



**EMPOWER**  
Rewarding Change

# EMPOWER GUIDANCE NOTES

## 6. Evaluation

Section 1: Evaluation overview	3
Section 2: Evaluation in a Living Lab setting	5
Section 3: Data sources in EMPOWER	6
Section 4: Typology of schemes and recommended data sources	7
Section 5: Key points, critical success factors and limitations	8
Section 6: Further resources	9
Section 7: References	9
Section 8: Annex 1 - EMPOWER KPIs	10
Section 9: Annex 2 - Sample Survey Questions	11
Section 10: Annex 3 - vulnerability index	16

EMPOWER is setting out to show the effect of positive incentives on urban transport behaviour, so evaluation of the effectiveness of these when implemented is of crucial importance to the project. Evaluations range from simply defining a scheme's effectiveness in meeting narrowly defined targets to defining its effect on society as a whole. More narrowly defined evaluations will tend to consider direct effects (number of users), while more holistic evaluations will require more elaborate data (e.g. users' behaviour before using the service). In the case of the EMPOWER project, the projects' Key Performance Indicators (KPIs) (see 2.2.2 Indicators) constitute the minimum criteria for evaluation, although a wider scope is recommended.

Evaluations:

- Allow for changes to be made (if the measure is still in operation), or next time
- Allow for lessons to be learned for future measures, locally or further afield
- Quantify the effect of the measure (as opposed to external factors)
- Provide arguments for a measure to be continued or expanded
- (Especially ex-post) provide the information and knowledge on which to base future ex-ante evaluations.
- Allow for a case to be made for the continuation or expansion of a policy

A solid evaluation requires many things to be defined or considered:

Aspect	Description
<b>Overall goals of the evaluation</b>	In general terms, what is trying to be proved by the evaluation?
<b>Indicators</b>	What aspects should the policy be measured by? What is considered success?
<b>Scope (geographic)</b>	Where should any effects be measured?
<b>Scope (temporal)</b>	After how long, and for how long should any effects be reasonably expected?
<b>Perspective</b>	From which (or whose) perspective should the effects be considered? This will affect the definition of the other aspects.
<b>Users</b>	Which users will be considered? Only those of the measure/app, or all transport users?
<b>Trip types</b>	Will all trips be considered, or only e.g. commutes?
<b>Cases (see over-leaf)</b>	The before (baseline), after and BAU (business as usual) cases for comparison of the behaviour of users with non-users. Helps establish the causality of any effects observed

Table 1. Aspects of an evaluation which must be defined.

## Evaluation cases

### Before case

This describes the behaviour of the whole sample group in terms of trips (modal split, number and distance/duration) before implementation, e.g. a week before, or the corresponding month in the previous year.

### After case

This describes the behaviour of people who have use of the measure ('users'), e.g. an app, after a certain period, or over distinct intervals, e.g. a week.

### Control or BAU case

This is the same as the after case, but for people who do not have the service/app. Thus this group can be thought of as exhibiting the behaviour that the whole group would have done if the service did not exist.

### Comparison: before vs. control

This shows how large the change in behaviour was without the measure, i.e. from external factors, e.g. weather or major events which affect people's transport behaviour. Figure 1 shows a theoretical graph of a high and low resolution of the number of trips made by bicycle per week, illustrating why it is important to know what would have happened.

### Comparison: before vs. after

This shows how much users' behaviour changed from before the measure to after/during it. This can be thought of as the effect of the measure in addition to the effect of external factors.

### Comparison: after vs. control

This shows the relative change in behaviour between the two groups. In other words, it is the effect of the measure corrected for the effect of external factors. Figure 2 illustrates the importance of a reliable BAU or control case by showing the actual effect of the measure (shaded blue) for different BAU/control (dashed red line) scenarios (assuming the same change in behaviour of users – blue line).

