Bastin, F. & Frejinger, E. (2016).
A decomposition method for estimating recursive logit based route choice models.
EURO Journal on Transportation and Logistics (pp. 1–23).

Mai, T., Fosgerau, M. & Frejinger, E. (2015a).
A nested recursive logit model for route choice analysis.
Transportation Research Part B: Methodological, *75*, 100–112.

Mai, T., Frejinger, E. & Bastin, F. (2015b).
A misspecification test for logit based route choice models.
Economics of Transportation, *4(4)*, 215–226.

Manaugh, K. & El-Geneidy, A. (2011).
Validating walkability indices: How do different households respond to the walkability of their neighborhood?
Transportation research part D: Transport and environment, *16(4)*, 309–315.

McCullagh, P. (1980).
Regression models for ordinal data.
Journal of the royal statistical society. Series B (Methodological) (pp. 109–142).

McMillan, T. E., Cubbin, C., Parmenter, B., Medina, A. V. & Lee, R. E. (2010).
Neighborhood sampling: how many streets must an auditor walk?
International Journal of Behavioral Nutrition and Physical Activity.

Millington, C., Thompson, C. W., Rowe, D., Aspinall, P., Fitzsimons, C., Nelson, N. & Mutrie, N. (2009).
Development of the Scottish Walkability Assessment Tool (SWAT).
Health & Place, *15(2)*, 474 – 481.

Montini, L., Antoniou, C. & Axhausen, K. W. (2016).
Route and mode choice models using GPS data.
Arbeitsberichte Verkehrs-und Raumplanung, *1204*.

Newson, P. & Krumm, J. (2009).
Hidden markov map matching through noise and sparseness.
In *Proceedings of the 17th ACM SIGSPATIAL international conference on advances in geographic information systems* (pp. 336–343). ACM.

Oyeyemi, A. L., Conway, T. L., Adedoyin, R. A., Akinroye, K. K., Aryeetey, R., Assah, F., Cain, K. L., Gavand, K. A., Kasoma, S. S., Kolbe-Alexander, T. L. et al. (2016a).
Construct validity of the neighborhood environment walkability scale for Africa.
Medicine and science in sports and exercise, *49(3)*, 482–491.

Oyeyemi, A. L., Kasoma, S. S., Onywera, V. O., Assah, F., Adedoyin, R. A., Conway, T. L., Moss, S. J., Ocansey, R., Kolbe-Alexander, T. L., Akinroye, K. K. et al. (2016b).

NEWS for Africa: adaptation and reliability of a built environment questionnaire for physical activity in seven African countries.
International Journal of Behavioral Nutrition and Physical Activity, *13(1)*, 33.

Pikora, T., Giles-Corti, B., Bull, F., Jamrozik, K. & Donovan, R. (2003).
Developing a framework for assessment of the environmental determinants of walking and cycling.
Social science & medicine, *56(8)*, 1693–1703.

Pikora, T. J., Bull, F. C., Jamrozik, K., Knuiman, M., Giles-Corti, B. & Donovan, R. J. (2002).
Developing a reliable audit instrument to measure the physical environment for physical activity.
American journal of preventive medicine, *23(3)*, 187–194.

Resnick, D. (2014).
Urban governance and service delivery in African cities: the role of politics and policies.
Development Policy Review, *32(s1)*.

Saelens, B. E. & Handy, S. L. (2008).
Built environment correlates of walking: a review.
Medicine and science in sports and exercise, *40(7 Suppl)*, S550.

Sepe, M. (2009).
PlaceMaker method: Planning 'walkability' by mapping place identity.
Journal of Urban Design, *14(4)*, 463–487.

Tainio, M., de Nazelle, A. J., Götschi, T., Kahlmeier, S., Rojas-Rueda, D., Nieuwenhuijsen, M. J., de Sá, T. H., Kelly, P. & Woodcock, J. (2016).
Can air pollution negate the health benefits of cycling and walking?
Preventive Medicine, *87*, 233–236.

Vanwolleghem, G., Ghekiere, A., Cardon, G., De Bourdeaudhuij, I., D'Haese, S., Geremia, C. M., Lenoir, M., Sallis, J. F., Verhoeven, H. & Van Dyck, D. (2016).
Using an audit tool (MAPS Global) to assess the characteristics of the physical environment related to walking for transport in youth: reliability of Belgian data.
International Journal of health geographics, *15(1)*, 41.

Wasswa, B. (2016).
The use of mobile phone technology in the collection of user perceptions of walkability along pedestrian routes from public transport in Cape Town and New Delhi.
Master's thesis, University of Cape Town.

Zandieh, R., Flacke, J., Martinez, J., Jones, P. & van Maarseveen, M. (2017).
Do inequalities in neighborhood walkability drive disparities in older adults' outdoor walking?
International Journal of Environmental Research and Public Health, *14(7)*.

