CiViTAS
Cleaner and better transport in cities

POINTER

Handbook on Evaluation for cities

Deliverable 4.4.2

Version: FINAL

Date: 5 August 2013

THE CIVITAS INITIATIVE
IS CO-FINANCED BY THE
EUROPEAN UNION

Deliverable No.	D 4.4.2	
Deliverable Title	Handbook on Evaluation for cities	
Dissemination level	Public	
Status and version	Final	
Written by	Stephanie Müller (POYRY), Katrin Dziekan, Veronique Riedel, Michael Abraham, Stefanie Kettner, Stephan Daubitz (TU Berlin)	
Verified by	Tariq van Rooijen (TNO)	
Approved by	Tariq van Rooijen (TNO)	
File name	D4.4.2-POINTER-PU-Handbook on Evaluation for cities-FINAL-POYRY-Aug 2013.pdf	
Issue date	August 5, 2013	

Evaluation matters

A practitioners' guide to sound evaluation for urban mobility measures

Katrin Dziekan, Veronique Riedel,
Stephanie Müller, Michael Abraham,
Stefanie Kettner, Stephan Daubitz

Evaluation matters

A practitioners' guide to sound evaluation for urban mobility measures

Waxmann 2013
Münster / New York / München / Berlin

Bibliographic information published by Die Deutsche Nationalbibliothek
Die Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the internet at http://dnb.d-nb.de.

Acknowledgement
Work on this publication was co-financed by the European Commission – Grant Agreement No. TREN/FP7TR/218953 and TREN/FP7TR/219026.

Legal Notice
The views expressed in this publication are the sole responsibility of the authors named and do not necessarily reflect the views of the European Commission.

CiVITAS Cleaner and better transport in cities MIMOSA

CiVITAS Cleaner and better transport in cities POINTER

THE CIVITAS INITIATIVE IS CO-FINANCED BY THE EUROPEAN UNION

Consulting for mobility and transport

Chair of Integrated Transportation Planning

ISBN 978-3-8309-2881-2

Postfach 8603, 48046 Münster
www.waxmann.com
info@waxmann.com

Cover and Layout: Inna Ponomareva, Münster
Print: Tesinska Tiskarna, A.S., Czech Republic
Printed on the age-resistant paper, acid-free as per ISO 9706

Table of Contents

Preface

All over Europe urbanisation has been a clear trend over the past decades and is expected to continue with the proportion of the European population living in urban areas increasing from 72% in 2007 to 84% in 2050 (UN Department of Economic and Social Affairs/Population Division, 2008). Accordingly, urban mobility is of growing concern to citizens and authorities. Cities need efficient transport systems to support their economy and the welfare of their inhabitants.

A relevant question in this respect is what an efficient transport system should look like and what positive impacts this may have on the economy and quality of urban life. In other words: "How can we achieve cleaner and better cities across Europe?" This question turned out to be the motto for the CIVITAS Initiative that the European Commission launched in 2002.

CIVITAS stands for CIty – VITAlity – Sustainability, an initiative co-financed by the European Commission and coordinated by cities as a programme 'of cities for cities'. Its fundamental aim is to support cities in the introduction of ambitious transport measures and policies towards sustainable urban mobility. The goal is to achieve a significant shift in modal split towards sustainable transport, an objective reached through encouraging both innovative technology and policy-based strategies.

So far there were CIVITAS I (2002-2006) and CIVITAS II (2005-2009). The third programme, CIVITAS PLUS (2008-2013), is about to come to an end and will be followed by CIVITAS PLUS II (2013-2017). In the present programme there are five so-called collaborative projects, namely, ARCHIMEDES, ELAN, MIMOSA, MODERN and RENAISSANCE with a total of 25 demonstration cities taking part, implementing over 300 measures. From the beginning of CIVITAS, evaluation played a key role for the European Commission. A specific element is the so-called framework for evaluation. The framework has set the working structures along which all local urban evaluations have taken place since 2002. For the development of this structure the European Commission established horizontal support action teams in each programme phase: METEOR (CIVITAS I), GUARD (CIVITAS II) and POINTER (CIVITAS PLUS). The CIVITAS framework for evaluation has been developed by representatives of these support action teams, complemented with valuable comments from the members of the CIVITAS Advisory Committee. Specific acknowledgement goes to Mike McDonald, Jinan Piao and Richard Hall from the University of Southampton (Transportation Research Group) and Martin van de Lindt from TNO. Many of the examples presented in this handbook have a direct or indirect link to the CIVITAS framework and related guidance notes.

Each of the five collaborative projects in CIVITAS PLUS had a work package installed which was responsible for the city specific coordination and support of the measures' evaluations. In MIMOSA this task lay with the Chair of Integrated Transpor-

tation Planning at the Technical University Berlin who coordinated the activities for the production of this book. Together with POINTER, the support action team for CIVITAS PLUS, who provided a wider perspective on evaluation and reflecting the interest of the European Commission to develop towards a learning society, the authors wrote this guide based on their practical experience gained in four years of CIVITAS. If you are interested to learn more about the CIVITAS initiative please go to www.civitas.eu. However, this handbook covers a broader spectrum of evaluation activities than just CIVITAS evaluation; therefore, in addition to the CIVITAS documents several other sources and guidance notes have been taken into account.

Regarding the production of this handbook we specially like to thank Christine Ahrend (Technical University Berlin), Hans-Joachim Becker (Technical University Berlin, CIVITAS TELLUS, CIVITAS MIMOSA), Kerstin Burggraf (city of Dresden), Dirk Engels (Transport & Mobility Leuven/TML, CIVITAS ELAN) and Isabela Velázquez (gea21, CIVITAS ARCHIMEDES) for intensively commenting the draft version of this publication and Nicola Moczek (PSY:PLAN) for coordinating its production. Special thanks go to the cities of Utrecht and Tallinn for providing the raw material for the evaluation examples that are used as examples throughout this book.

Hamburg and Berlin, January 2013


I THINK....
WE HAVE A PROBLEM
WE WANT OUR CITY TO BE A PLEASANT PLACE...
FOR THE PEOPLE
WITH CLEAN AIR AND TREES
SPACE FOR CYCLISTS AND PEDESTRIANS
WHAT IS STANDING BETWEEN US AND OUR OBJECTIVE?
TOO MANY CARS!
SO, WHAT ARE WE GOING TO DO?
MORE PUBLIC TRANSPORT!
© 2012 www.aukeromme.nl
LATER
IS THIS WHAT WE WANTED?
I THINK SOMETHING WENT WRONG
TO BE CONTINUED

1 Introduction

Have you read the preceding comic? Even if the pictured discussion might seem a little superficial to you, it does demonstrate one of the main concerns with cities aiming to improve their liveability. There is a lot of talk about sustainability, about reducing car dependencies, about making cities greener, quieter and a nicer place to live. But what's the evidence of the interventions in place? Many cities and organisations successfully implement measures within a given timeframe and budget and produce outstanding outputs. But policymakers are sometimes afraid of a systematic evaluation. However, evaluation is more then assessing their policy, proving that money was not well spend and then finding someone to blame. Evaluation can help to improve measures during their implementation by looking for ways to optimise the processes or identifying aspects to focus on. It can help to ensure that results are generated along the lines of what was intended and that mistakes will not be repeated in the future.

This book will help you in conducting such a sound evaluation. It will guide you through all the steps which are necessary to draw meaningful conclusions from your findings. But before we take you into the realm of evaluation, we are going to define the term and the purpose of evaluation. Then you will get an overview of all the steps involved in an evaluation before the more detailed chapters begin.

This handbook is not intended as an exhaustive instructional guide for evaluation. It provides a framework for thinking about evaluation of mobility measures and outlines the evaluation task, either independently or with the support of an external evaluator/consultant. For more detailed guidance on the technical aspects of evaluation, you may wish to consult the sources recommended in each section or in the bibliography at the end of the handbook.

1.1 What is evaluation?

Scientifically speaking, evaluation is a systematic determination of a measure's merit and significance, using criteria governed by a set of standards. It is part of a continuing management process consisting of planning, implementation, and evaluation; ideally with each substituting the other in a continuous and simultaneous cycle until successful completion of the measure. In other words: evaluation tells you what really happened in your measure – compared to what should have – why it happened and what you can learn from these deviations. On top of that, evaluation will determine if you have reached your intended goals.

To understand the essence of evaluation studies, it is necessary to emphasise that evaluation is not to be confused with audit or monitoring. These terms should not be

mistaken for evaluation, although they can be (in specific cases) a tool for updating the data collected during evaluation as well as for the needs of analyses carried out during evaluation. The differences are subtle, but they are nonetheless important and the terms are thus defined in this evaluation handbook too. In comparison to this, an audit is only the verification of compliance of the use of resources (mostly financial) with the binding legal regulations and specific standards. It is thus a tool for the internal control. Monitoring is usually conducted simultaneously with the implementation and is designed for verifying this process, particularly the achievement of assumed outputs and results of the measures undertaken as well as inputs mobilised for their implementation.

So you see audit and monitoring can be used as the source of information for the evaluation. But while the monitoring is checking specific values, your evaluation is drawing right conclusions in the whole perspective of a measure. This is why it employs its own methodology, which you will learn in the course of this book.

1.2 What is the point of evaluation anyway?

Why is it important to conduct an evaluation if a project is running well and everything seems fine? Surprisingly, this question often pops up at city-level evaluation. Ironically, the answer is the same that drives us to learn from less successful pilot measures. Evaluation is a natural thing and every one of us does it in our everyday lives without thinking much about it. Have you ever thought about what made you choose to buy one product and not another? Or did you ever think about telling your friend how well your new lawnmower cuts the grass compared to the expectations you had because of its advertisement? In a general sense, this is evaluation. In the context of transport projects, the number of variables and stakeholders increase and turns the evaluation task into something more complex. But frankly, evaluation is a powerful tool for learning what works, what does not, and the reasons for this. So basically, we evaluate because we want to:

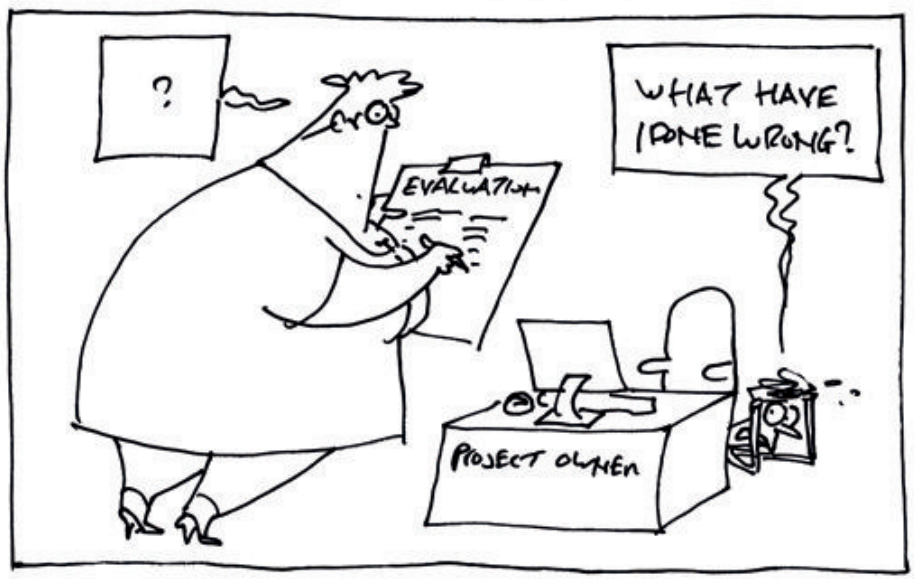
- measure the performance
- learn for future projects
- exchange experiences

Performance measurement means that through the application of proper experimental designs it is possible to quantitatively and qualitatively determine the implemented measure's effects on transport systems as well as on other related areas. This allows an appraisal of the measure's impacts. These conclusions can be used to legitimise the measure or to identify weak spots of its setup. For instance, through a time series

analysis of a traffic calming measure it could be discovered that an expected reduction in velocity and number of accidents did not take place. The aspired effects as well as the estimated economic benefits have not been achieved. Does this sound bad? Well, probably yes – but this is where the second function of evaluation becomes just as important. Through evaluation we want to learn for future projects. So, based on the identified weak spots the input parameters for future impact estimations can be adjusted. Or in other words, from the conclusions regarding the evaluation of the traffic calming measure other measures for the same street can be derived. Further, if we want to implement traffic calming measures elsewhere, we can better estimate their outcome because of this one measure which had unexpected results.

Since evaluation results should be made public, they also serve the purpose of improving measures for different places, different points in time and different stakeholders (see the chapter on up-scaling and transferability for more on this). For instance, the city of Utrecht, the Netherlands, is implementing a so-called Cargohopper, which is basically a small electric vehicle used to distribute goods in a very dense inner city. Their evaluation results – if made public – can serve as an example for other cities how they could replace heavy duty traffic in their city centre. And by showing the positive results, relevant stakeholders can be convinced to support the implementation of such a measure.

WHAT EVALUATION IS NOT



Evaluation thus delivers various benefits for everybody involved such as decision makers and (maybe most of all) for the citizens as it helps to:

- better understand public spending,
- better orient bundles of measures towards specific target groups,
- improve future planning and optimise the allocation of resources.

Remember that the human being is a judgemental creature. We tend to have an opinion on everything, all the more for what is publicly funded!

However, evaluation should never be conducted to primarily control people. Further, the initiation of 'sanctions' (e.g. payback of funding) must not be the goal of the evaluation. Such a focus would discredit the fundamental aim of evaluation and would impede innovative measure approaches.

1.3 How does evaluation work?

1.3.1 When should I evaluate?

Evaluation should run at all times of an implementation process: parallel to the planning, implementing and operating of your measure. It should be part of a continual development process by providing feedback about progress, encouraging reflection about outcomes and providing a basis for considering future strategies. Hence, evaluation is a set of interlinked activities. Each of these is an important part of the overall process and needs adequate time to keep the quality of the evaluation.

The figure 1-1 provides an overview with the common stages and key activities in project/programme planning, monitoring and evaluation. 'Common' stages because

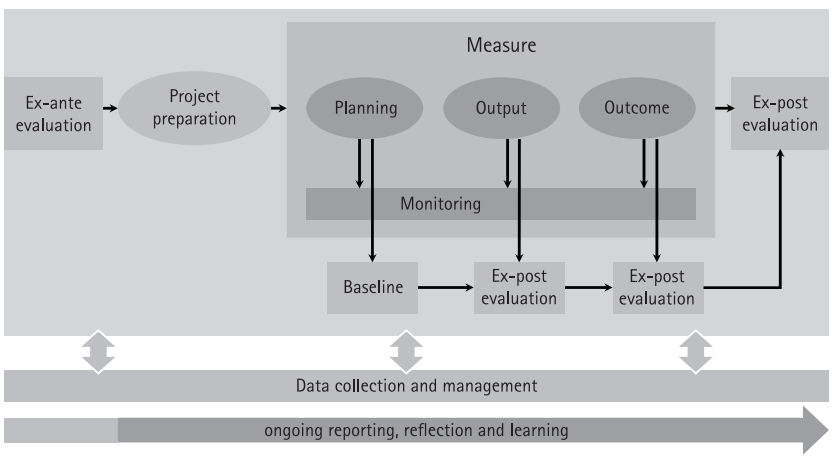
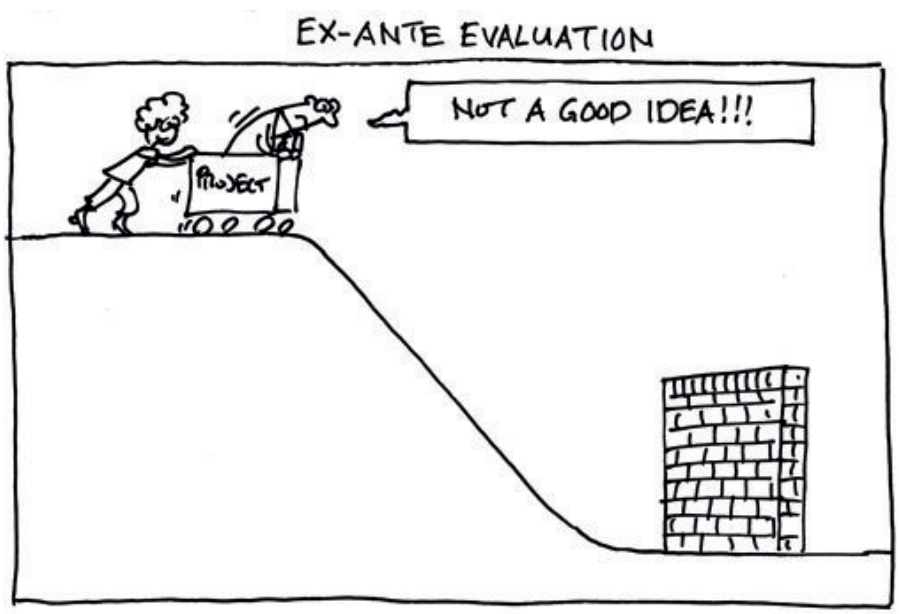


Figure 1–1: Arrangement of evaluation.

there is not one generic measure/project cycle, as each measure/project ultimately varies according to the local context and need.

The first stage in a measure cycle is an initial needs assessment. Ideally, this stage is already integrated in your transport policy formulation or urban mobility plan. This step is necessary to determine your needs and what could be done to improve the situation. This then leads you to a (selection of) measure(s) to which you attach certain expectations (in other words, how well the expected outcomes fit your problem) and from which you choose the measure that fits best. This is all part of the so-called ex-ante evaluation, the process of checking how well a scheme or strategy will perform. It helps you to make efficient choices between options. It is more a prediction or simulation of what you think will happen. For the purpose of this book we assume that these steps are already concluded. We will discuss mainly ex-post evaluation although some elements can be used for ex-ante evaluation too.

After your initial needs assessment you start the operational design of your measure/project and its objectives, indicators as well as means of (later) verification. This includes the identification of the purpose of the evaluation, your resources available and the determination of the appropriate evaluation design. Then, the baseline of data against which your improvement can be measured is compulsory. This is typically the end of the planning period and the beginning of the implementation. This baseline data will be the first real test of your data collection methods and give you an initial insight into the quality of your data assessment. In general your measure progress is accompanied by a monitoring process. This is an important reflection to assess and



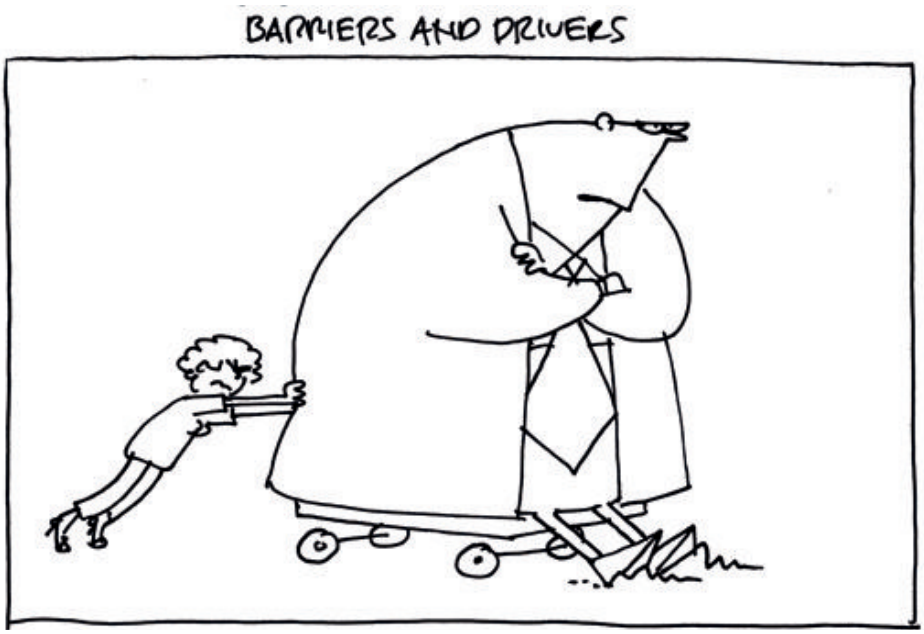
inform on the ongoing project/programme implementation. Often, the data you will need to provide for this can also be useful for the evaluation or vice versa. The final (ex-post) evaluation occurs after the measure/project is completed to assess how well the project/programme achieved its intended objectives. Then, the measure/project cycle is concluded by the dissemination and the results and lessons learned. However, the proper reporting, reflection and learning should occur throughout the whole measure/project cycle. As such, evaluation does not take place once or twice, but is a steady part to the measure's implementation. Do not be fooled, this does take a lot of resources (time, money and people involved). But it is a rewarding process.

1.3.2 Are there different types of evaluation?

You will not get very far in studying evaluation before realising that the field is characterised by enormous diversity. From large-scale, long-term, international comparative designs involving millions of Euros to small, short evaluations of a single measure in a city, the variety is vast. These can be categorised in a variety of ways, but for the urban mobility-related context, there are basically two fields of assessment – the impact and process evaluation.

The main goal of the *impact evaluation* is to draw a balance of the effects of the measure's implementation and the situation before the implementation. The purpose is to assess a mature project's success in reaching its stated goals. Impact evaluation is an appraisal of worth, or merit. Usually, this type of evaluation is needed for decision making as it presents 'hard facts'. The decision alternatives may include the following: disseminate the intervention to other sites (also called transferability); continue funding, increase funding, continue on probationary status, modify and try again, and discontinue. You can read more about this type of evaluation in chapter 2. However, this impact evaluation should not be confused with an output assessment. Think of it this way: if you have a herd of horses which are thirsty, you build them a water trough. If you consider the amount of troughs you build, you do an output assessment. If you lead your horses to the water and they drink, we talk about the outcome. The impact of your action would be the fact that your horses remain healthy because they are drinking water.

The *process evaluation* focuses on the means and procedures by which a measure is implemented. It begins during project development and continues throughout the life of the project. Its intent is to assess all project activities, negative and positive factors which are influencing the measure implementation process and thus provide information to monitor and improve the project. You can read more about this type of evaluation in Chapter 3.



Nonetheless, process and impact evaluation are to be seen as one. If we talk about the horses again: you can lead them to water, making sure that the tank is full, you can even monitor the quality of the water but without monitoring their health, you will not know if your measure had the desired effect. On the other hand, if you only monitor their health – in other words you focus on impact evaluation – and find out that they die anyway, how do you know it is not a result of bad water quality (assuming of course that they have been fed properly)? As you can see only a so called 'mixed evaluation approach' of impact and process evaluation can give real evidence for success or failure of measures.

1.4 Whom do I need to get involved?

Evaluation involves a broad spectrum of institutions as well as those people, whose actions are the object of the evaluation conducted. In other words, evaluation involves a lot of people with many different backgrounds who have different interests and motivation for an evaluation.

Politicians and decision makers – these can include administrations on various spatial and hierarchical levels such as the European Commission, national, state and local level authorities etc. For them evaluation constitutes the source of information about the project (its preparation, implementation and its results).

Managers of the measure – this is a group of people whose tasks include managing the different aspects of the project; the evaluation results supply them with the information about the effects of their work, about positive influences and difficulties.

People who implement the measure – these are the actual people who make your measure a success; the people belonging to this group may see the effects of their work in a wider context and see how it contributes to the overall improvement of liveability in their city.

Measure target groups – these are the potential beneficiaries or 'bearer' of a measure. The evaluation results enable this group to see what they may expect from the project (ex-ante) as well as what has been done within it (ex-post).

Other stakeholders – these are all other stakeholders which are not part of any of the immediate groups mentioned above, for them the evaluation results should be made available as well. This is necessary for transparency, which will then again improve the acceptance of a measure.

Experts, i.e. persons with technical and/or methodological knowledge as well as any person who can help in defining the evaluation questions or interpreting the evaluation results. The involvement of independent experts may be very crucial for supplying useful information to the evaluation team and during the debate of which the objective is to indicate more general lessons following the evaluation study.

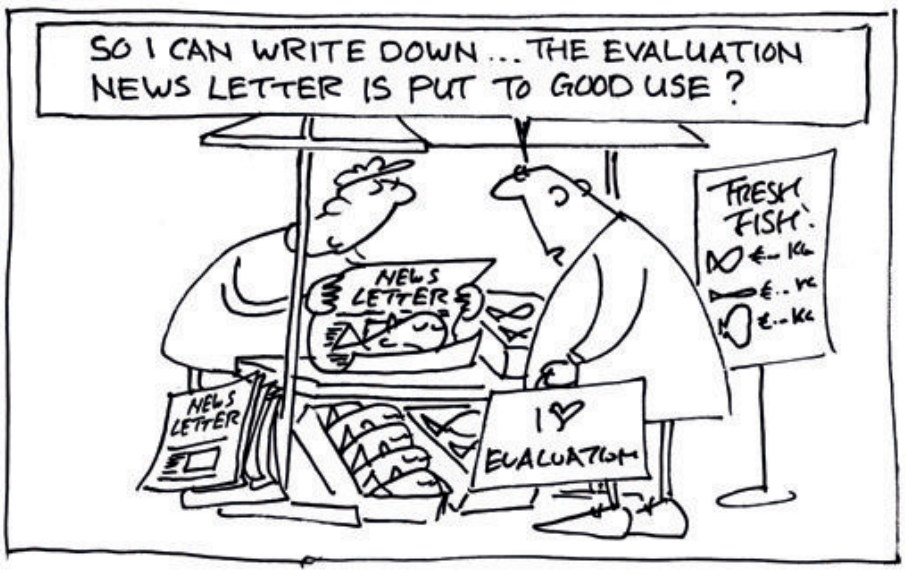
Thus the answer to the question of who should be involved is that everyone who is eventually positively or negatively affected by the measure should also be part of the evaluation procedures and result disseminations.

1.5 Why do we need this book?

Transport accounts for nearly a quarter of current energy-related carbon dioxide emissions with car travel constituting more than three quarters of all vehicles kilometres travelled. The effects of car travel – especially on cities – are numerous and mostly do not relate positively to a friendly living environment. Hence, cities are designing interventions – measures which aim, for example, at reducing car use and thus decreasing the negative impacts of travel mode choices. However, the evidence on the effectiveness of such measures found in the corresponding evaluation literature remains widely weak. Their evaluations usually vary in the methods they employed and the outcomes they reported. Frequently the evaluation methods have not been chosen in a way that they can deliver satisfying results. This is nothing to be ashamed of as the measure evaluation in the field of sustainable urban transport is a relatively young research area. In some cases evaluation seems unwanted by the people responsible as they mistake it for an audit; in other cases evaluation might be generally desired, but resource

restrictions or a lack of competence might produce results of minor value or hinder the evaluation entirely.

But why do we need another evaluation book? Well, our experiences in the CIVITAS MIMOSA and CIVITAS POINTER projects show that there is still a lack in evaluation literature that is specifically tailored to the urban mobility-related context. And this was confirmed when we received reactions to the first drafts of this book, they were all in line with: "When I began to read this book I thought, why did not I have this book before the project started". Now we do not want to brag too much about our product. Nonetheless, we hope this book will encourage managers to conduct evaluations when they might otherwise have viewed them as too expensive or time-consuming to be conducted to a high standard. Thus, the desired outcome is an increase in the quality and quantity of rigorous evaluations which are conducted.



This handbook is not designed to be read from cover to cover, it is a resource guide which can be used for reference as and when needed. For this reason readers will find that key points may appear to be repeated in different sections. The structure includes: case studies, key guidance notes, templates, checklists, and links to other methodological approaches, online documents, and sources of reference. You should not feel discouraged when you read this book. A sound evaluation is a complex task, but it is one anyone can manage with a little help, time and dedication.

In order to better understand the details of the evaluation process, three measures from within CIVITAS MIMOSA will serve as examples for this book. These examples are

What this book will do	will not do
demonstrate that evaluation is a useful task	substitute project specific guidelines
demonstrate that evaluation can be learned by everyone	provide information on how to assess European added value
present examples that illustrate the evaluation process through all stages	comment on specific evaluation structures within European projects
focus on small urban projects (CIVITAS-like measures)	consider peculiarities of infrastructure projects

Table 1–1: What this book can do for you.

the Road Safety Label and the Cargohopper from the city of Utrecht, as well as the 'Knitting Bus' campaign from Tallinn. We will describe them here briefly.

Utrecht Road Safety Label: The city of Utrecht awards Road Safety Labels to primary schools that proactively address road safety issues. The aim of the measure is to improve road safety in urban areas, particularly around schools where road safety problems increasingly occur. As more and more parents drop their children off and pick them up by car, children who cycle or walk to school are increasingly at risk, also because children often do not act safely in traffic. The situation has become such that the areas around many primary schools are no longer sufficiently safe. To this end, Utrecht aims to introduce the Road Safety Label in most of its primary schools. The Utrecht Road Safety Label is a quality hallmark for schools that include road safety in their school's policy. Key elements that schools will need to encompass are traffic safety education and the participation of parents. Commitment to the establishment of uniform school surroundings with regard to road signage etc. is also required. A specially trained Road Safety Label consultant indicates which criteria need the school's attention for improvement. The school appoints a road safety coordinator who is in charge of the project, and – if necessary – supports the implementation. If the school fulfils all criteria, it will be awarded the Road Safety Label. The consultant makes plans for maintaining the label and there will be checks every two years.

Utrecht Cargohopper: Freight traffic is a major contributor to deteriorating air quality and rising greenhouse gas and noise emissions. In July 2007, Utrecht introduced a low-emission zone that limited access for trucks with polluting engines. As a first step, Utrecht, in cooperation with a transport company, started a pilot with one electric mini-train (the Cargohopper) for goods distribution in the city centre. The city granted the Cargohopper various exemptions with regards to the entrance to the limited traffic

zone and it soon supplied 40 to 50 delivery addresses in the city centre per day. Later on solar panels were placed on the roof of the Cargohopper, which supply the train with solar power for eight to nine months a year. In the remaining time, it runs on green electricity. The Cargohopper has the capacity to do the same deliveries as five vans.

Tallinn Knitting Bus: Tallinn has identified low popularity of public transport and light transport as a problem that needs to be addressed. The city has already introduced innovative solutions and improved the quality of sustainable transport modes but these measures have remained largely unnoticed by the general public so far. The city has realised that it needs to develop a marketing strategy to promote its public transport service and to inform citizens of mobility options in the city. The first task for Tallinn in this measure has been to draw up a communication plan that includes specifications for a media campaign. Target groups have been defined (schools, work places, individuals) and practical interventions have been specified such as mobility plans, education and promotional activities. As Tallinn was European Capital of Culture in 2011, it was decided to take advantage of the many cultural events that were being held in the city. In conjunction with the Capital of Culture committee, Tallinn's Bus Company investigated the idea of 'Knitting Graffiti' or 'Yarn Bombing' which originated in the United States of America in 2006 as a mean of 'softening' and brightening up the urban environment. In Tallinn (where there is a strong tradition of knitting as a handicraft), it was decided to build on this trend and implement a 'Knitting Graffiti' campaign in one bus. Volunteers were recruited to knit, and the seats and also the hand rails were wrapped in knitting. The outside of the bus was covered in vinyl photos of knitting.

We will use parts of these measures as examples, while you can also find a complete evaluation report in Chapter 7. We will be using these three examples throughout this handbook to elaborate an ideal evaluation. For this purpose, some of the details about the examples are fictional, but others have been taken from real life experience. Remember, we do not wish to sell the measures, but show how uncomplicated and rewarding a proper evaluation can be.

2 Impact evaluation

There can be various influences that cause certain effects and side-effects. You want to know how to show the real impact of a measure in a structured way? So let us start with the basics of evaluating an impact of a measure.

Impact evaluation illustrates changes which are attributed to an intervention such as a project, measure or policy which was planned and implemented to reach a formulated goal. In contrast to outcome monitoring, which examines whether targets have been achieved, impact evaluation is structured to answer the question: How would outcomes such as participants' well-being have changed if the intervention had not been undertaken? Impacts of mobility-related measures can be determined in many different ways. Therefore, a variety of evaluation methods is offered in this chapter. Whatever method you chose, it is essential for the evaluation of any impact or effect to collect and analyse data before (baseline) and after implementation (ex-post). This enables you to compare both situations and draw conclusions. A central question to answer is: What was the situation before I implemented the measure and what changes can be attributed to the measure?

Ideally, it also puts the surveyed data in relation to control data which are collected where no measure has been carried out (the so-called business-as-usual, see Chapter 2.4). Except from collecting data before and after implementation, evaluation can even take place before implementation, the so-called ex-ante evaluation, which helps to decide which measure is best to solve your problem. Ex-ante evaluation takes place after you have identified a problem or issue that you want to tackle with an intervention. In this case you make assumptions about what impact certain measures are going to have and what the costs are. Based on this result you will then decide which measure to implement to reach your objective. This decision making process will not be part of this book which brings impact and process evaluation to light.

This chapter is about how to organise an impact evaluation and what you have to take into account from the start to avoid unwanted effects. These can arise by conflicting interests and can show up by having blinkers on as the persons responsible for the measure. To carry out a meaningful evaluation, it is necessary to think about some points before realisation of your measure: How do I clearly define the objectives? And how can I find the corresponding indicators to evaluate the impacts?

After having a closer look at the issues of objectives and indicators, we are going to tackle questions like "Which kind of survey or evaluation design is suitable and meets the requirements of a sound evaluation?" Furthermore, some issues regarding questionnaires and choice of control group as well as sampling and sample selection will be discussed more specifically. A survey, as one data collection method, can have various shapes and appearances. Hence, some basic concepts and relevant issues will

be outlined. And last but not least, we will introduce selected methods and hints for the actual data collection and analysis.

The following Figure 2-1 summarises the ideal steps of impact evaluation before and after implementation of a measure, which will be presented in detail afterwards.

It is recommended to analyse those data collected before the start of the measure shortly after its collection. So you still have the chance to adjust the measure according to your findings before its implementation, if necessary. Often this analysis and interpretation of data (steps 7 and 8) will be carried out later parallel to implementation or analysis of after-data due to lack of time or interest. However it is better to take the chance to get active before a measure fails.

As you have seen, before collecting any data, objectives you would like to achieve should be defined. So let us start with this issue.

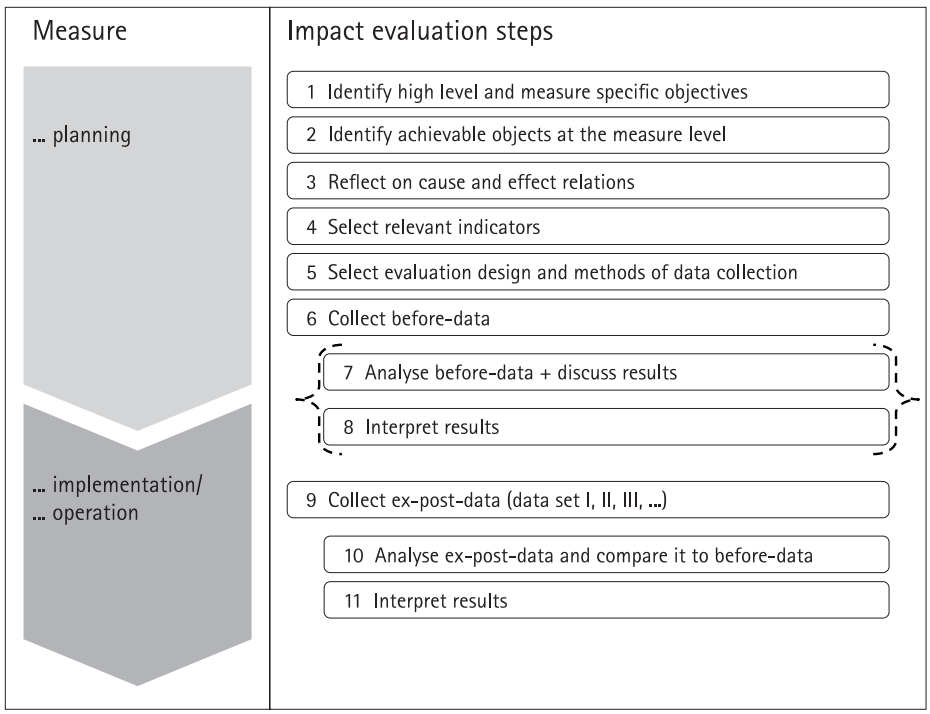


Figure 2–1: Steps of impact evaluation.

Further readings

European Commission, Directorate-General for Regional Policy Cohesion: Principal evaluation techniques and tools – MEANS collection: evaluating socio-economic programmes. 3rd volume, Lanham (Luxembourg), 1999.

Patton, Q. M.: Utilization-Focused Evaluation. 4th edition, Sage Publications, London, 2008.

2.1 Objectives of your measure

In order to evaluate the effects of an urban transport measure it has to be clear, which objectives should be addressed by the chosen measure and how the objectives relate to each other. The objectives should be defined by the evaluation team after consulting the steering group responsible for evaluation of the programme involved. Keep in mind that the output of your measure or project should always be attributed to the long-term strategies of political priority as formulated in a transport master plan of your city.

Regarding objectives to be achieved by a measure, there is always a hierarchy of objectives. So first of all view at the objective(s) at a higher, more long-term level, which has been assigned to the planned measure, for example improvement of quality of life or reduction of transport related emissions. Then phrase the objectives you want to specifically address with the measure, which are more short-termed and are small steps towards the final goal which is formulated in a high level objective. If those objectives were already predefined, check if they really fulfil the purpose to contribute to the evaluation of your measure (we will show you some criteria for good objectives later). It is better to renegotiate the objectives than to work with measure specific objectives formulated too general, unachievable or the like.

So there are two levels: high-level objectives and measure-specific objectives, which relate differently to the effects of your measure. The actions you undertake as input are followed by a direct output like constructed pedestrian crossings, or the event to promote sustainable mobility. The impact of the actions refers to the objectives on the measure level such as to achieve 5% fewer accidents with pedestrians involved in the area of implementation.