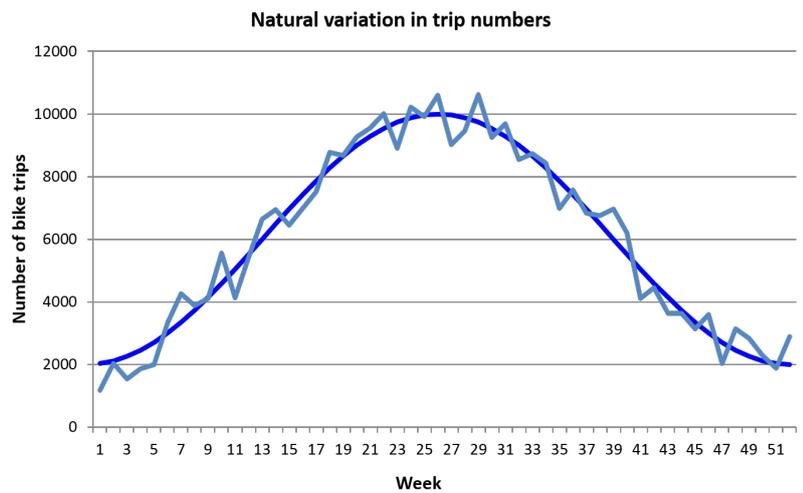


Figure 1. Illustrative example of the weekly variation in behaviour compared to the trend.

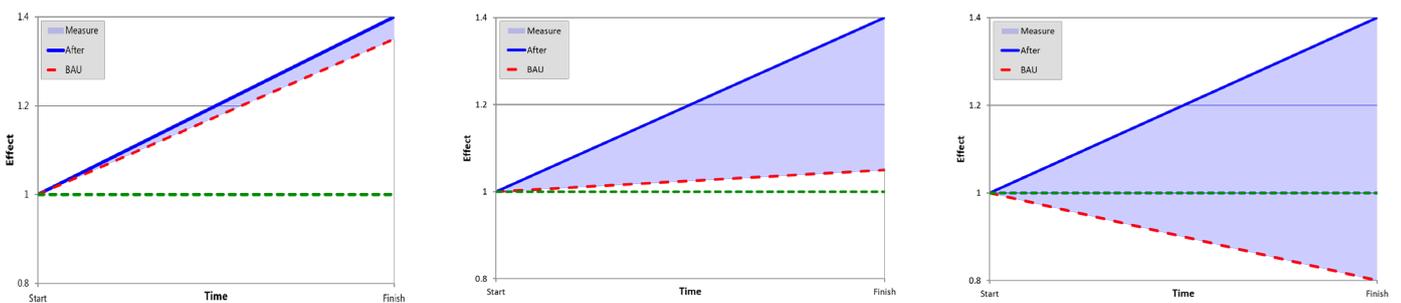


Figure 2. The effect of the control case on the apparent effect of the measure.

Working in a living lab setting has many advantages, such as the ‘real world’ relevance of any results. Unfortunately, there are also disadvantages stemming from the fact that researchers are very much constrained by practical considerations. Table 2 contains a (non-exhaustive) list of difficulties and sources of bias or errors posed by the living lab setting, specific to the EMPOWER project, based on a broad literature review (A.D.C. Automotive Distance Control Systems GmbH et al., 2014; Brög et al., 2009; Dziekan et al., 2013; Goodwin, 2004; Ker, 2008; Ortuzar et al., 2011; Richardson et al., 2004; van Rooijen and Nesterova, 2013) and personal experience observing the implementation and evaluation process of the EMPOWER schemes.