Zimmermann, M., Mai, T. & Frejinger, E. (2017).
Bike route choice modeling using GPS data without choice sets of paths.
Transportation research part C: emerging technologies, *75*, 183–196.

GUIDE TO DEPLOYING THE WALKABILITY APP

A.1   Use of the app: an overview

In this section the various functionalities of the app and the processes surrounding its use are explained. Please note, first of all, that the Walkability app was originally created and is owned by the University of Cape Town. All changes made to the app during this research were made with permission. Should you wish to use or change the app, either in its original or in the revised version, you should contact prof. Zuidgeest at the Centre for Transportation Studies at the University of Cape Town to obtain permission.

Figures 3.1, 3.2 and 3.3 show the design of the app and the content of the tool. How the user navigates through the various screens of the app is shown in a diagram in figure A.1. Before the app can be used, however, it should first be installed on the user's smartphone. Phones with operating system Android 4.0 (*"Ice cream sandwich", released October 2011*) or higher are supported. The researcher should have the Android app package (*.apk*) file prepared for transferring. This can be done from a laptop using a USB cable or from another phone using Bluetooth or Xender (personal wi-fi hotspot). The app itself weighs a modest 6 MB but the included maps are quite large, bringing the total to 20 MB which may take a minute to transfer using Bluetooth. Once the user has received the file, he or she can press install. They may be prompted to enable the system setting to `Allow installation from unknown sources`. On old versions of Android, the app will also ask for permission to access phone features like GPS and the camera - on newer phones these permissions will be asked case-by-case when needed.

When the user opens the app, they will go through the screens in figure 3.1. Additionally, they will be prompted to enable location services if GPS use is disabled; they should ensure to enable either `High accuracy mode` or `Device (GPS) only mode`. Furthermore, if the user has low storage space available (less than 25 MB), the app will instruct them to free up some space. Then, the **Initial Questions** screen is shown in which some data about the user is collected. On subsequent uses of the app, this screen will be skipped - a reset button is available on the beginning screen to allow phone sharing. It is mandatory for the user to enter a phone number, which is used as an ID to distinguish different participants. Any number of at least 10 digits will do, so inputting e.g. all zeros is an allowed alternative for those who do not wish to share their number. All the other questions are optional. The results are written to a CSV file called `personalCharacteristics.csv`.
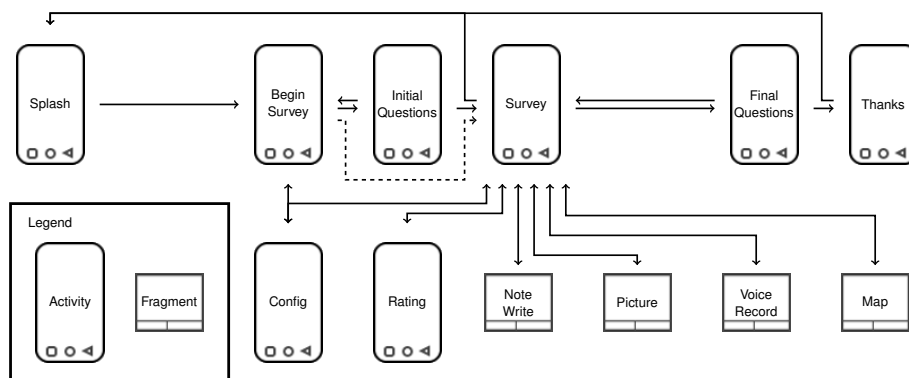

Figure A.1. Control flow of the new Walkability app, showing all the different components and the way users navigate between them.

While the user answers the questions, the app will try to determine the phone's location in the background. If successful, the survey can begin immediately after the answers have been completed. In most cases, however, the GPS fix will not be ready by that time and the user will have to wait - as shown in the final screen of figure 3.1. A GPS fix will be acquired faster when the phone's location services have been used often. During app development, the phone used to test changes would go through the process in just a couple of seconds. However, participants' phones could take up to five minutes to GPS fix the first time. Access to clear sky is also important: clouds, buildings and trees may delay the process.

GPS communication is handled through the `Android location API`, which works by allowing a component of the app to register for location updates and running a specified procedure every time the location estimate changes. The estimated location, which comes with an accuracy estimate, is inspected by that procedure. It has been programmed such that the new location is rejected if the accuracy is above a certain threshold. If two points in a row are below the threshold then the GPS communication is considered to have been sufficiently stabilised for starting the app. The user can configure the required accuracy in the **Settings** screen if there are problems reaching the threshold. However, a deviation of below 10 meters is desirable for proper reconstruction of the route. The current accuracy estimate is always shown on the survey screen, even if the estimate has been rejected - in that case the text turns red. This is also useful during the GPS fix acquiring procedure, as the user can see progress every time the accuracy display changes. If it stays constant for some time, moving around may trigger new location updates. On each successful location update, a line is written to `Route.csv` with a timestamp, the location, the accuracy and the number of connected satellites.