Objectives at a higher level aim to achieve an overall outcome such as the high-level objectives. Therefore, to show if the impact of your measure serves the defined purpose, you describe its outcome as contribution to measure-specific objectives as

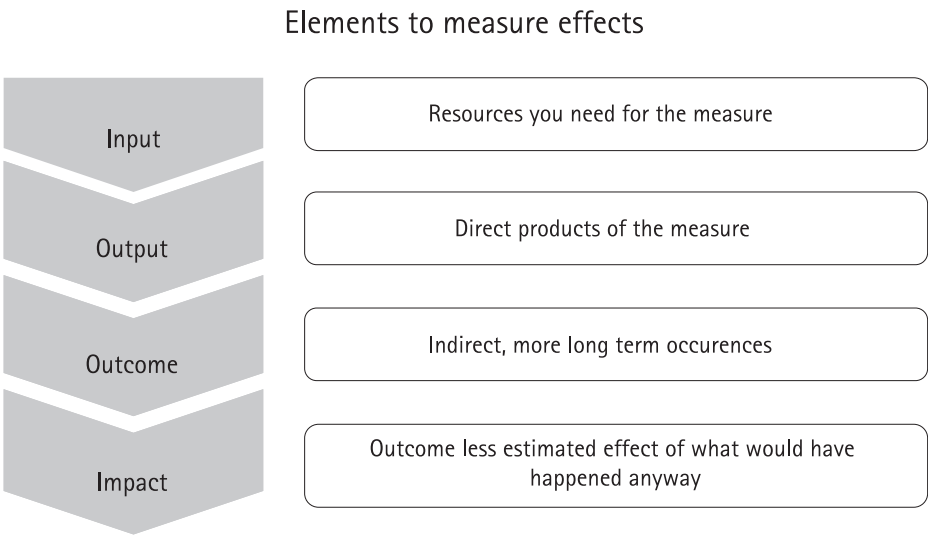


Figure 2–2: Definition of input-output-outcome-impact.

well as its estimated outcome, which is related to its high-level objectives. Finally, a measure's impact is described by the outcome less other influences which would have happened without the measure.

Let's take the example of the road safety label which was implemented in Utrecht. Zones around schools were equipped in this measure with common street signs and markings to achieve a uniform lay-out of the school surroundings. This should attract drivers' attention immediately to the fact of being close to a school. Increased safety and encouragement of pupils to go to school by bike or by foot is intended by this intervention. In the long run the implemented actions aim at increasing modal split towards sustainable modes and thus contributing to an improved quality of life. These long-term objectives correspond to the overall urban mobility plan which aims at strengthening sustainable modes of transportation and to increase road safety in the city.

For the evaluation of the measure it is important to formulate clear measure specific objectives, from which the indicators can be derived (see Chapter 2.3). These objectives help you to show the results of the measure and to see if the results mean a success or a failure of the measure. In case of our example at measure level "To reduce the share of home-school-trips made by car in favour of cycling and walking by 5%" was defined as an objective (see others in the table of indicators below or in Chapter 7). In order to set and formulate clear measure objectives within a certain period of time that allow for an assessment of the measure's success after its implementation, the SMART approach is helpful. We are using the example of the Utrecht measure's objective to show you what the SMART criteria mean for your choice of measure-specific objectives in Figure 2-3.

SMART dimension	Example Uniform School Surroundings (Utrecht)
Specific – Do the objectives spell out what to be achieved concretely and are therefore well-defined and understandable?	Taken the achieved influence on the share of home-school-trips, it is specified clearly which kind of trips between school and home are meant.
Measurable – Does the target make it possible to measure the success or failure of the measure? What is the evidence for success?	To be measureable, the concrete formulated objective should include what a significant increase means. Here for example 5%. The change in modal split can be measured surveying the current use of modes for school-home-trips before and after the implementation. So a changing share of the sample driving, cycling and walking to school delivers quantified results which fulfil this criterion.

Achievable – Are your set objectives achievable?	A decrease in share of home-school-trips by car of about 5% in favour of cycling and walking seems to be achievable if the available data collected before the start of the measure are in a format that allows this. After having a look at the existing data about share of car use for those trips before the measure, an increase which relates to a shift of half of the sample towards cycling and walking would be beyond reach and too ambitious.
Realistic – In a practical sense, is it really possible to achieve the objective with your available resources? (And does it fit to the overall objectives?)	Based on expert experiences in your city and available information about similar projects in other cities, think about a reasonable objective considering cause-and-effect relations of your intended measure. May it take longer or further measures to reach this specific objective?
Timely – Within which time frame would you like to achieve the objective? Is it feasible to meet the set time limit?	Within the period of the program Utrecht aims at a decrease of the share of home-school-trips by car about 5% in favour of cycling and walking. So the time frame is clear. The result may vary depending on the time of the survey after implementation.

Figure 2–3: SMART dimension of measure-specific objectives.

If necessary, the decrease of 5% in your objectives can be adapted after you have collected your baseline data. This is recommended, if the objective cannot be seen as achievable due to unexpected framework changes, for instance, if the zones could not be fully implemented as planned. Within a small frame, objectives are a flexible tool which can be adjusted to strong changes of project conditions and to the collected baseline data. However, this should always be reported in a transparent way. Otherwise your measure-specific objectives would seem to be set purely arbitrary.

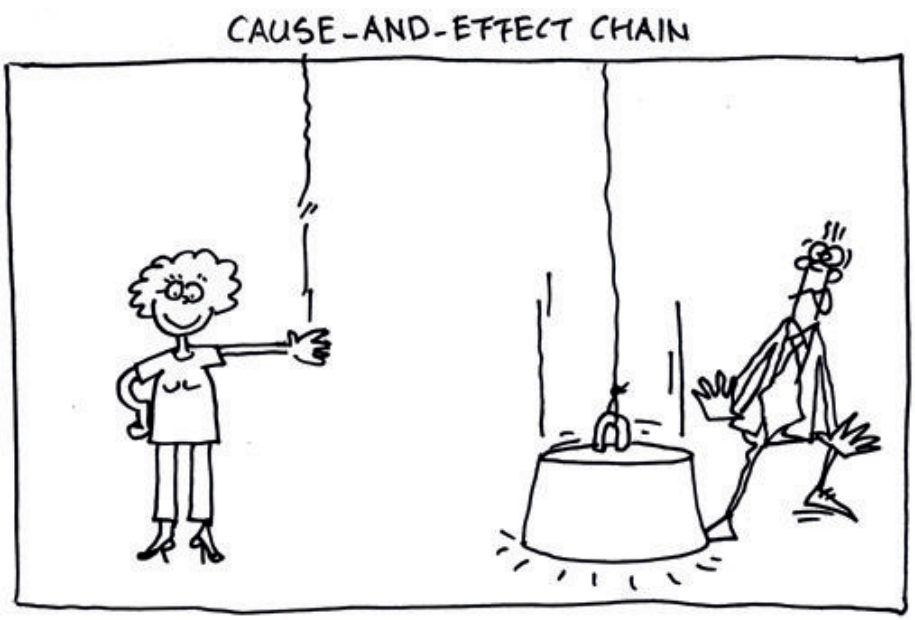
2.2 Effect analysis

The defined objectives tell you what you want to achieve with your measure. But how can these objectives be achieved and which spheres are influenced by the measure? This chapter offers support to recognise relevant cause-and-effect relations by planned actions to prepare the decision for the choice of indicators.

Obviously, the action of releasing the twine lets the weight fall on the foot as you may guess when you have a look at the cartoon below. But do we really know? Is it the same twine she releases as that one that is holding the weight? How can you show the impact of a measure and that this impact was really and directly caused by the actions taken?

We will introduce two methods that support cause-and-effect relations of your measure: the cause-and-effect chain and a summary of effects on stakeholder groups. We will refer to the example of effects of the Utrecht Road Safety Label again. You can find more detailed descriptions of structured cause-and-effect consideration of three examples at the end of this handbook (see Chapter 7).

To identify preferably all meaningful relations between your objectives and possible effects (output, outcomes and impact, see Chapter 2.1), a cause-and-effect chain is helpful. We are going to show you the benefits of it: Cause-and-effect chains show linkages between possible effects, both positive and negative, and the resources that are bonded through the implementation of the measure. Therefore, it will help you to



understand the wide range of possible and intended and unintended impacts of the measure and to consolidate your choice of indicators. It supports becoming aware of what causes might be responsible for an effect and that this respective effect at the same time represents a cause for another one.

The outcome of a measure can always be caused by a variety of effects which need to be considered, because impacts are often indirect, with several steps between an activity and its eventual impact. For instance, the twine of our female figure's character in the cartoon can be the twine of a balloon which is only out of sight now while something else is responsible for the dropped weight. Or a decrease in the number of traffic accidents is not necessarily caused by the implemented traffic measure. Impacts and outcomes – both transport- and non-transport related – as well as assigned costs and benefits constitute elements of the following cause-and-effect illustration, including also the direction of the effect relation. A cause-and-effect chain is provided for the Utrecht Road Safety Labels in Figure 2-4 to exemplify the conceivable effects on different spheres of action.

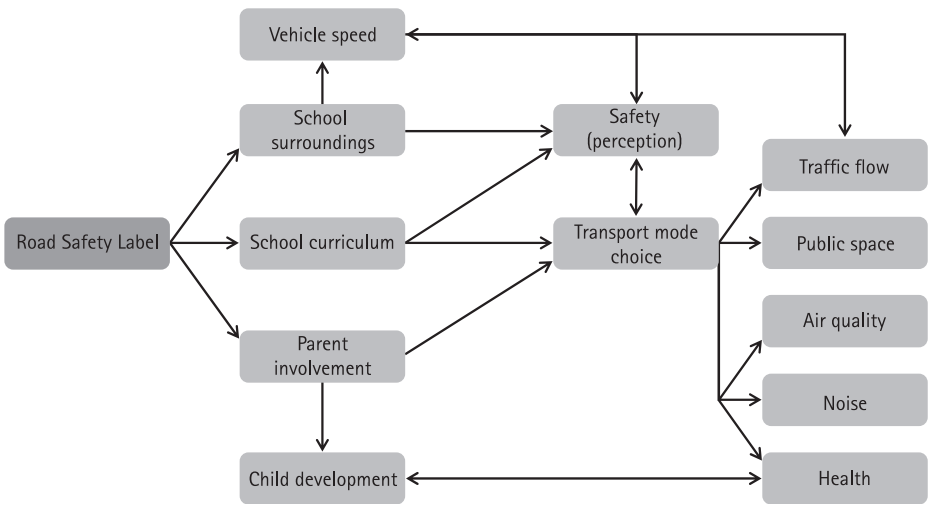


Figure 2–4: Cause-and-effect chain for the Utrecht Road Safety Label.

The city of Utrecht awards Road Safety Labels for primary schools that proactively address road safety issues. Thus this initiative has an impact on the school surroundings, the schools' curriculum and it encourages parental involvement in safety related issues. The unification of school surroundings over the city of Utrecht will have an effect on the average vehicle speeds in the school vicinity as drivers will be more aware

of the speed limitations due to the school nearby. This could also increase their attention and thus have an impact on safety and traffic flow.

As a second step you can decide about the main relations, which are considered to be most important relating to the objectives. The inclusion of traffic education in the schools' curriculum could have an impact on the transport mode choice as more parents might allow their children to go by bike or walk to school. This leads to an increase of safety (and the perception thereof), as well as changes in traffic flow, the use of public space, air quality, noise and children's health. When thinking about direct effects of your measure, take care in a second step that positive and negative side-effects do not escape your notice. The effect above could also be stimulated by an (increased) parent involvement. As positive side-effect parents are encouraged to teach their children to act responsibly on their own, which supports their personal development. A standard lay-out of road signs and markings causes not only the desired development but can also induce negative side-effects as less attention by school children when they move outside their school surrounding.

The level of noise, air quality, the use of public space and the children's health are a result of numerous factors. For instance, schools which have a canteen might already have 'healthier' students, or schools which focus on sport activities. As a result, these factors and their development were excluded from this analysis. For the same reason, the traffic flow will not be evaluated. The reduced cause-and-effect chain is shown in the next figure (Figure 2-5).

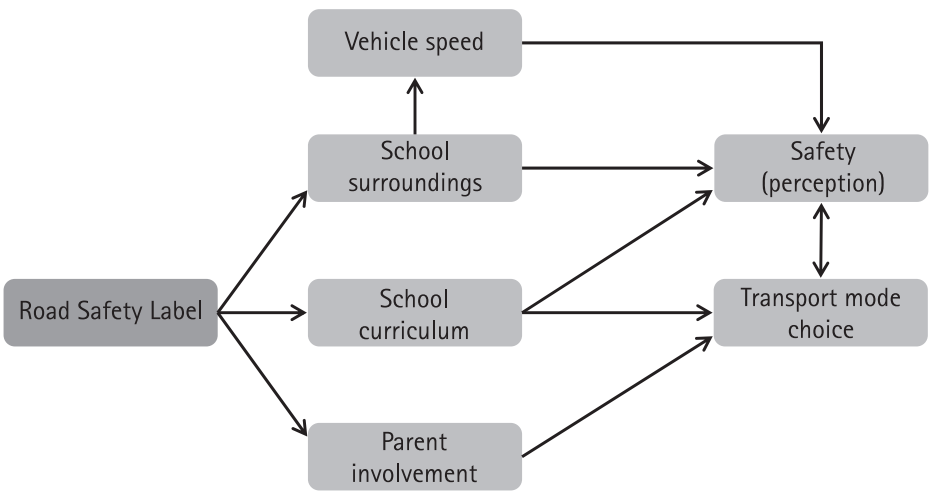


Figure 2–5: Reduced cause-and-effect chain for the Utrecht Road Safety Label.

Consequently, the evaluation will focus on the impacts on vehicle speeds in the schools' vicinities, safety of the children and their transport mode choice. The choice of indicators will be dependent on these findings. Make sure that you have identified at least one impact for each of the objectives. Revise objectives, if the cause-effect chain indicates that they are difficult to reach.

Another option to clarify the effects is a list of involved agent groups of the measure. Costs and benefits as well as negative and positive impacts by affected stakeholder groups offer a comprehensive view of the impacts (see Figure 2-6).

Pupils	+ less affected by accidents, lower accident risk, higher road safety + encouraged to do the school-home-trips on their own (by bike or foot), more responsibility for themselves, less dependent on their parents - assuming the habit that car drivers behave considerably outside the school surroundings
Teachers	+ perception of higher road safety (- more responsibility for transport education)
Parents	+ more reflective car use + time savings (for work and finally more free time for their children in the afternoon/evening) - less control over their children, uncertainty/fear, if their children got to school safely
Residents (school surroundings)	+ lower accident risks for themselves, especially for their children - more regulated, slower traffic - less affected by pollutant emissions, higher quality of life
Other car drivers (passing school surroundings)	+ pay more attention towards pupils, more awareness of school and pupils, less accident risks - confused about unknown, regional markings and street signs
Administration	+ improves its image by successful implementation - has to pay for the project funding

Figure 2-6: Effects by Stakeholder groups – Utrecht Road Safety Label.

When you first read about the Utrecht Road Safety Label measure, did you think about irritated car drivers who have never heard of these municipal solution and might have problems reacting to the street signs in an appropriate way without jeopardising somebody's health? This overview supports the awareness of effects towards groups other than the main target groups and shows further possible side effects of the measure. Based on these thoughts/preliminary considerations you can check, if the arrangement of the components of the measure and its impact on specific groups is considered sufficient in your set of objectives then use it for the selection of indicators. You can think about how to measure and react to possible negative side-effects which enables you to make adjustments of the measure, if necessary.

So considerations of relevant causal relationships are not a waste of time, but should be included as an integral part of your impact evaluation. There can always be unpredictable influences and changing conditions in innovative measures. In the end, after analyzing your results you should come back to this overview again, because it can also help you with the interpretation of results since they allow speculating where further improvements could be made.

2.3 Indicators

After having defined the objectives for your measure and understanding the cause-and-effect relations of your measure you need to select the most relevant indicators – those which show a possible impact of the measure best and can be assessed with the given budget. Ideally, your selection is based on the cause-and-effect considerations explained before. Indicators must closely relate to the objectives and thereby allow for statements about the degree to which the objectives have been achieved. So three basic requirements have to be taken into account when defining indicators:

- They must clearly reflect the performance or impact of your measure (see Chapter 2.2).
- Secondly, they must match the objectives (see Chapter 2.1).
- Thirdly, are capable of reliable assessment using the experimental tools and measurement methods which you chose (see Chapter 2.5).

There are various indicator sets developed in European transport-related programmes which may serve as a suggestion for an existing indicator set. Be careful with the choice of indicators, given sets are only useful for your orientation. If you use a defined indicator for your measure, it may need to be adapted to the specific circumstances of its application. This depends on the variety of measures involved in your city and in the programme and the availability of data. To apply indicators in a way which facilitates comparison of results of different or similar measures within your city or in different cities, it is necessary to agree on a definition of the indicator or use an already defined method to assess this indicator. Within a programme, indicators can be determined within their context and relevance for example in indicator-specific methodology sheets. As an example, if you try to promote and measure alternatives to individually used motorised vehicles in your city by modal shift, you should assess the *modal split* to measure a modal shift towards more sustainable modes by travel surveys. Data can be collected by the quantitative indicator *average modal split* (passenger, vehicles or trips), which could then described as a percentage of passenger- or vehicle-km or percentage of trips for each mode over the year. Modal shares of non-motorised modes

as walking and cycling can be contrasted to shares of public transport (bus, tram, metro, train) and private motorised transport by car or motorcycle.

Please note, to provide the best possible insight into the impacts of your measures they may require further interpretation and need possibly to be complemented by extra local indicators (see Box 1).

In our example of the Utrecht Road Safety Label, the objective of the measure and the cause-and-effect chain lead us to the following indicators:

Specific Objective	Indicator	Description of Indicator
"To increase the satisfaction about the road safety in primary school areas among children, their parents and teachers by 30%"	Perception of safety	Change in perception of road safety among teachers, pupils and parents
"To reduce the number of accidents with children involved in the surroundings of schools"	Safety	Numbers of accidents with children involved in the school area and the surrounding residential areas
"To reduce the share of home-school-trips by car in favour of cycling and walking by 5%"	Modal split	Average percentage for school-home trips for each mode

Table 2-1: Effects by stakeholder groups – Utrecht Road Safety Label.

A possible impact of the implementation of the school zones on road safety and accidents can be identified. Hence, a survey conducted among parents (with their children) could show their perception of road safety and the share of each mode for their children's trips to school. To evaluate the measure, the data (number of trips by mode) must be collected before the measure is implemented (the so-called baseline) and after (for the ex- post evaluation). The difference between the before and after shares of the modes less the estimated change which would have happened without any measure, describes the effect of the implemented measure of the Utrecht Road Safety Label (see Figure 2-7). You can find two other examples of indicator lists for the examples of Cargohopper and Knitting Bus in Chapter 7.

If you want to evaluate several transport measures in parallel, consider if there are indicators which might be affected by more than one measure. These measures are

Box 1: Common and local indicators

Especially larger EU-projects will often provide you with a list of so-called common core indicators (such as the Maestro indicator list). These are indicators you should consider for your evaluation and use if they are appropriate. They will help in the cross-site evaluation – meaning they will help to compare your results to those of another project partner. These lists will never cover all of your needs hence you are likely to make use of specific local indicators for your evaluation. Such indicators may be used:

- to make an assessment at a more detailed level;
- to assess the impacts concerning a particular local problem;
- to assess the impacts of the special characteristics of a local measure.

"bundled measures" with so-called "bundled indicators applicable for more than one measure (see Box 2).

Discussing and developing objectives, causes, effects and indicators before the implementation of your measure constitutes an essential prerequisite for your evaluation.

The next chapter will show you the evaluation design to choose, how to set it up and to collect the relevant data, which needs to be considered before the implementation of your measure (see Chapter 1.3.1).

Box 2: Bundled indicators

If you have a set of measures within your project and you compare the cause-and-effect chains of all of them you might want to select one or more indicators which are meaningful/applicable for more than one measure. The effect shown by this indicator may also be influenced by another measure, result of the combined impact of two or more measures. These are cross-measure indicators and not measure-specific ones. Ideally, the set of indicators for a measure should not consist of bundled indicators only. This would make it impossible to derive a clear message about the success or failure of this specific measure, because the effects would always belong to another measure too.

To refer to the Cargohopper again there is the indicator "freight vehicle reduction in the city centre" functioning as a bundled indicator with the beer boat measure, because this indicator can be influenced by both measures (see example in

Chapter 7). It is recommended to rank the measures related to a bundled indicator to see which measure influences a bundled indicator most.
A lot of other interventions – within your set of measures or external influences – can exert an influence on a peculiarity of an indicator (such as on the measure itself, see 'Business-as-Usual', Chapter 2.4). You cannot consider and measure all influences, of course, but at least refer to most important of them in the interpretation of results.

Further readings

Hensher, David A.: Performance evaluation Frameworks. In: Button, K.J.; Hensher, D.A. (Eds.): Handbook of Transport Strategy, Policy and Institutions. Elsevier ltd., Oxford, 2005, p. 83-96.

DISTILLATE – Design and Implementation Support Tools for Integrated Local Land Use, Transport and the Environment: Guidance on the Development of a Monitoring Strategy and the Selection of Indicators: Project C – Indicators. UK, 2008. (http://www.its.leeds.ac.uk/projects/distillate/outputs/products.php)

2.4 Evaluation design

The evaluation design is a plan for collecting and analysing evidence that will make it possible for you to answer whatever question you might posses. The choice for a particular design is frequently influenced by the need to compromise between expenditure and confidence. Generally, the more certain you want to be about your measure's outcomes and impact, the more costly the evaluation. It is part of your job to help make an informed decision about the evaluation design. This is an important task because when you evaluate transport-related measures, all factors which may change during the evaluation period need to be collected and presented. But, in order to draw conclusions, it is first necessary to identify what would have happened if the measure was not introduced. Only then can you ensure that the effects measured solely rely on your measure (see Box 3 for more detail). Possible ways to build this so-called 'business-as-usual' scenario include forecasting from historical data, modelling or monitoring a control group/site.

In an ideal situation, the group/area which was exposed to the measure (sample group, examination area) is compared to a group/area without measure implementation (to derive the business-as-usual situation). Under equal conditions, the results measured should be solely accounted for by the implemented measure. This concept is for

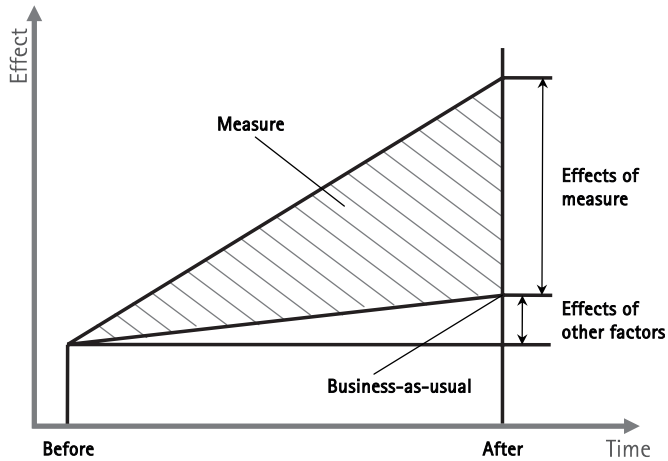


Figure 2–7: Baseline and business-as-usual scenario.

instance always applied in health studies. As an example, think about a treatment which is supposed to enhance the growth of children through hormone prescription – this will be your measure. How would you prove that a child's growth has only been caused by this hormone? You should do two things: First, you would forecast the average growth of 10-year-old males. From historical data you can safely assume that they will grow on average 5 centimetres per year. Then you would monitor this growth by comparing the growth of two identical twin brothers, where one was given the hor-mone, whereas the other receives a placebo. Your results show that the brother with the placebo grew 5.5 centimetres per year – this is your business-as-usual situation. The brother with the treatment grew 5.8 centimetres. With this information you can say that due to the hormone prescription, male children's growth can be enhanced by an average of 0.3 centimetres per year – this is the effect of your measure. The difference between the brother without the treatment and your forecasting is the effect of other factors which influenced both children.

How you account for these other factors which influence your measures can be the biggest quality hallmark for your evaluation. In other words: the more accurate you can describe the business-as-usual scenario, the more reliable are your evaluation results and your conclusions will be more valuable. Thereby, the choice of a so-called evaluation design usually does not depend on application level (EU, federal state, state, region, and municipality). The decision for a design is also independent on the complexity of the issue addressed. It is a question of your ingenuity to develop an evaluation design suitable to determine the net effects under the present technical and financial circumstances – regardless whether there is a thematically restricted individual measure or a complex bundle of measures.

In many cases adequate streams of data are available from various sources and can be included in your evaluation design. These need to be analysed and linked to your measure objectives and indicator choice (e.g. accident data from police, data regarding the density of traffic from road construction planners, data regarding air quality from environment agencies). Specifically, clear differences of evaluation designs can be made between:

- randomised control group designs
- experimental designs with a control group
- experimental designs without a control group

Box 3: Why do we need a control site?

Take a minute to consider this: A measure improves the surroundings of primary schools in order to enhance safety for school children. The indicators are number of accidents and the perceived safety by the school children as well as by their parents and teachers. A questionnaire was distributed before the implementation of the measure. After the respondents (children, parents and teachers) had given their ratings, the same procedure was repeated after the implementation of the measure. The rating of the second survey is higher. First, the conclusion would be that here was a positive effect of the measure on perceived safety. But what about other factors that influence the perception of safety? What if petrol prices had been rising so much during the two surveys that more and more parents and teachers stopped driving a car? Further, a city-wide debate on safety around primary schools focused the attention on this topic. So, how can the effect of the measure (redesigning the surroundings) actually be described? This question could be answered by using a control-group. So in this case, another primary school with similar characteristics should have been selected without any infrastructural changes regarding the traffic to be implemented. At this control site the same questionnaire will be distributed at the same time as the questionnaire at the case site (hence, it will be distributed before and after the measure implementation). The difference between the score from the measure implementation site and the control site is the real effect of the measure, since the other influences like less car use or shift in attitude would have occurred in the control site as well.

Those will be described in detail in the next chapter. Apart from the evaluation design, it is first of all necessary to define a baseline, which is a set of factors or indicators used to describe the situation prior to an intervention. Therefore the baseline acts as a reference point against which the progress of your measure can be assessed and a

comparison to the results can be made. To receive meaningful evaluation results, baseline data needs to encompass all indicators that may change because of the project or measure. It is also important to have baseline data of a sufficient scale to enable expected changes to be judged statistically, if appropriate and possible (see Chapter 2.5.7). Consider the Utrecht Road Safety Label again. It is awarded to schools, after they have adopted the uniform school surroundings and the awareness raising campaigns. Its objectives are to increase the safety of school children, to promote active transport modes for the children's school routes and to increase parents' perception of safety at schools. For each of the sites, where the uniform school surroundings are intended to be implemented, data needs to be gathered on parents' perception of safety, on actual accidents and modal split of school children. To eliminate influences of the planned measure on baseline data, application of a baseline requires that these respective data are collected before any actions are taken and also before parents are informed about the coming changes to the school surroundings. At a later stage of our evaluation, these baseline figures can be compared to the data gathered after restructuring of sites and related campaigns.

Now, do you remember the twin example from above? How can you separate the impacts of the measure from other effects that would influence parents' perception on safety? This is where the control site comes into play (see also Box 3). So do not forget to collect your baseline data for the control site, too.

Further readings

Bhattacherjee, Anol: Research Design (Chapter 5). In: Bhattacherjee, Anol: Social Science Research: Principles, Methods, and Practices. Open Access Textbooks, Book 3, 2nd edition, 2012. (http://scholarcommons.usf.edu/cgi/viewcontent.cgi?article=1002&context=oa_textbooks)

2.4.1 Randomised control group designs

The strongest and thus preferable evaluation design is to compare a group or area exposed to the measure to a group or area without specific measure implementation (control group/site). Both groups are 'tested' before and after the measure implementation with the same data collection methods (i.e. surveys). The inclusion of a control group/site allows for a profound assessment of a measure's effectiveness because in this case, the pre-test/post-test comparison of your control group corresponds to the business-as-usual scenario.

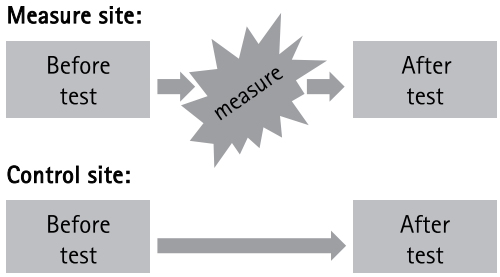


Figure 2–8: Before-and-after data collection in the measure site and in a control site.

Ideally, in order to ensure the comparison of equivalent groups (e.g. regarding age, sex, and labour) the allocation of participants to either the experiment group or the control group is at random (by a coin toss or using a table of random numbers). Hence, any outcome differences observed between those groups at the end of a study are likely to be due to measure and not to differences between the groups that already existed at the start of the study. These features of experiments are so highly priced that the randomised experiment is often referred to as gold standard. So for our twin example further above, you choose the twins randomly, instead of choosing those that applied to this special treatment who might have been a lot smaller than their classmates.

2.4.2 Quasi-experimental control group design

There are a number of methodological advantages of evaluation designs in which subjects are randomly assigned to the project and control groups prior to the project as this avoids systematic differences between those two groups. However, in many operational settings random assignment is not possible so the two groups will be matched as closely as possible. This might be the case for schools of similar size within the same region or city as for the Utrecht Road Safety Label.

But, if you are not going to randomise, you still have to demonstrate that you will be able to make valid conclusions from your evaluation. Quasi-experimental control groups may differ from the measure group in many systematic ways other than the presence of the measure itself – they could be, for instance, already very susceptible to changes before your measure starts. Many of these ways could be alternative explanations for the observed effects, and so you have to worry about controlling them. As such, it is important that you measure all relevant variables at baseline and investigate baseline differences among the groups.

Taking the Utrecht Road Safety Label again, schools freely participate in this programme, while the control group chosen did not want to. If the measure group does

better, can this be attributed to the measure itself, or to the possible fact that teachers and parents decided to participate because they had already been more aware of the subject even before the measure began? Additionally, an alternative explanation might be that the non-random control group included more socially disadvantaged children of which fewer have a bike for cycling to school.

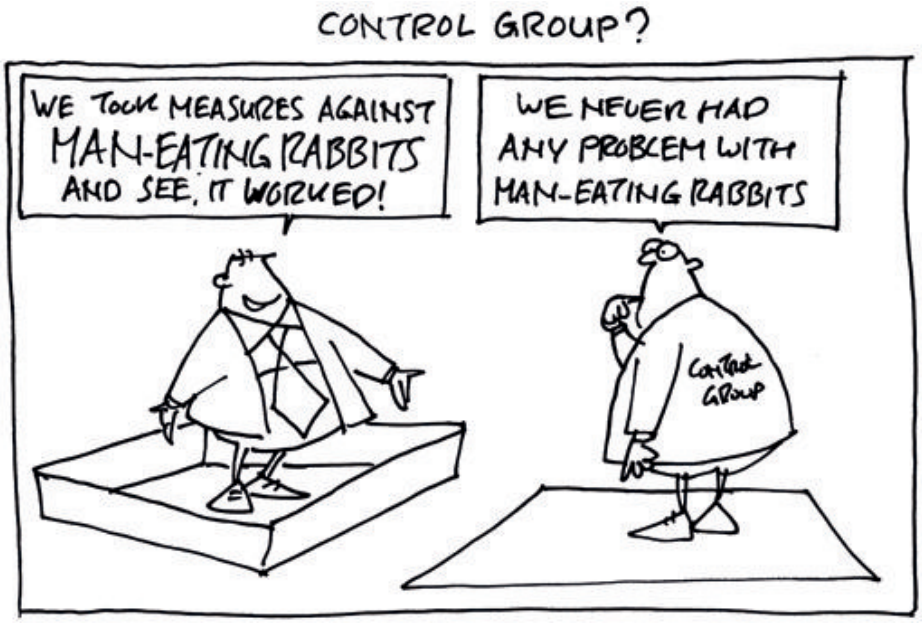
Your baseline data collection should also ensure that your chosen control group experiences the same or at least similar problems which could be addressed with your measure. If you think of the Road Safety Label again, you should not choose a school as control site which is situated in a pedestrian zone. Their problems with children's safety on the road will most likely differ from the schools that participate in the label. Just bear in mind that it is neither feasible nor desirable to rule out all possible alternate interpretations of a causal relationship. Instead, only plausible alternatives constitute the major focus and your cause-and-effect-chain may support you to find those (see Chapter 2.2). This serves partly to keep matters tractable because the number of possible alternatives is endless.

Box 4: Quasi-experiment?

The term quasi-experiment refers to a type of research design that shares many similarities with the randomised control trial, but specifically lacking the element of random assignment. With random assignment, participants have the same chance of taking part in a measure. As such, random assignment ensures that both the experimental and control groups are equivalent. In a quasi-experimental design, assignment to a measure is based on something other than random assignment. In the Utrecht Road Safety Label, random assignment would imply that the participants (children, parents, and teachers) would have been randomly assigned to a school which is (or is not) applying the Road Safety Label – clearly this is not the case for school classes. You need to work with the already pre-defined groups. Simply choosing the schools at random would still be quasi-experimental since you do not 'control' factors like the area the school is located. As a result, most research designs applied in transport measures are actually quasi-experimental.

2.4.3 Designs without control groups

If control groups cannot be identified or can only be included in the evaluation at high expenditure, merely 'before-after-comparisons' and 'panel-designs' are acceptable from a methodological point of view. In those cases it will no longer be possible to



identify any net effects. Designs of time series analysis normally provide more reliable results but potentially would require higher survey efforts. A panel means that the same persons are asked at least twice, in the before and the after test. In this way the person variables stay constant and do not bias the effects. However, the panel must be planned very carefully; it should include persons from all relevant target groups in a sufficient number (see Chapter 2.5.7). Additionally, the dropout rate of the follow-up survey(s) can be minimised by offering incentives for the participation in the after test (for example a lottery). But most likely, the dropout rate will not be zero and should therefore be taken into account in advance. There are a lot of benefits of a panel such as high data quality and validity of data, but also the high costs for conducting a panel have to be considered.

2.4.4 Choosing the evaluation design

When choosing the evaluation design the existence and character of a control group as well as the frequency of measuring are important criteria. This is true regardless of the question whether a new baseline investigation is conducted or available material is analysed.

In this step, you will choose an evaluation design according to the framework of your measure and the available budget. In all cases, you will need to have at least

one measurement before and one measurement after the measure implementation. Otherwise, the evaluation will have no validity. Random experiments allow for causal analytical conclusions with highest validity (see Figure 2-9). But especially in complex measure bundles (e.g. traffic development plan) control group designs are difficult to implement due to high expenditure and sometimes are even impossible because there is no group available.

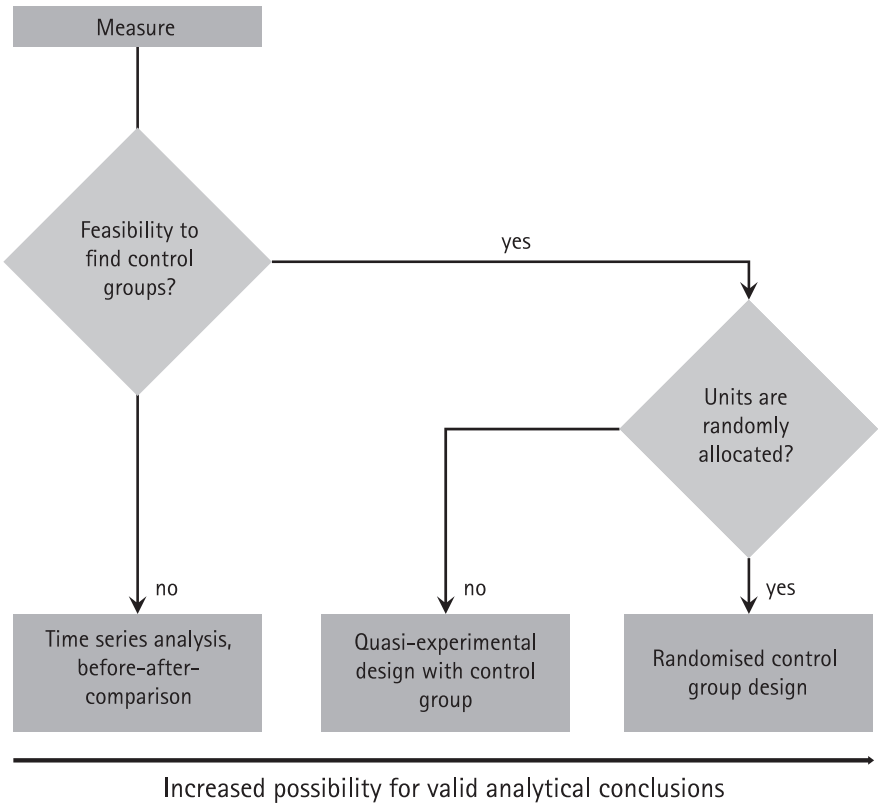


Figure 2–9: Choosing the evaluation design.

Data collection often represents more than half of the cost of an evaluation and is a time-consuming task. Hence, this section presents ways to simplify the evaluation design and discusses both the impact on the quality of the evaluation and the validity of conclusion. In fact, some of the most economical designs (designs 4 and 5 discussed below) cannot be considered high quality impact evaluations, although they may produce operationally useful findings.