Item	Description
Low number of users	With a small number of users overall, many sub-groups will be populated by too few users to draw significant conclusions.
Difficult to get before/control cases	Without the scheme’s incentives, users have no obvious benefit to participate, making the before and control tracking data difficult to collect.
Difficult to organise LL in the first place, so evaluation not top priority	If the organisers of a scheme are not a key stakeholder themselves (i.e. a municipality), the schemes are especially difficult to implement. Starting the scheme at all becomes a natural priority (if there is no scheme, there is nothing to evaluate).
Data access from external sources	External stakeholders may not be able to secure access to crucial data (bus ridership figures, traffic counters).
Survey questions and length	The longer a questionnaire, the fewer responses can be expected, so a balance must be drawn between obtaining a sufficient number of responses and sufficient answers per response.
Being observed effect (socially desirable or undesirable behaviour)	If users know about the purpose of a scheme, they may wish to over-report the behaviour they think is expected. This may affect the control and before cases and any questionnaires.
Extraneous factors	The weather, drastic changes in fuel of public transport prices and holidays (amongst others) should be controlled for to ensure that the effect of schemes is correctly attributed.
Technical problems gathering data	Technical tools may not distribute questions at the correct time or at all.
Shifting user groups (i.e. can’t track users across all conditions)	The individuals which make up the group ‘users’ will inevitably change over time. As each user will be different in their behaviour and the tendency to change, so changes in the behaviour of the ‘users’ as a group will be affected, to a degree, by the difference in behaviour of the individuals, rather than the individuals changing their behaviour.
Imperfect recollection of behaviour	People have difficulty accurately recalling their past behaviour regarding trip number, distance and mode, and may recall time differently whether moving or stationary.
Imperfect (unrepresentative) user groups/self selection	In schemes which recruit users by passively making potential users aware of the schemes, and only those who are interested sign up, there is a significant, probably large element of self-selection bias. Any results are thus not automatically transferrable to the general population.

Table 2. List of difficulties in evaluation in a Living Lab setting.

To minimise the errors and biases introduced in Table 2, and to maximise the general accuracy of the evaluation, Figure 3 overleaf shows a conceptual ‘onion model’ of data reliability and applicability used to guide the evaluation in EMPOWER, with the reliability or applicability increasing toward the centre.

### App tracking data

The data provided by the apps used in EMPOWER is very fine-grained, allowing analysis of individuals on a trip-by-trip basis. This contrasts with conventional evaluation of transport measures – especially for walking or cycling measures – which revolves around travel censuses taken at intervals of a year or longer, and/or (fairly few) traffic counters scattered around a city. Meshing these two data forms is not easy, and no precedents of comparable processes could be found.

### App usage data

Usage of the app is the ultimate sign that users are satisfied with the service. Further insight can be gained by examining the patterns of usage, measured in downloads, active use, frequency of use, length of use, challenges accepted etc.

### Traffic counters

If the city (or other entities) operate traffic counters, these can provide valuable information on the traffic situation.

### Surveys, interviews/focus groups

Surveys can be in the form of paper questionnaires, or the same via email. Additionally, particular questions can be delivered in-app, for example just before or after a user accepts or completes a particular challenge. The questions can also be used to cross check the evaluation of behavioural change, especially in the case of imperfect before or BAU data. Interviews/focus groups can be considered a more in-depth and personalised version of surveys. A set of common possible questions (see 0 Appendix 1: sample survey questions) covering all possibilities of data availability was developed for the project.

### Employee/group information

For schemes which deal primarily with employees rather than the wider public, the employer may be able to provide basic demographic or travel behavioural information on their employees.

### Travel censuses

(Travel) censuses (such as Germany's Mobility in Germany) are wide-ranging, often national reports on people's travel behaviour performed at fairly long intervals, in Germany's case 5 years.

### Local external events or conditions

Information on local conditions, such as the weather, special events or holidays can provide valuable information on why traffic behaviour has changed. Local public transport companies may also have useful local ridership information.

### Data/results from other cities or projects

Other cities or schemes and projects may have results which can be used to bolster the evaluation in EMPOWER-type schemes.

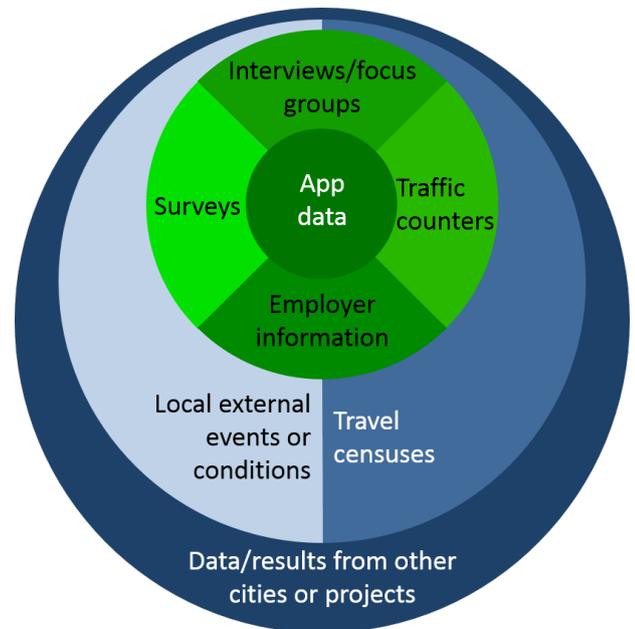


Figure 3. 'Onion model' of reliability of data sources (greatest reliability or applicability in the centre).