As soon as the GPS is stable, the assessment can begin in earnest through the screens shown in figure 3.2. While the participant walks their route, they will be prompted to assess their location every so often as the **Ratings** screen pops up. For each rating category, a value of 1-5 (or blank) can be given. There are question mark icons that, when pressed, show a more detailed explanation of the criteria used for assessment (see also table 3.1). The user may also dismiss the screen if they do not want to answer the questions there, simply by pressing the back button. Normally ratings will be asked every 150 meters, as long as there are no other screens currently shown. The phone will notify the user by sending vibration when it is time to answer the questions. The assessment interval can be changed on the fly in the settings screen; if the user wishes to answer the questions in a specific location they can press the star button in the survey screen. Each time the questions are completed, a line is written to `Ratings.csv` with the specifics.

Besides using the rating screen, the participants can also record information about their location by writing **text notes**, taking **pictures** or recording **voice notes**. All these features are accessed through their respective buttons on the survey screen. The user may need to give permission for the app to use the microphone and/or camera if they are running a relatively new version of Android; if they refuse permission then the respective features are not available. The results are collected in `Media.csv`, where each line describes one location where users added media. Text notes are reproduced in full; for pictures and voice notes only the path to the file is recorded.

The final feature of the survey screen is the **Map**. The map shows the route that has been walked so far during this assessment and there are markers in each location where the participant gave ratings or added media. This feature gives important feedback to the users that their assessment is currently going well, and can visually indicate problems when they arise (e.g. location issues). The route is overlayed over OpenStreetMaps (OSM). A section of OSM, including Kampala at a four different zoom levels, has been pre-loaded into the app. This ensures that there is no internet connection needed for viewing the map. For use outside of Kampala there is an option in the Settings screen to turn on data connectivity which will procure map sections from OSM as needed.

When the user has reached their destination, they can proceed to the final screens (the first two in figure 3.3). They are asked to answer a few **Final Questions** about the route they have used and the destination they are going to. Then the assessment is complete and the **Thanks** screen is shown. All the files that have been written so far, including all CSV files and any pictures or audio recordings, are archived into a single zip file. The archive is named by putting the user's ID (phone number) together with a timestamp. This is then stored in the phone's internal storage. The user is recommended to turn off their location services. If the researcher is around, the zip archive can directly be transferred to their phone or laptop in the same way the app was originally sent; for convenience the thanks screen also includes a button for directly transferring through Bluetooth. Otherwise the user can close the app and do the transfer at a later date. In the meantime they can do as many assessments as they like, as long as their phone has enough storage space available.

During testing two issues with phone storage were encountered; one in which the phone refused to write any files at all (giving error messages), which was solved by rebooting the phone. A second phone (a Huawai model with a very new Android OS) did not allow USB access to the internal storage location where the zip file is written; this is easily solved by either copying the file to another location or by using a different transfer method.

The app is currently available in three languages: English, French and Luganda. The settings menu includes an option for switching the language. For illustration purposes, the note writing screen is shown in Luganda in figure 3.3 (rightmost).

A.2 Instructions for adapting the app for other research

Those who have received the source code of the app can use Android Studio to build new apk files adapted to the circumstances for their research. A large portion of the content of the app is described by a few XML files which are relatively easy to change. By far the most important of those is: `app/src/main/res/values/strings.xml` which contains all texts displayed to the user (and also printed to the CSV files). To translate the application, one would copy the strings XML into a new folder (e.g. `values-lg` for Luganda, `values-fr-rCA`, for French Canadian, etc.) then translate the texts. If the user has their device set to the corresponding language, the app will automatically use that language; otherwise a new option can be added to the settings menu for changing to the respective language (read on for more on this).

It is trivial to change the rating categories: in strings.xml there is an xml element of type `string-array` called *rating_categories*. One can change its `items` and the ratings screen will automatically add or remove categories as needed. When adding a category, a corresponding explanation should also be entered in the *rating_categories_hints* string-array. In the same XML file, the questions for the initial and final question screens can also be adjusted. Adding, removing or otherwise changing the answers can also be done in the same manner as with the rating categories.