All the 'simplified' designs described eliminate one or more of the following observation points: the baseline (pre-test) control group, the baseline project group or the

post-test control group. It is not possible to eliminate the post-test project group as this is always needed to measure the project effect. But be aware that by eliminating the baseline you greatly reduce the validity of the evaluation. The five simplified design options are summarised in the table below. Because of their popularity, the two designs which do not qualifying as sound impact evaluation are nonetheless included.

Bear in mind, if the technical or financial frameworks do not allow any use of control group designs and thus make it necessary to fall back on a 'weaker' experiment design, the validity of the results will be very restricted. It must be stated clearly that the utilization of a simple time series analysis or a before-after comparison does not allow for precise conclusions regarding the measure's exclusive effects. So take the time to choose an appropriate design and consider your decision carefully; Table 2-2 might help you with that.

Further readings

Bonate, Peter L.: Analysis of Pretest-Posttest Designs. Chapman & Hall / CRC, Boca Raton (Florida), 2000.

Flick, Uwe: Qualitative research design. In: Flick, Uwe: The Sage Qualitative Research Kit – Designing Qualitative Research. Sage Publications, London, 2007, p. 36-50.

Shadish, W.R.; Cook, T.D.; Campbell, D.T. : Experimental and quasi-experimental designs for general causal inference. Boston, Houghton Mifflin, 2002.

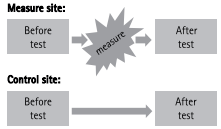
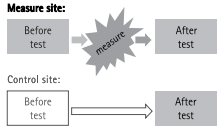
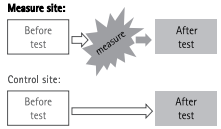
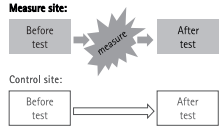
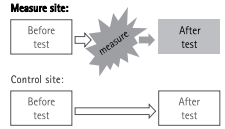
☺ ☺ ☺ ☺ ☺	☺ ☺ ☺	☺ ☺ ☺	☺ ☺	☺
Design 1 – the best	Design 2	Design 3	Design 4	Design 5
Pre- /post implementation project /control group design with (non-) randomised assignment.	Pre- /post-implementation group with only post-implementation control group.	Post-implementation project / control group with no baseline data.	Pre- /post-implementation project group analysis with no control group.	Post-implementation project group with no baseline data and no control group.
				
Strongest design in most cases where it was thought on evaluation from the beginning, starting it at the same time as the project planning.	While weaker than the previous design this is relatively strong as it permits comparison over time and post-project transversal analysis.	A common design when the evaluation begins late in the project cycle or when the project has ended. The lack of baseline data makes it difficult to control for initial differences between the two groups.	A common design when data are only collected on the project group. Methodologically weak, because using before data as the counterfactual requires assumptions about time varying effects and individual unobservable variables.	Weakest design but one which is commonly used when evaluation comes in late in the project with very limited time and money.

Table 2-2: Design options for the reduction of the evaluation design.

2.5 Data collection

In general and as already mentioned, there are two different kinds of data you can use for impact evaluation: Firstly, data that is already available and secondly, data that you collect yourself for the impact evaluation. The acquisition of data is called *primary data collection* since the data is collected by the evaluators themselves (that is you or your subcontractor). If you re-analyse or use data for the impact evaluation that has already been collected, it is called *secondary data analysis*. It is always advisable to look for available data, because it could save you a lot of time and money. This data could be for instance: statistic on companies, ticket sales numbers, accident statistics, statistics on issue of fines as well as statistic on purchased new vehicles, periodic traffic counts and speed measurements, public transport passenger surveys and/or mobility surveys.

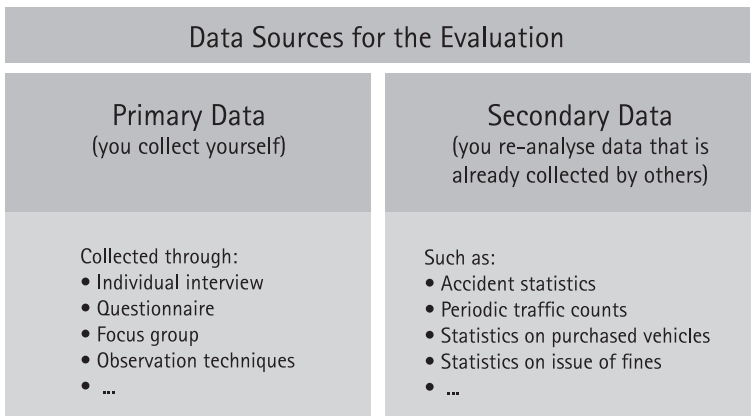


Figure 2–10: Possible data sources for the evaluation.

Secondary data is important to consider because it can save considerable time and expense. It can also be applied to support triangulation of (see Box 5) data sources and to verify primary data analysis collected directly as part of the measure. However, it is critical to ensure that secondary data is relevant and reliable. As secondary data is not tailored specifically for the needs of your measure, it is important to avoid the trap of using irrelevant secondary data just because it is available. However, secondary data analysis is often not enough to monitor the effects of a measure for all selected indicators (see Chapter 2.3 how to do that). Also, keep in mind that secondary data is not always free of charge. So before you buy data from somewhere else, be sure that it is useful to your evaluation. If not, you will be better off collecting your own data. In this primary data collection, surveys can constitute an important part and is a frequently used method.

Box 5: Triangulation

Data used in the evaluation usually originates from multiple sources and results of data analysis are mutually set against one another or compared. This procedure is called triangulation and is used to ensure reliability of the data gathered and to define logically and methodologically proper conclusions. Triangulation can be used for data collection methods (diversity of methods applied), but also for information resources (collecting information from different respondent groups). By combining multiple methods and empirical materials, evaluators can overcome the weakness or intrinsic biases and the problems that result from application of single method, single-observer and single-theory studies.
Taking the Utrecht Road Safety Label again, you could use road safety statistics, answers from the questionnaires to teachers, parents and students and inhabitants around the school.

In the collection of primary data, you can also use a subcontractor. In doing so, make sure that you specify directly what information you want. The more specific you are, the better the results will be. And secondly, check the data or part-delivery of data before you do a (partial) payment. Check it carefully for completeness and accuracy before confirming that they are ok.

In general, mobility data can be gathered for passenger transport (people) and commercial transport (e.g. goods). For these two types of transport the collected mobility data can be either behaviour-related and/or traffic-related. Behaviour-related data can be collected by asking people (e.g. interviews or questionnaires) but also by observing the behaviour (e.g. behavioural observations of cyclists at crossings). Collecting traffic-related data can be done by counts (e.g. vehicle counts). The following chapters will not focus on this traffic-related data collection (including emission measurements) as they differ very much in the technology capacity available in your city. So we will rather focus on using interview methodology – both individual interviews and questionnaires. This means in general talking to selected people in a more or less structured way either face-to-face or via communication channels such as telephone or internet. The following chapters will describe the different forms of surveys and discuss some sampling issues.

2.5.1 Individual in-depth interview

This method can be used in all types and at all stages of evaluation. Its objective is to gather qualitative information and opinions of persons involved in a particular programme – those in charge of designing programming documents, programme implementation, and its direct or indirect beneficiaries. There are several forms of interviews differing in structure and purpose: the informal conversation interview, the semi-structured/guide-based interview and the structured interview. With individual in-depth interviews you have the possibility to learn about all aspects of the researched project. On the one hand you can touch upon complicated and detailed issues. On the other hand it gives the interviewee the possibility to express his or her opinion in his or her own words and to talk about things that are important from his or her point of view. So you can come across unexpected perspectives on an issue which you have not considered before. By the way, this information can also be used for the process evaluation. Weak points of this method are high expenses per interview compared to the questionnaire, as it requires laborious work including thorough preparation transcription as well as a complex and time-consuming analysis of the qualitative data afterwards. So this research method is not suitable for examining hundreds of respondents but is useful to investigate measures dealing with very innovative or complex issues with a small sample size involved and with little or no existing secondary data.

For instance, these individual interviews were used for evaluating the Utrecht Road Safety Label. Evaluators met with teachers from the participating schools for 30 to 45 minutes. This was only a semi-structured interview whose central questions covered the basic stakeholders (students, parents, mobility managers). All conversations were recorded and transcribed. The following text analysis revealed in all schools that it was difficult to reach parents who have not been involved from the beginning. Secondly, the data showed that the program runs more successful, if the teacher and principals involved did not change. While the first result (parents are hard to reach) might be something you come up with yourself and could find out in a survey, the problem with staff rotation during the implementation of the measure was 'discovered' through the interviews. You find information about how to analyse interview data in Chapter 2.6.3.

Further readings

Bhattacherjee, Anol: Survey Research (Chapter 9). In: Bhattacherjee, Anol: Social Science Research: Principles, Methods, and Practices. Open Access Textbooks, Book 3, 2nd edition, 2012. (http://scholarcommons.usf.edu/cgi/viewcontent.cgi? article=1002&context=oa_textbooks)

Kvale, Steinar: The Sage Qualitative Research Kit – Doing Interviews. Sage Publications, London, 2007.

Rapley, Tim: The Sage Qualitative Research Kit – Doing Conversation, Discourse and Document Analysis. Sage Publications, London, 2007.

2.5.2 Questionnaire survey

This tool can be applied to a larger group of respondents than interviews, and it is quite easy to undertake and analyse. It is conducted by the use of a categorised list of questions and possible answers, which are presented in the same form and order to all respondents. This kind of interview is used to decrease the differences in questions asked to various persons, and thereby increase the comparability of answers between the participants of one survey and surveys conducted in different time periods. The more standardised a questionnaire is, the larger number of closed questions it contains. Closed means that the interviewee is being given predefined answer categories such as 'yes', 'no' and 'do not know' or a scale is provided, on which a value for the answer should be rated (see the example below). There are different kinds of scales and the most appropriate scale for the purposes of the questionnaire must be selected. In case of a less standardised questionnaire, the respondent is free to formulate his or her answers as he or she wishes to one or more questions of the given set of questions. In this case you cannot or do not wish to restrict the answers and you are able to receive answers beyond the expected ones. A mixture of standardised and open questions can be recommended exploiting the advantages of both types of questions.

Box 6: How to code the questionnaire ID in dependent samples

You want to ask the same person before the implementation of a measure and after (a panel survey). For instance it is a good design to look at modal splits of people that were treated with an individualised smart travel campaign and compare them to people without the campaign. Questionnaires need to be anonymous due to legal restrictions of data security and privacy. It is no problem to build in a question that allows you to identify people with and people without the smart travel counselling. But how could you link the persons before questionnaire to the after questionnaire filled in by the very same person? Here comes a clever solution that is easy to implement and also addresses the privacy issues.

You need a code or ID that is anonymous and also not forgotten by the respondent (he or she needs to fill it in the after questionnaire again). Therefore let them create their own unique six digit ID for example in the following way:

1. First Letter of the birth name of your mother
2. Last Letter of your own birth name
3. Day and month of birth day (form: ddmm)

In this way you will get from a person named Smith (birthday 20th September) whose mother was born as "Picek" the following ID: PH2009. You can be sure that a person asked in the second wave of the questionnaire survey will remember this code if you reveal the coding-procedure.

A questionnaire can be undertaken by postal mail, telephone, e-mail/online or face-to-face (see also Chapter 2.5.6). This method is, however, characterised by only small flexibility of answers by predefined categories. There always remains a risk of overlooking one of the most important issues or most probable answer categories if the questionnaire contains no questions or categories referring to these particular ones. So the list of questions and answer categories need to be prepared carefully and pre-tested to reduce this error. The questionnaire survey is suited to the observation of the results and impacts of a programme and is therefore to be reserved for evaluations of simple and homogenous measures. Nonetheless, these surveys can be more effective than interviews since the data analysis is faster and makes a larger sample size possible.

Questionnaires for primary data collections were used for the evaluation of the Tallinn Knitting Bus campaign. Onboard the concerned bus line, the users were asked several questions about their perception of comfort and the overall image of public transport. The same questions were asked before and after the campaign (see Figure 2-11 for an excerpt, the complete questionnaire can be found in Chapter 7).

1. What is the purpose of your journey?

☐	Work	☐	Medical reasons
☐	Education	☐	Home
☐	Social/Recreational	☐	Other

2. How often do you use the city bus services?

- ☐ 5–7 days a week
- ☐ 3–4 days a week
- ☐ 1–2 days a week
- ☐ 1–3 days a month
- ☐ Less frequently

3. How satisfied are you with the following elements of the bus survice?

	Very dissatisfied	Dissatisfied	Neither dissatisfied nor satisfied	Satisfied	Very satisfied	Don't know
Lighting	☐	☐	☐	☐	☐	☐
Ventilation	☐	☐	☐	☐	☐	☐
Seat comfort	☐	☐	☐	☐	☐	☐

Figure 2–11: Excerpt from Tallinn bus survey.

Further readings

Bhattacherjee, Anol: Survey Research (Chapter 9). In: Bhattacherjee, Anol: Social Science Research: Principles, Methods, and Prac-tices. Open Access Textbooks, Book 3, 2nd edition, 2012. (http://scholarcommons.usf.edu/cgi/viewcontent.cgi?article=1002&context=oa_textbooks)

Transportation Research Board (TRB): A Handbook for Measuring Customer Satisfaction and Service Quality, TCRP Report 47. National Academic Press, Washington D.C., 1999. (http://onlinepubs.trb.org/onlinepubs/tcrp/tcrp_rpt_47-a.pdf)

2.5.3 Focus groups

The focus group is an established method of social inquiry, taking the form of structured discussion, moderated by the evaluator who supplies the topics or questions for discussion. Focus groups provide a facility to bring together different stakeholders in a measure for mutual discussion and confrontation of opinions. It is especially useful for analysing themes or domains which give rise to differences of opinion that have to be reconciled, or which concern complex questions that have to be explored in depth by various perspectives. The technique makes use of the participants' interaction, creativity and spontaneity to enhance and consolidate the information collected and make use of possible group synergy effects.

Focus groups can be used at every stage of the evaluation process and in all evaluation types. But before holding focus groups it is essential that you know your target audience. In some cultures people of different income groups, gender or age can be mixed. In others this would be a recipe for disaster since either one group would hog the discussion or even worse nobody talks at all. For example, focus groups are typically used when feedback on new interior designs for trains or airplanes is needed. It is common to present the drafts to a mixed group of (possible) users such as business people, mothers with children, mobility impaired, and elderly people. Often, they each judge the designs differently. Through a guided moderated discussion among all of them, which supports mutual understanding, it is possible to accomplish a broad consent or compromise. Hence, the focus group is a good evaluation technique for innovative pilot measures.

Further readings

Barbour, Rosaline: The Sage Qualitative Research Kit – Doing Focus Groups. Sage Publications, London, 2007.

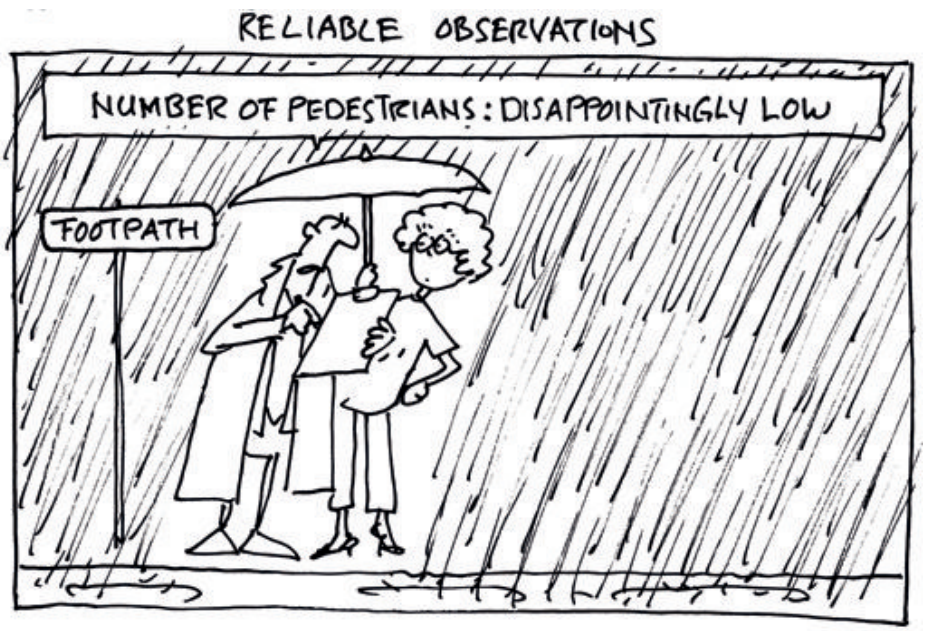
Krueger, Richard A.; Casey, Mary Anne: Focus Groups: A Practical Guide for Applied Research. 3rd edition, Sage Publications, London, 2009.

2.5.4 Observation techniques

Observation assumes that evaluators collect data by direct participation in the measures undertaken or observation from an external point of view. The idea is that you are going to the place where the measure is implemented and thus better understand

the context of the measure and how the user/target group experience it. A trained evaluator may also perceive such phenomena that escape others' attention, as well as issues that are not tackled by participants in interviews (like conflicts). Observation can enable you to exceed participants' selective perception. With this technique it is possible to present a versatile picture of the measure, beyond use of questionnaires and interviews. However, this technique should only be seen as a supplement to surveys and/or interviews. It is particularly useful for studies from a disabled person's perspective.

Besides going to the implementation site yourself and filling a more or less standardised report/protocol including your observation criteria and space for further impressions, it is also possible to rely on video surveillance, which needs to be analysed afterwards. The latter is often used in anti-vandalism measures. However, for certain information it is advised to go there yourself. As an example, if you want to know, if passengers are reacting differently to the information system actually telling them their train would arrive shortly. Do they run to catch it? Will they walk as before because the next train will arrive in 10 minutes? In this example, you will need to look at the information table as well as at people the (re)acting. However, if you would like to know why they behave one way or another or if there are other decisive factors than the information table you need to do a survey afterwards. As a result, camera use would not be advisable since you would likely miss information. In any case, you need to make sure that you protocol your observation in detail.



2.5.5 Basic guidelines for conducting a survey

As mentioned above, the format of an interview guideline can range from a list of open questions with the respondent asked to talk more or less freely ('guided interview') to a list of questions with given answer categories, where the respondent's answers are categorised directly into the possible answers. Open questions producing more complex answers have the advantage that more qualitative data can be collected, revealing thoughts and providing deeper insights of personal reasons. But the data analysis is time-consuming since the categorising has to be done afterwards. Already categorised answers have the advantage of effective data analysis but entail the risk that the selected categories do not reflect the real answer structure. Usually, a combination of open and closed question formats is applied in surveys.

Box 7: How do I know that I have a good survey?

A survey is only as good as the clarity of your indicators, the directness of your questions and the relevance of your answer choices. This is where asking for the "best" of something fails. Best how? We all define it differently, and in different contexts. Someone might consider "best" the shortest, whereas somebody else might consider "best" the fasted. The question is: are you asking the kind of question which will give you the answers needed?

You should spend considerable effort in developing smart questions. To take you a little off the subject: When it comes to movies, instead of asking in a survey "who is the best actor," you would need to ask something such as, "which of the actors below has consistently demonstrated their skill in character portrayal?" It's still an opinion, but you're at least asking their opinion about a very specific thing. Also, your answer choices are huge, and you need to list options which are meaningful. If you list a bunch of actors and ask people to choose their favourite, you're invariably going to get folks who do not see their favourite actor, or who don't see any differences in your list.

The following statements might help you in designing a suitable good survey:

- Be clear in the statement of the objective of your survey.
- Collect only relevant data – avoid collecting irrelevant data, this would be a waste of resources and could in the case of overlong questionnaires influence the quality of the data collected.
- Think about the required precision of the result of your survey before you make measurements; this is further considered later in this section.

- Remember that your staff needs special training for administering the survey. Adequate supervision is required and early checking of the quality of the collected information is invaluable.

Also, carry out a pilot/pre-test of your survey to determine if there are problems of understanding/interpretation of the questions and of the method of conducting the survey. You may simply ask your colleagues or friends to pretend to be survey persons. Especially under tight budgets, it is always better to have a smaller sample rather than to skip the pre-test. Be sure to ask the same questions before and after the measure implementation. In other words, you design one survey, but use it twice. Except, of course, for the introductory passage and modifications due to the fact that – the second time – you are asking after the measure implementation.

Box 8: Choosing the right scale

A rating scale is a set of categories designed to elicit information about a quantitative or a qualitative attribute. They are either unipolar or bipolar. A unipolar scale prompts a respondent to think of the presence or absence of a quality or attribute: "never", "rarely", "sometimes", "often", "always".

Where a unipolar scale has that one 'pole', a bipolar scale has two polar opposites. A bipolar scale prompts a respondent to choose between two opposite attributes, determining the relative proportion of these attributes. A common bipolar scale is the Likert scaling. The format of a typical five-level Likert scale is: "strongly disagree", "disagree", "neither agree nor disagree", "agree", "strongly agree". Sometimes an even-point scale is used, where the middle option of "neither agree nor disagree" is not available. This can be called a "forced choice" method; since the neutral option is removed and the tendency is measured (a option is considered better or worse). The neutral option can be seen as an easy choice to take when a respondent is unsure, so that it is questionable, whether it reflects a true neutral attitude. But both types of Likert scale provides the metric data necessary for proper quantitative data analysis (see Chapter 2.5 and figure 2-8 for an example).

Over time, there have been many discussions and disagreements focused on one central question: What works best with the Likert scale to give you the most accurate responses? The following are points that most scholars agree on:

- More than seven points on a scale are too much. Studies show that people are not able to place their point of view on a scale greater than seven. So go for seven or less. What is the perfect number? Studies are not conclusive on this, most commonly mentioned are five- or three-point scales.

- Numbered scales are difficult for people. For example, scales that are marked "1 to 5, with 5 expressing the highest degree of something result in less accurate results than scales with labels such as "good" or "poor". If numbered scales are used, explanatory notes are recommended (e.g. put "poor" above 1, "satisfactory" above 3 and "excellent" above 5).

Further readings

Baltes, Michael R.: Surveying for Public Transit: A Design Manual for Customer On-Board Surveys. National Center for Transit Research (NCTR), University of South Florida. (http://www.ctre.iastate.edu/pubs/midcon2003/BaltesSurvey.pdf)

Brace, Ian: Questionnaire Design – How to Plan, Structure and Write Survey Material for Effective Market Research. 2nd edition, Kogan Page, 2008.

Brink, T.L.: Questionnaires: Practical Hints on How to Avoid Mistakes in Design and Interpretation. Heuristic Books, Chesterfield, 2004.

2.5.6 Advantages and disadvantages of data collection methods

Selection of an appropriate method requires careful consideration of many factors, not in the least of which is coverage of the target population. While the method of data collection might be largely determined by the population and sample frame, other common determinants include survey costs, response rates, and data quality issues. Method selection can also be influenced by the complexity and length of the survey and timeliness needed. But also your country specific data protection and privacy regulations can have an influence on your choice. This section provides a summary of methods of data collection along with associated features of each.

Face-to-face data collection typically yields the most complete coverage of the topic, achieves the highest response rate, and produces the best quality data. Not surprisingly, in-person interviews are also the most expensive of the methods. For this reason, telephone and mail methods are more commonly used despite well-recognised trade-off in data quality. Other obstacles to personal interview include personal security and access, such as to gated communities, busy families with children or businessmen and the like.

In a *telephone questionnaire*, respondents are called by survey teams to answer a series of questions which are recorded during the survey. Depending on the scale of the survey, it can be costly to set up the appropriate systems to conduct telephone surveys, although professional agencies may provide a suitable solution. Compared to postal questionnaires, telephone surveys can get higher response rates, as such they are more representative of the population, provided possible bias in the telephone number sampling frame is addressed (e.g. young people using mobile phones). It may also be difficult to obtain a sample within a defined geographical area.

The main advantage of *postal questionnaires* is that they are reasonably priced, and they can have a wide geographical distribution. However, postal questionnaires take a long time to send out and get back. Low response rates and incomplete forms are common problems with such methods.

Box 9: main(){cout << "Hello World!" << endl; return 0;}

In recent years, it has become very popular to use the internet for surveys. Often, it is easier and faster than the 'traditional' paper questionnaire and – with more and more people who have access to the World Wide Web, the response rates are reasonable. There are, however, some things that you will have to remember when using the internet. Distribution and access are different among countries and cities. While you can safely assume that in Iceland there is 100% coverage of the people having access, in Italy it is only about 50%. Hence a true random sample is difficult to achieve so the results may be biased to higher educated groups and younger people who do have access to the internet and miss out other groups.

On the other hand, there are also major benefits. The results are usually sent directly into a database allowing you to access the responses immediately. They are also relatively cheap to conduct as there are already numerous platforms for presentation of online questionnaires including: survey monkey, esurveys.pro, NetQ and QuestionPro. Especially through their individualised approach you can adapt the survey for your specific use by avoiding such common phrases like "if no, continue with number ..."

2.5.7 Sample size and sampling issues

It is important to give proper consideration to the size of the sample required: The more completed surveys you get, 1) the greater the confidence and 2) the greater the cost. So, a trade-off exists between two objectives in a survey project: Maximise confidence but minimise costs. Four factors determine the statistical confidence:

Size of the targeted population: The population is the group of interest for the survey. A sample is drawn from the population and the survey is administered to the sample. Some percentage of the sample responds to the survey invitation. That percentage is the response rate. For instance, if you invite 400 people to take part in an online survey, but only 250 participate, then your response rate is 62.5%.

Segmentation analysis desired: Typically, we analyse the data set as a whole, but we also usually analyze the data along with some segmentation, for example, number of participants of the measure or car users compared to public transport users. Each segment in essence is another population. If the conclusions will be drawn on the analysis of a segment, then statistical confidence must be focused on the segment, not the population. By the way, if you want to analyse men and women separately, you have two segments.

Degree of variance in responses from the population: This factor is the hardest to understand for the statistically challenged. If the respondents' responses tend to be tightly clustered, then we do not need to sample as many people to get the same confidence as we would if the responses range widely. Imagine you polled your office colleagues, and the first five people gave the same answer. Would you continue polling? Probably not; what if you got five different responses? You would probably keep polling. Therefore, more variability requires larger samples. But until we do some surveying and analyse the data, we do not know anything about the variance! So, initially, we need to employ conservative assumptions about the variance.

Tolerance for error: How accurate do you need the results to be? If you're drawing conclusions from the evaluation which could result in a multi-million Euros investment decision, then you probably have less tolerance for error. You can find more information on this aspect in the next chapter under confidence interval.

Box 10: How many is enough?

Often, the only things people recall about their statistics class is the Central Limit Theorem and remember that a sample size of 30 or more is considered to be large enough for the theorem to take effect. 30 responses would provide acceptable accuracy only if a) you have a very small target group, b) you have very little vari-

ance in the responses, or c) you are willing to accept very low accuracy. As a very rough rule of thumb, 200 responses will provide fairly good survey accuracy under most assumptions and parameters of a survey project – except for analysis within each segment! 100 responses are probably needed even for marginally acceptable accuracy.
Under the assumption that the population to be sampled is approximately normally distributed and the total population is large, a first approximation of the sample size n is given by:

$n_0 = t^2PQ / d^2$, where
t = the abscissa of the normal distribution for probability α
P = expected population value of the proportion
Q = (1-P)
d = margin of error

For example: A company wants to know how its employees feel about company-provided car parking at its head office and at each of its regional offices by asking the employees to fill in a questionnaire. To keep costs down, they only want a small sample to complete the questionnaire. The head office has 5,000 employees and they want to know that the worst case answer (i.e. when 50% of respondents give a particular answer, for instance to whether the employee would prefer to have a free bus pass or a parking space at the office) will be correct within $\pm$3%, with only a 5% chance that it will be outside this range (i.e. 95% certainty), so:

$d = 0.03$, $p = 0.5$, $\alpha = 0.05$, $t = 1.96$
and thus: $n_0 = (1.96)^2(0.5)(0.5) / (0.03)^2 = 1067$

Thereby, n_0 is only an approximation, as a general rule of thumb, the coefficient of n_0/N should be sufficiently small. In this case $n_0 / N = 1067/5000 = 0.2134$ is not sufficiently small, so you use the equation:

$n = n_0 / [1 + (n_0-1) / N]$
$n = 1067 / [1+1066 / 5000] = 880$

to compute the 'final' sample size.
This means that 880 respondents were needed – just less than a fifth of their head office work force.

Hence, when deciding a sample size, n, you should find an appropriate formula for linking n with the desired precision. Formulae for the estimation of n from a simple

random sample are provided in the Box 10. There are also a number of online sample size calculators available. One such example can be found at http://www.surveysystem.com/sscalc.htm. Please remember that the accepted standard varies between 90% and 95% for the confidence level and between 5% and 10% for the margin of sampling error.

In designing a questionnaire survey, it is easy to become overburdened by trying to generate a perfect random sample whereas in reality a perfect random sample will never be achieved. Whilst measures can be taken to improve the random nature of the sample there will always be some people who will be more inclined to respond to a questionnaire than others. For example, retired people will have more spare time disposable to filling in the questionnaire. (But in some cultures they may not be able to read or have bad eyesight.) It also can happen that a certain measure deals with quite an emotive issue – those more concerned about transport-related issues will be more inclined to fill it in. It is therefore important to choose sample sizes large enough to have enough respondents within certain sub-samples of interest (e.g. young people compared to old people).

It should also be noted that sample sizes are the numbers required to be returned, and this can differ quite drastically depending on the subject of the questionnaire, incentives for reply and the target group. Local information on response rates from previous questionnaire surveys can be very informative. This response rate will depend on your survey method (e.g. postal, e-mail, face-to-face, handed out).

Further readings

Bhattacherjee, Anol: Quantitative Analysis: Descriptive Statistics.
In: Bhattacherjee, Anol: Social Science Research: Principles, Methods, and Practices. Open Access Textbooks, Book 3, Chapter 14, 2nd edition, 2012. (http://scholarcommons.usf.edu/cgi/viewcontent.cgi?article=1002&context=oa_textbooks)

2.6 Data analysis

Numerous books have been written on the 'correct' analysis of data. There are also many experts on this field, but this should not scare you. Just bear in mind that the following subchapters can only provide an introduction to this vast discipline. The following Chapters 2.6.1 and 2.6.2 describe the quantitative data analysis, while Chapter 2.6.3 gives you an introduction to the qualitative data analysis. They are pointing out key elements and a terminology you should familiarise yourself with. However, if this is the fist time that you are concerned with an evaluation, you might want to think about getting one of those experts involved.

2.6.1 Data preparation and back-up

Experts in statistics like to boil everything down to one single figure. They are fond of the possibility to formulate statements regarding the distribution of characteristics of interviewees or the effect of a measure based solely on one measured value. In order to do so, the monitored objects are matched with numbers. Nonetheless, a mere matching of numbers to objects is not a feasible definition for 'measuring'. That would mean that grades at school might just as well be assigned by rolling dice. The measured values have to display interrelationships among each other that correspond with the relations of the measured objects (i.e. that can actually be monitored). Looking at a simple length measurement, this becomes obvious: Here the largest measured value has to match the longest object while the smallest measured value matches the shortest object. In this example an amount of objects sorted by length corresponds to an amount of figures sorted by magnitude of number. The more characteristics of an object are expressed in the magnitude of the corresponding number, the more information the measured values contain. In this context we also speak of scales. There are four different types of scales which are important to know for all following statistical analyses. The different levels of scales determine which mathematical operations or statistical methods are allowed to be used when working with data.

The nominal scale allows us to differentiate only between equality and disparity of the examined dimension. The matching of a number in this case is a mere denomination. The ordinal scale allows us to rank objects. The assigned numbers have to match the ranking. In case of an interval scale the differences between two consecutive objects are of the same size. Since the intervals' sizes are identical, the assigned numbers must always have the same difference. The ratio scale contains a 'natural point zero'. The measured value 0 states the actual absence of the measured characteristic. This scale allows us to make statements regarding the proportion of two objects. You can find illustrating examples in the table below.

Types of scales	Allows statements on	Examples
Nominal scale	equality disparity	Sex (e.g. male, female) choice of traffic mode (e.g. walk, bike, car, public transport etc.)
Ordinal scale	larger than/smaller than relations	customer satisfaction (e.g. very satisfied, satisfied, neither nor, dissatisfied, very dissatisfied)
Interval scale	equality of differences	temperature (e.g. °C)
Ratio scale	equality of proportions	measuring velocity (e.g. 30 km/h), measuring length (e.g. 100 m)

Table 2–3: Examples for different types of scales.

In order to reasonably structure the data that is to be collected and to depict it in a table, a so-called code plan is necessary. It is a list of all gathered variables containing all their possible manifestations and always has to be documented. Thereby, every possible manifestation is assigned with a specific numeric value (code). Then the information gathered, has to be encoded according to this plan (see table 2-4).

Following excerpt from the Tallinn example where people travelling on the bus were surveyed shows the setup of a possible code plan.

Take a look at the question: How often do you use the city bus services? The five possible answers are assigned with numbers 1 to 5. Regarding the question of sex, there are two manifestations matched with the numbers 1 and 2. In a chart, the variables should be arranged in the same succession as the answer options in the questionnaire. The code plan should start with a so-called identification number, which is individually assigned to every interviewee. The identification number serves control purposes. This way completed questionnaires can be compared to the data entry and possible mistakes when entering the data can be corrected.

3. How often do you use the city bus services?	1. 5-7 days a week 2. 3-4 days a week 3. 1-2 days a week 4. 1-3 days a month 5. Less frequently
4. Do you have a car?	1. Yes 2. No
5. Could you have used the car today but instead chose the bus?	1. Yes 2. No
6. Gender	1. Male 2. Female

Table 2–4: Code plan for Tallinn Knitting Bus example (partial).

All the actually surveyed data (e.g. answers of the interviewee) can be depicted in a chart. The data originating from our example is reconstructed using spreadsheet software. Rows represent the surveyed persons and columns show the answers given to certain questions, which were entered according to their code plan. In our Tallinn example, as shown partially in Table 2-5, the answers on a scale from 1 to 5 to the assessment of certain items regarding the evaluated bus are listed.

ID	Lighting	Ventilation	Seat Comfort	Cleanliness of the seats	Cleanliness of the floor
39	4	2	4	3	3
40	5	2	4	3	3
41	5	3	4	4	4
42	4	4	4	4	4
43	4	2	4	4	3
44	5	4	5	5	5
45	4	4	3	4	3

Table 2–5: Data chart for Tallinn Knitting Bus (partial).

The direct input in a tablet can be handled in different ways. Ideally, there is a programmed input screen (use for instance software like Data Entry or Microsoft Access) that looks exactly like the questionnaire used. When using an input screen, unreasonable inputs (e.g. age: 200) can be prevented through corresponding error checking. For very extensive questionnaires or large samples programming an input screen is highly recommended. An input screen can also be equipped with automated filters which guide the person entering the data through the questionnaire. The disadvantage is the extra

time the programming of an input screen takes. Furthermore, you have to potentially consider high acquisition costs for the software.

Another possibility of data entry is provided through scanners. This way, questionnaires can be scanned and read in according to their code plan. This method of data entry saves time and effort. It makes sense when dealing with a high number of cases while the questionnaires have only few pages. The questionnaires have to be filled out properly and reliably. Otherwise there might be problems in the scanning process which increase the effort of manual rework. The requirement of system-compatible filling in of forms can be a disadvantage, if many of the questionnaires are not completed properly. In this case, the data entry has to be done individually after all. Also another disadvantage is that the design of the questionnaire has to be system-compatible with the scanner software. Thus the design options are limited by the software and the produced questionnaire might not be very appealing. A nicely designed survey form is often the key to motivating potential interviewees to complete the survey.

Eventually, we can feed the data to spreadsheet software (Microsoft Excel) or statistics software (e.g. SPSS, SAS, Statistica etc.). Here the numbers are mostly entered into a prepared data matrix. The advantage is that there is hardly any preliminary work for data entry and no further technical prerequisites are necessary. Ultimately, this way of data entry is recommended for a minor number of cases and short questionnaires (up to 300 cases with 30 variables). Unfortunately, the process of data entry is very time-consuming and labour intensive and also very prone to error. This is because data entry has to be done highly concentrated. Let us face it: Both financial and technical restrictions in many projects leave no other choice but to feed the data e.g. into Microsoft Excel. In these cases it makes sense to carry out data entry in a team of two people. One person reads the numbers aloud the other makes the entry using the numeric keypad.

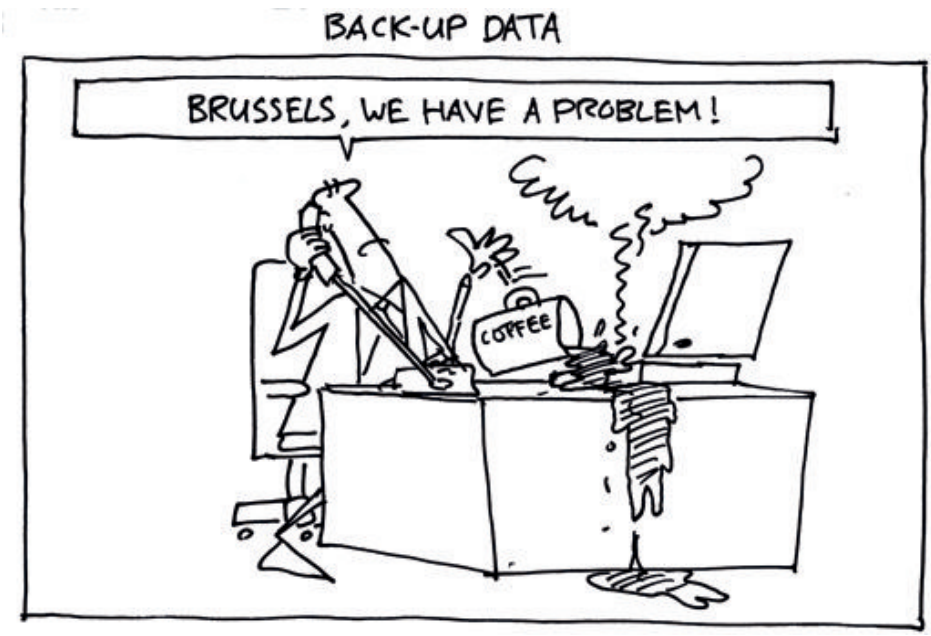
When entering the data, you should stick to some rules. You should enter everything except for obvious mistakes just the way it is stated in the questionnaire. When you enter something different from the answer in the questionnaire, you should document it (e.g. on the questionnaire). Stick to these rules in order to prevent manipulation or falsification of data. If many answers are missing or the answers were not given seriously, the questionnaire is not taken into account for the analysis and interpretation later on.