The first consideration is whether the scheme can make use of data from a tracking app or not, and within that, the availability of ‘off-incentive’ tracking data – periods in which users’ travel behaviour was being recorded, but they were not being incentivised to change that behaviour - to provide the control/before case. The second point of differentiation is whether the projects point of contact is with employers or end-users directly.

		Contact with users	
		Active/personal	Passive/Indirect
Tracking data	With off-incentive data	Type 1	Type 2
	Without off-incentive data	Type 3	Type 4
	None	Type 5	Type 6

Table 3. Typology of scheme-type regarding available data and evaluation.

#### Common recommended data source

Indicator	Source
CFV use: reason for change	Surveys Interviews/focus groups Employee/group information Travel censuses Local external events or conditions Data/results from other cities or projects
Accessibility & attractiveness Vulnerable groups	Surveys Interviews/focus groups Employee/group information
User satisfaction	App usage data Surveys Interviews/focus groups Social media feedback/app ratings
Impact and legacy aspects	Personal judgement of project members Interviews with relevant stakeholders Media reports of developments

Table 4. Recommended primary data sources for all city types.

- Agree on data collection early on – baselines are especially difficult to define.
- Formulate distinct cases and boundaries.
- Record when and how the scheme is altered (e.g. change the incentives).
  - New questionnaires may be needed.
- Be aware that there might be a trade-off between the amount of data collected and the rate of participation.
- The evidence provided by the evaluation is only as good as:
  - The data upon which it is based.
  - The definition of the before, after and control cases.

City type	Before & BAU cases	After case
Type 1	Surveys Interviews/focus groups Employee/group information Travel censuses Local external events or conditions Data/results from other cities or projects	Surveys Interviews/focus groups App tracking data Employee/group information
Type 2	Surveys App tracking data Traffic counters	Surveys App tracking data
Type 3	Surveys Interviews/focus groups Traffic counters Employee/group information Travel censuses	Surveys Interviews/focus groups App tracking data Employee/group information
Type 4	Surveys Traffic counters Travel censuses	Surveys App tracking data
Type 5	Surveys Interviews/focus groups Traffic counters Employee/group information Travel censuses	Surveys interview/focus groups Employee/group information
Type 6	Surveys Traffic counters Travel censuses	Surveys
All	Local external events or conditions Data/results from other cities	

Table 5. Type-specific recommended data for indicator 'CFV use: measured change'

## Section 6.

## Further resources

**TIDE** ([www.tide-innovation.eu](http://www.tide-innovation.eu)): Impact assessment methodology for urban transport innovations - A handbook for local practitioners <http://www.tide-innovation.eu/en/Results/Impact-assessment-methodology-for-urban-transport-innovations-A-handbook-for-local-practitioners/>

**FESTA V model FESTA V model.** Handbook: [http://fot-net.eu/wp-content/uploads/sites/7/2014/06/FESTA\\_HBK\\_rev5\\_ref3-2.pdf](http://fot-net.eu/wp-content/uploads/sites/7/2014/06/FESTA_HBK_rev5_ref3-2.pdf)