However, to add or remove questions altogether is just a tad more involved. The question screens are implemented as app preference screens which are described in their own XML files: `app/src/main/res/xml/initial_questions.xml` and, in the same folder, `final_questions.xml`. The Android documentation on PreferenceScreens has the full information on how to edit these files, but it should be fairly easy to figure out using the existing code. Questions can take the form of multiple choice (`ListPreference`), checkboxes (`CheckBoxPreference`) or text input (`EditTextPreference`). A number of values should be provided, including e.g. the question title, a summary or hint and the answer options. Rather than immediately writing these out, directives on where to find the texts are used; these look like `"@string/public_space_use"` or `"@array/public_space_use_options"`. Then, corresponding entries should be included in strings.xml (for all languages). A complete example is shown in figure A.2. First there are two EditTextPreference elements, one with normal and one with phone inputType - see the phone input screen in figure 3.1. Then there are four ListPreference elements, the first three of which have been folded up in the editor since they are all very similar to one another. The final one is shown in full and for each of its values, the corresponding part of the screen is indicated. Note the difference between `entries` and `entryValues`: although these are the same in English, only the entries are supposed to be translated to other languages. The entryValues are what is read by the code and what is written to the CSV files.
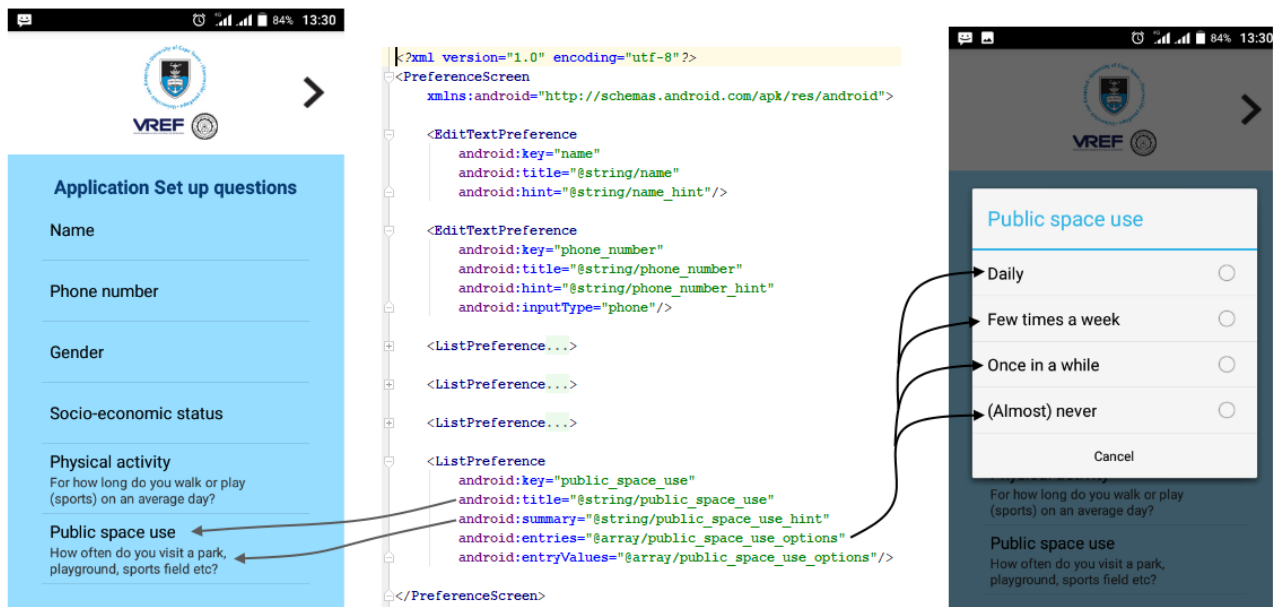
Figure A.2. Example of a set of questions and its definition. All `@string/` directives are place-holders for localised texts.

Returning to the matter of translations: adding language options to the Settings screen works in a similar way to adding answers to questions. As previously stated, first a `values-LOCALE/strings.xml` file should be created and filled with the localised texts. The text LOCALE should be replaced with the two-letter ISO-639-1 code[1] for the specific language,

---

[1] `http://www.loc.gov/standards/iso639-2/php/code_list.php`
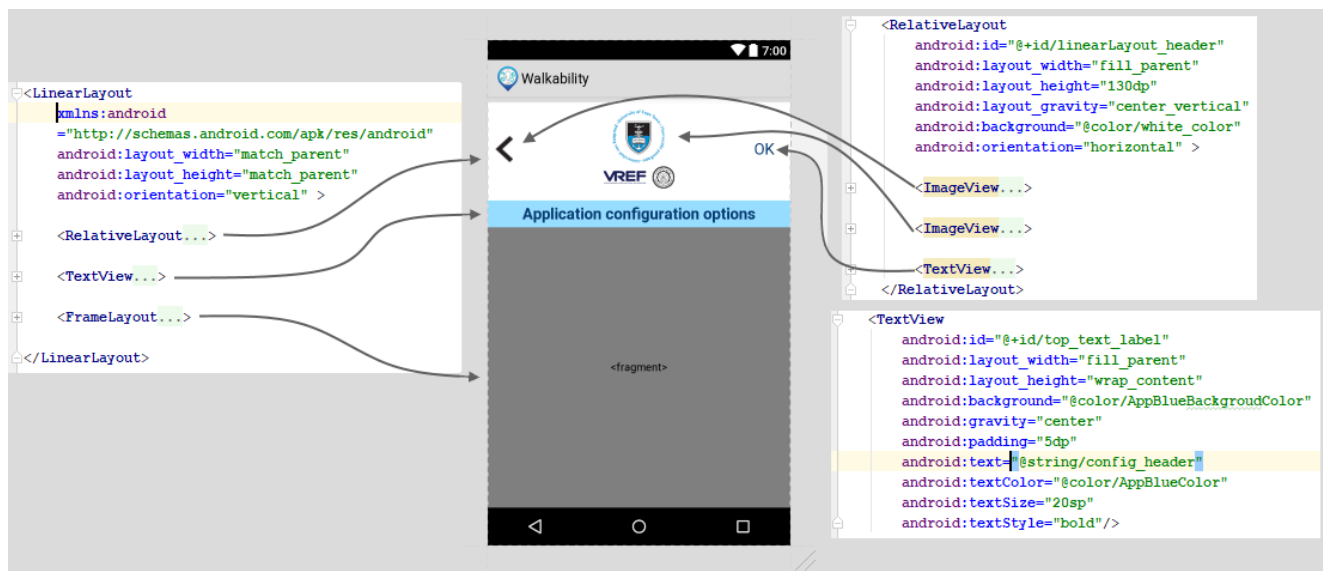
Figure A.3. Example of how an XML layout file translates into a screen. The '+' sign indicates code sections that have been folded. On the right side, the RelativeLayout and the TextView are shown unfolded one level. The FrameLayout is a container where a fragment will be placed.