When entering the data, mistakes can occur. Especially, the data entry to charts of spreadsheet software is often very prone to error. Typing errors can occur easily or an undocumented code is entered (i.e. a number is entered that is not stated in the code plan).

Because of the possibility of mistakes that can occur in the process of data entry, we need to check for completeness and outlier values. A simple enumeration of the variables and the creation of a contingency table – a simple cross–table providing an

	Chart	Input screen	Scanning of questionnaires
Advantages·	• No preliminary work • No technical prerequisites	• Error checking • Filter regulations	• Minor amount of time and work
Disadvantages·	• High amount of time and work • Humans as source of errors	• Additional work • Creation of input screen • Possibly high acquisition cost for software	• Dependency on hardware • Design of questionnaires has to be system-compatible • Questionnaires have to be completed system-compatible
Recommended application	• Very small number of cases • Short questionnaires (up to ca. 100 cases and 30 variables)	• High number of cases • Extensive questionnaires • Questionnaires are filled in untidil	• High number of cases • Small number of pages • Questionnaires are filled in tidly and reliably

Table 2–6: Comparison of different data entry methods.



overview – are the first approach to the data. To begin with you should take a look at the data and see how feasible it is. A five-year-old child holding a drivers license or a car owner not holding one are both hardly plausible. If the data contains such obviously wrong information, you should check on it by referring to the questionnaires and the electronic data record and subsequently correct it or if impossible delete the wrong cases.

Since the gathered data is the foundation for evaluation, publication and decision making and thus the most important element for the following work in general, it has to be stored on durable and secured storage mediums (such as: CD-ROMs, USB-sticks, external hard drives, etc.). The data backup should be conducted regularly, as to always have the latest data secured. Having at least two spatially detached backups increases the reliability of data recovery. The data should regularly be checked for completeness and integrity. As the data is most valuable we have to secure it from unauthorised access.

2.6.2 Data processing and analysis

Some of the first tasks in the course of data analysis are to arrange the empirically gathered data in proper order, to depict it graphically or in tabular form and to sum it up using key parameters such as average value, grand total, minimum, maximum or spread. That is how the data is arranged in a certain pattern.

The question "How often do you use the city bus services?" of our example from Tallinn has been summarised and processed into the following chart.
So the first step of descriptive statistics is the counting of measured values. These counts give you an idea of the characteristics of the objects under investigation. They are possible for all types of scales. Besides absolute frequencies there are cumulative (i.e. the successively summarised frequencies of the categories) and percentile frequencies. The latter allows you to compare distributions with different total extents.

	Frequency	Percent	Cumulated percent
5-7 days a week	472	58.1	58,1
3-4 days a week	186	2.9	80.9
1-2 days a week	86	10.6	91.5
1-3 days a month	40	4.9	96.4
Less frequently	29	3.6	100.0
Total	813	100.0	

Table 2–7: Example chart for a first analysis of frequencies.

In general these frequencies should always be stated with the according number of cases surveyed. When working with scales creating categories or classes, it is often necessary to better illustrate the distributional properties. The number of classes should be chosen, so that the reader can keep an overview. Usually, classes should have the same width.

Tables convey an impression of the overall distribution of a certain characteristic. By contrast, statistical parameters serve the purpose of giving summarised information regarding the specific traits of the distribution of characteristics. In this context, we are most interested in those values that represent a kind of summary of all measured values. Therefore statistical parameters serve the purpose of summarising, while at the same time reducing the information from the distribution of frequencies. There are three kinds of parameters: location parameters, measures of spread and design parameters. Which parameters can be computed depends on the level of scales. Dividing sexes obviously makes no sense. Therefore you have to be sure about which scale (see Table 2-3) the variable under investigation belongs to.

Location parameters are also termed as measures of central tendencies. They allow for statements about the distribution of characteristics regarding the question: Which number best represents the overall distribution? Important location parameters are modal value, median and arithmetic mean (see Box 11).

Box 11: Location parameters

The *modal value* or *mode* is the value of a distribution that occurs most frequently, i.e. the most often occurring category. Modes can be determined on every level of scales.
The *median* divides the frequency distribution into two parts of the same size. In a cumulated depiction you can easily make out the 50%-value as the median.
The *arithmetic mean* is the quotient of the total sum of measured values divided by the amount of measured values.

If two distributions are similar concerning their central tendencies, they can still be very different regarding the spread of individual values. The measure of central tendency (location parameter) states which value best represents a given distribution. By contrast, the measure of spread informs us about the disparity of measured values. The measure of spread thus serves a more precise description of the distribution and also quantifies the differences detected in the survey. Range, variance and standard deviation are common kinds of measures of spread (see Box 12).

Box 12: Kinds of measures of spread

The *range* (or range of variation) comprises the overall range of variation of all measured values. It is defined as the difference between the maximum and the minimum value.
The variance and the standard deviation are the most common indicators of a distribution. In order to compute them we need data on an interval scale.
The *variance* is a number expressing the disparity among the values of a given distribution. It is computed by dividing the sum of the squared deviations from the arithmetic mean by the number of measured values.
The *standard deviation* is defined as the square root of the average squared deviation (variance). It is the most important quantitative figure regarding the variability of the data and is essential for further statistical methods (e.g. statistical tests).

Mere counting is of only limited informative value. The compression of information is mostly reached through statistical processing of information. Their validity however greatly depends on which characteristics have been processed. In this context the reasonable linkage of properties is vital.

In our example from Tallinn we want to correlate the interviewees' sex to the question: How often do you use the city bus service?
Through linking properties we can gain further information. For example, it can be of special interest whether there is a correlation between two variables. Regarding this problem, the statistical data analysis offers us the possibility to calculate measures of correlation. The variety of correlation measures are categorised according to the level of scale of the variable under investigation.

So far, we have described characteristic sample values. But the part which is probably really exciting for you is to make inferences concerning a whole population, in this case the inhabitants of Tallinn. This is the topic of inference. Inferential statistics takes a conclusion regarding the whole population as the starting point and verifies its validity for the sample. A prerequisite for such inferences is that the sample was generated by a procedure of random sampling (see also Chapter 2.4). For people concerned with statistics random sampling is vital. That is because it allows us to calculate probabilities for the realisation of samples. In random samples the statistical units of a population – like balls in a non-transparent container – have a calculable probability of being chosen as part of the sample. It is necessary that every element from a population has the same chance of being included in the selection procedure. The most important criterion of random sampling is the independence of the investigator. The selection is made by chance only.

			Gender		total
			Male	Female	
How often do you use the city bus services?	5-7 days a week	count	176	296	472
		percentage	37.3%	62.7%	100.0%
	3-4 days a week	count	64	122	186
		percentage	34.4%	65.6%	100.0%
	1-2 days a week	count	33	53	86
		percentage	38.4%	61.6%	100.0%
	1-3 days a month	count	12	28	40
		percentage	30.0%	70.0%	100.0%
	Less frequently	count	11	18	29
		percentage	37.9%	62.1%	100.0%
total		count	296	517	813
		percentage	36.4%	63.6%	100.0%

Table 2–8: Example of contingency table "How often do you use the city bus services?"

Within the framework of inferential statistics we mostly use statistical testing methods of which there is a great variety. Choosing one requires knowledge about their adequacy. And, of course, the choice also depends on the level of scale. But some testing methods require for example a specific distribution (e.g. Normal / Gaussian distribution) while others help answering specific problems. Take our Tallinn example. Here we are looking at a before-after-comparison. To choose the adequate testing method we need to know, whether the survey was done using paired or independent samples. But do not lose your head about this. Inferential statistics is not magic, but can be rather complex in parts. So ask for competent professional advice, if you are not sure about details.

For mobility-related evaluation methods, the correlation between two or more groups is of greatest importance. The most commonly used testing methods in this context are the t-test and the chi-square test; whichever test you choose they always follow a similar pattern (see also Figure 2-12).

1. Formulating the null hypothesis and the alternative hypothesis

The null hypothesis is the counterpart to the especially formulated alternative hypothesis. The former is formulated as a competing hypothesis dependent on the alternative hypothesis. It includes the not being true of the circumstances described in the alternative hypothesis, i.e. the expert was wrong with his or her hypothetically formulated statement. The null hypothesis states only the opposite compared to the alternative hypothesis but not circumstances building on an opposing theory. So content-wise, the statement of the null hypothesis is not of any special interest. In the classical test

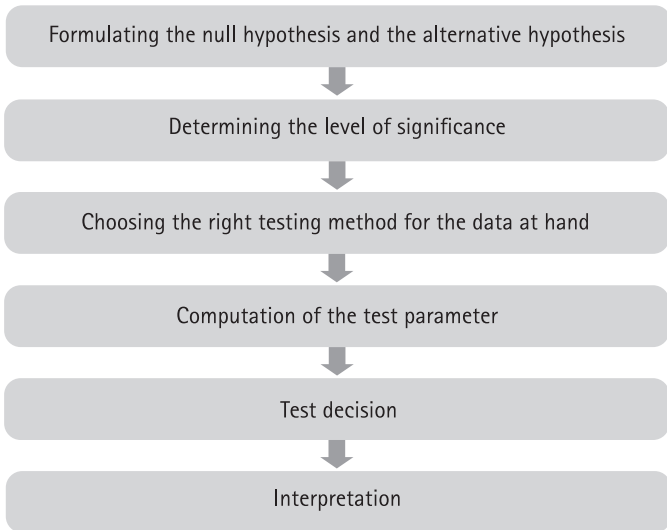


Figure 2–12: General steps to take for statistical testing methods.

statistics the null hypothesis forms the basis on which we decide whether to accept or reject the alternative hypothesis. Only if reality cannot be explained by the null hypothesis, it can be discarded in favour of the new alternative hypothesis. As described above, the null hypothesis (H_0) is opposed to the alternative hypothesis (H_1) in test statistics, while the former necessarily has to originate from the latter.

If the statement in a hypothesis is formulated with a direction, we speak of a directed hypothesis. An undirected hypothesis is the case, if there is just any difference stated. This differentiation does not include any valuation of the two possibilities. Here are some examples regarding the formulation of hypotheses.

Knowing which kind of hypothesis you are testing in your special case is essential for applying the correct statistical test (you need for instance to check a box in your statistic program whether you want to test a directed or an undirected hypothesis – it is often called one-side or two-side testing).

The null hypothesis is the foundation of test statistics. If there are distinct contradictions between the null hypothesis and the measurement results, we discard the former in favour of the alternative hypothesis.

2. Determining the level of significance

The determination of a level of significance is important in this context. It tells us in a percentage, in how many cases we are willing to discard the null hypothesis although it is true for the population surveyed. A null hypothesis is only rejected, if the probability of error is less or equal to 5% or 1%. This limit is called alpha error or level of

	Alternative Hypothesis H_1	Null Hypothesis H_0
Directed hypotheses (one-side testing)	Perceived safety around schools with installed uniform school surroundings is higher than around schools with unchanged school surroundings.	Perceived safety around schools with installed uniform school surroundings is the same or lower around schools with unchanged school surroundings
	Frequent bus users experience higher seat comfort in the knitting bus than low frequent users.	Frequent bus users experience the same or lower seat comfort in the knitting bus than low frequent users.
Undirected hypotheses (two-side testing)	There is a correlation between perceived safety and the kind of school surroundings.	There is no correlation between perceived safety and the kind of school surroundings.
	Frequent bus users ratings of seat comfort in the knitting bus differ significantly from low frequent user's ratings.	Frequent bus users ratings of seat comfort in the knitting bus do not differ significantly from low frequent user's ratings.

Table 2–8: Example of contingency table "How often do you use the city bus services?"

significance. If a survey leads to a conclusion well-founded in the manner described, we speak of a significant (α=5%) or a very significant (α=1%) result. Still a non-significant result does not prove the validity of the null hypothesis. In these cases no final statement is possible besides that our alternative hypothesis could not be proven correct. Looking at the hypotheses mentioned above this question arises: How high is the probability of detecting such a difference when comparing two samples from the same population? If this probability is less than a level of significance defined at the beginning, we reject H0 in favour of the alternative hypothesis.

3. Choosing the right testing method for the data at hand

In this step we choose the adequate computation method based on the underlying problem, level of scale, etc. Here we just would like to name two typical tests – chi-square and t-test – please refer to statistics books or experts for further explanation. The chi-square test has the least requirements regarding the level of measurement. So it is suitable particularly for data on a nominal scale. A typical case for a chi-square test is a comparing analysis between men and women concerning our question: How often do you use the city bus services?

The t-test is a method to validate hypotheses regarding the disparity of two average values (e.g. before and after satisfaction with a public transport service). In this context it is especially interesting to investigate whether a measure has for example led to different behaviour in traffic or increased customer satisfaction. A prerequisite for this test are variables on an interval scale. The t-test can only be used, if average

value and variance of a series of measured values can be reasonably interpreted based on the level of scale.

4. Computation of test parameter

Using the data we determine a one-sided or two-sided exceeding probability (see also directed and undirected hypothesis). Depending on the test method this parameter is named t, U, c^2 etc.

5. Test decision

The determined exceeding probability is compared to the defined significance level. If we have a significant result e.g. test parameter < 5% the Null Hypothesis is rejected in favour of our alternative hypothesis.

6. Interpretation

We now interpret our result (significant or not). But be aware of overhasty conclusions! It is possible that you have calculated a significant result although the prerequisites for this were not even met. For example the number of cases can be too little to validate a result that would otherwise be significant. So be careful on details here.

An example of how to present this kind of data analysis can be found in Chapter 7 as part the Tallinn Knitting Bus evaluation.

Further readings

Griffiths, Dawn: Head First Statistics. O'Reilly Media, Sebastopol (CA), 2009.

Rumsey, Deborah: Statistics for Dummies. 2nd edition, Wiley Publishing, Indianapolis, 2011.

Free collection of micro lectures on video: http://www.khanacademy.org/ (accessed 4th October 2012).

Free online tutorials on statistics and matrix algebra: http://stattrek.com/ (accessed 4th October 2012).

2.6.3 Qualitative data analysis

Analysing and interpreting qualitative data such as individual and focus group interviews, photos and video recordings, published articles and other documents require other expertise/know-how than dealing with numbers and frequencies of survey data. So this chapter is dedicated to point out important basics to treat qualitative data, which come along with your evaluation work especially when you want to analyse recorded interviews. In those cases your aim is to discover patterns and explain them. But you should also remember to collect and state so-called metadata such as your notes about the context of interviews (date, place, specifics of the interview situation, initial ideas for analysis etc.) and relevant information about the interviewees (name, biographical information etc.).

So before you can explain the results from (guided) interviews that were held with, for example, the teachers at schools where the Road Safety Label was implemented, the recording needs to be transferred into text (transcription) and has to be sorted (coding).

1. Transcription and excerpt of recorded data

Before you start analysing your qualitative data you need to bring them into a written format. This transfer can be done either very detailed word by word (called transcription) or by summarising the main content of the answers to the questions in your own words (excerpt). For a complete transcription we suggest to use a professional service, since transcribing an interview of for instance 1 hour would take many more hours to be transcribed correctly. With information about the topics and context of your interviews and the required level of transcription, the professional typist is well-prepared for transcribing the audio data.

A transcript is useful, if the issues addressed are very complex and a detailed analysis is necessary. An excerpt however consumes fewer resources in writing it based on the taped material. But you miss details and context because you have already summarised the interview content. Hence you no longer separate the analysis from the interpretation. This is also the case if you analyse the interviews on the basis of notes directly taken during the interview. An excerpt of recorded data is appropriate, when the interview was well-structured along some questions; the answers to the discussed issues were not convoluted and when you only want to draw general conclusions. To prevent errors in transcript as well as excerpt you should go back to the recordings from time to time and ensure that you have correctly summarised. Also, do not forget to make the data such as names and places in the transcripts anonymous but take care to have a list with all the items changed.

2. Coding of texts

Remember that for the analysis of qualitative data you want to identify patterns. Henceforth, its analysis – be it from a previously prepared excerpt or transcript – requires organising of the text according to topics and ideas addressed. This so-called thematic coding entails sorting text parts to categories (codes and sub codes) to combine all text parts of the same subtopic, idea or explanation which occur in the interviews. This can be done in two ways: You define categories in advance or you derive the categories from the material while analysing it. Both represent a result of analytic thinking about the texts and its interpretation. To reconstruct your ideas and guarantee a consistent coding process it is necessary to attach a note to each of the developed categories with its definition and further information for its application. You should always take notes for possible interpretation of text parts during the process of coding. This coding can be done manually using spread sheet software like Excel or tables in Word or using professional software of qualitative data such as ATLAS.ti, NVIVO or MAXQDA, which facilitate the organisation of textual data and prepare your further data analysis and interpretation. An example of a set of categories, a so-called code tree is visualised in the figure 2-12.

Please note that the software offers support but cannot free you from thinking yourself. Analysing qualitative data has to be conducted in a structured way following the same rules as the quantitative data analyse: It must be well-documented in notes/memos – each step in your analysis should be traceable and comprehensible. As already mentioned, the task of coding itself is an iterative process; you will have to revise the coding of one interview more than once to relate all relevant parts of all the interviews conducted to the categories of your final set of codes and sub codes. To ensure objectivity and validity it is strongly recommended that at least two people independently

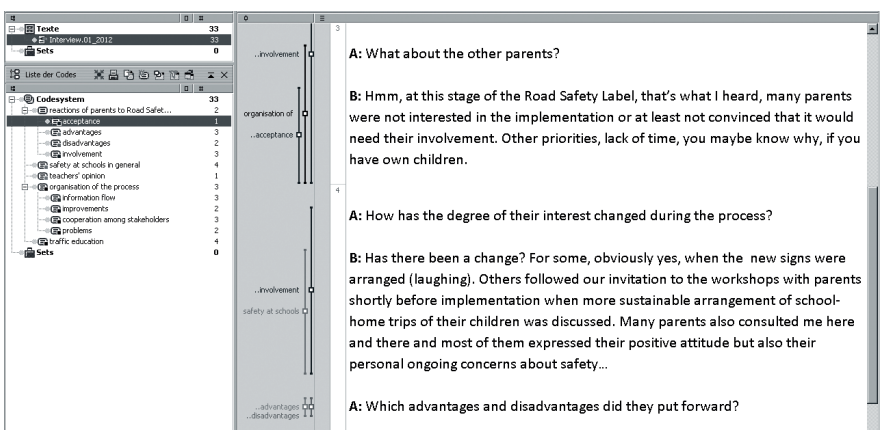


Figure 2–12: Visualisation of coding a transcribed interview.

from each other work with the texts and do the coding. If the coding varies strongly among the involved coders, they should discuss their different assessment to harmonise their kind of coding.

Box 13: Coding of guided interviews

A practical way for you can be to prepare a first rough draft of categories, a so-called code tree, which covers those codes/topics that you consider as important for your evaluation task. It should therefore be oriented towards your interview guideline. However after having coded the first transcribed interview texts, you reorganise the structure according to the content of answers. Insert sub-categories (sub-codes) according to the sub-topics of the answers given and the reasoning you recognise in the text material. It is necessary to handle the code tree flexible because an important new (sub-) category may arise when you do the coding of the second or third interview which may require a new (sub-)code. With this further knowledge gained during the process of coding, you should also check again if the interviews already coded include relevant information for this category. So this coding is part of analysing the material and reflects your thinking process.

The depth and accuracy of your coding is always dependent on the purpose and depth of your interviews. Guided interviews where you request expert knowledge about a specific process or occurrence require less depth of coding than narrative interviews trying to reconstruct individual behaviour and concepts of thinking. If you ask yourself what to code in your text, here you find some suggestions which help you to define categories:

- Strategies elaborated or applied to achieve a goal
- Meanings, interpretation and relevance of phenomenon and events (influence the actions of interviewees)
- Relationships or interactions between people or institutions
- Events which show what (and how) something happened
- Acts and behaviors representing what is said or done
- Activities undertaken within a specific setting (also with other people involved)

3. Comparison and generalisation

The next step is to look at the combined material in the assigned categories which were applied for all the interviews, summarise it and relate it to your evaluation questions again. You can compare your complete set of coding on different levels. Within

a category you see how a phenomenon is dealt with in different interviews. On the level of a single case (interview) you can find out if the reasoning of one interviewee is consistent throughout the interview. And finally you can compare the answers to one question or topic between cases (interviews). Write down your findings and interpretations as soon as possible and also use the notes/memos you have produced during the process of data analysis.

For the compilation the answers a table might be helpful to be integrated in the report with descriptions and main results. Illustrate your summary of each category with some references to the original text (citations). This eases understanding for the reader. The last step is the verification by the person who was interviewed to avoid misinterpretations of the statements. In the majority of cases those people are also interested in the results of your evaluation and therefore it is polite to send them your report in advance and ask them for comments. Make sure that you have made the document completely anonymous to protect the privacy of your data source.

Further readings

Bhattacherjee, Anol: Qualitative Analysis. In: Bhattacherjee, Anol: Social Science Research: Principles, Methods, and Practices. Open Access Textbooks, Book 3, Chapter 13, 2nd edition, 2012. (http://scholarcommons.usf.edu/cgi/viewcontent.cgi?article=1002&context=oa_textbooks)

Denzin, Norman K.; Lincoln, Yvonna S.: The SAGE Handbook of Qualitative Research. Sage Publications, London, 2005.

Flick, Uwe: The Sage Qualitative Research Kit – Designing Qualitative Research. Sage Publications, London, 2007.

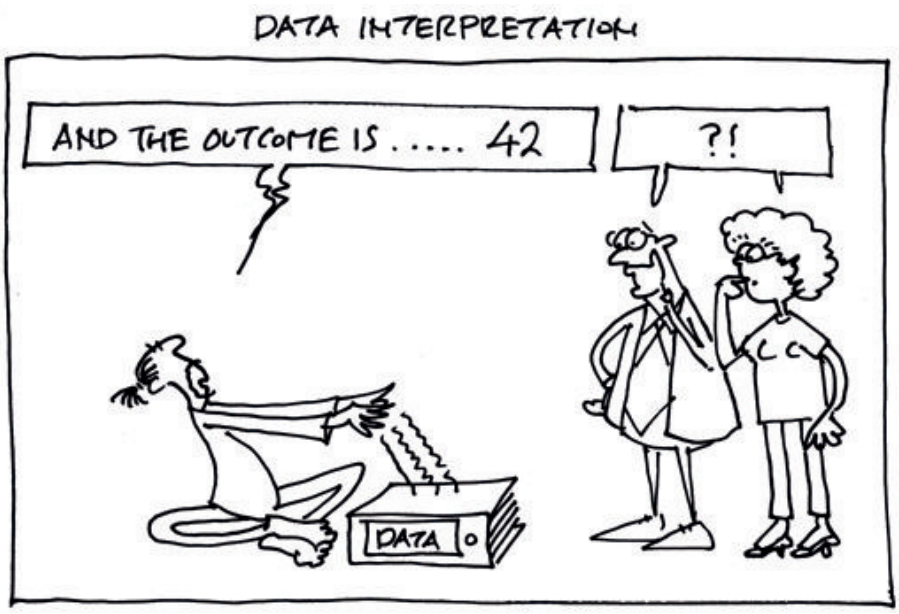
Gibbs, G.: The Sage Qualitative Research Kit – Analyzing Qualitative Data. Sage Publications, London, 2007.

Kvale, Steinar: Analyzing interviews. In: Kvale, Steinar: The Sage Qualitative Research Kit – Doing Interviews. Sage Publications, London, 2007, p. 101-119.

2.7 Discussion of results from data analysis

After all this data collection and analysis it is time to think about your data interpretation and what does the data mean for your measure and your city. It has become good practice to separate the chapter data presentation/analysis and the discussion of those results. In the data presentation chapter the facts are objectively documented in figures and texts explaining them (e.g. number of accidents in front of schools per year for the last 6 years). In the discussion chapter a more subjective assessment and appraisal of the results takes place including an answer to which extent the objectives were reached and how this can be attributed to the measure. The reason to always separate data presentation and interpretation is that a reader should be able to draw his own conclusions from your data material. That is why the data needs to be presented in a way that it is understandable what kind of data were collected and which results came out in the end for an outside reader. We highlighted that during the previous chapters.

In the interpretation chapter you draw conclusions from the comparative data for the measure as a whole and for your city in general. A well-elaborated and critical interpretation of your results avoids misunderstandings by inside and outside stakeholders and makes your data less vulnerable to reviewers/critics. So take the chance to frame a conclusion of your results that makes clear, what you have reached by the measure and what the results mean for your city. Make clear that your arguments are coherent and logical. This helps to estimate the value of this specific measure and this



kind of measure in general, as well as making it easier to communicate the value of the measure towards stakeholders and edit it for other dissemination purposes.

For drawing appropriate conclusions, it is important to relate back to the measure objectives. What was your initial high level and what were the measure specific objectives you wanted to measure by your indicators chosen (see also Chapter 2.2 and 2.3)?

The data for the indicators have been collected and analysed. Now the question needs to be answered, whether the measure had some impact. Is there a difference in the indicator between the after measurement and the business-as-usual scenario? In the ideal case the business-as-usual scenario is determined by a control site measurement as in the case of our Utrecht Road Safety Label, where the schools with activities were compared to schools without activities. In other cases the business-as-usual scenario needs to be derived with some assumptions from the before measurement as was the case for the Tallinn knitting bus measure. The kind of business-as-usual scenario needs to be taken into consideration when drafting the interpretation of your data.

It is recommended that you collect data for more than one indicator to be able to apply the method of triangulation (look for results with different methods) and to ensure quality/valid data basis. You need to weigh in different indicator results that you found out by different methods into a general result. Some findings might weigh stronger than others. For instance, numbers of tickets sold provided by the transport operator's ticket system are hard facts while interviews with a too small sample size (see Chapter 2.5.7.) of public transport customers might provide interesting insights, but could not be interpreted as setting the general trend.

What should you write, if the data indicates that your objective e.g. increased satisfaction with public transport service, have not been reached? Well, you simply need to state that and discuss in the interpretation section, what the reasons for that results could be. An example for this could be that a strike of bus drivers happened during the period in question and might have influenced the satisfaction with public transport negatively. In other cases, your data might show clearly that you reached the objectives. Consider carefully, whether the result was only due to your measure and share your thoughts, including doubts, with the reader of your report. It might be that the freight transport volume transported by light duty vehicles in the inner city of Utrecht – as stated in the measure objective of the Cargohopper – decreased significantly during the last years. But this might be also a result of the financial crisis and that less freight was transported in total.

Many hints to which other influences should be considered and discussed within the interpretation section can also come from your described cause and effect reflections and process evaluation results.

3 Process evaluation

In the last chapter we have shown how you can evaluate the verifiable effects of the measures implemented in your city. You now are able to demonstrate for example that with the Utrecht Road Safety Label the perception of safety among parents differs significantly between schools where the measure was implemented and the control site. But evaluation also has a second role in helping improve measures as, or after they are implemented. Many measures have to deal with a host of problems: management weakness, cultural issues and the failure to take into account the sometimes enormously complex systems that contribute to your city's mobility-related problems.

However, to have information on the planning and implementation process of your measures is crucial as it helps to make processes more efficient and avoids making the same mistakes again. Consequently, in the following chapter you can read about what the aim of process evaluation is and which approaches and methods are applicable to fulfil this task.

3.1 Aim and outcome

Let's be clear about something: implementing mobility-related measures is a time consuming and sometimes tedious task. But why is it so difficult? Part of the answer seems to lie with how measures are legislated and planned. Policymakers seldom seem to analyse the feasibility of implementing their ideas during decision making. As a result, the implementation process can sometimes be incremental – meaning that the measure takes shape slowly and adaptively in response to the emerging situations and early experiences. Even where planning includes a trial period, what gets finally adopted often varies from what was tried out in the pilot effort. As bizarre as this might sound: this is a natural and at times desirable process. It could easily be that without any adaptation some measures might simply be stopped. But what is the evaluation part in this?

Process evaluation focuses on the internal dynamics and actual operations of a measure in an attempt to understand its strengths and weaknesses. Consequently, process evaluation asks for example: What do stakeholders experience when they take part in a mobility-related measure such as the Utrecht Road Safety Label? What are the strengths and weaknesses of the day-to-day operations? How can the process of acquiring and keeping the label be improved? This approach derives its name from an emphasis on looking HOW an outcome is produced, rather than measuring its impact.

Thereby, it is less interested in the formal activities and anticipated outcomes, but it investigates informal patterns and unanticipated consequences in the full context of the measure implementation and development. To make this more clear: the Utrecht

Road Safety Label was implemented in different steps. The schools were approached, contracts were signed, baseline data was acquired, curriculums were changed and the school surroundings were redesigned (for more information see Chapter 7.1). Thereby, when these outcomes were achieved and if they were achieved according to the prior timetable is part of monitoring and thus it is not an issue in this book (see Chapter 1.3). The process evaluation searches for explanations on the delays, changes, failures but also success of the measure. Therefore, if process evaluation is conducted during the measures development phase as well as later it can provide useful information for improvement. In this context it is also crucial for follow-up projects or can help if justification at political or management level is required. It also helps to avoid making the same mistakes again.

Finally, process evaluation usually includes perceptions of people close to the measure about how things are going or went. A multitude of perspectives should be sought from people inside and outside the measure. For example, the perspective on changes induced through the Utrecht Road Safety Label can differ between parents, teacher, students and administrators. These different viewpoints can provide unique insights into the measure's process as experienced and understood by different people.

These examples are elucidated that the main objective of the process evaluation is to get insight in the 'stories behind the figures' and to learn from them. This objective applies to the practitioners who are intending to increase the efficiency of the measure processes, and also for other cities which might choose to replicate a measure based on your good example. Together with the results of the impact evaluation all stakeholders involved have the chance to effectively learn from the experiences made.

3.2 Barriers and drivers

Process evaluation searches for barriers and drivers of a measure. Under real conditions, people and unforeseen circumstances shape measures and modify initial plans in ways that are seldom trivial. Process barriers are the events or the overlapping conditions that get in the way of the process to reach your measure's objectives. But also welcoming effects are identifiable. Process drivers are events or overlapping conditions that stimulate the process to obtain measure objectives. If for example a high rank politician such as the Minister of Environment decides to act as the Ambassador for the Cargohopper this will, surely help to promote the project. If the Cargohopper for example cannot deliver goods at night times because there is no staff present at this time of the day in the shops to receive the goods, this is a clear barrier to an efficient operation of the Cargohopper.

For the identification and also for the derivation of activities directed at these barriers and drivers it is helpful to cluster them into different categories (see Table 3–1). Thereby, the same category of possible influencing factors can turn out to have a positive or a negative impact on the measure process.

The categories identified in Table 3–1 are not an exclusive list, there are also other categories. Spatial conditions could change during the measure process in a way that they offer your measure new opportunities to expand a service (i.e. for Park & Ride) or limit the area necessary for the implementation (i.e. Park & Ride lot has been re-designated to a residential area). Also your project could suddenly become part of a city-wide programme for sustainable transportation and thus benefit from this new 'positional' driver.

Remember, as different as cities are, as different and complicated can drivers or barriers be, thus it is always helpful to provide some sort of clustering, but the barriers and drivers presented in Table 3–1 are only abstracted examples. In any way, you will have to describe what really happened in the city. Key questions to be answered can be:

- What exactly happened?
- How did it occur?
- Which impact did the barrier or driver have on the process of the measure?

This description should be as brief as possible but at the same time give enough information on what has happened so that even 'the outside reader' (the one without any knowledge of the measure) will understand what was going on. This is an ambitious task! But the following description of a barrier which occurred during the implementation process of the Utrecht Road Safety Label demonstrates that this is indeed attainable:

Field	Examples of barriers	Examples of drivers
Political/ strategic	Opposition of key actors based on political and/or strategic motives, lack of sustainable development agenda or vision, impacts of a local election, conflict between key (policy) stakeholders due to diverging believes in directions of solution	Commitment of key actors based on political and/or strategic motives, presence of sustainable development agenda or vision, positive impacts of a local election, coalition between key (policy) stakeholders due to converging (shared) believes in directions of solution
Institutional	Impeding administrative structures, procedures and routines, impeding laws, rules, regulations and their application, hierarchical structure of organisations and programs	Facilitating administrative structures, procedures and routines, facilitating laws, rules, regulations and their application, facilitating structure of organisations and programs
Cultural	Impeding cultural circumstances and life style patterns	Facilitating cultural circumstances and life style patterns
Involvement, communication	Insufficient involvement or awareness of key (policy) stakeholders, insufficient consultation, involvement or awareness of citizens or users	Constructive and open involvement of key (policy) stakeholders, constructive and open consultation and involvement of citizens or users
Planning	Insufficient technical planning and analysis to determine requirements of measure implementation, insufficient economic planning and market analysis to determine requirements for measure implementation, lack of user needs analysis: limited understanding of user requirements	Accurate technical planning and analysis to determine requirements of measure implementation, accurate economic planning and market analysis to determine requirements for measure implementation, thorough user needs analysis and good understanding of user requirements
Organisational	Failed or insufficient partnership arrangements, lack of leadership, lack of individual motivation or know-how of key measure persons	Constructive partnership arrangements, strong and clear leadership, highly motivated key measure persons, key measure persons as 'local champions'
Financial	Too much dependency on public funds and subsidies, unwillingness of the business community to contribute financially	Availability of public funds and subsidies, willingness of the business community to contribute financially
Technological	Additional technological requirements, technology not available yet, technological problems	New potentials offered by technology, new technology available

Table 3–1: Examples for barriers and drivers identified through process evaluation.

Organisational barrier: change in school management – The time period between the school starting the process to obtain the label and the moment of really getting the label is in some cases around 3 years. In the same time it was found that in about 35% of the schools a change of management takes place within such a period. A change of management means in some cases that the process and advantages of safety label have to be explained again. More generally, this means that often the process with the respective school is temporarily delayed, or even stopped.

This is a good description because it explains which barrier category has occurred (organisational); it describes what happened at the schools (a change of management in the schools took place which had agreed to participate in the label) and what negative impact it had on the implementation process (process with the respective school is temporarily delayed, or even stopped).

Let's refer to an example of the Tallinn Knitting Bus, to show you an example for good driver description:

Positive feedback from stakeholders – The positive feedback from public transport users, city officials and other CIVITAS cities on the knitting bus has increased the motivation of the initiators to plan new activities to promote sustainable transport. As a result, it was decided to defragment the different marketing activities in the city by providing a single strategy framework for further campaigns.

This example shows which driver category has occurred (feedback from stakeholders), it describes what happened in the city (positive feedback has been received) and what positive impact it had on the implementation process (increase of motivation and planning of additional activities). This driver is also a good example because it demons-

trates the different 'layers' in which you can look at barriers and drivers. If you look more closely at this driver you might want to know why the people of Tallinn were so enthusiastic about the Knitting Bus. Was it because of the long tradition of crafting in Estonia? Or was it because of the generally low quality of Tallinn public transport services and with all this knitting, the bus looked new? In the end, it is up to you to decide how deep you want to dip into those barriers and drivers. In all cases, make sure that the barriers and drivers are stated in a way that they can be acted upon.

3.3 Activities related to barriers and drivers

What happens if you realised that something in your measure is not progressing as expected? Say, for example, the Road Safety Label in Utrecht has only been awarded to a very low proportion of schools among those that initially signed the contracts to join the scheme? As you now know about process evaluation, you would try ways and means to find out what the exact barrier was (see Chapter 3.4 for more on that). Let's assume for the moment that the problem is a significant delay in the restructuring of the schools' surroundings and your evaluation has concluded that this is in terms the result of too little staff involved. Then, if you still have time left before the end of the project, you would do something about this barrier – you would adapt the measure. In our example, you could, for instance, force your subcontractor who is responsible of this restructuring to employ more people to speed up the process.

As already stated above, the aim of process evaluation is to describe what really happened and why. Consequently, you need to address these activities to obtain a comprehensive picture of a measures' process. Additionally, the activities you undertook to make use of the drivers or to overcome process barriers identified during the implementation process are probably the most interesting part for the reader of your report. It shows how problems have been solved and how positive factors have been utilised for carrying out the measure implementation process more efficiently.

However, there are two things to keep in mind. First, you will need additional time for these new activities to be carried out. So if you only evaluate your measure at the very end, there will be no time left. This is one of the strongest arguments for measure-accompanied evaluation (see Chapter 1.3.1). Second, your process evaluation should not stop after the new actions have been identified. This means for our example, after you have obliged your subcontractor to employ more people you realise that the proportion of schools with the label – in comparison to those that initially signed the contracts to join the scheme – still do not increase significantly. You should ask again why this has happened and maybe then you will find out that the problem lies in the long approval procedures in the administration. This problem was obscured before.