**NISTO 2013.** New Integrated Smart Mobility Options NISTO Evaluation of Approaches and Recommendations.

## Section 7.

## References

A.D.C. Automotive Distance Control Systems GmbH, Association of French Motorway Companies (ASFA), BMW, Chalmers University of Technology (SAFER), Daimler, Delphi France, European Road Transport Telematics Implementation Coordination Organisation (ERTICO), Federal Highway Research Institute (BAST), FIAT Research Centre (CRF), Fraunhofer Society for the advancement of applied research, French National Institute for Transport and Safety Research (INRETS), Galicia Automotive Technological Centre (CTAG), INFOBLU SPA, International Automobile Federation (FIA), ITS Bretagne, Loughborough University, Ministry of Transport, Public Works and Water Management (RWW), Netherlands Organization for Applied Scientific Research (TNO), Orange France, Promotion Of Operational Links With Integrated Services, Association Internationale (POLIS), Renault, Robert Bosch GmbH, Swedish National Road and Transport Research Institute (VTI), Technical Research Centre of Finland (VTT), Technical University of Aachen (IKA), University of Koln, University of Leeds, Volvo Car Corporation, Volvo Technology Corporation, 2014. FESTA Handbook Version 5.

Brög, W., Erl, E., Ker, I., Ryle, J., Wall, R., 2009. Evaluation of voluntary travel behaviour change: Experiences from three continents. *Transp. Policy*. <https://doi.org/10.1016/j.tranpol.2009.10.0003>

Dziekanski, K., Riedel, V., Müller, S., Abraham, M., Kettner, S., Daubitz, S., 2013. Evaluation matters: a practitioners' guide to sound evaluation for urban mobility measures. Waxmann, Münster.

Goodwin, P., 2004. Valuing the small: counting the benefits.

Ker, I., 2008. Measuring travel behaviour change. *Road Transp. Res. J. Aust. N. Z. Res. Pract.* 17, 71.

Ortuzar, J. de D., Armoogum, J., Madre, J.-L., Potier, F., 2011. Continuous Mobility Surveys: The State of Practice. *Transp. Rev. Taylor Francis Routledge* 31. <https://doi.org/10.1080/01441647.2010.510224>

Richardson, A.J., Seethaler, R.K., Harbutt, P.L., 2004. Design issues for before and after surveys of travel behaviour change. *Transp. Eng. Aust.* 9, 103.

van Rooijen, T., Nesterova, N., 2013. Applied framework for evaluation in CIVITAS PLUS II (Deliverable No. 4.10), CIVITAS WIKI – Coordination, evaluation and dissemination of CIVITAS PLUS II.

The primary indicators for the EMPOWER evaluation are the Key Performance Indicators (KPIs) as listed in the project's Description of Action (in italics), listed below.

### **Conventionally-fuelled vehicles (CFV) use**

*15%-50% reduction in the use of conventionally fuelled vehicles in cities, using packaging and synergies. (Measured by reductions in Vehicle Kilometers travelled by Conventionally Fuelled Vehicles).*

This can be quantified based on users' behaviour: the number, distance and duration of their trips, differentiated by mode, whether trips are at peak/off-peak times.

### **Urban accessibility and attractiveness**

*30% increase in travellers' self-reported positive evaluation of urban accessibility and attractiveness (measured through feedback mechanisms including social media and questionnaires, disaggregated to establish impact on vulnerable groups and gendered effects).*

Accessibility refers to the ease with which people in a city can travel to the places they need to go about their day-to-day tasks. EMPOWER is interested in whether its schemes change users impression of accessibility and attractiveness; the best and easiest way to define that out is to ask users directly.



Figure 5. Graphic of Transport for London's PTAL accessibility indicator (red areas have the best accessibility) .

### **Customer/user satisfaction**

*75% Customer/user satisfaction with the EMPOWER mobility service (measured through feedback mechanisms including social media and questionnaires, disaggregated to ensure representation of vulnerable travel groups and gender).*

Possible indicators for satisfaction include the number of downloads of the app or active users, the level of engagement with the services (general frequency of use, number of challenges accepted/achieved, rewards gained etc.) or app-store ratings, however the latter may be at the extreme ends of the opinion spectrum. Measurement via their engagement on social media platforms is also possible, via retweets, comments, replies, likes and the number of participants can be analysed via social media analytics tools (such as Facebook Insights, Twitter Analytics etc.).

Unfortunately for the project, from a users perspective, 'the service' consists of several separate aspects, not all of which are within the project's control, and regardless it would be helpful to know which parts users (and non-users) liked or disliked. In order to capture this detail, surveys are useful, either in-app (e.g. post registration or after completing a challenge) or more conventional paper or online/email versions.