e.g. `de` for German, `es` for Spanish, etc. It is also possible to add a ISO-3166-1-alpha-2 region string for languages with varying spellings; e.g. `en-rGB` for British English and `fr-rCA` for French Canadian. Now, simply adding the language name as an item to the default `strings.xml`'s `language_options` array will make it show up in the menu. In the `language_options_values` array, the corresponding locale code should be entered. The app will automatically switch to that locale after the user presses the OK button in the settings menu.

One more part of the app is important to consider, namely the map. When the Walkability tool is used in an area where wifi is commonly available, or mobile data is not very expensive, or when it is not known exactly in which areas the users will be walking, it is probably easiest to set the *allow map downloading* option on by default. This is done in `app/src/main/res/xml/config.xml` (which is where the Settings menu is defined, similarly to the Questions screens). There is an entry there (a CheckBoxPreference) for this option and one can change the `defaultValue` to true.

To include pre-downloaded map sections in the app is a bit more complicated; it is however much appreciated by users and so should be strongly considered. The map itself just consists of a number of PNG images, called *tiles*, which are stitched together in a grid (for each zoom level). These are located in `app/src/main/assets` in a folder called Mapnik (which is an alternative name for OSM). Because of the large file size of these maps, one would do best to remove the existing Mapnik folder and replace it with a new one. However, it is certainly possible to add to the existing map - two Mapnik folders can just be merged together. The tiles can be downloaded from OpenStreetMap using a number of different tools. For the Kampala map, OSMMapTilePackager was used. One needs to give the desired the desired zoom levels, 13-16 in this case, and coordinates for the bounding box. The program then produces a Mapnik folder with the tiles which can be copied into the app assets; these will then be recognised automatically and the new map will be shown to whomever uses the Walkability tool in that area.

A.3 A brief explanation of the inner workings of Android apps

Any app consists of static elements, which are called *resources* and dynamic behaviour specified by *application code*. The resources notably include images and text strings (potentially with translations) as well as *layouts*. Each layout describes the look of a portion of the app using an XML format. In the simplest (and most common) case each layout corresponds to one screen that the user may encounter. All kinds of *Views*, like buttons, texts, input fields etc. can be declared and placed on certain locations of the screen. Since phones may have different display sizes, the locations should not be hard-coded coordinates but instead be measured relative to each other. Commonly, one would use boxes (e.g. `LinearLayout` or `GridLayout`) that take care of the placement of their assigned display elements. An example of a simple layout file and its corresponding screen in the app are shown in A.3.

The layout files are then glued together by the code, which handles the dynamic side of the app. This code is normally written in Java, though any language that runs on the Java Virtual Machine can be used (attractive alternatives are Scala and Clojure). Apps can make use of the Java standard library, which includes functions for common functionalities

like file reading and writing, internet communication and database handling. Furthermore, extensive Android libraries are available which help with declaring the behaviour of app components. The code is organised into files that each describe a certain part of the app; either an *Activity* or a *Fragment*. An Activity is a self-contained part that fills the entire screen, meaning that only one can be visible at a time. Fragments, on the other hand, are components that clearly have their own behaviour but occupy only part of the screen - dialog windows are a prominent example. For each Activity or Fragment, the programmer should write the corresponding XML layout file and then start implementing the behaviour of the Views. This may involve setting the actions for the various buttons and specifying what happens when the component is created, destroyed or made invisible.