3.4 Methods for process evaluation

As you can guess from all of the above, process evaluation is developmental, descriptive, continuous, flexible and inductive. It therefore needs a methodology that enables you to collect information from and analyse the processes during all measure phases preferably in a consistent manner. To give the full picture of what has influenced the process it is also necessary to gather the information on a regular base and during all phases of the measure. The frequency of your data collection relates to these phases but does not solely rely on it. It certainly depends on the total lifetime of your measure and unfortunately, you have to judge in advance what a good timing will be.

Not only the frequency but also the depth and the type of data gathering for process evaluation depends to a high degree on the availability of personnel and financial resources. However, in any case the goal should be to analyse the whole process for a measure and document all relevant activities linked to the measure process from preparation to implementation and operation.

There are several methods to observe and assess these processes – mainly differing in how detailed they are observing them. Two of these (standardised forms and learning history workshops) will be discussed in more detail in the following chapters.

3.4.1 Standardised forms

Using standardised forms to collect the data relevant for process evaluation has three obvious advantages: Everyone involved in the measure process can do it by himself/herself, only one or some few persons are necessary to complete the forms, and because they are standardised it is easy to compare the results with other measure processes in your own city. Thus, this is a method which has been proven as practicable. The form should contain a part with general information such as the name of your measure, the time period in which data has been collected, targeted groups and partners involved in the measure implementation. Additionally, do not forget to indicate who has compiled the form and who is the main contact person dealing with the measure. This is important for you as the evaluator to have someone to address questions if that might arise later. Figure 3–1 shows a rough structure of such a standardised form and Chapter 7.3 shows the standardised form that was used in the Tallinn Knitting Bus process evaluation.

In any case, the core part of the form should be the documentation of the process barriers and drivers as well as of the activities undertaken to deal with the identified problems. Thereby, showing tables in the form listing categories of abstracted fields of drivers and barriers (as presented in Chapter 3.2) could be a help especially to

General information	• General (administrative) information such as objectives/ target groups/partners, stakeholders/measure phase
Barriers	• Brief and clear description of the most important barriers • What happened? How did it occur? Which negative impact did it have on the process of the measure?
Drivers	• Brief and clear description of the most important drivers • What happened? How did it occur? Which positive impact did it have on the process of the measure?
Activities	• Brief and clear description of the most important activities you undertook to make use of the drivers or to overcome process barriers
Risks	• Estimation of risk to reaching the objectives
Any other comment	• Brief and clear description of any other relevant information to explain the measure process: extraordinary conditions, pictures, methods, etc.

Figure 3–1: Scheme of a standardised form at a glance.

those who are not familiar with the concept of process evaluation. Additionally, you could also ask for the identification of the most (second most, third most ...) important barriers and drivers.

Consequently, if you do not collect the information at the end of the measure lifetime it is also advisable to ask the person filling in the form to what extent he or she thinks that the barriers identified could risk the measure-specific objectives. The estimation of the risks are an important step in the process evaluation: It helps your project management to estimate if the measure implementation is endangered and if further steps have to be undertaken to correct impeding conditions or to 'save' the project. As such, the standardised forms are strongly linked to the monitoring.

The assessment of the development of a measure with standardised forms is a simple and effective method and it can be achieved by mobilising minimal resources. Nevertheless, it is limited in its scope of information delivered and thus limited in its capacity to communicate and transfer experiences to others. Especially if the information collected on the form covers a time period of e.g. one year, drivers, barriers and activities are not recorded by the date when they occurred. Accordingly it is difficult to comprehend the entire story of the events and their consequences.

The standardised form has, however, one significant advantage: the barriers and drivers are already written down and can be put directly into your final evaluation report. Of course, you should check for duplicates but other than that, you are done.

3.4.2 Learning Histories

The concept of a Learning History is based on the idea to learn from 'story telling' involving different perspectives and stakeholders. In fact, a Learning History is a process that results in a jointly told tale in multiple narratives, with illustrations and reflections on all events that occurred. Story telling or listening to stories is an interactive and enjoyable activity. If a story is told well and in an exciting way it makes you think about your own situation. And moreover, it stays in your mind for a longer time. These stories are structured along a time line visualising the strategies, noticeable results, what happened why and in which way. Thus, it gives insight into the organisational dynamics and the internal logics in dealing with change.

The Learning History workshop is the core of the learning history approach, which was developed by researchers at the American research institute MIT in the late 1990s. In this workshop stakeholders get together to discuss how and why the measure process has evolved. The participants jointly collect the process information on barriers, drivers and learning experiences and bring it in coherent clusters of themes. The Learn-ing History requires multiple narratives and thus brings together a more comprehensive picture of what actually happened.

A Learning History should be carried out in the same frequency as other methods to assess the measure process – several times and at different stages of the project. Since a Learning History delivers detailed information on the reasons why the project has been suffering from problems, delays or even of cancellation, it is advisable to apply this tool

whenever these severe problems have occurred. Thus, it offers the chance for you to avoid the same or similar problems in future. The approach of a Learning History is depicted in the Figure 3–2.

Good preparation is the key to a successful Learning History Workshop. As with workshops in general, this can be a very time consuming task and you need to have a good understanding of what will happen during the workshop. Consequently, if you are planning this for the first time, you might want to discuss your timetables et cetera with a college. The question of who you will invite has to be answered: In general, there should be a variety of participants – people who are formally or informally involved in the measure implementation processes – all those of whom you think they could have had an influence of any kind on the process of your measure. These could be politicians, representatives of private companies or respective associations. But keep in mind that in a Learning History workshop every participant is actively involved – everyone needs to

BEFORE	Preparation	• Three weeks before the workshop: invitation, identifying moderator, setting up a draft time-line, booking room and catering: approx. two working days
WORKSHOP	Step 1 Setting up a time-line	• Presentation of draft chronological event time-line, then discussion and completion with input from the participants approx. 30 – 60 min
	Step 2 Introduction	• Explanation of the workshop goals and procedures; what are barriers, drivers, activities in process evaluation, which 'rules of the game' approx. 20 min
	Step 3 Inventory of barriers	• Plenary inventory of barriers encountered in reaching the objectives; discussion and clustering by moderator approx. 45 min
	Step 4 Inventory of drivers	• Plenary inventory of drivers encountered in reaching the objectives; discussion and clustering by moderator approx. 45 min
	Step 5 Inventory of actions	• Plenary inventory of actions taken to overcome the barriers and to make use of the drivers to reach the measure objectives; clustering by the moderator. approx. 45 min
	Step 6 Discussion	• Reflective discussion on learning experiences approx. 45 min
AFTER	Documentation	• Summarizing and compiling the results of the workshop in standardized forms – ideally directly after the workshop approx. 3 – 4 hours
	Feedback, revision, dissemination	• Sending draft form to participants and incorporation of their feedback two weeks after the workshop • Distribution of final form 1 hour

Figure 3–2: Scheme of a Learning History process.

have sufficient opportunities to speak out! Thus the number of participants should not be too high (at maximum 10). Additionally, when you choose your location ensure that you will have enough space to move around between your equipment.

Although analysing processes within a Learning History are performed collectively, the success of the workshop depends highly on the skills of the moderator. It is his/her task to steer the discussions as well as to cluster and rank the barriers, drivers and actions.

Box 14: Qualifications of a Learning History workshop moderator

She/he should be:
- Experienced with moderating techniques
- Able to create an atmosphere of trust, in which participants feel respected and free to speak openly
- Well informed about the measure which is subject of the workshop: objectives, stakeholders, and possible delicate aspects and problems
- Well informed about the roles and responsibilities of the participants within the measure.

It might also be contributing if there are two moderators or if you involve an external moderator who brings in ideas from outside. In each case, the moderator should keep in mind:
- Time slots and aimed results
- Balanced contribution from participants (not only of those with the loudest voice)
- Doing a first rough analysis (clustering) of results during the workshop to steer the discussion

The workshop itself starts with setting up a time-line of main events (Step 1, Figure 3–2) that occurred during the period of the measure process which will be observed. Only events which had an influence on the measure process have to be taken under consideration and should be visualised in some way. Accordingly there might be gaps in the time-line where nothing happened that affected the measure process.

Depending on the people invited, it might be useful to sent a draft timeline to the participants prior to the event. This could then be discussed and finally completed with input from the participants. This will result in a shared perception of what actually happened in the observed period. After defining the time line, the moderator tells the participants what will happen during the workshop (Step 2, Figure 3–2).

The participants are then asked to write down on sticky notes or moderation card barriers they experienced during the period of assessment (Step 3, Figure 3–2). The mo-

derator collects all the notes/cards and puts them on the paper at the wall or flipchart. This will result in a collection of anonymous sticky notes/cards, with all kinds of process barriers faced by different project partners. After this, the moderator roughly clusters the sticky notes/cards with a common logic and starts facilitating a discussion with the central questions why these barriers have been perceived as barriers and what have been the impacts of them on the measure process. This discussion will probably result in a range of different beliefs why these (clusters of) process barriers were barriers and what the impact was on the process and the objectives. Together with the participants the moderator will rank the most important barriers (see Figure 3–3). There are different techniques that you can use to 'rate' the clusters, at this point it might be good to refer to moderation guidelines.

These steps are then repeated for the barriers and for the activities, meaning that you do this same process three times (see Step 4 and 5 in Figure 3–2).

The last step is a reflective discussion on learning experiences; looking back at the time-line of events that occurred, the barriers and drivers that were faced and actions that were taken (Step 6, Figure 3–2). For the facilitation of the discussion it is very

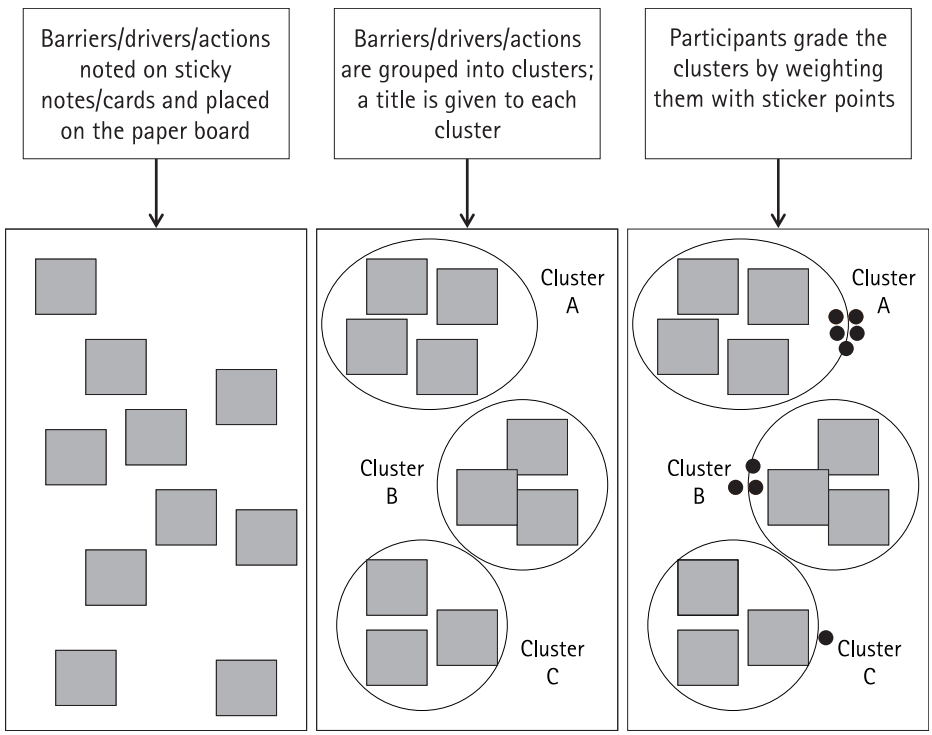


Figure 3–3: Example on how to work with sticky notes and cards.

helpful for the moderator(s) to refer to the corresponding papers with the various clusters. Central questions to structure the discussion are:

- Which of the actions can be regarded as a success and which as a failure and why?
- What have we learned? What are the do's and don'ts in terms of the process and actions?

Based on the outcomes of this Learning History workshop a report needs to be assembled for documentation and dissemination. This report will then be distributed to all participants of the workshop to give them a chance to check if their opinions and experiences are being represented in a 'true' way and are thus validating the report. Then it can be made available to politicians, civil servants, practitioners and other relevant stakeholders in your city or even beyond.

A Learning History workshop is not magic if the described steps are followed. The major advantage is that a thorough analysis of the measure processes can be achieved. It reflects the stakeholders' perceptions of facts that have occurred during the lifetime or a period of a project or measure. The method is also well suitable to be applied for learning from a finalized measure. It focuses on why things have happened as they happened, and what can be learned from this for future measures or projects. But you should also use it to learn during the implementation of a measure or project. It focuses on how to intervene during the measure or project to reach its objectives.

The CIVITAS cities have benefitted very much from this type of process evaluation. They reported that this method highly contributes to increase the motivation of the measure stakeholders – simply by bringing them together and facilitating struc-tured discussions to jointly find ways to avoid barriers and to more efficiently reach the measures objectives. But the cities also learned that participants of a Learning History are sceptical when they are asked to join the activity. The name of the method can be misleading. It sounds rather playful and non-technical. Especially in transportation and administration surroundings people are more used to hard and technical sounding working methods. Thus, it could be a good idea to communicate by referring to a more technical term such as 'process optimising event'.

Also, the data gathering by using sticky notes seems to be unpopular in some cultural backgrounds or working environments. Some cities who have applied this method reported that this way of data gathering and categorising has a somewhat playful character which makes it appear 'less serious'. However, solutions of this problem are easy: Sometimes it already helps to enter the information of the participants in a computer and cluster it digitally together with the stakeholders. But the simplest way is to convince the participants of the method and break the ice by just starting the activity.

Even if there are some disadvantages when conducting a Learning History – intense preparation and documentation, will provide evidence to sceptical participants

that this method is an excellent and effective tool to elaborate detailed findings on the processes of measures. These findings thereby are produced not only by one or two people involved in the measure implementation – it reflects the collective experience of a group of important stakeholders while having a positive effect on the spirit of your team.

3.4.3 Other methods

Apart from the standardised forms and the Learning History Workshop, there are other good methods available for data gathering in your process evaluation such as the focus group or interviews (see Table 3–2). You might even choose to employ the same methods that you use in your impact evaluation (see Chapter 2.5). Often, there are

	Standardised form	Learning History	Focus group	Interview
How?	Data gathering by completing a standardised form	Systematic in-depth gathering of information with all involved participants following the Learning History method	Moderated group discussion of 5-10 selected participants for a specific topic/ focus	Stakeholder interview steered with guidelines by phone or face-to-face
Data gathering easy and efficient	+	+/-	-	+
Involvement of stakeholders sufficient	-	+	+/-	+/-
Comprehensive picture obtainable	+/-	Frequently influenced by moderator	Frequently influenced by moderator	Frequently influenced by interviewer
Possibility to learn from the 'stories' told	+	++	+/-	+/-
Reasonable resources consumption	+	+/-	-	-

Legend: ++ very good; + good; +/ – sufficient; – poor

Table 3–2: Comparison of methods for data gathering in the process evaluation.

only minor differences between these methods, but they can be those that make one method better than another for your situation. Whichever method you chose to employ, it is essential that it clearly addresses barriers, drivers and actions taken to indicate the lessons learned. Moreover, to ensure that this information does not get lost, it should be recorded and the findings should be written down so it will be available for your final report.

All these alternative methods have one aspect in common: The quality of the outcome strongly depends on the effort you put into a preparation of the data gathering activity. In any case the method chosen should, in addition to the above, try to fulfil these important criteria:

1. It should address as many people involved in the measure process as possible – the outcome should reflect an objective picture.
2. The person in charge of moderating or interviewing should be able to avoid influencing the answers given by others.
3. The effort for preparing, conducting and reporting of the activity should be reasonable. Otherwise the risk of failure increases.

Often the data quality in these methods relies heavily on the persons involved in gathering them. Consequently, whichever method you choose should also depend on the resources and knowledge you can build on.

Process evaluation delivers valuable information on the internal dynamics and actual operation of a measure. As such, it can offer reasons why a measure became a success or a failure. It is an assessment tool to boost learning from the experiences made and gives indications on how to increase efficiency. Unfortunately, process evaluation is frequently seen as an additional task without any extra outcome for those who are conducting it. Due to its nature it is neither mere monitoring of projects nor a judgment of quality of the work people have accomplished within the course of a project but a natural linkage to the impact evaluation. Its ultimate aim is to get insight in the 'stories behind the figures' and to learn from them.

Further readings

Kleiner, A.; Roth, G.: Learning Histories – A new tool for turning educational experience into action. 1997. (http://ccs.mit.edu/lh/21CWP002.html)

Learning Histories – participative change by storytelling. www.learninghistories.net.

4 Information reporting and utilisation

Now, if you have gone through the entire evaluation procedure, and your measure or group of measures shows to have been successful in achieving the objectives, then what can you do with these positive results?

Give a moment's thought on disseminating your results. And then there are two options: if it is applicable, the measure could be scaled up in your own city to have an even larger impact or the success could be repeated in another city. The following three chapters will show you how to present results, how to approach up-scaling and how to assess transferability of a measure.

4.1 Result presentation

1. What do we mean by results presentation?

There are different aspects of presenting your results. Usually, there is some kind of reporting requirement to fulfil, either internally in your organisation or to a (co-)funding organisation as the European Commission. But especially in urban mobility contexts results presentation also means the dissemination of information to the public and to other interested organisations. What we are not talking about here is, if a public campaign related to some transportation issue (e.g. informing people about a new ticketing system with flyers) is the actual measure. It is then part of a measure package and subject to evaluation (how differently would people have reacted to the new ticketing system if they had not been informed about it through the flyer?).

The results presentation or dissemination of results we are talking about here is a follow-up action of the evaluation, i.e. it is what you do when you have your results at hand (you might even be doing it during the ongoing measure implementation process).

2. Why is dissemination important?

The ultimate goal of every evaluation procedure is to provide useful information. The purpose of reporting or any other way of presenting results is the utilisation of information (by stakeholders, decision-makers etc.). In order for the information to be used it needs to be well prepared and – disseminated, i.e. distributed. Due consideration should be given to the dissemination and use of information right from the start.

3. Dissemination can have several objectives, e.g.

- Making project results available to interested people and key decision makers
- Enabling technology transfer
- Stimulating community participation at all levels

How do you disseminate and what should you take into account?

You should put up a dissemination plan for your measure right at the start, clarifying the following:

A. What do you want to tell?

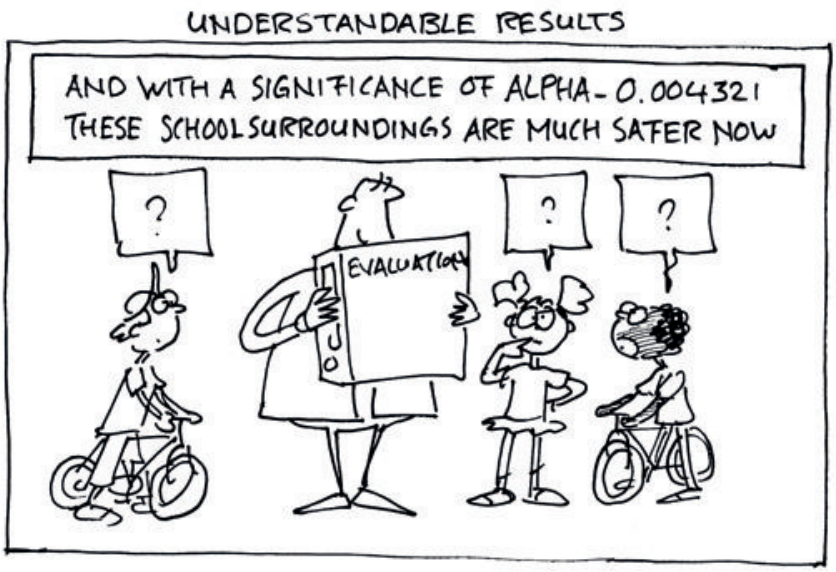
That your measure implementation is going well, that you have encountered some problems and how you have overcome them, what lessons you have learned etc. Depending on when you want to disseminate, it can be process information from the ongoing implementation or results from a measure that is up and running and has been evaluated, and you are now disseminating the evaluation findings.

B. What's your aim?

Is it just because you need to fulfil a requirement or are you aiming at providing valuable data for decision making? Do you want to contribute to the organisational learning and knowledge-sharing? Or do you report for accountability and to show your project's compliance with the plan? Do you want to inform the public of any changes (e.g. in a system) that have an influence on their everyday lives? Do you seek the acceptance or passive support from somebody for what you are doing? Or do you simply want to receive some reward for all the hard work you have been doing lately?

C. Who is the target group and what are their informational needs?

Are you talking to the project management, project partners, stakeholders, donors or the general public? If you want them to use your information, what do you think they



would need to know (not want to hear)? Your dissemination material should be concise and easy to understand; it should clearly address the target group and not contain any unnecessary or redundant information.

D. How do you want to tell?

How do you want to get your information across? This depends largely on your resources and what is suitable for the audience. Here are some suggestions:

Printmedia	Multimedia	Face-to-Face
Newsletters	Website	Meetings
Flyers	CD/DVD	Seminars
Project fact sheets	Web-based forums	Workshops
Press releases	Teleconferences	Panel discussions
...	Webinars	...

Table 4–1: Suggestions for dissemination channels.

Think of the pros and cons of each option and bear in mind the aim of the dissemination, i.e. the use of the information. Whichever mode you choose, it is important how you present the data (Chapter 4.1.1).

4.1.1 Data presentation

Besides depictions in table form, often graphics illustrate our results to the reader. People not too involved in the evaluation especially have a hard time to comprehend all the information transported through charts and statistic indicators. Graphics serve the purpose of conveying this information. There is a variety of possibilities to present and visualise data. We will pick out some examples here.

If you are looking at a classified variable (see Chapter 2.6.2 if you do not know what this is) like in the question: "How often do you use the city bus services?" a histogram is most appropriate.

The frequences are depicted by vertical bars. The individual rectangles are established above the classes of characteristic values listed on the x-axis.

To still better illustrate a frequency distribution that is depicted through a histogram we often add the so called polygon. This is done by connecting the upper centres of the classes (i.e. the middle of the rectangle's upper side) through straight lines. Pie charts are well suited to illustrate frequencies on a nominal scale.

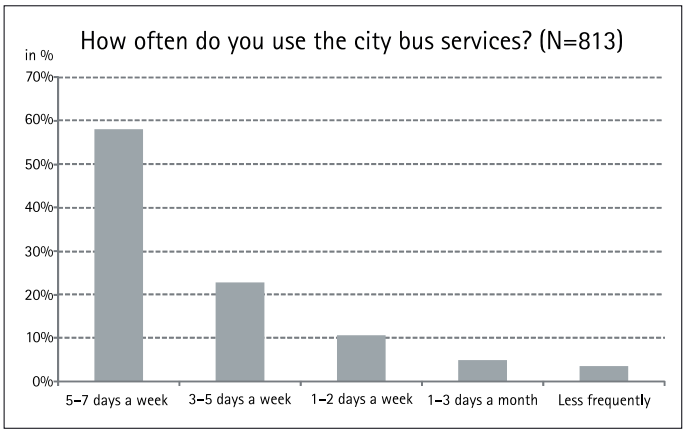


Figure 4–1: Example for results presentation with a histogram.

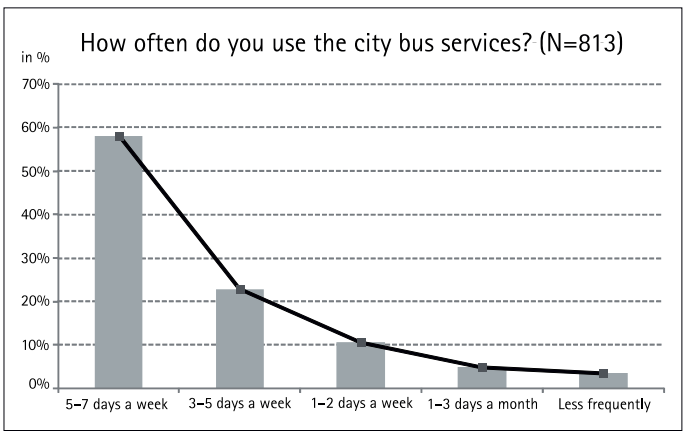


Figure 4–2: Example for results presentation with histogram and polygon.

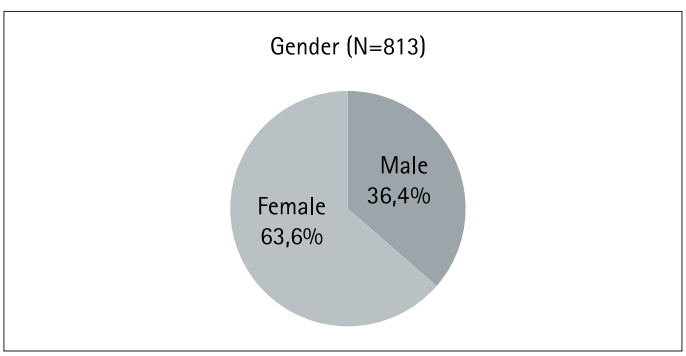


Figure 4–3: Example for results presentation with pie chart.

Time series can be depicted descriptively using line charts. For this purpose we use a system of coordinates: On the horizontal axis we put the time units while the corresponding characteristic values are on the vertical axis.

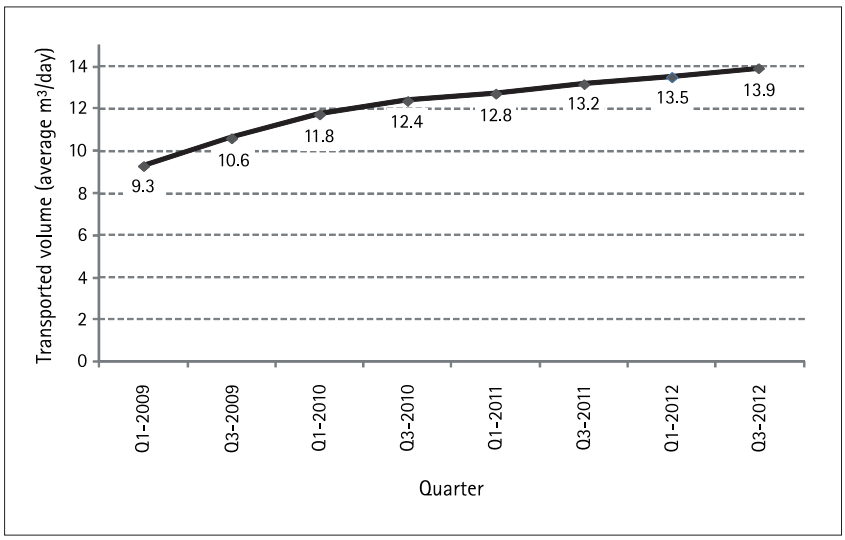


Figure 4–4: Example for results presentation with line chart.

Another interesting schematic illustration of a frequency distribution is the so called box plot (Figure 4-5). It consists of a box that is established between 25% and 75% of the values. So 50% of all the observations fall within this area. The lines projecting up and down from the box illustrate how far the other 50% of the values scatter. The following box plot shows the ratings of the Tallinn citizens regarding cleanliness of seats. (The number 5 states a high level of content.)

Box plots are well suited to compare different distributions. A single box plot needs less space than a histogram. That makes it possible to arrange a couple of (horizontal) box plots on top of each other or to arrange them side by side (when dealing with vertical box plots like the one depicted above).

When using a scatter plot (Figure 4-6), we can illustrate two statistic variables at the same time. This mode of depiction helps to detect a possible statistic correlation between two variables as well as this correlation's intensity. The values of one variable are stated on the x-axis; the values of the other variable are found on the y-axis. Every statistical unit (i.e. every person surveyed) then is a point within our system of coordinates. Usually, the variable on the x-axis is the one we expect to influence the other. Consequently, the variable on the y-axis is the one we believe to be influenced. The following example illustrates body height and body weight in a scatter plot.

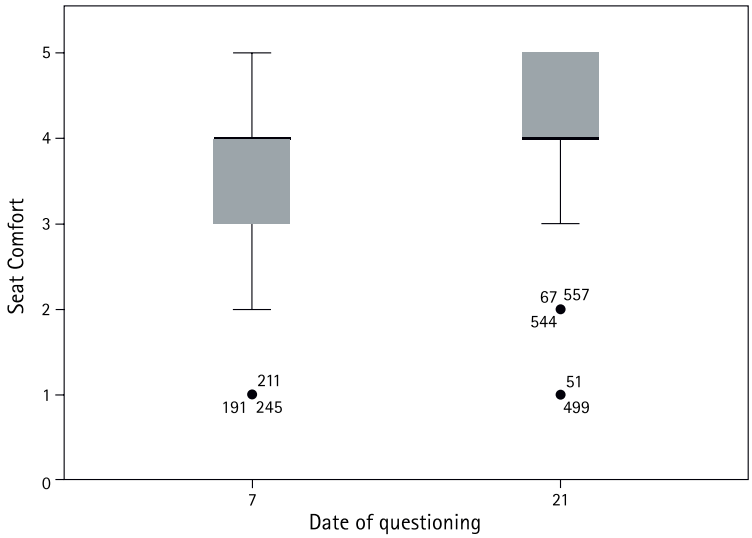


Figure 4–5: Example for results presentation with box plots.

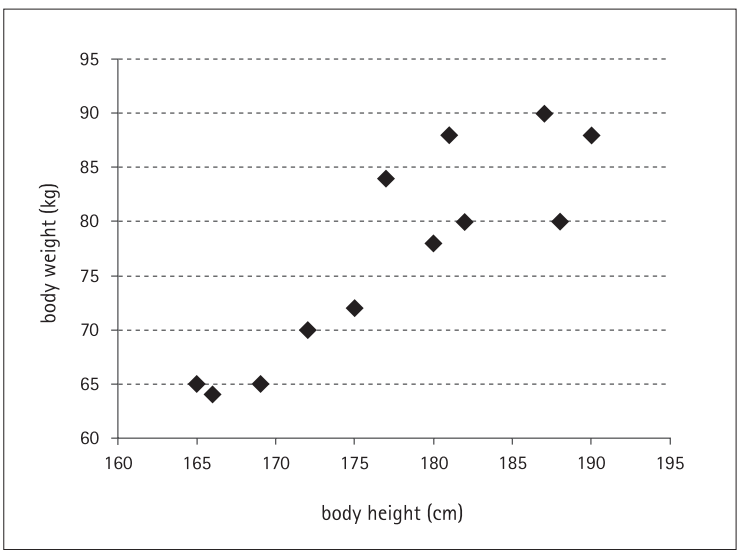
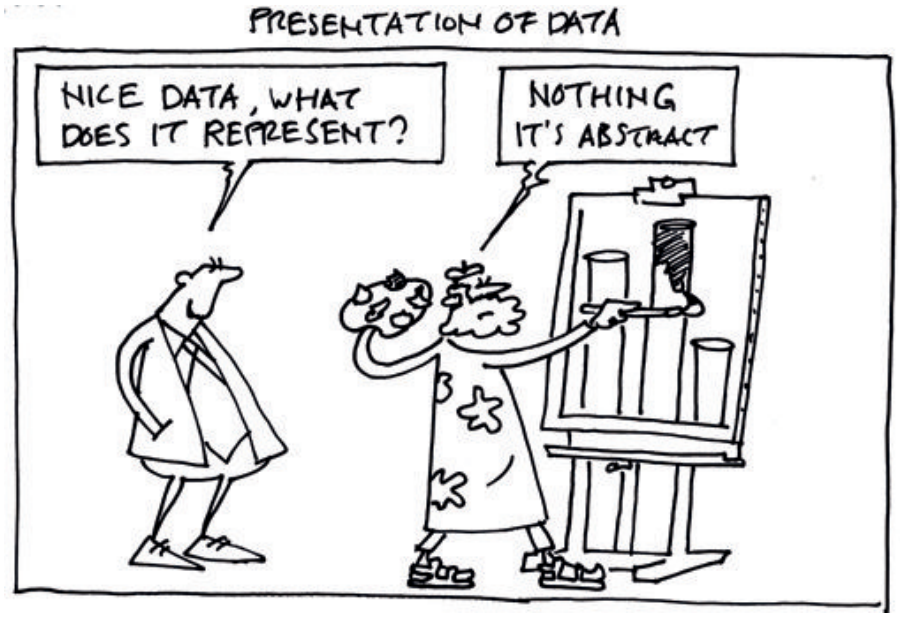


Figure 4–6: Example for results presentation with scatter plot.

There are many kinds of diagrams and ways to present your data. All can be created easily with standard software such as Microsoft Excel. You should be aware now that you need to choose carefully the way of presenting the data.



Further readings

Maxwell, Joseph A.: Research Proposals: Presenting and Justifying a Qualitative Study. In: Maxwell, Joseph A.: Qualitative Research Design – An Interactive Approach. 2nd edition, Sage Publications, London, 2005, p. 117-137.

4.1.2 How not to present your data

"Do not trust a statistic which you have not falsified yourself." This quote is commonly used in statistics books to underline the possibilities of manipulation or conscious as well as unconscious misinterpretations. Unfortunately, there are a number of such possibilities.

Statistics and illustrations suggest objectivity and precision. But especially illustrations can persuade the reader of a statement which could not have been validated through a thorough investigation (using for example statistical tests). A simple reductionof the y-axis can strongly emphasise the depiction of effects.

The perceived seat comfort of the bus service in Tallinn has increased from 3.76 to 4.02 after measure implementation. The illustration's y-axis on the left side of Figure 4-7 starts at 3.7. This way the diagram suggests a strong increase in the perceived seat comfort.

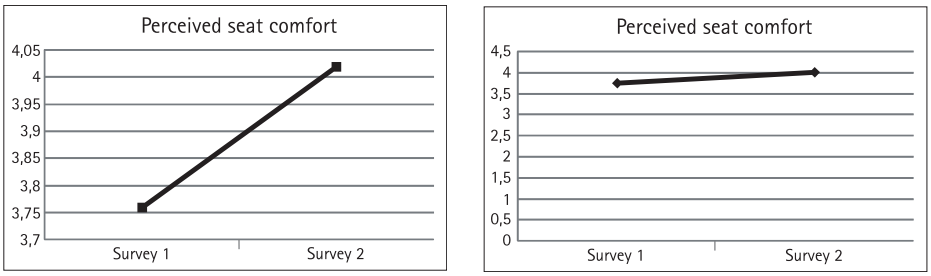


Figure 4–7: Illustration of a y-axis shifting.

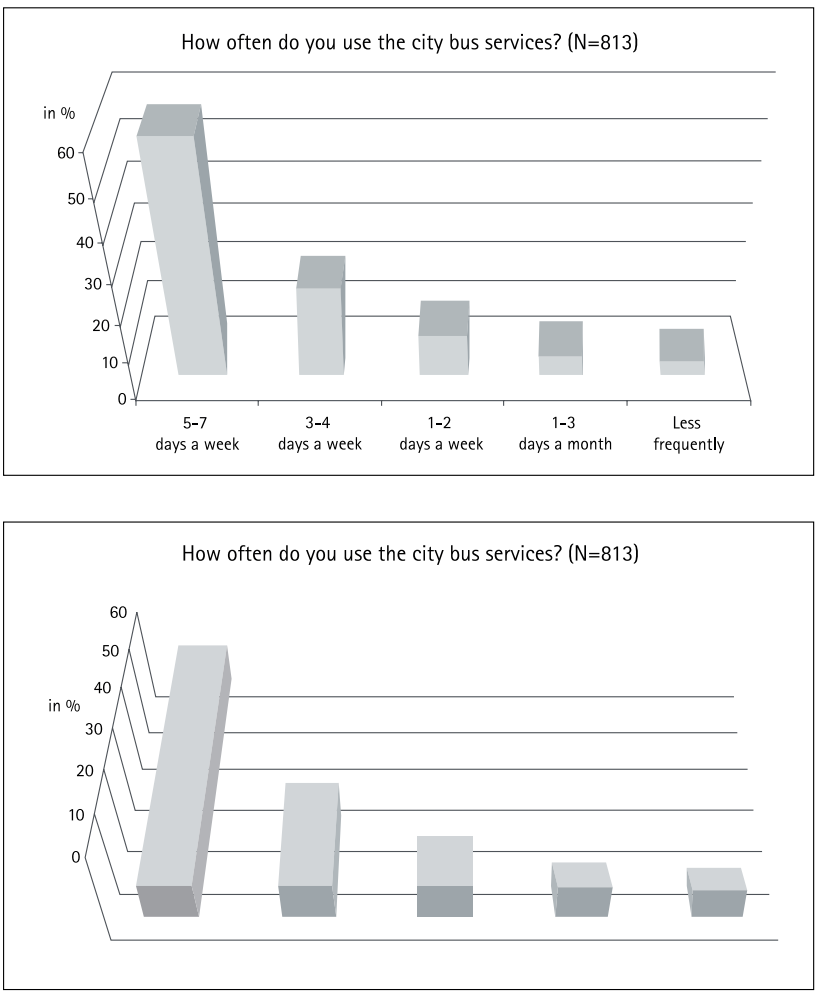


Figure 4–8: 3D graphic presentation of a histogram.

Starting the y-axis at zero, as done on the right side of Figure 4-7, we can hardly visually a difference between the two appraisals. In addition, when using bar charts, 3-D depictions are commonly used. On the one hand they are rich in variety; they offer many possibilities for manipulation. Every new perspective leads to a different impression. A tall column seen from far above seems rather flat while a short column seen from below can seem rather tall (see Figure 4-8).

In general, if you wish to discuss your observed proportions of figures, you should also publicise all specific figures of your evaluation (absolute and relative figures). This way you allow the reader to make up her/his own impression. Keep in mind that data presentation through illustrations is also a major source of error. (For example the scales on the x- and y-axes should each be partitioned consistently, etc.). Additionally, when depicting quantitative data you should consult specialised literature or skilled people.