## Vulnerable groups

10% response rate for vulnerable travel groups (we will actively approach vulnerable groups and aim for a minimum active contribution from this approached group of 10%)

For EMPOWER, a social or accessibility-based approach to vulnerability was adopted, as opposed to the common definition of pedestrians and cyclists as vulnerable.

To assess vulnerability across aspects, an indicator for the overall vulnerability of users was developed to give a single answer of whether a user is vulnerable or not, but also because it accounts for the intersectionality of vulnerability.

Impact and legacy aspects

EMPOWER-style incentive schemes may have a benefit for sustainable transport in forms which are intangible or indirect, but which still make a significant contribution to making the transport system in a city more sustainable or equitable.

Category	Impact and legacy aspect
Instrumental	Influencing development of policy Influencing practice Influencing service provision Shaping legislation Altering behaviour
Conceptual	Contributing to the understanding of policy issues Reframing debates Introduction of new concepts
Capacity building	Technical development Skill development Partnership building

Table 6. Impact and legacy aspects.

## Section 9.

## Annex 2 - sample survey questions

Following are tables containing the set of all suggested questions to be included in surveys. The questions cover all scheme types, and a selection would need to be made based on the data available from other sources locally. Providing these questions served two purposes: providing guidance to the scheme organisers, and in doing so, ensuring the maximum possible commonality in the questions asked across the schemes.

Notes: all appropriate questions had a 'Rather not say' option, omitted here to save space. #### should be replaced with the local scheme name.

The questions are provided in three different levels, depending on the level of detail required, as a trade off with the time required to complete - and thus probable response rates. The green-coloured questions are the minimum required to address the KPIs, while the orange-coloured questions go into much more depth in the explanation of why users answered or acted as they did.

**Most basic and quick, most important questions**

Moderately basic and quick

Most detailed and time consuming

Vulnerable groups, gender & demographics

<b>What is your gender?</b>
Female
Male
Other

<b>What is your age?</b>	<b>(Alternative) In what stage of life are you?</b>
14 to 17/18 (local driving age)	School
19 to 30	University
31 - 40	Work
41-50	Unemployed
51-60	Retired
Over 60 (or local retirement age)	

<b>Do you have any mobility impairments?</b>
Yes
No

<b>What is your current living situation?</b>
I live with my parents/relatives
I live with friends
I live with my partner without kids at home
I live with my partner and kids at home
I live with my kids without a partner at home
I live alone
Other

<b>Which of the following describes your travel budget situation best?</b>
Comfortable - it is not an important factor in my modal choice
A bit tight - it influences my modal choices
Quite tight - I try to use low-cost mobility options if possible

**Please indicate the importance of different factors on your modal choice?**

Travel costs	Likert scale from very relevant to not relevant
Convenience to travel with children or to transport goods	
Comfort	
Security concerns (e.g. while travelling at night)	
Traffic safety concerns	
Environmental considerations	
Health considerations	
Other	

CFV usage

**What is your average travel distance by car on weekdays in km?**

Open field

**Has the use of #### use affected your use of cars?**

Increased/No effect/Decrease

**Did your travel behaviour change since you use #####?**

Car/motorbike/moped as driver with no passenger	Increased No effect Decrease
Car share as driver with one or more passengers	
Car share as passenger	
Public transport	
Bicycle	
Walk	
Other	
My travel behaviour did not change	

**Do any of the following apply regarding your access to a car?**

Do you have a driver's licence?	Yes	If yes > What type of car do you have access to?	
Do you own a car?		No	Petrol car
Do you have access to a car (e.g. a car is available in the household)?			Diesel car
	Hybrid car		
	Electric car		
	CNG car		
	Other		

**What ##### functionality influenced your decision to travel differently?**

My mobility information (overview of CO2, calories, etc.)	Likert scale for each: 1 = most influence, 3 = least influence
Route planner	
Challenges	
- If yes, which challenges?	
##### points	
Traffic info (heatmap of traffic situation)	
Other	
My decision to travel differently was not influenced by #####	