It should be noted that Activities and Fragments are placed on a stack. Whenever a new component is started, it is placed on top of the stack and if it obscures other components, those are *paused* and the user cannot interact with them. When the component *finishes*, e.g. because the user pressed the Back button on their phone, whatever's below it on the stack can be *resumed*. An example of use of this stack in the Walkability tool is in the Config screen, which is an Activity that can be reached by pressing a button either on the BeginSurvey screen or the Survey screen. When the button is pressed, Config goes on top of the stack and the calling activity is paused; when the user goes back Config is popped off the stack and BeginSurvey or Survey is shown again (refer back to figure A.1).

GUIDE TO DEPLOYING THE GPSTIONNAIRE APP

Using the app

The GPStionnaire app consists of four key components which the user can interact with. First, the Splash screen, which will show a button that allows users to start a new survey. If a previous survey was left uncompleted, there will also be a button to continue from there. The splash screen also performs some initialisation in the background and will ask the user for permission to write to the phone's storage and to access location information, as well as prompting the user to turn on location services if they have not yet done so.

Next up is the Survey Start screen. Here, basic information about the current location should be entered. This includes the surveyor name, the location name and survey ID (the last two of which need to be unique and not used before on this phone). On the bottom of the screen a map is shown (OpenStreetMap). This map will be centred on the phone's current location estimate (either through GSM- or GPS-based information) and will move when more accurate estimates become available.[1] In the map centre, a marker is shown which indicates the current position. If the phone is unable to pick the right location, users can manually move the marker to the correct position of the survey (which will also stop the map from recentring itself). The map also shows markers with the locations of previous surveys. Showing the map **requires an active data connection**, although previously shown map portions will be cached for the next time.

The third screen is the Survey menu. Here, all components of the survey will be shown as separate buttons. The first button always allows for taking of pictures, while the rest will give access to the different parts of the questionnaire. Finally, the green check-mark button will complete the survey and write all responses to file.
Every time one of the survey component buttons in the menu is pressed, the user is taken to a screen showing the questions corresponding to that component. There are four types of questions:

1. Open questions *(input any text)*
2. Yes/no questions *(tick a checkbox)*
3. Choice questions *(tick one from a list of options)*
4. Multi-choice questions *(tick all that apply from a list of options)*

Furthermore, questions can be grouped into categories, which when pressed take the user to a new screen which again can include the above question types. Categories can be nested arbitrarily deeply.

Retrieving the data

All information relating to the app is stored on the phone in the following directory: `/Android/data/com.gpstionnaire`. There are three sub-directories: `files`, which is where the survey responses are written to, `pictures` which is where any photos are stored and finally `questions` which is where the survey questions are read from (as will be explained in the next section).

For each component of the survey, there is a separate file of responses. These files are written in CSV file-format, in which values are separated by semi-colons. CSV files can easily be opened in Excel or similar programs for data processing and analysis.[2] For each survey, the results (chosen options) are written as one or more rows in the

---

[1] if the map is showing only a blue area, it put your location at (0,0) which is in the middle of the Atlantic ocean. Move to an outdoors area for better GSM/GPS connectivity.

[2] Use Data→Text to Columns in case the data doesn't look good in your spreadsheet program.

spreadsheet where the values are recorded in the corresponding cell. When the questionnaire included multi-choice questions, the results for a survey may need multiple rows to reflect all the chosen options for those questions. It is now easy to get counts or percentages of answers chosen by selecting the column and creating a pivot table.

Survey ID	Location name	1.1: Example open	1.2.1: Example yes/no	1.2.2: Example choice	1.3: Example multi-choice
A	Simple example	We typed something	TRUE	Second option text	
B	Second example	Something else	FALSE	Something typed in an open option	First item selected
B	Second example				Second item selected
B	Second example				Third item selected

Figure B.1. Example output of the tool

In addition to the CSV files with survey responses, the app will also write a file containing the locations of all assessments: `locations.kml`. This can easily be opened in Google Earth or loaded into ArcGIS (using the KML to Shape tool) and will give a quick overview of the geographic distribution of survey locations.

Changing the survey questions

The survey questions are specified in a set of XML files which can be found in the questions directory. Survey components can be added, removed or changed by respectively, creating a new XML file, deleting the file or editing its contents. The easiest way to get started on editing questions is to copy the default files over from the phone to your laptop and to start changing the existing questions, copying the files back when you are satisfied. [3] You can always reinstate the default question sets by deleting the questions directory in its entirety; then, the next time you start up the app it will be remade with all the default questions in it.

The file format for describing the questions is simple and uses standard XML. That means that the file consists of a number of nodes, which look as follows:

```
<node attribute1="value" attribute2="value">
    ...
</node>
```

Here, ... represents a space where further nodes can go, which would be this node's children (e.g. options to a choice question, or questions inside a category). In many cases, a node will not have any children and in this case we can use a short-hand notation:

```
<node attribute="value"/>
```

The survey XML files should always start with a `questionnaire` node, which has all the questions as its children:

```
<questionnaire title="Canis-tionnaire">
   ....
</questionnaire>
```

Adding open questions or yes/no questions is easy:

```
<yesno-question title="Do you own a pet dog?"/>
<open-question
   title="Which is your favourite dog race?"/>
```

Presenting the user with a list of options works as follows:

```
<choice-question
   title="Which dog toy is the most fun?">
 <option title="Tennis ball"/>
 <option title="Chewing stick"/>
 <option title="Pulling rope"/>
</choice-question>
```

Change `choice-question` into `multichoice-question` to make this a 'tick all that apply' question, rather than a 'choose one'. Also, you can add: `<other-option/>` to allow the user to type in another answer.