Now, in order to find the best way to communicate and disseminate your information there are a few more things you should take into account: Beware of confidentiality regulations for specific data. Beware of any corporate design related to the measure, to your city or any larger project you may be involved in. What is the scope of your dissemination (local, regional, national, European, global)? This may not only have an implication for the content you are producing, but for the language in which you are addressing people. In order to exploit the material more widely you should consider publishing it either in English or in several languages.

Think of information dissemination as a multi-directional activity, i.e. you provide information which is potentially valuable to others, and through the same channels you will receive information from other sources which might be valuable for your city or your measure implementation. In the beginning of this chapter we said that results presentation is a follow-up of the evaluation activities. But it is also a prerequisite for further action, such as up-scaling or transferring a measure to another city.

Further readings

Alkin, Marvin C.; Christie, Christina A.; Rose, Mike: Communicating Evaluation. In: Shaw, Ian F.; Greene, Jennifer C.; Mark, Melvin M.: Handbook of Evaluation – Policies, Programs and Practices. Reprint, Sage Publications, London, 2007, p. 384-403.

4.2 Up-scaling of results

4.2.1 Introduction

A measure can be applied at small scale, as a pilot, to test its effects and find out if there are any unforeseen side-effects. Moreover, in the process of implementing the measure as pilot you might encounter obstacles that need to be overcome. However, the reason why your city has chosen to take action in favour of sustainable urban mobility was that they want to bring about change. So if, after all, the measure turned out to be successful, you probably want to further deploy it and bring about even bigger change! This is when you consider realising your measure at a larger scale. In order to not only scale up the costs, but first and foremost the effects of it, you need to have a careful look at your measure whether or not it qualifies for up-scaling.

Up-scaling refers to the estimation of the effects of a measure (or group of measures) if it/they were applied at a larger scale. For the Utrecht Road Safety Label for example up-scaling would mean to change the surroundings of every school in the city of Utrecht or even in the entire region.

Up-scaling is not limited to the city level, it can also take place within the region, e.g. if a public transport network covers not only a city but extends to the surrounding region, implementing a new ticketing system might be tested at city level and after successful implementation be scaled up to be used for the entire region.

For a sound estimation of the effects of the measure at larger scale a number of aspects have to be taken into account; thus, up-scaling is closely related to the issue of transferability (see next chapter) – it is a form of *vertical transfer.*

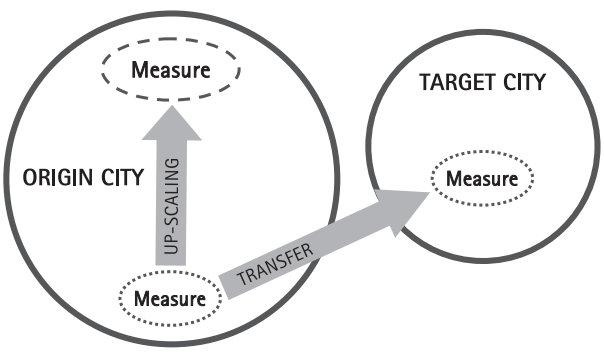


Figure 4–9: Distinction between up-scaling and transferability.

In this chapter we will show you why and when up-scaling is relevant and what the preconditions are. We will draw your attention to different scales and to what you have to consider before we tell you where to get your information from.

Box 15: What up-scaling is not

- Changing the target group, e.g. first addressing the blind and later extending the measure to address the elderly
- Adding an enforcement strategy to a measure which has shown little success before.
- It cannot be considered as up-scaling if doorstep interviews are replaced by approaching people through social media in a next step.
- And neither is the transfer of a measure from one corridor of a city to another one a matter of up-scaling.
- Up-scaling can be achieved through extrapolation, but extrapolation, e.g. in ex-ante evaluation or cost-benefit analysis is not the same as up-scaling!
- Up-scaling is not if you connect a local service website to an existing regional or national service platform (e.g. journey planner).
- Up-scaling is not about simply calculating the resources that would be needed to implement a measure throughout the entire city.

You may already have grasped why up-scaling is being dealt with in this evaluation handbook. The assumptions and estimates made for up-scaling a measure are being taken primarily from the results of impact and process evaluation.

With the variety of measures it is difficult to tell you what exactly you should do to scale up your measure as it largely depends on the characteristics of the individual measure in question. But there are clearly limits as to what up-scaling is.

4.2.2 Why and when is up-scaling relevant?

An up-scaling assessment provides you with guidance concerned about the potential for further deployment of a measure or group of measures. The question the up-scaling is trying to answer is: will more measure have more impact of the same quality?

If you look at the evaluation cycle/measure process (Chapter 1.3) then up-scaling will in most cases be an issue in the ex-ante evaluation of a measure. When the responsible people are defining the problem they want to tackle and look at the present situation and what seems feasible to undertake in order to achieve the objectives set. They will give some consideration as to what scale their measure should have and if it might be scaled up later on. Once the scale of a measure has been defined people are busy with implementing and up-scaling is usually not an issue anymore. However, the grounds for scaling up a measure are laid during implementation and ex-post evaluat-

ion, which is why data collection and the logging of information are crucial! During the planning and implementation process you can gather valuable information which you need for your estimates and assumptions for the up-scaling later.

4.2.3 Preconditions for up-scaling

Before you start making assumptions and providing estimates of what the effects of your measure would be at a larger scale, there are a few things that you should think about.

For various reasons, grounded in the characteristics of your measure, up-scaling might not be applicable at all; if it is or not is usually a question of common sense.

In some cities, some measures will be applied in a sufficiently coherent manner and widely enough that the effects will not need to be scaled up to a city level. However, most measures will not be of such a scale, and the effects of wider application must be estimated.

It seems trivial, but in fact is not: nobody wants to scale up a measure that failed. Therefore, success of a measure is a precondition for its up-scaling. So at first, you need to be clear about whether your measure has been successful or not. But how is success determined? It can only be determined on the basis of a sound evaluation where you compare the results with the business-as-usual situation and take into account additional socio-economic factors. For marketing and public campaigns the simple feedback from users can be encouraging for an up-scaling of the measure in order to spread the effect more widely.

But beware, a measure can be successful in terms of the desired effects, but it might not be economically viable, e.g. a local government could be paying rewards to automobilists for avoiding driving in the city centre during the rush hours. This might be a suitable measure for a short period when major road works are ongoing; however for a longer period or at a larger scale this is likely to become too expensive. However, if your measure was successful and expensive, but you expect there to be economies of scale, then up-scaling can well be an option.

Vice versa, scaling up a measure that showed economic viability in its pilot stage is not guaranteed to pay off at large scale. There might be cases where a scaled up measure needs some extra funding, at least in the beginning.

Regardless of whether a measure is successful at small scale there might be cases where the scaled up measure even destroys the success at small scale. Enforcement measures for example have a high risk of showing this effect.

4.2.4 Defining the scale

If up-scaling is applicable to your measure, you need to define the extent to which you want your measure to grow. There can be minimum and maximum sizes. If the measure is a pilot of an application with e.g. a limited target group or in a limited area, then you would expect there to be something like a full implementation throughout the city. But depending on the characteristics of your measure, the maximum scale you might be aiming at is what is technically possible or what is practically possible and may well be affected by what would be politically acceptable. There are likely to be geographical/location constraints and perhaps capacity limitations. With other measures the up-scaling might be possible gradually, e.g. by increasing the number of cameras, ticketing machines, bus stops, bicycles etc. involved in the measure. Or by enlarging, step by step, the geographic area that serves as test field or simply by providing more and more information with a service that you are developing. In any case it is essential that you define what exactly will be done at what scale and that the assumptions made concerning the up-scaling are clearly understood.

Box 16: Examples for scaling up measures

- If the implementation of a measure was a single event (e.g. training for eco driving), then up-scaling could mean introducing such training as permanent, mandatory part of the bus drivers' professional routine and you should define how often this training will take place (once a week or once a year?).
- If the measure was to implement a low emission zone, scaling it up could mean: a) enlarge the area (to which administrative border?), b) make the access rules stricter (e.g. by vehicle class, emission category, fees), or c) include heavy goods vehicles into the measure (that had not been part of the pilot measure).
- If the measure is about bio fuels, up-scaling could involve a) the type of vehicle using it (heavy goods vehicles, busses, cars) or b) the blend of the diesel fuel (percentage of bio components).
- For soft measures like educational or awareness-raising campaigns up-scaling could simply mean to have a larger target group.
- Up-scaling a smart (i.e. multi-modal) travel card system could mean to include other services such as e.g. car club and city bike membership.

The more aspects of a measure you change – apart from the scale – the more complex it'll get to make assumptions and estimates on the implications of the up-scaling. If the indicators for measuring success of the measure will be different ones at a bigger scale, you have probably changed the measure too much to call it the same measure at larger scale.

4.2.5 Considerations for up-scaling

The heart of up-scaling is to take into consideration all the factors that will change if you implement your measure at a larger scale and what implications this will have for the impact of the measure.

The more technical approaches are extrapolation and modelling. In order to apply one of these methods, or simply make assumptions and estimations, you need to know what factors are involved, what aspects are likely to change and in what direction. Here are some of the fields that might have an implication for the impact of your measure:

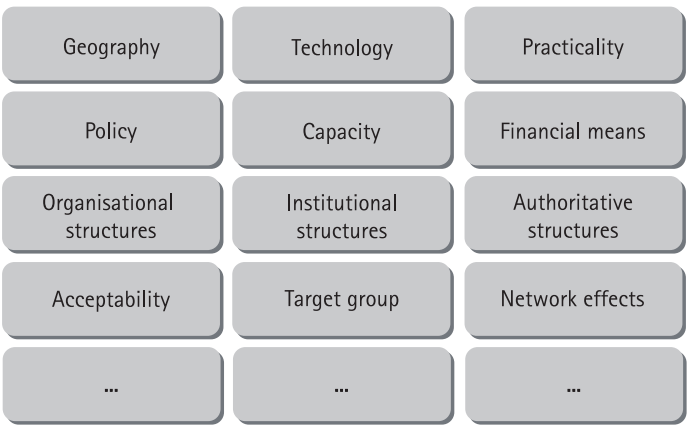


Figure 4–10: Possible fields with an implication for the scale of a measure.

For example, if you realised a shared space area in your city, and after a while people got used to it and you can claim overall acceptance. Still if the concept is too radical or too innovative to apply it to entire cities; further experience will show whether it is practical at larger scales at all (acceptability, practicality).

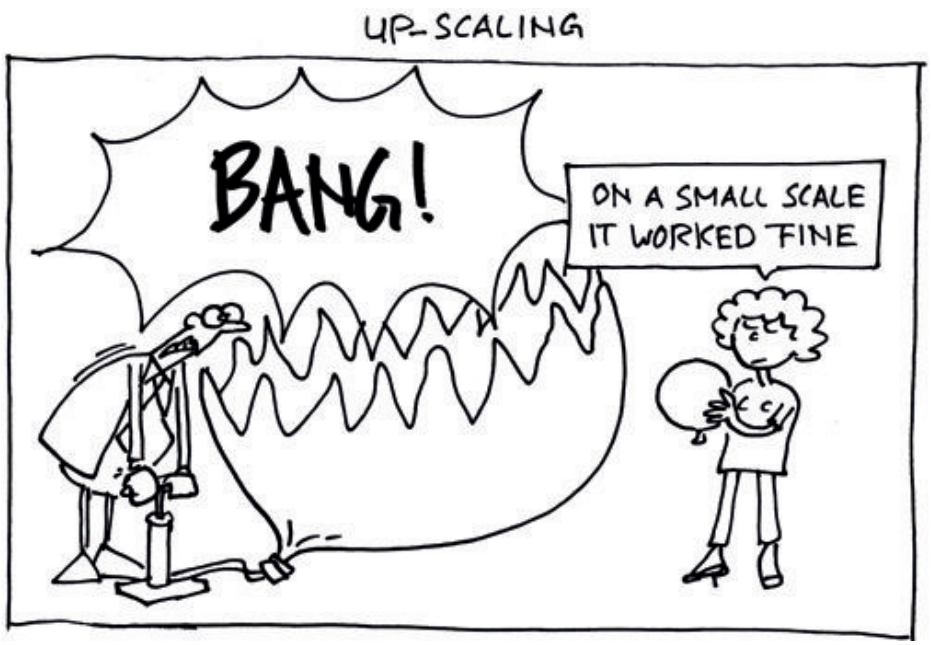
When considering up-scaling your measure you should be aware of potential effects on the transport network that have not come to existence at small scale. For instance, the introduction of a single bus lane and a reduction of road capacity on that section may have limited impact, since cars will probably divert from their regular routes and the travel time gains for the bus are limited. If, however, bus lanes are constructed full-scale as a network, diversions will be much more difficult, while the bus travel time gains are likely to be more substantial. Therefore, the impacts on the modal split and congestion are different (network effects).

Time is another factor which may become relevant for up-scaling. Some measures will have impacts which take time to develop and these should also be estimated for the larger scale.

For the Cargohopper example up-scaling could be the introduction of more vehicles or a higher frequency of delivery services or more shops having their goods delivered by the Cargohopper. One of the main criteria for assessing its potential for up-scaling is certainly the overall capacity of the service, but the wider acceptance of the measure in the target group is also important.

Another critical factor is behaviour, because it is difficult to predict. The users of a pilot may not be used to the new technology or service; the use or impact in the demonstration project may thus be lower than in a full-scale implementation, or when time has been allowed for adjustment. This brings us to the issue of synergies of measures which is why they are often grouped. If you plan to introduce a new technology or service it can be wise to combine it with an information campaign to spread the news and increase acceptance. With regard to up-scaling, if the general public already knows about the technology or service, the accompanying campaign might not be necessary any more. On the other hand, if the campaign was only locally restricted, then up-scaling would include a campaign at city level too, in order to bring about the same success of the measure it is aiming to promote as had been achieved at small scale.

Take the Cargohopper. It was introduced in the city centre of Utrecht and the shop keepers have been informed about the new service. Now, if they planned to enlarge the catchment area of the vehicle, another promotion campaign would be necessary to reach the new target group.



The level of evaluation which provides most relevant information for up-scaling is the city level, i.e. which effects the measure or group of measures has for the entire city. But you might also consider your lessons learnt from the process evaluation, because it might get more complicated at a bigger scale, e.g. involving more authorities and thus more stages to get planning permission. However, the information you get from the process evaluation requires more knowledge transfer and abstract deduction: will the drivers and barriers to the implementation process be the same as those you encountered when implementing the measure at small scale?

Surveys, studies and statistics from the impact evaluation will provide some hard facts that go into modelling or extrapolation. Cause-and-effect chains also provide valuable input if you look at them with the larger scale in mind. In the up-scaling scenario external data (e.g. historic data, data from previous surveys) might be included to get a more reliable picture of the measure's impact at a larger scale.

If prior to the implementation of a measure studies have been carried out at small scale in order to assess the feasibility of it, the same investigations are necessary for scaling up the measure.

Up-scaling for promotional measures will first and foremost involve a review of good practice in engaging with the target group. If the measure explores a variety of initiatives and events involved in promoting sustainable mobility some may have had more success than others, and in some locations but not others. For a successful up-scaling an appraisal of what works best, and where, will be of the highest importance. Evaluation results should therefore include detailed case studies which demonstrate drivers and barriers of the interventions and suggest methods for future work.

If a user survey is part of the evaluation process of a measure, the survey can be designed so that users or potential users are being questioned whether they like the service and would like to see it operate throughout the entire city. This would deliver valuable information for up-scaling of the measure.

All relevant considerations drawn from the various sources eventually go into an up-scaling scenario that you draw up and that will help you or any other person in charge to make a decision on whether it is feasible and sensible to scale up a measure.

4.3 Transferability

4.3.1 Introduction

What if your city has implemented measures which have been shown to be very successful – not only in terms of positive impacts on the environment and on the quality of living in your city, but also in the effectiveness of the planning, implementation and operation process? Then other cities will surely consider copying the measures in their city. And

they should copy them! So, do not keep your achievements a secret. And even if you encountered difficulties in implementing your measures, your experience can still be valuable for others.

You may have guessed already – the idea behind all this is to learn from each other, not only what works, but also what did not work, in order not to make the same mistake twice. This includes a transfer of a measure or a measure bundle from one city to another. But each measure in each city is different. The example of the Cargohopper in Utrecht has a number of positive impacts on the city environment (reduction of noise and air pollution, more energy efficient freight transport, improved accessibility of the city centre) and would not it be nice if other cities could benefit from this experience? Introducing the Cargohopper in their city might be an option for others. But only because it was successful in Utrecht does not necessarily mean it is going to be a success in any other city in the world.

Transferring a measure from one city to another one is not an easy task. We have to consider the specific conditions under which the measure has been implemented successfully (enabling context) and thus determine the transferability of a measure – and that is what we are going to look at more closely in this chapter.

What do we mean by transferability? It refers to the degree to which the effects of a measure or group of measures can be transferred to other contexts or settings (see Figure 4-9). When we talk about transferability we try to identify measures which could be implemented successfully in other cities and under which circumstances this would have to take place. Transferability does not simply refer to individual technical or operational features, but how a measure corresponds to the city context and how the measures are interrelated, i.e. the conditions under which and combinations in which a specific measure or package of measures can be applied with a comparable degree of success elsewhere.

Does this sound familiar to you? Well, in fact assessing a measures' transferability is not very much different from assessing its up-scalability. If you like you can say that up-scaling is a vertical transfer (see Figure 4-9). Therefore, a lot of the information needed is also the same. If you have considered up-scaling for your measure, then the transferability assessment is going to be peanuts!

4.3.2 Framework for transferability assessment

If you simply compare the city where the measure has already been implemented (the origin city) with the city which would like to implement the measure (the target city), you will not get any significant predictions as to whether a transfer is likely to lead to success. In fact, transferability depends to a large extent on the characteristics of the

Box 17: Cross-site evaluation

European cross-site evaluation is a distinct exercise as compared to the evaluation undertaken by the individual cities and will provide the European Commission with information necessary to disseminate the notion that changes in urban transport technologies and policies can indeed be highly beneficial to European citizens at large, and to replicate similar practices elsewhere.

In CIVITAS the horizontal support actions for evaluation produced so-called cluster reports in which they evaluate measures of a specific thematic area (e.g. bio fuels, walking and cycling, freight logistics) across cities. You can find these cluster reports on www.civitas.eu.

measures themselves and the prevailing conditions for implementation in the target city.

This means that – in a first step – all related information which can contribute to explaining the success/failure of a measure need to be collected and disseminated (e.g. presented in a report or fed into a common database).

Looking at the implications of cross-site evaluation, i.e. comparing the results of measures between cities or generalise results across several sites, can provide addi-tional clues.

The success of a certain measure may be related to the impact of another measure too. Packaging of measures often leads to synergies with mutually reinforcing effects.

When you set up your framework for a transferability analysis please bear in mind the following general transferability principles:

- The transfer of experience should involve all the policy makers and practitioners in the cities considered to have the power to take decisions affecting a given policy context.
- The most significant barriers to transfer a measure relate to political frameworks and public acceptance, underlining the vital importance of obtaining political and public support.

For assessing the transferability of a measure, you can use the information collected during the impact and process evaluation activities, looking at all levels of its objectives (high-level and measure-specific), and combine these with external aspects describing the context or specific circumstances under which the measure has been implemented. This complex information is put in relation to the setting and the aim of the city to which it is going to be transferred.

There are two main actors in a transfer process – the origin city and the potential target city. So when we speak about transferability analysis in this handbook there are two ways of looking at it:

- From the perspective of an origin city
- From the perspective of a potential target city

Corresponding to these perspectives, the responsibility and tasks are shared between the city that exports its measure and its experiences and the city that imports the measure and hopefully the success of it. The origin city provides with its evaluation results of the measure on one hand, plus some extra information on the context of the successful implementation on the other, the basic input for the ex-ante evaluation of the potential target city which bears the responsibility for making the judgement of how sensible the transfer would be in their own context.

And accordingly, there are two different analysis procedures. However, for a better understanding we recommend you to read the entire chapter (it's not that much) – no matter if you are an origin or a target city.

The role of the origin city

Cities that implement measures to make urban transport cleaner, better and more sustainable should disclose their experiences which are valuable to other cities. The demonstration cities will enable transfer of knowledge as well as of effects by preparing for transferability analysis which will then be carried out by potential receptor cities and would otherwise not be feasible.

If you want to become an origin city you should keep an eye on transferability from the outset, i.e. when you start planning your measure. The transferability assessment will be based on your city's evaluation results of the measure. This requires that you plan, implement and evaluate the demonstration measures taking due account of the requirements for a possible transfer. This includes using consistent evaluation methodologies across sites and collecting the necessary information through the process evaluation to help explain the success and failure of the measures.

There are certain requirements and prerequisites for a transferability assessment which only you as an origin city can provide:

1. Define the success of the measure in your city

Nobody wants to transfer a failure. Therefore, a precondition for transferability is the successful implementation of a measure. The definition of success will naturally depend on the objectives set. It is desirable that the success dimension of a measure be translated into a predefined quantitative scale, i.e. success criteria should be measurable (see also SMART Table 2-3).

In a transferability assessment you need to define success of a measure by describing the problem that your city was facing, that called for action, by determining what your objectives were, by identifying what the effects of the measure are and how these relate to each other. The reason for this being that if you have been successful in implementing your measure it does not mean that the results would be considered a success

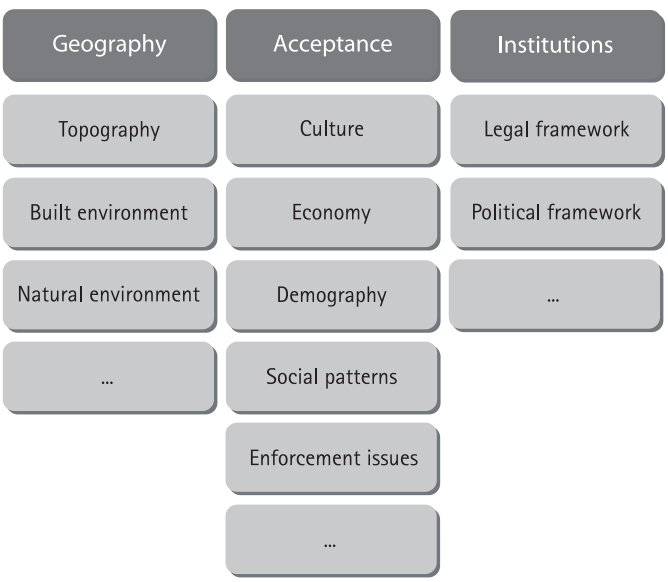


Figure 4–11: 'Things' to keep in mind.

everywhere else! Therefore, you need to define your success in terms of objectives achieved and what indicators you have used to measure the impact (see also Chapter 2.3). This is part of the impact evaluation, so you can take this information from there. In addition to this quantitative approach, qualitative statements e.g. the opinion of experts can contribute to the definition of possible correlations.

2. Describe in detail the settings of your city

A detailed identification of the characteristics of the city environment and urban structure is essential for assessing transferability of the measures and permit comparison with conditions in other cities. As no two cities will have exactly the same conditions in reality, and not all settings and characteristics have an effect on the outcome of the measure, it is important to identify the potential effects of differences in city settings on the replication of a measure, especially the differences in key conditions for success or failure. Figure 4-11 shows some aspects you should look at in relation with your measure.

3. Build cause-and-effect relationships between measure impacts and city settings

Understanding the cause-and-effect relationships between the impacts and the city settings is necessary for the identification of key factors. This is the essence of success for transferring practices into another city (see more on cause-and-effect chains in Chapter 2.2).

It is important to understand which of these factors have prevailing importance in the success of the measures. The weight of the factors contributing to the city settings will be used to determine the conditions for applicability. The causal relationships between socio-economic, environmental or institutional aspects and the measure or packages of measures in question lead to conditions for applicability of a measure at the various levels (measure level, city level).

4. Identify the interdependencies between measures

In CIVITAS cities realised the need to adopt combined measures, assuming that not only measures considered alone but their coherent bundling with other measures (packaging), will ultimately determine the overall degree of success of a set of measures within and across policy fields and clusters.

5. Describe in detail the process that the implementation has gone through

This includes the identification of barriers, how they were overcome, and drivers that have pushed the implementation of your measure in the right direction.

The evaluation of the measure processes can reveal very helpful results in this context: Just think about the political, legal, financial and organisational structures in your city. What influence did they have on your measure?

6. Define the lessons learnt

Do not let the others make the same mistakes as you did. All this information will go into a database of measures, such as the databases developed in Eltis or CIVITAS for further use for decision makers in potential target cities.

The role of the potential target city

If you are transport practitioner at city level you are probably in the best position to screen measures, on the basis of your knowledge of the local setting and your city's local needs! This underlines why the city that wishes to import the effects of a measure or group of measures from another city to a different context in their own city is mostly responsible for making the judgement of how sensible a transfer would be, i.e. whether it will bring about the intended results and help to achieve the objectives.

Box 18: Where can I find case studies for urban mobility measures?

Selected sources for case studies, feasibility studies and ideas for measures promoting sustainable urban mobility can be found on various European websites, including:

- www.civitas.eu
- www.eltis.org
- www.epomm.eu
- www.mobilitymanagement.org

By the way, have a look at the presented initiatives and organisations. Maybe they will be able to offer help and information beyond providing measure examples.

From the perspective of a potential target city the transferability assessment of other cities' measures is particularly interesting in the ex-ante evaluation where the most suitable measure for an existing problem is being selected.

4.3.3 Conclusions to transferability

The difference between what you need to do for an up-scaling of your measure and for assessing its transferability is very small, in fact it's mainly the direction (vertical – horizontal) and the scale, plus the perspective. Thus, the motivation is different. If you are an origin city: your interest in a proper evaluation of the measures and dissemination of results is uppermost in your mind. It is probably not primarily motivated by providing useful information for potential target cities. First of all you may think in terms of political strategies and city marketing – but in order to get reliable and verifiable results out of your evaluation you should include all that is necessary for the transferability assessment in your evaluation anyway! So think of other cities that could benefit from your experience with the implementation of sustainable transportation measures – evaluate and report the impacts and processes in a way so that other cities can learn from your experience.

If you are a receptor city: you can be glad that other cities have already made an effort to tackle problems that might be similar to yours and that, by communicating their experiences, they help you a lot already. Your task is to clearly define the problem in your city and its settings in order to find a suitable measure that can be successful in your city! If it was – you will find out by doing a sound evaluation which can then be the starting point for yet another city for their way to cleaner and better transport.

To conclude this, the concept of transferability has a broader perspective and is part of an approach of mutual learning. It requires you to look beyond the rim of your teacup and think outside the box, but we promise you – it's all worth it!

Further readings

CIVITAS, METEOR Consortium: Final cross-site evaluation report. Rijswijk (Netherlands), November 2006. (download possible from www.civitas.eu)

5 And now: back to reality

The last chapters talked a lot about what you should do in your evaluation – it should have given you a good idea of how you can accomplish a good quality evaluation. But let's be honest, there are still some myths that stick around and there are many thinks that can go wrong. This chapter will talk about these issues and give you some recommendations how you can strengthen the validity of your evaluation approach. It will help you to plan ahead and to be prepared for the challenges of real life evaluation.

5.1 Some myths about evaluation

In a very broad sense, the word myth refers to any story that mankind uses to tell himself how things came to be. It often turns into a popular belief that has become associated with a person, institution, occurrence or sometimes a task such as evaluation. There are many myths when it comes to evaluation. The three most popular seem to be: evaluation is too complicated, it is too expensive and it will automatically determine whether or not my program (and me, who is in charge of it) will be eliminated. As with most myths, there is some truth about them. But let's look at them a little closer.

Evaluation is way too complicated.

Well no, while you are sleeping you will have a vision and the next morning you will write down the objectives, indicators, results and interpretation for the evaluation. Everybody will congratulate you on your good job and the fabulously positive results. You will receive awards; you will get a raise and live happily ever after. No, let's be honest. Evaluation can be a complex task. But it is one that can be broken down into small and easy to handle tasks. Then the 'huge' evaluation work will be a lot easier to handle. But you will have to make your first baby steps to get things started and ready to go. The idea of evaluation is often rejected because measure/program personnel do not know how to evaluate. So, why not ask for help? Although some aspects are complex, the evaluation process is a practice in which most people have already engaged in an informal way.

Evaluation is too expensive.

Many have an aversion to evaluation because they say it costs too much. With funding for measures/programs in short supply, leaders would rather devote resources to services than to evaluation projects. They believe that professional evaluators are people who speak an inaccessible academic jargon and produce costly and unwieldy, tele-

phone book-sized reports that are of little use to anybody. Evaluation should be considered an investment in your measure/program. The benefits of a properly conducted evaluation can have an invaluable effect. Although evaluation does use up time and resources, there are a wide range of inexpensive options at your disposal. Especially if there are concerns about the cost of evaluation, it is possible to plan very low cost evaluation. For example, you might reach out to a local university and find a graduate student who can execute a small-scale evaluation project in exchange for internship credit. In addition, EU-programs often also fund evaluation activities. Remember that you can always start small by evaluating a single measure or discrete set of activities. You can always ramp up your evaluation efforts over time. By then, you will have experience and with all things, practice makes (almost) perfect.

Evaluation automatically determines whether or not my program will be eliminated.

Evaluation is for measure improvement not its potential elimination. A common misconception is that your evaluation needs to illustrate no need for improvement when, in fact, each program has room for improvement. Ongoing feedback to analyse, understand, and refine your program is essential. So do not interview just the successes. You will learn a great deal about the program by understanding its failures, dropouts, etc.

Further readings

Datta, Lois-Ellin: The Practice of Evaluation: Challanges and New Directions. In: Shaw, Ian F.; Greene, Jennifer C.; Mark, Melvin M.: Handbook of Evaluation – Policies, Programs and Practices. Reprint, Sage Publications, London, 2007, p. 419-438.

5.2 'Things' which can mess up your evaluation

For most of us it is very difficult to look into the future and to see where circumstances arise that try to mess with your evaluation. However, it is helpful to anticipate potential problems for the evaluation in order to prepare and eventually begin actions that could prevent these problems. The issues mentioned here are reflecting selected problems for the evaluation in the CIVITAS project. This list is by no means complete; you might encounter very different problems. But this chapter is supposed to sensitise you and all actors involved for the task ahead.

The first and probably most common problem with urban transport measures is that your measure changes sometimes significantly as you progress. There are numerous reasons for that; the financing does not allow a continuation as planned, the political support for your measures increased/decreased or ceased, the legal framework changes and so on. But let's not focus too much on the problem. Fact is, your measure implementation changes and most things that you laid out in your evaluation plan are no longer true. Face it, this can happen every day, but do not stick your head in the sand and stop the evaluation, rather see the change of implementation as something you can use to improve your measure. For the evaluation, it might mean that you no longer have a 'before' measurement (because the data you acquired before the implementation has no relation to the 'new' measure anymore). Try to do the best you can to keep your evaluation design and if you change your indicators or design, make this part of your process evaluation.

Another common problem is that delays frequently occur – you cannot escape it. They will also affect your evaluation. With delays the period between implementation, operation and evaluation might be too short and the measure can not fully unfold its effects. It could mean that the impact evaluation can not be applied completely and the results determined stands on shaky ground. However, it should be pointed out that the evaluation concept with impact and process evaluation also deals with measure performance quality and their causes such as drivers and barriers. Hence, delays might lead to shifts within the evaluation concept and then again the process evaluation will play a strong part.

Even though it might seem unlikely to you at the moment, but it does happen that you actually try to acquire too many indicators. You would not do that on purpose, but often figures are included in the report because 'they are available anyway'. But assessing measures with a great number of indicators doesn't only mean that it will be assessed in every detail. Foremost, it means a lot of resulting data that has to be analysed, interpreted and written down for reporting. This could – in some cases – lead to an additional and difficult to calculate need of working time. If this need will not be taken under consideration sufficiently, some data gathered might not be analysed and interpreted in the end and will therefore be useless.

But even if you only have very few indicators, you will get into trouble somewhat if the data you thought was available is not. Especially, sensitive data – data of which stakeholders are suspicious about how the data is used – can be hard to get. Try to assess from the beginning which data will be available to you and have others commit to the delivery in a contract-like form including a date by which they will provide it. Especially, if you pay for this data, make sure that you are very specific about what you want and at which time you want this.

If things go wrong and you have no baseline data available, respondents can be asked to recall their earlier situation with respect to use of transport mode, or time and cost of travel. Recall data are subject to potential biases due to problems of memory, under- or over-reporting, and the distortion of socially desirable or undesirable behaviour. Results are also very sensitive to the time period covered and how questions are formulated. But it is still useful where survey data is not available and far better than having nothing at all.

Further readings

Walker, Robert; Wiseman, Michael: Managing Evaluations. In: Shaw, Ian F.; Greene, Jennifer C.; Mark, Melvin M.: Handbook of Evaluation – Policies, Programs and Practices. Reprint, Sage Publications, London, 2007, p. 360-383.

5.3 Strengthening the quality of your evaluation

While there are many challenges for designing and conducting quality evaluation, there are also proven strategies that will strengthen the quality and point ways to overcoming future obstacles. These points refer to future evaluations you might be involved in and result in many cases from the examples discussed above.

First, it is important to define realistic expectations for the evaluation of your measure. Ensure that the evaluation is considered alongside measure design and a plan is put in place which clearly articulates how and when evaluation will occur throughout the measure management cycle.

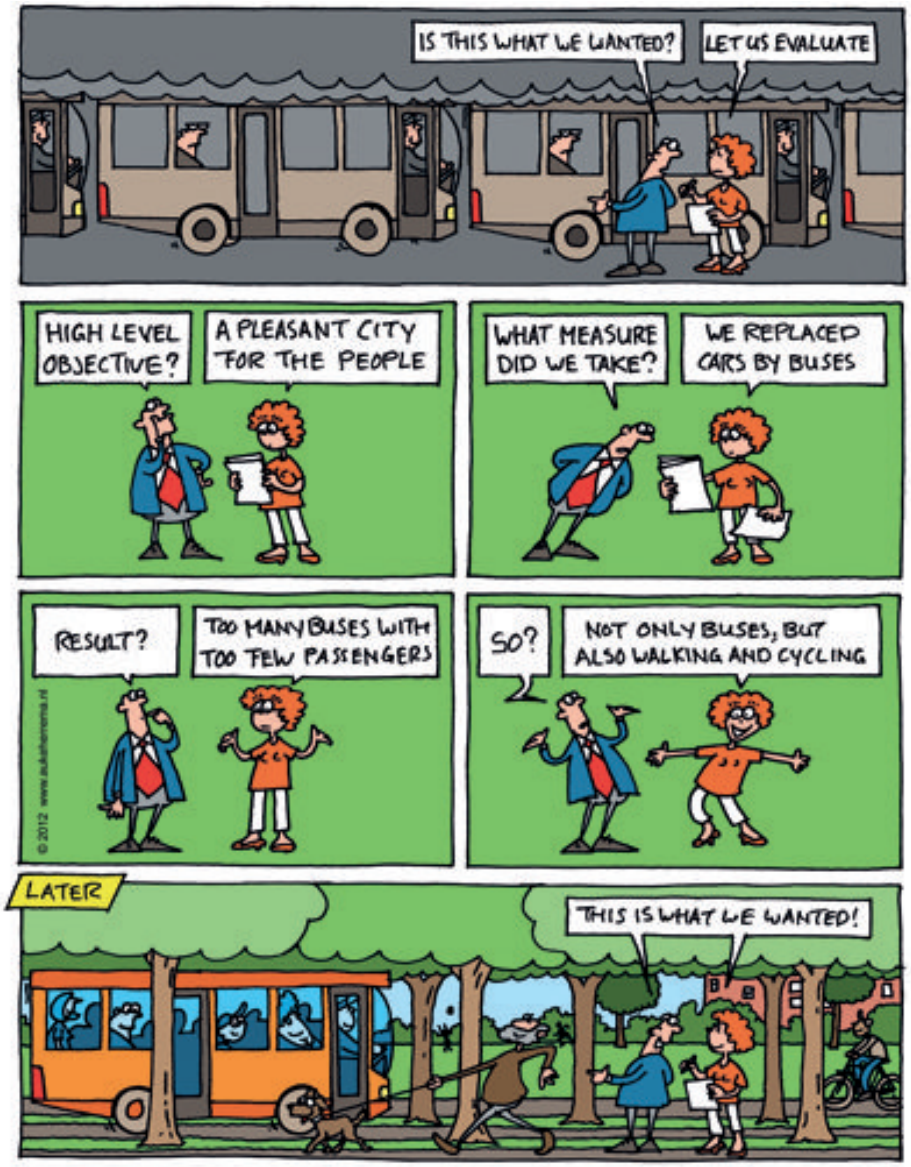
In the same context, always search for feedback. You and your team should be reflecting, seeking feedback and providing feedback on a continuous basis. This feedback can be both formal and informal and can come from your team, other staff members or other stakeholders. A good method is peer review where you also give feedback to someone in the same situation as yourself.

Especially, in transport-related measures, changes can only be measured over a long period of time. Most municipalities have computers, typically used for word processing, but few have the software and/or skills needed to manage outcome data usefully. Developing computer capacity to store, analyse, report and track outcome data are essential to such evaluations. Make this a priority in your working field. Remember: Don't throw away evaluation results once a report has been generated. Results do not take up much room, and they can provide precious information later when trying to understand changes in the program.

Further readings

Whitmore, Elizabeth; Guijt, Irene; Mertens, Donna M.; Imm, Pamela S.; Chinman, Matthew; Wandersman, Abraham: Embedding Improvements, Lived Experience, and Social Justice in Evaluation Practice. In: Shaw, Ian F.; Greene, Jennifer C.; Mark, Melvin M.: Handbook of Evaluation – Policies, Programs and Practices. Reprint, Sage Publications, London, 2007, p. 340-359.