Do any of the following apply?	Walking	Cycling	Public Transport
It's too expensive to use	Y/N	Y/N	Y/N
It's too difficult to transport children	Y/N	Y/N	Y/N
I feel unsafe travelling alone, especially at night	Y/N	Y/N	Y/N
Difficult to use because of mobility impairments	Y/N	Y/N	Y/N
Difficult to use because of atypical travel times, origins or destinations	Y/N	Y/N	Y/N
I don't use it in bad weather	Y/N	Y/N	Y/N
I don't use it because of insufficient facilities or infrastructure	Y/N	Y/N	Y/N
I don't use it because of lack of information	Y/N	Y/N	Y/N

**What is your average weekly distance in car?**

Open field
------------

**Please order your three main modes of transport of your daily travel activities?**

Car/motorbike/moped as driver with no passenger	Likert scale for each: 1 = most influence, 3 = least influence
Car pooling as driver with one or more passengers	
Car pooling as passenger	
- If yes, which challenges?	
Public transport	
Bicycle	
Walk	
Other (open field)	

## Attractiveness & accessibility

### Did the use of ##### change the following for you?

Attractiveness of public transport	Increased No effect Decrease
Attractiveness of walking/cycling	
Your ability to reach the places you travel	

## User satisfaction

### Installing and using ##### was simple

Likert scale from Completely agree to Completely disagree

### Are you planning to keep using ##### in the future?

Yes, please feel free to provide suggestions for improvement below

No	If no -> If you plan to stop using the app, why?	Yes No
I don't know	It's irrelevant to me	
	The incentives aren't appealing	
	Excessive battery use	
	It's too difficult to use	
	Privacy concerns	
	I ride/walk/use PT anyway	
	Other	

A definition of vulnerability has been formulated based on vulnerability regarding transport accessibility, and from that an index has been developed to measure the overall vulnerability of respondents to questionnaires. Based on individuals' responses to questions regarding their income, mobility budget, physical mobility, age, gender, living situation, nation of birth and education, respondents are allocated a vulnerability score. This accounts for the individual aspects, but also for the 'intersectionality' of vulnerability, i.e. individuals may be vulnerable because of one specific aspect (e.g. disability), they may also be vulnerable because of multiple aspects, if assessed in isolation, wouldn't classify the individual as vulnerable. Practically, the definition is considered to include the following groups:

- Low-income groups
- People with stretched mobility budgets
- Children, youths and the people caring for them
- Women (travelling alone)
- The elderly
- People with disabilities
- Migrants
- Lower education groups
- Other possible groups (indicative)
  - Workers with atypical travel needs
  - Particularly late/early hours
  - Remote workplaces – needing to travel to and from, but especially between remote workplaces
  - Those who are poorly served by PT services and/or AT infrastructure

The index applies the points, on the basis of the descriptions, to answers given in surveys, and sums the points for each respondent, giving their 'vulnerability score'. Scores of 3 and above would classify a respondent as vulnerable.

Points	Definition
0	Not vulnerable
1	Somewhat or potentially vulnerable
2	Quite vulnerable
3	Vulnerable

Table 7. Description of points allocation of vulnerability answers.

Table 8, overleaf, gives some examples of the allocation of points for various answers to survey questions. Variations would have to be made for local conditions, especially the income level which would constitute vulnerable.

Questions	Possible Answers	Points
How would you describe your mobility budget?	Ample: It does not play a role in the choice of my transport mode Tight: It affects the choice of my mode of transport Very tight: It makes me choose the cheapest mode of transport	0 1 3
Do you have some physical limitations that make it difficult for you to cycle or walk properly?	Yes No	3 0
What is your gender?	Female Male	1 0
What is your age?	>10 11-15 16-18 All 19-64 65-70 71-75 >76	3 2 1 0 1 2 3
What is your country of birth?	Open	1 for non-local
What is the highest education diploma you have obtained?	None Secondary Vocational/trade school University	3 2 1 0
What is your current living situation?	Single with children Couple with children All others	2 1 0
What is your gross monthly income?	<€999 €1-1499 €1500-1999 >€2000	3 2 1 0
Can you comfortably reach your workplace/school on foot, by bike or with public transport at the times to need to?	Yes No	0 1

Table 8, Vulnerability-point allocation for specific responses to survey questions.