Finally, you can group questions together into a screen using the `category` node:

---

[3] open the XML files in Notepad, or e.g. Notepad++; do not use Word!

```
<category title="Dog food">
    ...
</category>
```

Note that the app will automatically number the questions for you. A category will also show a small ">" mark just before the question title, to show the user that they will navigate to a new question sub-screen. Let's look at a small example with four questions, corresponding to the output shown in figure B.1:

```
<questionnaire title="Example">
    <open-question title="Example open"/>

    <category title="Example group">
        <yesno-question title="Example yes/no"/>
        <choice-question title="Example choice">
            <option title="First option text"/>
            <option title="Second option text"/>
            <other-option/>
        </choice-question>
    </category>
    <multichoice-question title="Example multi-choice">
        <option title="Never select this!"/>
        <option title="First item selected"/>
        <option title="Don't even think about it!"/>
        <option title="Second item selected"/>
        <option title="Third item selected"/>
    </multichoice-question>
</questionnaire>
```

Note that you are free to leave white space or blank lines wherever you wish. The only restriction is that you are not allowed to use semi-colons (;), backslashes (\) or ampersands (&) anywhere. Further, each (multi-)choice question can only have one `other-option`.

Extra options

Should you find yourself doing a survey that has the same answers most of the time, you can make the app remember your answers (rather than having them be cleared when you complete the survey). To do this, add the `sticky="true"` attribute to your `questionnaire` node (after the title and before >). This can cause confusion on which questions you have already answered, so use this option sparingly.
In cases where a single question set needs to be answered repeatedly in a single survey (e.g. multiple persons need to be interviewed), add `repeatable="true"` to the questionnaire node. When you open the question set, the app will ask you whether to continue with the existing answers or to move to a new set.

The most advanced option relates to the map screen shown at the beginning of each survey. It is possible to show markers (small orange dots) in a pre-determined set of locations on this map, which can be useful to guide the surveyor on which locations need to be visited. In the main directory of the app (`/Android/data/com.gpstionnaire`) create a file called: `intended_locations.csv`, with the coordinates of the locations in the following format:

```
latitude;longitude
12.345;0.000
78.9012;-1.1111
    ....
```

The final option is to change the picture shown when the user just opens the app. You may want to change the picture to something that relates better to your research; feel free to add any symbols and logos. The image is located in the `questions` directory and should be named `background.jpg`. Images should not be too large, certainly not larger than 2000x4000 pixels; otherwise they will not render on small phone screens.

Note that the GPStionnaire app is Free Software and is released to the public under the GNU General Public License.