Stake, Robert E.; Schwandt, Thomas A.: On Discerning Quality in Evaluation. In: Shaw, Ian F.; Greene, Jennifer C.; Mark, Melvin M.: Handbook of Evaluation – Policies, Programs and Practices. Reprint, Sage Publications, London, 2007, p. 404-418.


IS THIS WHAT WE WANTED?
LET US EVALUATE
HIGH LEVEL OBJECTIVE?
A PLEASANT CITY FOR THE PEOPLE
WHAT MEASURE DID WE TAKE?
WE REPLACED CARS BY BUSES
RESULT?
TOO MANY BUSES WITH TOO FEW PASSENGERS
SO?
NOT ONLY BUSES, BUT ALSO WALKING AND CYCLING
© 2012 www.aukehoeve.nl
LATER
THIS IS WHAT WE WANTED!

6 The very last chapter

Now, it seems, everything has been said about evaluation! But of course we all know that there is much more. In this very last chapter we would like to say a few words about evaluation in general.

6.1 Some words to the evaluator

From our experience with the evaluation support for cities within the CIVITAS initiative we have to admit that evaluation is mostly treated as a bureaucratic contractual obligation to be fulfilled with as little effort as possible. But don't we all evaluate every day? In fact, we start doing so when we start to exist, making our way through life. Little babies grow and draw a lot of what they need to survive from trial and error, a method of problem solving, repair, tuning, or obtaining knowledge. And learning doesn't happen from failure itself, but rather from analysing the failure, making a change, and then trying again. The same holds true in our complex world, where the evaluation of a measure is fundamental not only in order to demonstrate the activities undertaken, but also to identify and share the results – good ones and bad ones. Thus, evaluation can lead to a learning community.

If you want to understand what you and your measure are doing, you need to systematically evaluate. Without evaluation you will only see fractions, but you will not be able to see the effects of the measure as a whole. The obvious solution for solving your city's mobility and environmental problems might not always turn out to be the right one. Evaluation can help to better identify your problem, find better solutions and make the right choices in selecting and designing measures. Evaluation appears to be seen as a difficult task and surely, it's easy to find reasons why you couldn't do it. But think about what you can do!

A good reason for getting up and start is that there are numerous positive side effects, e.g. the data needed for impact evaluation often leads to more structured data-collection in general. You may discover synergies with other data. Evaluation can help you to save a lot of money by optimising measures during their implementation and operations. And evaluation can produce a good case for your politicians demonstrating to the citizens that their policy shows the desired effects.

With this handbook we hope to contribute to the formation of a culture that seeks to increase the amount of reliable information that is made available to the public, against the still prevailing feeling that it is done because of a mere contractual requirement. Obviously, this also depends on how much support you get – financially and morally – from your politician. So ...

6.2 Some words to politicians

This book is targeted at staff in city governments who are faced with having to carry out the evaluation of small urban transport measures. But all that was said about learning from success and failure applies to politicians too. So, if you are a politician or have any kind of political influence and read this, we would like to tell you about our vision:

> *Evaluation should be an integral part of every project, provided with a budget that enables the evaluator to carry out the tasks necessary for a sound evaluation.*

Moreover, evaluation should be as natural as it is for every human being in everyday life. It should receive every due support from you at every stage and long-term commitment in order to collect the necessary data.

The evaluation process must be planned simultaneously to the measure planning. In order to have reliable data and results that you can sell, the correspondence between measure and impact (through indicators) must be defined at the beginning. There is a high risk of identifying outputs only, but no results which show the correlation between action taken and impact. Only with systemised data and significant results, it is possible to direct future decisions. It is human nature to believe what we see and seek conclusions from 'experts.' We cling to the obvious for many reasons: It is justifiable. It is popular. It is familiar. It is comfortable. And it is often the path of least resistance. But what is obvious is not always what is true.

Evaluation ensures transparency, and that is what the public wants. If you can back up your political decisions with results from an evaluation, your voters will appreciate that, no matter if the measure has been a full success. It is more important to show openly what happened with the money from public funding.

We would like you to share our vision and to be an active part of a learning society. Today's world is characterised by an abundance of information. The problem is not to get some information – as it was for our grandparents –, but to get the information that we need and in the desired quality. Our contribution to manage this information society should be to provide reliable information of good quality and no trash. This includes providing good quality information about one's failures, so that others need not make the same mistakes. And you might be the other one who benefits from the evaluation results from others. But at this moment, we are asking you to make the first step. Assign extra budget for evaluation of your next urban mobility measure or establish a standardised data collection process and database for mobility related data in your city.

Keep in mind: Evaluation is not meant as a threat for you. It helps to justify your policy with valid data and in the case of ill-effect; you can show why it happened and can learn from it for your next political decisions. Your citizens like proof for successful

measures but they appreciate even more an honest and transparent evaluation of what went wrong. You will be surprised to learn how much hope and trust can be put in a politician who is able to manage failure just as much as he can manage success. In the end, it is this that you will be elected for.

7 Evaluation examples

The following subchapters include evaluation reports for the measures which have served as examples throughout this book. It is the intention of the authors that the reader of this book can see the measure evaluation in its context. Also, it is meant as an inspiration for reasoning and a structural approach in the evaluation of urban transport measures. Therefore, it focuses primarily on the impact and process evaluation and mostly neglects the statement of measure outputs – such as the energy consumption of the Cargohopper for example. For the purpose of this book it is assumed that these 'technical' details are part of a general research and development activity which is reported separately. This is, in fact, the case in many European funded programs. If you conduct your own evaluation, be aware of your target group and other deliverables that are mandatory in your program. If you are not required, or do not wish to have a separate technical deliverable, you should include more details about this in the evaluation report. You could include a chapter 'measure outputs', to separate it clearly from your impact evaluation. But now, enjoy reading the examples. Please note – as mentioned in the introduction – some details about the examples are fictional to emphasize good-practise, while others have been taken from real life experience.

7.1 Utrecht Road Safety Label

7.1.1 Problem description and measure context

Every year, approximately 40,000 people are killed on the roads in countries of the European Union and this is the leading cause of death among children. Speeding and the fact that drivers do not adapt to the conditions of the road and its surroundings is amongst the main causes of fatal accidents.

Utrecht is the fourth largest city in the Netherlands and 13% of its 300,000 inhabitants are between 0 and 11 years of age. As such it is not surprising that road safety problems increasingly cluster around schools. As more and more parents drop their children off and pick them up by car, children who cycle or walk to school are more and more at risk, even further because children often do not act safely in traffic. The situation has become such that the areas around many primary schools are no longer sufficiently safe. This is emphasised by the fact that studies show that the average vehicle speed around schools is well above the allowed 30 kilometres per hour.

The city of Utrecht was thus keen to bring up the issue of road safety in education and to improve road safety around schools. This is inline with the fact that schools and parents nowadays expect the government to guarantee the safety of their children

around the school and on the route from and to the child's home. Often road traffic education, however, does not receive priority, because the school curriculum is already filled with other courses which are seen as a priority.

Preceding this project, the city of Utrecht did not have a standard lay-out for the arrangement of road crossings for pedestrians and bicycles near schools. The majority of teachers had very negative views about how safe roads are. In 2008 only 22% of them thought the pupils never had to deal with unsafe traffic situations around the schools. Most solutions to increase road safety are created for just a few short periods during daytime (such as school crossing patrols) and differ from school to school and have become progressively divers.

7.1.2 Objectives

As road safety problems are clustered increasingly around schools, the city of Utrecht aims to contribute to an improvement to the quality of life. The long-term objectives which correspond to the Utrecht urban mobility plan are:

- Increase road safety.
- Increase of modal share of sustainable modes of transport.

To reach these objectives, this measure has the following specific objectives:

- Reduction of accidents around schools in the school areas and in the surrounding residential area.
- Improve the satisfaction with the road safety in primary school areas among children, their parents and teachers.
- Reduce the share of home-school-trips by car in favor of cycling and walking by 5%.

7.1.3 Preparation and implementation stages

Stage 1: Action plan preparation

During the years 2007 and 2008 the city of Utrecht decided to improve their efforts regarding traffic safety of school children and to take action in three different areas:

- Creating uniform and recognisable school surroundings.
- Traffic education for the pupils.
- Influence the traffic behavior of parents.

The city discussed different options with stakeholders (parents, teachers and student representatives). In April 2008, the city council decided to award primary schools with

a so-called Road Safety Label when they deal actively with traffic safety. They were approached and informed about the Road Safety Label. They get help for setting up and implement an action plan that indicates which criteria will need the school's attention for improvement. Thereby, the city fully covers the costs for this support. If a school fulfils all criteria, it will be awarded the Road Safety Label. Examples of these criteria are:

- Traffic education is part of the school's policy and the school guide contains a para graph about road safety and traffic education.
- The school has a traffic team with teachers and parents that meets regularly.
- Teachers of all grades spend reserved time on traffic education in the class.
- All schoolchildren must take part in the already existing national traffic exam (the theoretical and the practical part).
- The school children practice their behavior and the traffic rules at the school yard, the street or the traffic garden at least once a year.
- The school takes care of a safe school exit and a safe accessibility by bicycle or on foot.
- The school informs the parents frequently about road safety and traffic education, makes agreements with the parents on how the children come to school (bicycle, public transport or on foot).

In return, the city will provide managerial and safety education support in the form of a subcontractor and rewards schools that meet the criteria with a compensation of up to 75% of the costs of traffic education. In addition, the city implements uniform and recognisable school surroundings at all schools that are working on this Road Safety Label.

Stage 2: Pilot testing (2009)

- In February a plan was developed to make the school surroundings uniform and recognisable; schools were approached to participate in the pilot.
- In March from eight schools that volunteered to participate in the pilot, three have been chosen as pilot sites for an implementation.
- For each of the sites where the uniform school surroundings are were intended to be implemented, data has been gathered on parents' perception of safety, actual accident data and the modal split of school trips (baseline data).
- In the late spring, these school surroundings of the pilot schools have been redesigned to match the new street furniture. Parallel, the schools included road safety trainings in their schedule (as part of the physical education courses).
- The implementation has been evaluated after the summer.

Figure 7–1: First school zone before (on left) and after.

The positive feedback has led to the decision to continue with improving all the other surroundings of primary schools in the city that are willing to obtain the Road Safety Label.

Stage 3: Further implementations (2009 on-going)

Starting after the pilot test the process for the measure implementation is as follows:

- Step 1: The educational support service approaches primary schools that do not already participate in the Road Safety Label scheme. They telephone directors, visit them and present the concept to the schools directory board.

If the school decides to participate in the label scheme:

- Step 2: Baseline data will be collected for one out of 10 schools where the uniform school surroundings are intended to be implemented. Data is gathered on the average modal split for home-school journeys, satisfaction with the road safety of the schools' area and accident data through a questionnaire distributed among parents.
- Step 3: Schools are responsible for making traffic education part of their policy and the general curriculum. They appoint a road safety coordinator who is in charge of the project at the school, and – if necessary – supports the implementation.
- Step 4: New school zones will be implemented in the school surroundings.
- Step 5: Road Safety Label is awarded if the schools fulfills all criteria.
- Step 6: For those schools which were selected for the baseline data collection, after testing is taking place after implementation.
- Step 7: The city continues to support schools that are working towards getting the label or already have it. The schools will be audited every two years to verify their fulfillment of the Road Safety Label criteria.

If the school does not wish to participate:

- The educational support team that was in charge of the contacting the schools stored the various reasons for the rejection of the scheme in a database. Approximately two years after the initial rejection, the schools are approached again and invited to a site visit to schools which have been awarded the Road Safety Label.
- In 2009 a sample of three schools that did not want to participate was randomly selected for a control site data collection, those were asked twice (in 2009 and a second time early 2011) about the average modal split for home-school journeys, satisfaction with the road safety of the schools' area and accident data. The information was collected through a questionnaire among parents. In return, the schools received extra equipment for their physical education facilities.

7.1.4 Cause-and-effect relationships

The city of Utrecht awards the Road Safety Labels to primary schools that proactively address road safety issues. Thus this initiative has an impact on the schools surroundings, the schools' curriculum and it encourages parents' involvement in safety-related issues. The unification of school surroundings over the city of Utrecht will have an effect on the average vehicle speeds in the school vicinity due to speed limitations, but also due to a raised awareness for children's road safety amongst drivers. This could also increase their attention and thus have an impact on safety and traffic flow.

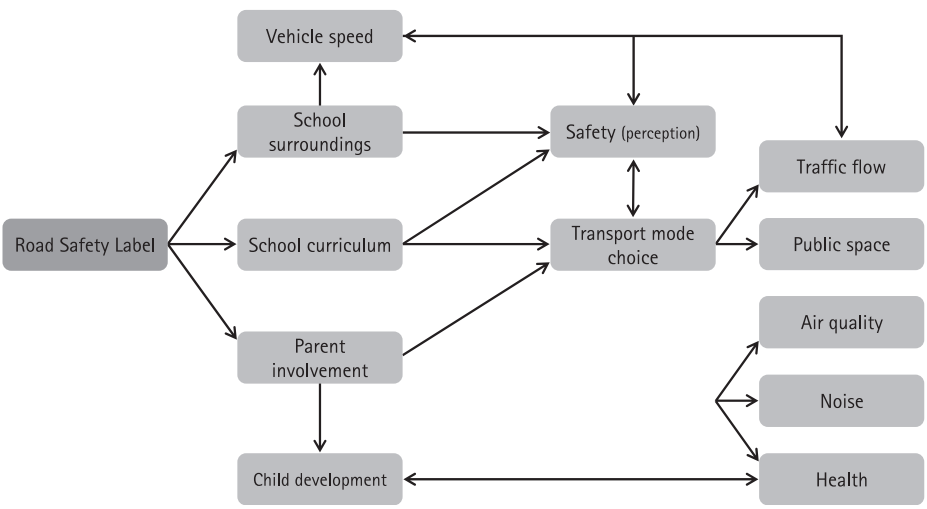


Figure 7–2: Cause-and-effect chain for the Utrecht Road Safety Label.

The inclusion of traffic education in the schools' curriculum could have an impact on the transport mode choice as more parents might allow their children to go by bike or walk to school. At least in the higher grades it can be expected that this would be more and more influenced by the children themselves. This leads to an increase in safety (and the perception thereof) as well as to changes in traffic flow, the use of public space, air quality, noise and children's health. The above-mentioned could also be stimulated by an (increased) parent involvement. As a positive side-effect parents are encouraged to teach their children to act responsibly on their own, which supports their personal development.

The level of noise, air quality, the use of public space and the children's health are a result of numerous factors. For instance, schools which focus on sports might already have 'healthier' students. As a consequence, these factors and their development were excluded from this analysis. For the same reason, the traffic flow will not be evaluated. Consequently, the evaluation will focus on the impacts on vehicle speeds in the schools' vicinities, safety of the children and their transport mode choice.

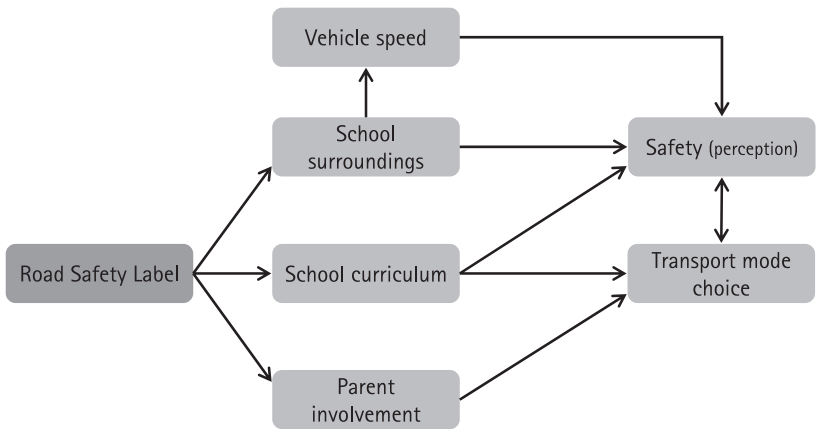


Figure 7–3: Reduced cause-and-effect chain for the Utrecht Road Safety Label.

7.1.5 Indicators and data collection

The cause-and-effect relations shown above and the objectives demonstrate the central aim of this measure: the impact on children's safety in the schools' surroundings. Therefore, the indicator with the highest priority should be the number of registered accidents with children involved in the school areas. However, there is a serious problem in the Netherlands regarding the traffic accidents data: the data from 2010 is not comparable to the former years due to a different registration system. Above that,

accident statistic from the police is not divided into different city areas. It would be very time consuming to re-analyse the accident data and to allocate it to the different regions. Consequently, it was decided to measure the safety impact of the measure through the perception of safety by parents and the average speed of vehicles that pass the school.

No.	Impact	Indicator	Description	Source
1	Transport mode choice	Modal Split	Average modal split for home-school journeys	Questionnaire among parents
2	Safety	Perception of safety	Satisfaction with perceived road safety in the school surrounding area	Questionnaire among parents
3		Vehicle speed	Average vehicle speed in the school surrounding area	Measurement by the police

Table 7–1: Impact evaluation indicators for the Road Safety Label evaluation.

The indicators 1 and 2 were measured with the same questionnaire. Therefore the same written questionnaire was handed out to parents at the control sites and at the measure sites. Also, the same parents are asked in the after data collection. This ensures the comparability between their answers. For the measure implementation evaluation, every 10th school that agreed to the participation to the program (in terms of signing the necessary contracts) received a questionnaire both before and after the implementation of the uniform school surroundings. With this method it was ensured that the data would be sufficient and quasi-random sampling is thus applied. The indicator methodologies are described as follows:

1. *Modal split* – through a questionnaire among parents with children in the 3rd grade (then they are between 7 and 9 years old) they were asked how their children go to school most of the time, given the choice of car, bicycle, public transport (PT), on foot, or other.
2. *Perception of road safety among primary schools* – Through a questionnaire among parents with children in the 3rd grade they were asked how they rate the road safety in the vicinity of the schools given the choice of safe, reasonably safe, unsafe and dangerous.
3. *Vehicle speed* – To measure the speed of vehicles near the schools a radar is placed along the road for one week. These measurements were done before the implementation of the new school surroundings and approximately two weeks afterwards. The locations of the radar were directly before the schools and were the same before and after.

7.1.6 Business-as-usual scenario

The business-as-usual scenario describes the situation in which schools do not participate in the Road Safety Label scheme. In 2009 a sample of three schools that rejected the participation was randomly selected for data collection. In early 2011 they were asked the same questions a second time. In 2009, more than half of the parents with school children in the 3rd grade in the control group rated the road safety to be at least reasonably safe. This could be one of the reasons why they rejected this measure in the first place.

7.1.7 Data analysis and results

By the end of 2012 84 schools worked towards receiving the Road Safety Label, which is 81.5% of the total number of schools, while 42 of them have already received the Road Safety Label (RSL). For this analysis only those questionnaires were included in which the before-and-after measurements were already completed. In 2012 data from 5 different schools is available. The exact sample sizes (n) are presented in the table with the results below.

No.	Indicator	Answer category	Before RSL		After RSL	
			Control site (n=171)	Measure site (n=265)	Control site (n=159)	Measure Site (n=281)
1	Modal split	On foot	24.8	22.8	26.0	21.3
		Bicycle	65.2	66.8	63.6	68.4
		Car	8.5	7.2	8.4	6.3
		PT	1.1	2.3	1.2	2.3
		Other	0.4	0.9	0.8	1.7
2	Perception of safety	Safe	1.5	1.4	1.8	6.9
		Reasonably safe	56.3	49.6	46.5	68.7
		Unsafe	35.0	39.8	43.2	22.2
		Dangerous	7.2	9.2	8.5	2.2
3	Vehicle speed	0-25 km/h	17.6	15.2	17.8	18.5
		26-35 km/h	27.7	27.3	28.8	30.1
		> 36 km/h	57.5	57.5	53.4	51.4

Table 7–2: Comparison of before-and-after questionnaire results in per cent.

The Table 7-2 shows that there were no significant differences between the control site and the measure site before the implementation with respect to the modal split. More than 65% of the children in the 3rd grade go to school by bike, roughly another 25% walk. The average vehicle speeds in the vicinities of the schools are similar. Around 35% of vehicles drive too fast. There are, however, differences in the perception of safety in the two groups. In the control group approximately 58% of the parents rated the road safety to be at least reasonable; this figure is lower in the measure sites (51%). All in all, parents whose schools' are participating in the label scheme rate the safety worse compared to the control site prior to the implementation.

After the implementation the differences between the control and measure site are more obvious. At the measure site more children walk to school. At the same time, roughly the same percentage of children no longer ride their bicycle. Other than this, there are only minor changes in the use of the other transport modes. These changes are even less when they are compared to the before measurement. Moreover, there is no noteworthy difference in the average vehicles speed. On the other hand, the perception of safety differs significantly – confirmed with a X^2-test for this categorical data. While fewer than 50% of the parents in the control group believe that the school surroundings are at least reasonably safe, this percentage rises to over 75% in the measure site.

Interestingly, the data also shows a difference in the perception of safety from the before-and-after measurement in the control group. While 56.3% of the parents rated the schools environment to be reasonably safe in 2009, after the Road Safety Label has been introduced in other schools, this number decreased by 10 percentage points.

7.1.8 Impact result interpretation

As could be expected, parents in the measure site rate the safety in the schools surroundings worst than the control site before implementation. The participation in the Road Safety Label consists in a series of actions involving the schools teachers and staff, parents and the children. They would have only committed to this if they thought it would be necessary. The control sites usually rejected the initiative with the reason that the additional burden on the children's curriculum is disproportionate to the schools safety problems. The results however show the impact on parents' perception of safety at the schools. With respect to the before measurement and the control site parents rated the safety higher after their school was awarded with the Road Safety Label. This is a good result, one which could be expected considering the major changes in the schools surroundings and the children's curriculum. Nonetheless, it is not possible to say how the control site might have been influenced by the intensive media coverage of the Road Safety Label implementation.

The changes in the modal split – more children walking to school in the measure site – could be an outcome of this increase in perceived safety. However, at the same time, roughly the same percentage of children no longer rides their bicycle and in the net sum, the percentage of children who use active modes for their home-school journey remains the same. In addition, it has to be considered that the parents were not asked during the same season (i.e. summer or winter). The timing of the questionnaire solely depended on the time of signing of agreement to the program and the awarding of the Road Safety Label. Hence, the data cannot provide more information and a discussion on the reasons for this shift would be purely speculative.

The evaluation results show no significant decrease in the average vehicles speed in the schools' surroundings, neither in relation to the control site nor the before measurement. It was hoped that the visibility of schools and the signage in their environment would remind drivers that children are in close proximity and that they would slow down – at least below the legal threshold of 30 km/h. The results can only strengthen the argument that drivers, in general, do speed regardless of their surroundings. It appears to be a problem with the enforcement which is generally low in minor roads on which schools are usually located. On the other hand, it could well be that the drivers are aware of the children but because of the increased visibility they feel that they can see children approaching and thus do not need to slow down.

7.1.9 Process evaluation

The process of the Road Safety Label implementation has been evaluated with individual interviews. The evaluators met with teachers from the participating schools for 30 to 45 minutes for a semi-structured interview the questions of which were targeted at (students, parents, see annex). All interviews were recorded and written-down afterwards. The teachers' statements were organised in categories later and the following barriers and drivers were deduced from this:

Barrier: difficulty to get parents involved – It is difficult to get the parents involved, which is one of the requirements of the Road Safety Label. This is especially true at schools where the parents have a lower level of education. There are many reservations among the parents with regards to the extra workload for the children.

Barrier: change of local framework – In some of the cases the schools are in a situation in which change is envisaged, or has just taken place. This could be change of the school building, or in the educational programme. This means that the school is in need for temporary solutions, which is sometimes difficult to manage. Consequently delays occur, since decisions making takes longer. Questions arise such as: Does it still make sense to install the new school surroundings if the school will soon move to another building?

Barrier: change in school management – The time period between the school starting the process to obtain the label and the moment of really getting the label is in some cases around 3 years. In the same time it was found that in about 35% of the schools a change of management takes place within such a period. A change of management means in some cases that the process and advantages of safety label have to be explained again. More generally, this means that often the process with the respective school is temporarily delayed, or even stopped.

Driver: good collaboration among stakeholders – There is a good and continuing collaboration between the measure management team (department of traffic and transport), the department of education, and the subcontractor. This is considered as one of the main drivers for a successful implementation of the measure.

Driver: political and media attention – There was a lot of political and public attention for the issue of road safety around schools, which assured the availability of the necessary budget and increased the willingness of schools to implement the Road Safety Label.

Driver: manager comes 'from the inside' – The subcontractor who is managing the educational support service is a former local school director and thus well experienced with the needs and environment of the schools in Utrecht. This had a positive influence on the cooperation of the team implementing the measure.

7.1.10 Result exploitation

By the end of 2012 84 schools worked towards receiving the Road Safety Label, while 42 of them have already received the Road Safety Label. It is the cities intention to involve all Utrecht schools in this initiative. The involvement of more schools following the pilot implementation can be considered as up-scaling of the measure.

Transferability, however, could be an interesting concept to exploit, meaning that every school in the Netherlands would have the same (or at least similar) surroundings and all schools would have traffic safety in their curriculum. This is reasonable if similar activities have not yet taken place in other cities in the country. Nonetheless, the expectations should not be too high. This evaluation was not able to demonstrate the measure's actual impact on children's safety. Merely the safety as it is perceived by parents with children in the 3rd grade increased. Adding to this, an overall traffic calming in the vicinity of the schools could not be verified.

7.1.11 Appraisal of the evaluation approach

For this evaluation a control group design was chosen to follow the highest standards for the evaluation of the Road Safety Label. Schools freely participate in this programme, while the chosen control group chosen did not want to. Since parents in the measure group rate the safety higher after its implementation, the question is, is it because of the measure itself, or are they more aware of the subject even before the measure began? Maybe they already were before the measure began and this is why they chose to participate in the first place? All parents which participated in the survey were – prior to this – approached by the educational support service to join the initiative. As such, there has already been a focus on road safety before the first data collection. This kind of response bias could have only been prevented if the parents were asked to fill in the questionnaire before they were briefed about the Road Safety Label.

With regards to the objectives of this measure two peculiarities have to be pointed out. The increased satisfaction with safety has only been surveyed among parents; teachers and students – as was stated in the objectives – were not considered. This blur in the evaluation has been accepted from the beginning to keep the number of questionnaires at a manageable level. The teachers' perspective was included in the process evaluation. Another shortcoming in this evaluation is the lack of accident data. As argued above, the statistics from the police only exists as an average over the entire city of Utrecht. However, this initiative might have encouraged those responsible to change the general reporting scheme in the future.

In addition, the information on the reasons why schools choose not to participate in the Road Safety Label scheme could be evaluated. Some data is already available in the educational support service. Eventually, this will become necessary if the city of Utrecht wants to involve all 103 primary schools in this initiative.

7.1.12 Conclusions

Children are a particularly vulnerable group in road traffic and need to learn how to act safely on their own. Before 2008 road traffic education was given low priority in Utrecht because curriculums are typically already chock-a-block with other important subjects. The Road Safety Label offers schools the opportunity to set up and execute a structured traffic education plan at their own pace. In exchange, the schools get money and support for traffic lessons and small infrastructure improvements in the direct surroundings of the schools.

By the end of 2012 84 schools worked towards receiving the Road Safety Label, while half of them have already received the Road Safety Label. The evaluation with

a control group design was not able to demonstrate the actual impact on children's safety due to lack of proper data. However, the safety as it is perceived by parents with children in the 3rd grade increased. Also, an overall traffic calming in the vicinity of the schools could not be verified. This demonstrates the importance of other needed measures which directly counter this problem. But it also means that the Road Safety Label should be re-evaluated when proper accident data is available. Until then it 'only' is a well-accepted quality mark for primary schools that put effort into improving the road safety around the school and incorporating traffic education into their curriculum.

7.1.13 Annex

Interview guideline Road Safety Label (version for teachers)*
Introduction (aim of the interviews, information about recording and data privacy)

1. What was your first impression of the idea of implementing a Road Safety Label for school surroundings?
    - **flashback, attitude, acceptance

2. What do you think about the introduced Road Safety Label today?
    - acceptance, advantages and disadvantages, its impact

3. How have the parents got informed about the planned Road Safety Label?
    - information process, involvement

4. How were their reactions?
    - feedback/acceptance, advantages and disadvantages, engagement

5. How did the children deal with the changes concerning the Road Safety Label?
    - involvement, transport education, lay-out of school surrounding
    - acceptance, traffic behaviour

6. How do you see your own role in the Road Safety Label?

7. What is your opinion on the cooperation among stakeholders?
    - communication, interests, conflicts

8. How would you improve the Road Safety Label?
    - wishes, proposals

End of the interview (thanking for the interview, interviewee will receive the transcript of this discussion to confirm their statements, information will be handled confidentially)

*guideline – to be handled flexible, depended on the process of conversation

**notes for more detailed questions/enquiring to encourage talking of the interviewee

7.2 Utrecht Cargohopper

7.2.1 Problem description and measure context

Urbanisation is one of the fastest growing global mega trends in the 21st Century. According to Eurostat over 40% of the population of the EU-27 lived in cities in 2011. Naturally, all these people must logically be supplied with goods. In all countries across Europe the growth of domestic freight is tremendous and only slowed down due to the economic and financial crisis in recent years. The combination of these two trends results in increasing road freight transport within cities. This causes three major problems:

- A progressive deterioration of air quality due to exhaust emissions such as CO_2, NO_X compounds or particulate matter (PM2.5 and PM10).
- A strain on the roads and infrastructure through a variety of vehicles, especially in the morning and evening rush hours, with the result of traffic congestion.
- Increased nuisance to local residents from noise caused by heavy trucks.

As a reaction to these developments, various measures such as truck bans or low-emission zones have been established in many places. The desired effects have, however, been very limited.

The city of Utrecht, as many other cities, faces these problems due to the increased freight traffic on the streets. This problem is enhanced through a growing population within the past years which goes hand-in-hand with an increase in the number of shops and stores which need to be supplied. In addition, the city has a dense, historic city centre which imposes special transport-related restrictions and needs to be preserved. As a consequence, in July 2007, the city introduced a low-emission zone that limited access for trucks with polluting engines. The objective had been to reduce the PM10 level by 2.6 µg/m3 (in the city centre) and by 1.1 µg/m3 (in the entire city) to get below the EU-threshold of 40 µg/m3. An evaluation of the short-term change after the introduction of environmental zone showed that major changes had only been achieved in the displacement of vehicles of class Euro-0 and Euro-1. However, the regulation was limited to vehicles over 3.5 tonnes. An unintended side effect of the introduction was thus a replacement in favour of lighter, but still relatively 'dirty' trucks and vans. Consequently, the results were not satisfying.

7.2.2 Objectives

By establishing a low-emission zone in July 2007, Utrecht tried to limit the access to the centre for trucks with 'dirty' engines with ambiguous results. But the city of Utrecht continues to aim at improving air quality by encouraging and supporting a modal shift towards more sustainable modes of freight traffic. Thus *the high level objective* of this measure was:

- Improvement of air quality through the use of more energy efficient freight distribution.

In order to further reduce the PM10 and NO_X emissions from road freight traffic they made a plan to improve accessibility of the Utrecht city centre for cleaner and quieter freight transport. The main aim of this measure was to give benefits to transport companies that use 'super clean' (cleaner than the EURO5/EEV/EEV+ norm) vehicles to stimulate purchase of these types of vehicles. Thereby measure *measure-specific objective* was:

- Reduction of the CO_2, PM10 and NO_X emissions from road freight traffic.

7.2.3 Preparation and implementation stages

Stage 1: Measure preparation (2008-2009)

As part of the CIVITAS MIMOSA project, the city of Utrecht was able to direct funds to the research and implementation of sustainable freight transport solutions. In a round table discussion with stakeholders (city representatives, transport companies and shop owners) the idea of an electric-powered mini-train was born.

The private transport company Hopper transport[1] was responsible for covering the production costs and to convince shopkeepers to use the service of this new vehicle (the so-called Cargohopper) to be delivered with goods. In return, the city guaranteed the Cargohopper the right to deliver goods outside of the delivery time-windows as well as driving on bus-lanes, bicycle paths and through pedestrian areas. In this respect, the Cargohopper has many advantages to the conventional delivery van. In addition, Hopper has also been supported by finding a suitable and affordable transfer location near the city centre.

1 Name changed for privacy reasons.

Figure 7–4: Picture of the Utrecht Cargohopper.

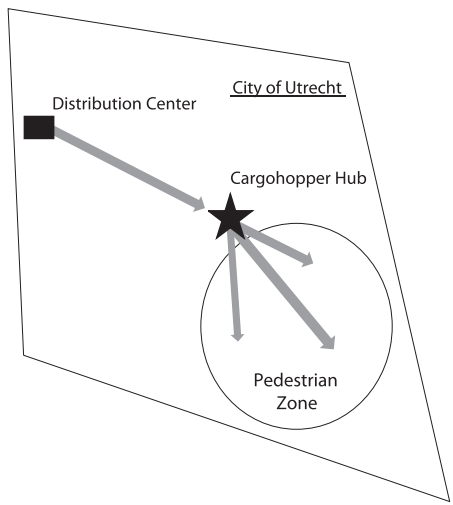


Figure 7–5: Schematic of new logistic concept with the Cargohopper.

Stage 2: Implementation of Cargohopper (April 2009)

The Cargohopper came officially into service. This vehicle is able to tow up to three trailers in a line by means with an electric engine. Its maximum speed is 20 kilometres per hour which is sufficient as it is exclusively driving in the inner city of Utrecht. The trailers are steered on both axles which gives the vehicle good manoeuvrability.

With its 1.25 metres, the Cargohopper is smaller than normal transport vehicles, it does not block other traffic in the city centre. The containers are separate boxes

which can be put on and taken off the undercarriages by means of a forklift. Eight of those boxes fit on a Euro trailer of 13.6 meters. The boxes are preloaded outside the city in the Cargohopper's distribution centre and towed to the boarder of the inner city by means of a regular truck. There is another transhipment point (Cargohopper Hub) where the boxes are put on the Cargohopper and rolled into the pedestrian zone: from there the deliveries to the shops start (see Figure 7–5 below). Once emptied, it collects dry carton, paper and empty packaging from shops for recycling so it never runs empty. The reloading with goods for the next round only takes 10 minutes.

Stage 3: Solar panels for the Cargohopper (August 2009)

Solar panels were placed on the roof of the Cargohopper. This allows the vehicle to drive eight to nine months per year on solar power. With this the Cargohopper switched to self-produced solar power, making it a CO_2 neutral form of freight transport. The other months it drives on green electricity.

7.2.4 Cause-and-effect relationships

The following cause-and-effect chain was designed representing the potential effects of the Cargohopper. The Cargohopper enables further bundling of trips, as Hopper sends a light truck to a loading point, 300 meters outside the city centre where goods are being transferred onto the Cargohopper. From there, the Cargohopper drives into the city center and makes deliveries. With 'conventional' delivery more light vehicles (e.g. Sprinters) are driving directly from the distribution center to the city centre. Thus, the Cargohopper contributes to the decrease of inner city freight transports.

In combination with its fuel savings due to the electric engines, it can have a favorable impact on the traffic flow and the noise level as well as safety and air quality. In

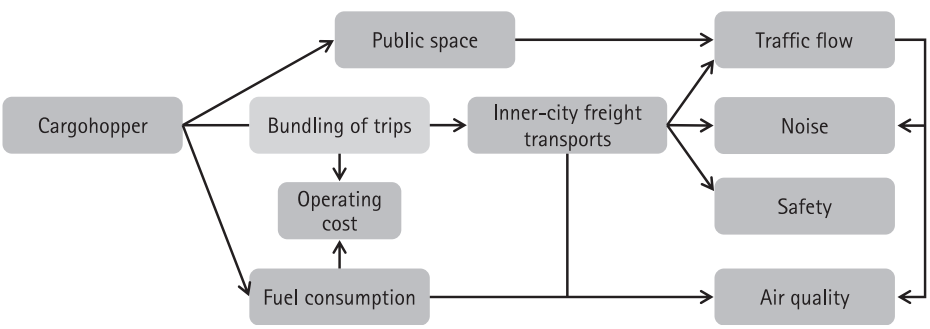


Figure 7–6: Cause-and-effect chain for the Utrecht Cargohopper.

addition, the vehicle is better suited to Utrecht's narrow, cobbled streets than conventional light vans, so the city's roads are no longer blocked by oversized delivery vehicles. This increase in traffic flow further makes this vehicle favourable for the environment.

The level of noise, safety and traffic flow in the city centre are a result of numerous factors, including the presence or absence of events, conventions or festivals for instance. It is also highly fluctuating during the time of the year. During holiday and vacation time for instance, the city of Utrecht attracts tourists from all over Europe and the world. As a result, the impact of the Cargohopper and the development of these factors were excluded from this analysis.

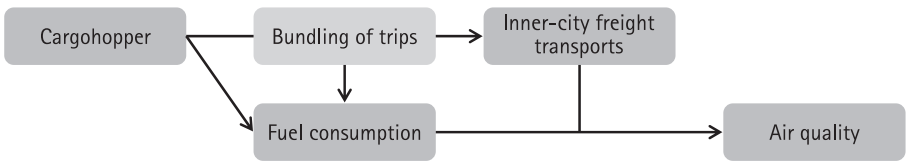


Figure 7–7: Reduced cause-and-effect chain for the Utrecht Cargohopper.

It was decided to focus the impact evaluation on the air quality impact resulting from both the direct fuel savings due to the electric engine as well as the bundling of trips. The impact on air quality is further specified through the identification of transport-related emission savings such as CO_2, NO_X and PM.