FULL QUESTIONNAIRE OF THE BUILT ENVIRONMENT INVENTORY

```xml
<?xml version="1.0" encoding="utf-8"?>
<questionnaire title="Built environment inventory">

        <category title="Walking surface">

                <choice-question title="Dedicated sidewalk">
                        <option title="No sidewalks"/>
                        <option title="Berm only"/>
                        <option title="Sidewalk one side"/>
                        <option title="Sidewalks both sides"/>
                        <option title="This is a footpath"/>
                </choice-question>
                <choice-question title="Sidewalk pavement (best of both sides)">
                        <option title="No sidewalk"/>
                        <option title="Unpaved sidewalk"/>
                        <option title="Pavement in bad condition"/>
                        <option title="Pavement in decent condition"/>
                        <option title="Pavement in good condition"/>
                </choice-question>
                <choice-question
                        title="Side-walk continuity (best of both sides)">
                        <option title="No sidewalk"/>
                        <option title="Small sections only"/>
                        <option title="Longer sections"/>
                        <option title="Full-length"/>
                        <option title="Full-length including crossing aids at intersections"/>
                </choice-question>

                <open-question
                        title="Path width (m)"/>

                <yesno-question
                        title="Drainage channel presence"/>

                <choice-question
                        title="Drainage channel covers">
                        <option title="Not applicable (no channel)"/>
                        <option title="No covers"/>
                        <option title="Intermittent"/>
                        <option title="Full length"/>
                </choice-question>

                <choice-question
                        title="Slope">
                        <option title="Flat"/>
                        <option title="Mild"/>
                        <option title="Medium"/>
                        <option title="Heavy"/>
                </choice-question>

                <choice-question title="Nr. of permanent path obstructions (e.g. utility poles, fencing)">
                        <option title="NA / there is no footpath"/>
                        <option title="No obstructions"/>
                        <option title="Some obstructions"/>
                        <option title="Unavoidable obstructions"/>
                </choice-question>

                <choice-question title="Nr. of street shops obstructing the path">
                        <option title="NA / there is no footpath"/>
                        <option title="No obstructions"/>
                        <option title="Some obstructions"/>
                        <option title="Unavoidable obstructions"/>
                </choice-question>
                <choice-question title="Nr. of parked cars obstructing the path">
                        <option title="NA / there is no footpath"/>
                        <option title="No obstructions"/>
                        <option title="Some obstructions"/>
                        <option title="Unavoidable obstructions"/>
                </choice-question>

                <choice-question
                        title="Crowdedness">
                        <option title="Empty"/>
                        <option title="Low people presence"/>
                        <option title="Medium presence, some obstruction"/>
                        <option title="High presence, big crowds"/>
                </choice-question>

        </category>

        <category title="Street + Traffic safety">

                <open-question
                        title="Street width (m)"/>
                <choice-question title="Road surface condition">
                        <option title="Unpaved road"/>
                        <option title="Paved road in bad condition (dusty, large holes)"/>
```

```
                        <option title="Paved␣road␣in␣decent␣condition␣(small␣holes)"/>
                        <option title="Paved␣road␣in␣good␣condition␣(no␣holes)"/>
                        <option title="Paved␣road␣in␣excellent␣condition␣(with␣road␣markings)"/>
                </choice-question>
                <yesno-question
                        title="Speed␣bumps"/>
                <open-question
                        title="Traffic␣speed"/>
                <yesno-question
                        title="Traffic␣island␣/␣crossing␣aid"/>
                <yesno-question
                        title="Presence␣of␣minibuses"/>
                <yesno-question
                        title="Presence␣of␣motorbikes"/>
        </category>

        <multichoice-question
                title="Road␣features␣either␣side␣of␣segment">
                <option title="Roundabout"/>
                <option title="Dead␣end"/>
                <option title="Three-way␣intersection"/>
                <option title="Four-way␣intersection"/>
                <option title="Inferior␣road␣merging"/>
        </multichoice-question>

        <multichoice-question
                title="Destinations">
                <option title="Motorbike␣taxi␣stand"/>
                <option title="Minibus␣taxi␣park"/>
                <option title="Street␣market␣or␣kiosks"/>
                <option title="Enclosed/Strip␣mall"/>
                <option title="Other␣shop␣buildings␣(e.g.␣supermarket)"/>
                <option title="School␣or␣church"/>
                <option title="Field,␣park␣or␣court"/>
                <option title="Water␣tap␣point"/>
        </multichoice-question>

        <category title="Personal␣safety">

                <yesno-question title="Street␣light␣presence"/>

                <yesno-question title="Stray␣dogs␣presence"/>

                <choice-question title="Liveliness">
                        <option title="Empty"/>
                        <option title="Scattered␣people"/>
                        <option title="Groups␣of␣people"/>
                        <option title="Crowds␣of␣people"/>
                </choice-question>

                <yesno-question title="Children␣presence"/>
                <yesno-question title="Beggars␣or␣homeless␣people"/>
                <yesno-question title="Police␣or␣private␣security␣guards"/>
        </category>

        <category title="Streetscape␣+␣views">
                <choice-question title="Litter">
                        <option title="None"/>
                        <option title="Low:␣scattered␣and␣avoidable"/>
                        <option title="Medium:␣piles␣of␣garbage"/>
                        <option title="High:␣big,␣unavoidable␣heaps"/>
                </choice-question>

                <yesno-question
                        title="Presence␣of␣trees␣or␣greenery"/>

                <yesno-question
                        title="Presence␣of␣smells␣or␣heavy␣air␣pollution"/>

                <yesno-question
                        title="Presence␣of␣disturbing␣noises"/>

                <yesno-question title="Presence␣of␣overhead␣contact␣wiring"/>

                <yesno-question title="Abandoned␣buildings␣or␣construction␣works"/>
                <yesno-question title="Advertisments␣or␣signage"/>

                <open-question title="Building␣height␣(stories,␣largest)"/>

        </category>

        <multichoice-question
                title="Buildings">
                <option title="Offices"/>
                <option title="Detached␣dwellings␣(compounds)"/>
                <option title="Single-story␣dwellings␣(not␣walled␣off)"/>
                <option title="Multi-story␣dwellings␣mixed␣with␣commercial"/>
                <option title="Multi-story␣dwellings␣(flat/apartment␣block)"/>
                <option title="Dedicated␣commercial␣or␣services␣(e.g.␣hospital)"/>
                <option title="Industry"/>
        </multichoice-question>

        <multichoice-question title="Area␣characterisation">
                <option title="Slum"/>
                <option title="Sub-urban"/>
                <option title="Urban␣medium␣density"/>
                <option title="Urban␣centre"/>
                <option title="Industrial"/>
                <option title="Upper␣class"/>
        </multichoice-question>
</questionnaire>
```