7.2.5 Indicators and data collection

This measure aimed at increasing the use of more energy efficient freight distribution and decreasing the resulting CO_2, NO_X and PM10 emissions in the city centre. However, emissions which were measured by fixed stations in the city centre could show a reduction which also includes the effects of other transport-related measures in the city, or might not show a reduction, as these measurements are highly dependent on the surroundings (i.e. weather conditions). Thus, the emissions saved through the use of the Cargohopper and the bundling of trips involved will be estimated based on the saved kilometres compared to the business-as-usual scenario 'no Cargohopper'. These saved kilometres are estimated for the years 2009 to 2012 based on the actual transported volumes by the Cargohopper. Table 7–3 summarises the impact evaluation indicators.

Transported volumes – Transported volume by the Cargohopper on average in cubic meters per day (m3/day). The transport company logged the volume transported by the Cargohopper and reports them in week 20 and week 40 from 2009 to 2012.

No.	Impact	Indicator	Description	Source
1	Inner-city freight transport	Transported volume	Transported volume by the Cargohopper	Hopper transport
	Air quality	CO_2 emissions	Emission based on average roundtrip distance	Calculated
3		NO_x emissions	Emission based on average roundtrip distance	Calculated
4		PM emission	Emission based on average roundtrip distance	Calculated

Table 7–3: Impact evaluation indicators for the Cargohopper evaluation.

PM10, NO_X, CO_2 emissions – The calculation of emission relies on the average roundtrip distance. These emissions are calculated for the business-as-usual scenario and compared to the Cargohopper scenario.

7.2.6 Business-as-usual scenario

The business-as-usual scenario (BaU) describes a situation in which all goods transport towards the inner city is done with Mercedes Sprinters. In 2008 – before the implementation of the Cargohopper – Hopper used 3 delivery vans per day to deliver goods to clients in the city center. It is assumed that without the Cargohopper this number would be proportional to the transported volumes. This would result in 3.5 Sprinters per day on average in the year 2009 and 5 Sprinters per day in the years 2010 to 2012. The average roundtrip distance (from the distribution center to the clients) is 25 kilometers and it is assumed that this figure does not change with the increased volumes. The number of customers has increased, but since they are all within the dense pedestrian zone, the mileage of the transporter would not need to increase. For the calculation of emission, the following conversion factors have been used:

Emission	conversion g/km (based on Sprinter-type)
CO_2	301
NO_X	0.645
PM	0.065

Table 7–4: Conversion factors for the emission calculation in the BaU scenario.

Assuming that the Sprinters operate 6 days a week during 51 weeks per year, they emit 11 tonnes of CO_2 in 2009 and 16 tonnes in 2011 (for more results see next section).

7.2.7 Data analysis and results

The average daily transported volumes with the Cargohopper are presented in the figure below and show a steady but slowing increase. It shows an increase of goods deliveries of 150% from 2009 to 2012. According to Hopper Transport the Cargohopper made 32 stops per day on average in 2012, serving well over 100 different customers per week.

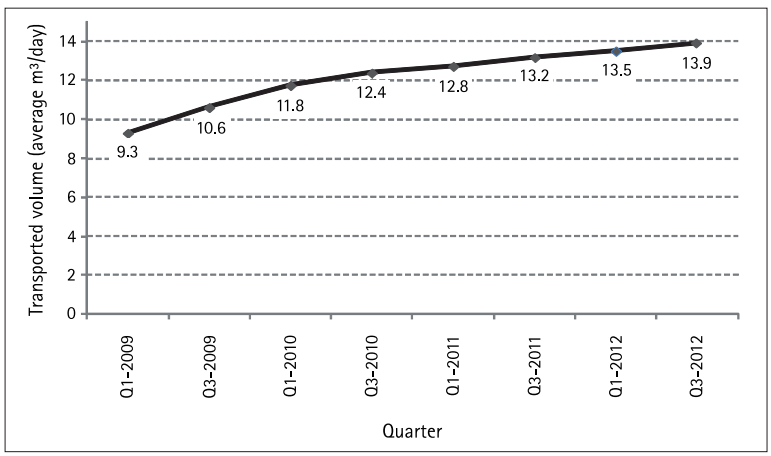


Figure 7–8: Transported volume with the Cargohopper (2009-2012).

With the Cargohopper the goods are first transported from the distribution center to the Cargohopper hub. This is done once a day with a light van and its average roundtrip distance equals 16 km. For the calculation of emission from the light van, the following conversion factors have been used:

Emission	conversion g/km (based on Actros-type)
CO_2	511
NO_X	2.95
PM	0.18

Table 7–5: Conversion factors for the emission calculation with the Cargohopper.

The impact in terms of a reduction of emissions due to the Cargohopper (CH) can be determined through the comparison with the business-as-usual (BaU) case. The Table 7-6 summarises the impact (CH-BaU) for the years 2009 to 2012 under the assumption that deliveries are made on 6 days per week during 51 weeks per year.

		2009			2010		
		BaU	CH	CH-BaU	BaU	CH	CH-BaU
Number of Sprinters		3,5			5		
Number of light vans			1			1	
Average roundtrip distance (km)		25	16		25	16	
Emission	CO_2	8.059	2.502	-5.557	11.513	2.502	-9.011
	NO_X	0.017	0.014	-0.003	0.025	0.014	-0.010
	PM	0.002	0.001	-0.001	0.002	0.001	-0.002

		2011			2012			2009–2012
		BaU	CH	CH-BaU	BaU	CH	CH-BaU	CH-BaU
Number of Sprinters		5			5			
Number of light vans			1			1		
Average roundtrip distance (km)		25	16		25	16		
Emission	CO_2	11.513	2.502	-9.011	11.513	2.502	-9.011	-32.592
	NO_X	0.025	0.014	-0.010	0.025	0.014	-0.010	-0.034
	PM	0.002	0.001	-0.002	0.002	0.001	-0.002	-0.006

Tables 7–6: Comparison of yearly emission.

The table concludes that through the implementation of the Cargohopper, 32.6 tonnes of CO_2, 34 kilograms of NO_X and 6 kilograms of PM were saved from 2009 to 2012.

7.2.8 Impact result interpretation

The volumes transported with the Cargohopper could indicate its popularity among the shop-owners in Utrecht. While the freight transport volume of light duty vehicles in the inner city of Utrecht decreased from 2009 to 2012, the volumes transported by the Cargohopper increased steadily. Thus, while the overall decrease could be an indicator of the financial crisis it seems to have passed the delivery with the Cargohopper. However, many of the new customers of Hopper transport were shops that are located in

extremely narrow streets and which cannot be reached with a conventional light truck. Thus, an increase in the customer base could be expected as their delivery is eased. It might have been induced by the Cargohopper. With regards to the pricing schemes, there has been no change for the customers in these four years.

The results from the emission calculation and comparison show the positive impact of the Cargohopper goods distribution on the environment. However, the magnitude of these emission savings only presents a rough estimated due to the limitations in the business-as-usual scenario. Hence it can only be concluded that there is a positive impact, but whether – in reality – the savings in CO2 is 30 tonnes (as stated by Hopper transport) or 32.6 tonnes for four years (as concluded in this report) cannot be determined.

7.2.9 Process evaluation

The process of the implementation has been evaluated using two different methods. First, in 2010, a questionnaire was distributed to different stakeholders to ask for their opinion on barriers and drivers to the implementation as well as recommendations for further exploitation of the vehicle. Second, a Learning History Workshop was conducted one year later. The questionnaires – which were also used for the considerations of the results exploitation – revealed the following barriers and drivers:

Driver: the vehicle is not the main factor, the logistical solution is – Some of the participating shops agreed to the mini-train distribution centre as it ensured that there was only one delivery per day and not various deliveries at different times. This ensured that they could use their time primarily to sell goods rather than wasting time receiving multiple deliveries.

Driver: dedicated private partner combined with fast legislation – the local transport company involved is run by people who live in the city and know the city well. As such, they are aware of its problems and had an intrinsic motivation to come up with a solution. The Cargohopper was the idea of Hopper transport and the developments and implementation has been a financial risk for the company. Their commitment combined with short approval periods from the legislation's side was crucial during the first years.

Barrier: difficult maintenance – The electric van can not be repaired in a normal garage. It has to be brought to the importer of the vans who has to order spare parts in Italy and thus maintenance takes some extra days. Learning how to handle the Cargohopper was an experimental task as no manual is available.

Barrier/Driver: client base – It is better to have a company with an already existing client base. It is not recommended for a new company to be built on the hopes of finding new customers as – in competition to other companies – this could prove to

be quite difficult. The best way for a city to have this sort of work done is through an established transport company. The company then needs to convince others.

A Learning History workshop was conducted in mid-2011. Together with one city representative and the managing director of Hopper transport, several shop owners and one affiliate from another logistic company was present. They acknowledged and discussed some of the barriers and drivers which were taken from the former questionnaire. However, they had two other insights and came up with their solutions:

Slow speed of the Cargohopper – the participants of the workshop agreed that the delivery speed of the Cargohopper was still relatively slow because it has to use the same streets as conventional vehicles. Increasing the speed of the delivery would, in addition, allow more deliveries. The solution proposed was to allow the vehicles to use bus lanes. After the meeting the city amended the legislation to make it possible for the Cargohopper to drive on bus lanes and thus reduce delivery times.

Delivery at night times difficult – it has been a problem to deliver the shops at times of the day where no one is available to receive the goods. This is typically in the times between 6 and 9 am in the morning. This limited the distribution of goods and did not make full use of the benefits which are given to the Cargohopper and which Hopper would like to exploit. The shopkeepers, on the other hand, argue that they cannot ask their personnel to be in the shop from 6 a.m. onwards, receive the delivery and then wait until the shop opens at 10 a.m. In the workshop all stakeholders have discussed how this could be solved and they came up with the solution that shop owners can give an extra key to their stores or restaurants to the drivers of the Cargohopper who will then deliver the goods.

7.2.10 Result exploitation

With the Cargohopper up-scaling could be the introduction of more vehicles or a higher frequency of deliveries or more shops having their goods delivered by the Cargohopper. One of the main criteria for assessing the projects potential for up-scaling is certainly the overall capacity of the service, but also the wider acceptance of the measure in the target group is important. It is, however, difficult to say from the impact or process evaluation how this will turn out.

For transferability purposes it is strongly recommended to have a close look at the legal framework before implementation. In many cities 'cleaner than average' vehicles are already exempt from restrictions or they can use reserved lanes. In these circumstances, there is no extra incentive for a company to buy an electric mini-van. Moreover, the size and build up of the Cargohopper were specifically tailored to match the Utrecht setup with its narrow and cobbled streets. There are parts of the city which can

only be reached by the small Cargohopper. This might not be the case for other cities. If it isn't, some of its potential for the logistic companies (in terms of more customers) is gone.

7.2.11 Appraisal of the evaluation approach

The impact evaluation is based on assumptions regarding the kind of vehicles used in the business-as-usual scenario and their average round trip distances. Also, all emission factors were averaged. As such, it is difficult to say whether the results are close to the real emission saving or an under-/overstatement. In the beginning, it might have been possible to follow a 'control site' approach. This would have been possible if the Cargohopper distribution from Hopper transport had been directly compared to the goods distribution of another private company which is operating in the city and has a similar client distribution. This approach, would be very difficult. Private operators are very reluctant to provide any information on top of which they are obliged to. Nonetheless, the concept itself and the evaluation showed that there was a positive effect for everybody involved. The answer to the question about its magnitude is only estimated. Therefore, the results of this study on the Cargohopper should not be compared to other technologies.

The process evaluation, in its two parts, has given good results. Especially the Learning History Workshop among the stakeholders has fueled the positive spirit for the Cargohopper. However, in further evaluation steps, and especially if extra rights like the bus lane use are to be expanded, the public transport providers need to be involved since they will be affected by such a shared usage.

7.2.12 Conclusions

The Cargohopper, whose story began in 2008, has several advantages. First, it is better suited to Utrecht's narrow, cobbled streets than conventional 'light vans', so the city's roads are more open to other types of traffic. Second, the Cargohopper is more favorable to the environment by reducing the CO2 emissions when compared to 'conventional' delivery. In addition, the Cargohopper can make quicker deliveries and it can be expected that delivery times for trucks will get even longer in the future than delivery times of the Cargohopper. This is because the Cargohopper can use bus-lanes, bicycle paths and can deliver outside of the delivery time windows, which makes the service less sensitive to an increase in congestion levels. All in all, the Cargohopper is a good example of a public-private partnership providing benefits to all sides.

7.2.13 Annex

Selected answers from the managing director of Hopper transport to the questionnaire used in the process evaluation.

When you think of the entire implementation and operating process, what were the most important drivers of this innovation until today?

"It has been a good collaboration with the city council. We invested in the hardware and the equipment, and worked everything out. But without the city's cooperation, it wouldn't have happened."

What advice would you give other cities that are looking to introduce the Cargohopper?

"Personally I am living in the city of Utrecht. I know the city very well and I know what the problems are. Then it is only a matter of logical thinking to find a solution. Every city has its own problems and character. You have to find the best system/solution to solve these problems. Don't copy Cargohopper 1 to 1 but use the system as a base."

Where do you see your company in 5-10 years?

"You can imagine that we attract a lot of attention. It helps us in our market to find new customers. Nowadays we know a lot about inner city distribution and we are seen as experts. And we are. We want to develop further and in 5-10 years in a lot of cities in Holland you will find a system such as Cargohopper or look-alikes."

7.3 Tallinn Knitting Bus

7.3.1 Problem description and measure context

An economic downturn and then rapid economic growth have imposed large structural changes on the city and its transport system during the past 20 years. The number of private cars has increased rapidly and this has turned the modal share towards private car dominance. The overall quality of public transport has increased during this period, but this has not affected the overall trends in the modal split. The reputation of active transportation modes as everyday alternatives to private car use is low in Tallinn. Public transport is generally considered to be slow, dirty and expensive. Walking and cycling possibilities are not considered possible or reputable alternatives to using private cars. It is common understanding that public transport users are just using it for everyday commuting because they cannot afford private cars. Public surveys confirm this bad perception. In addition to the low popularity of public transport Tallinn has identified

a small modal share of walking and cycling as a problem that needs to be addressed. People are not aware of the advantages of the active transportation modes and they are not regularly and systematically promoted.

The city has already introduced innovative solutions and improved the quality of sustainable transport modes but these measures remain largely unnoticed by the general public. The city has realised the need for a marketing strategy to promote its public transport service and to inform citizens of mobility options in the city. One of the main tasks for Tallinn in this measure has been to draw up a communication plan that includes specifications for a media campaign. Target groups have been defined (schools children, commuters, etc.) and practical interventions have been specified such as mobility plans, education and promotional activities.

7.3.2 Objectives

The city aims to deal with the identified problem by increasing the modal split towards more sustainable modes such as cycling, walking and public transport. The infrastructure and public transport schemes in Tallinn have been improved to accommodate these modes but this hasn't lead to the desired changes in modal split. Thus the *high level objectives* for this measure are:

- Improvement of the quality of life in urban areas and
- Reduction of transport related pollution.

The Knitting Bus is specifically intended to increase the modal split towards public transport use. Hence, this measure has the following *measure-specific objectives*:

- Promoting the attractive and high quality public transport service.
- Increasing the share of public transport in the modal split at the expense of car traffic.
- Raising the satisfaction with the public transport service and improving the overall image of public transport in the urban area.

The Knitting Bus campaign was part of an integrated awareness raising strategy for public transport which pursued the goal to increase the modal share of public transport and active transport modes by 10%. This evaluation will only cover the Knitting Bus contribution to this broader objective.

7.3.3 Preparation and implementation stages

Step 1: Identifying target groups

To identify target groups, surveys were conducted during the Smart Traveller Day in Tallinn in May 2010 by a subcontractor. This was an open air event held on the main square to raise awareness of mobility management. There was a special focus on traffic safety addressing all age groups involved in traffic (from kindergarten children to elderly people). For instance, video clips relating to traffic safety, how to behave correctly in traffic, were shown on a huge LED display in the city centre. Surveys were carried out to measure peoples' awareness of sustainable mobility issues, to gather suggestions for improvement and to measure current travel behaviour and willingness to change. This was done by handing out paper questionnaires to citizens at the event.

The results of the survey revealed a perception of passengers' bad behaviour across all transport modes, bad driver- and bad public transport passenger behaviour. This perception of bad behaviour on public transport was reflected in low satisfaction with cleanliness and comfort of public transport, especially in trolley buses and trams.

Many of Tallinn's measures focus on public transport such as the implementation of bus lanes, central control system, integrated ticketing, and eco driving education for bus drivers and they have set objectives to improve the quality and image of public transport and raise satisfaction. These 'hard' measures can be supplemented with 'soft' communication campaigns to encourage a change of behaviour. The survey findings suggested that communication campaign(s) aimed at improving travellers' behaviour should be the first step to improving the image of public transport in Tallinn.

Analyses of the surveys revealed a segment of respondents who indicated that they were likely to change to more sustainable modes of travel. Further investigation of the data revealed that individuals of this segment see themselves as creative and fashionable people, interested in arts and culture and new technology. They are less materialistic, less conservative and more likely to take risks (than those who say they will not change their travel behaviour).

Step 2: Selecting an appropriate measure for the target group

As Tallinn was European Capital of Culture in 2011, it was decided to take advantage of the many cultural events that were being held. The Tallinn Bus Company investigated the idea of 'Knitting Graffiti' or 'Yarn Bombing' which originated in the USA in 2006 as a mean of 'softening' and brightening up the urban environment. Knitting as a craft and a hobby is seeing a rise in popularity. The increase of the number of wool shops, incorporating cafes and classes are demonstrating this development. Knitting is no longer a solitary activity; it is now engaging, interactive and communal. The internet helped fuelling the boom in knitting providing a source of supplies and online knitting

Figure 7–9: Tallinn Knitting Bus from the outside and inside.

communities share, blog and showcases their work. Luxury yarns are used and works of art are produced.

It was therefore believed that the Knitting Bus would yield very positive responses from the target group.

Step 3: Setting a Baseline

Prior to the implementation of the campaign, surveys were conducted with passengers on the bus to measure their satisfaction with the service quality, covering issues such as cleanliness, comfort and safety (see questionnaire in the appendix to this report).

Step 4: Implementation

Volunteers were recruited among the Tallinn population to knit and the seats and hand rails were wrapped in knitting in one bus. The outside of the bus was covered in vinyl photos of knitting and the campaign was launched in June 2011. It was operational with the knitting inside until the end of the summer, because the knitting work wore out. However, the bus still has the stickers on the outside.

Step 5: After testing

A follow up survey was conducted during the second week of the knitting installation and the results were compared to the before survey.

7.3.4 Cause-and-effect relationships

The following full cause-and-effect chain was designed representing the complete potential effects of the campaign.

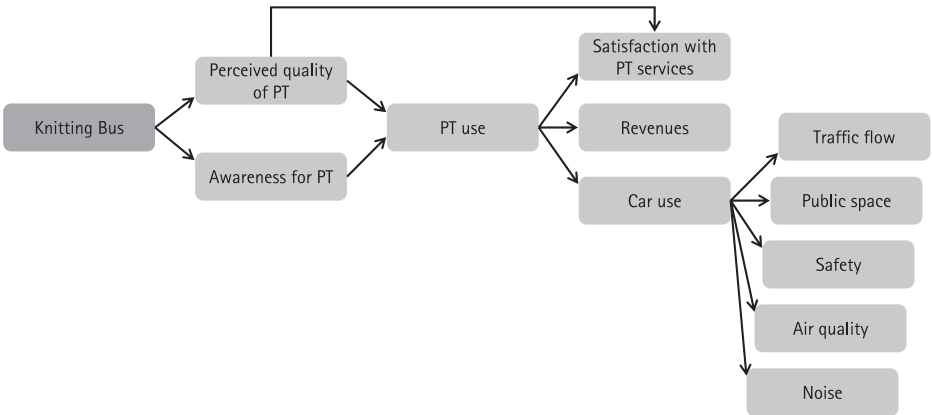


Figure 7–10: Complete cause-effect chain for Knitting Bus campaign.

The Knitting Bus is a marketing campaign that imposes costs on the municipality and aims at increasing awareness of public transport. Increased awareness hereby aims to increase the use of and satisfaction with public transport (PT). It is expected that increased satisfaction with public transport services will lead to even more awareness and as a result to an even higher public transport use. This is a self-enforcing process. Besides this, increased public transport use implies additional revenues for public transport companies and a reduction of car use. The latter has positive effects on traffic flow, frees public space and results in health benefits such as increased safety, air quality and noise reductions in the urban environment.

Although there are a large number of potential effects of the Knitting Bus campaign, evaluation is only focused on measuring the increased satisfaction and awareness for public transport. Increased awareness and satisfaction can be measured directly as a result of the knitting bus campaign, while other indicators represent the result of the overall implementation of hard and soft measures. The effects are therefore difficult to assess as a direct result of the knitting bus campaign. It may be that the measure's time frame is too short and the implementation size is too small to collect reliable data on secondary effects. This also includes the evaluation of increased use of public transport. The cause effect chain is therefore reduced to the following.

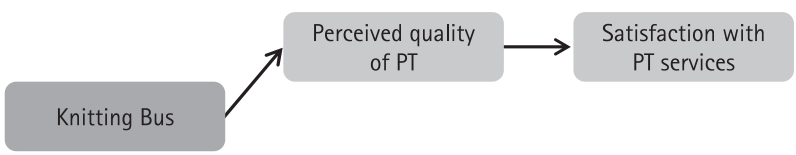


Figure 7–11: Reduced cause-effect chain for Knitting Bus campaign.

7.3.5 Indicators and data collection

The reduced cause-and-effect chain illustrates that in the evaluation only the primary impacts of the measure which lead to a more positive perception of the quality of service are considered. The impacts were surveyed on the bus before and after the Knitting Bus campaign. For this survey a questionnaire was used onboard the respective bus line in which the users were asked several questions about their perception of comfort and the overall image of public transport. The same questions were asked before and after the campaign. No control groups were selected which made the survey a time series analysis with different before-and-after groups. The before-sample was 408 passengers and the after-sample 405 passengers with two weeks interval between surveys. The first questionnaire was given out on 7 July 2011 and the second on 21 July 2011. The survey was based on a self-filled questionnaire which was given to all passengers on the selected trips with 6 fields about the passenger and 20 fields for rating aspects regarding this particular bus and bus service in Tallinn in general. The questionnaire is presented in the annex.

7.3.6 Data analysis and results

As might be expected, the majority of people taking part in the survey were on their way to or from work or school. In the after-questionnaire the percentage of passengers using the service to pay a visit equalled those of the commuters. In both samples the share of male (35%) and female (65%) users remained stable. In more than 80% of the cases the customers used the bus service at least 3 times per week (for more details see annex of this report).

The table 7-7 shows the before-and-after data which were collected from the questionnaires. The questions were asked on a five-level Likert scale and coded from 1 to 5, with 1 representing "very dissatisfied" to 5 representing "very satisfied" (see also questionnaire in the annex). Hence an average value of 4 means that the passengers are – on average – satisfied with the respected component of the bus service.

If applicable, two sample location test (t-test) of the null hypothesis – the means of two distributions are equal – was conducted. The respected P-Value is given in the last column.

The number of respondents differs per indicator and per survey. For every indicator a number of respondents answered "don't know". These respondents were not included in the analysis. The before-and-after data illustrates that almost all indicators are rated higher after the implementation of the campaign. However, only the differences for three indicators ("Ventilation", "Seat comfort" and "Overall satisfaction with the bus")

No.	Indicator	Average	Average	Difference	ND?	var?	t-sig.
1	Lighting	4,02 (N=334)	4,07 (N=393)	0,05	Y	Y	,658
2	Ventilation	**2,80** (N=388)	**3,11** (N=397)	**0,31**	Y	Y	,002*
3	Seat comfort	**3,76** (N=388)	**4,02** (N=401)	**0,26**	Y	Y	,000**
4	Leg room and space to move around	3,50 (N=378)	3,65 (N=401)	0,15	Y	N	---
5	Cleanliness of the seats	3,46 (N=404)	3,59 (N=399)	0,13	N	N	---
6	Cleanliness of the floors	3,54 (N=406)	3,66 (N=398)	0,12	Y	N	---
7	Cleanliness of the windows	3,58 (N=404)	3,63 (N=399)	0,05	N	N	---
8	Behavior of other passengers	3,63 (N=398)	3,67 (N=398)	0,04	Y	Y	,342
9	Sound	3,58 (N=394)	3,65 (N=382)	0,07	N	N	---
10	Overall satisfac-tion with the bus	**3,88** (N=407)	**4,10** (N=403)	0,22	Y	Y	,004*
11	Price-quality ratio	**3,05** (N=338)	**2,75** (N=378)	-0,30	Y	Y	,000**
12	Bus routes	3,62 (N=380)	3,67 (N=386)	0,05	Y	N	,342
13	Frequency of services	3,18 (N=402)	3,23 (N=401)	0,05	Y	N	,520
14	Bus-stop locations	3,89 (N=401)	3,91 (N=398)	0,02	N	N	---
15	Information on sche-dules (at stops etc.)	3,88 (N=405)	3,91 (N=400)	0,03	N	N	---
16	Adherence to schedules	3,68 (N=401)	3,80 (N=397)	0,12	Y	Y	,071
17	Level of customer service on the bus	3,61 (N=351)	3,70 (N=377)	0,09	Y	Y	,163
18	Adequate se-lection of ticket types available	3,69 (N=315)	3,62 (N=335)	-0,07	Y	N	---
19	Security and safety on buses	3,73 (N=399)	3,76 (N=397)	0,03	Y	Y	,535
20	Overall satisfaction with bus company	3,79 (N=396)	3,84 (N=393)	0,05	Y	N	---

ND: standard normal distribution, tested with Kolmogorov-Smirnov test
Var?: same variance tested with Levene

**significant on 1%-level
*significant on 5% level

Table 7–7: Survey results with before-and-after comparison and t-statistics.

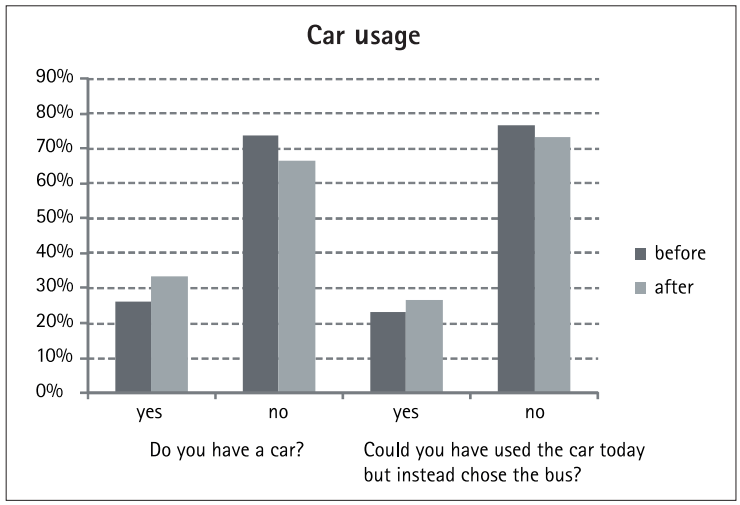


Figure 7–12: Distribution of car users among bus users in the before and after sample.

were significant. The service component "price-quality ratio" was rated significantly lower (on 1%-level) for the knitting bus. For all other variables it cannot be shown that the changes in the responses are not due to chance.

An aspect which did not change during the two weeks between the questioning was the number of customers who could have used a car but used the bus instead. This figure remains stable at roughly 25%. The slight increase as seen in the figure above is not significant and can be explained with the increase of passengers who own a car in general. Nonetheless, it should be remembered that this is not representative for all Tallinn bus customers, as the survey was only conducted on one line.

7.3.7 Impact result interpretation

The measure was embraced in Tallinn. The key result is that promotion campaigns have a clear influence on public transport-users perception of the service. The Knitting Bus campaign shows that passengers care about the environment in busses and it is possible to raise customer satisfaction significantly and change customer mindsets with a relatively inexpensive method.

One important aspect must be considered when interpreting the results. During the two weeks between the before-and-after surveys, the Tallinn city government announced that the price of a monthly ticket will be reduced and the quarterly ticket (the cheapest available season ticket) will be discontinued. The consequence for everyday users is a 30% price increase plus the need to buy a ticket 12 times a year instead of only 4

times. This is likely to affect public opinion on public transport but it is difficult to take into account when evaluating this measure. Still, this information is important because it can be a reason why fewer people were satisfied with the price-quality ratio and the supply of ticket products.

Also, there were significant changes regarding the satisfaction levels of bus ventilation. It is however difficult to defend that the Knitting Bus campaign influences ventilation. An explanation could be that the weather conditions were different on the days the questionnaires were answered. The same applies to leg room and space; it is well possible that a new interior design and higher seat comfort in buses leads to a feeling of a more spacious environment, but on the other hand the number of passengers might just have been lower on the day of the after questionnaire.

7.3.8 Process evaluation

The following measure barriers and drivers were discovered during the evaluation of the implementation process evaluation of the Knitting Bus campaign. They were compiled with a standardised form which was distributed to the measure stakeholder after the campaign in August 2012 (see annex).

Barrier: Insufficient organisational arrangements – The lack of a common understanding and strategic vision on prioritising sustainable transport in Tallinn on a higher political level hinders implementing the measure on a broader level reducing the potential impacts of the measure

Barrier: Insufficient involvement and awareness of policy key stakeholders – This reduces the impact of the measure, because strong support from politicians and officials would have helped to increase the visibility of sustainable transportation by operating the knitting bus. This has led to delays in signing agreements important for measure implementation.

Driver: Involvement and communication – The positive feedback from public transport users, city officials and other cities on the knitting bus has increased the motivation of the initiators to plan new activities to promote sustainable transport. As a result, the idea came up to develop a marketing strategy which compiles all single marketing activities in a document.

The barriers and drivers lead to the following planned activities:
Organisational arrangements and involvement – Following the delay in creating a marketing and communication strategy for sustainable transport additional members were added to the team preparing the document. This speeded up the process of drafting the strategy paper.

Involvement and communication – New campaigns and activities were planned based on positive experience and feedback from the Knitting Bus campaign. Special attention was placed on involving key stakeholders more thoroughly in the implementation process.

7.3.9 Results exploitation

The objective of the Knitting Bus campaign was to draw the attention of people to public transport and to make public transport more attractive in general. Since the campaign was only implemented in one bus it is difficult to make statements on these objectives.

The questions would therefore be: If people's attention was only drawn to this one bus, can we say that the measure was successful? Certainly, the measure could be up-scaled and applied to more busses, but would this attract more people to public transport? How much would it cost to furnish more busses and would the costs be justified by greater acceptance of public transport as an effect? Moreover, what are the long term effects of these and similar campaigns? Will they result in lasting increased acceptance, or will the effects only be temporary?

Further exploitation of the campaign requires careful deliberation on these questions. As such, it is difficult to say if and where this campaign might be successful in terms of transferability or if an up-scaling would have the desired effects.

7.3.10 Appraisal of the evaluation approach

The change in attitudes on ventilation and leg room show the benefits of having a control group. A before-and-after survey of passengers on busses without knitting on comparable bus lines in the same region might reveal explanatory factors for yielded results of the 'Knitting Bus' campaign and put results in the right perspective.

Besides this, awareness of public transport has not been measured properly because only bus-passengers have been addressed with the questionnaire. A logical step would have been to ask non-public transport users if the new design of the outside of the bus changed their mind about using public transport services. It also might have been interesting to interview those who were knitting for the buses. Their affiliation to public transport is likely to increase in a larger extend than the ones who were just using and seeing it.

It should be remembered that the Knitting Bus campaign was part of an integrated awareness strategy for public transport. The goal was to increase the modal share

of public transport and light transport choices by 10%. As this evaluation only covers the Knitting Bus and only those on the bus were surveyed, it is not possible to say how much or in which way this measure has contributed to achieving this general goal.

7.3.11 Conclusions

The Knitting bus campaign has received a lot of positive feedback from voluntary knitters, passengers and guests of Tallinn including members of the CIVITAS network. It encourages organising similar campaigns on different public transport modes, in other cities and countries.

The challenge for Tallinn is to capitalise on these results, understand where change is possible, and follow-up with messages to reinforce the improved perceptions of the bus environment: cleanliness, comfort, space, etc. It proved that improvements on these aspects can change the perception of the bus services in general.

This type of campaign can be more successful if it is part of a larger strategy to change behaviour. Tallinn should build on this success with a follow-up campaign that maintains the momentum and reinforces the image of the bus company as one that understands and cares about its customers and is not afraid of doing something out of the ordinary. This requires additional focus on organisational elements and stakeholder involvement and communication and a deliberation on exploitation questions.

7.3.12 Annex

Supplementary information to the evaluation

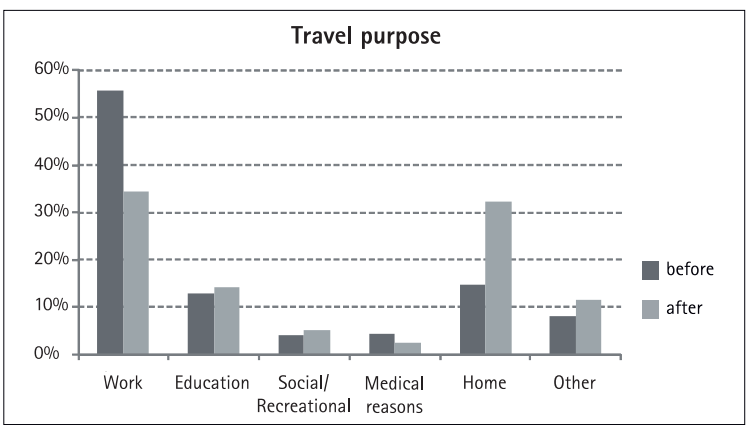


Figure 7–13: Distribution of travel purpose among bus users in the before-and-after sample.

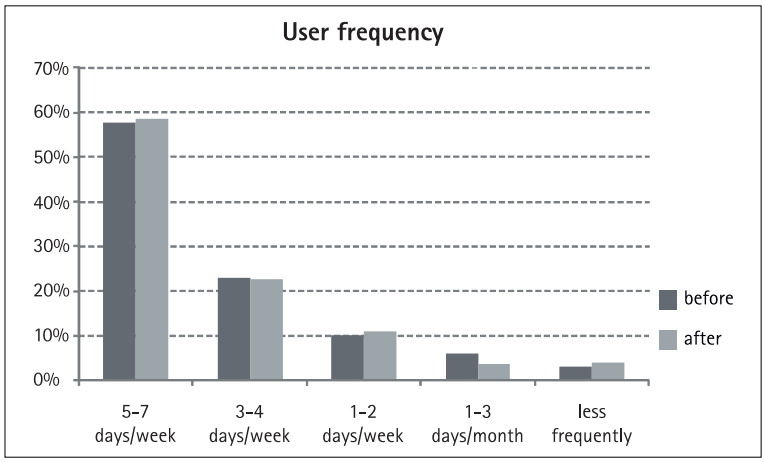


Figure 7–14: Distribution of user frequencies among bus users in the before-and-after sample.

Questionnaire used

Public Transport Customer Survey

This survey is aimed at finding out your options on the public transport provision. Please help us surve you better by telling us about the service you have received. We would really value your views and hope you are able to spend two or three minutes completing this survey. Please return the completed survey to a crew member.
The information collected is completely confidential and is used for statistical purposes only.

1. What is the purpose of your journey?

☐ Work ☐ Medical reasons

☐ Education ☐ Home

☐ Social/Recreational ☐ Other

2. How often do you use the city bus services?

- ☐ 5–7 days a week
- ☐ 3–4 days a week
- ☐ 1–2 days a week
- ☐ 1–3 days a month
- ☐ Less frequently

3. How satisfied are you with the following elements of the bus survice?

	Very dissa-tisfied	Dissa-tisfied		Satisfied	Very satis-fied	Don't know
Lighting	☐	☐	☐	☐	☐	☐
Ventilation	☐	☐	☐	☐	☐	☐
Seat comfort	☐	☐	☐	☐	☐	☐
Leg room/ space to move around	☐	☐	☐	☐	☐	☐
Cleanliness of the seats	☐	☐	☐	☐	☐	☐
Cleanliness of the floor	☐	☐	☐	☐	☐	☐
Cleanliness of the windows	☐	☐	☐	☐	☐	☐
Behavior of other passengers	☐	☐	☐	☐	☐	☐
Noise level	☐	☐	☐	☐	☐	☐

	Very dissa-tisfied	Dissa-tisfied	Neither dissatisfied nor satisfied	Satisfied	Very satis-fied	Don't know
Security und safety on buses	☐	☐	☐	☐	☐	☐
Overall satisfaction with the bus	☐	☐	☐	☐	☐	☐

4. How satisfied are you with the following elements of Tallinn city buses in general?

	Very dissa-tisfied	Dissa-tisfied	Neither dissatisfied nor satisfied	Satisfied	Very satis-fied	Don't know
Price-quality ratio	☐	☐	☐	☐	☐	☐
Bus routes	☐	☐	☐	☐	☐	☐
Frequency of services	☐	☐	☐	☐	☐	☐
Bus-stop locations	☐	☐	☐	☐	☐	☐
Information on schedules	☐	☐	☐	☐	☐	☐
Adherence to schedules	☐	☐	☐	☐	☐	☐
Level of custormer service on the bus	☐	☐	☐	☐	☐	☐
Types of tickets available	☐	☐	☐	☐	☐	☐

Security and safety on buses	☐	☐	☐	☐	☐	☐
Overall satisfaction with Bus Company	☐	☐	☐	☐	☐	☐

Please be so kind as to answer a few questions about yourself.
All information will be treated strictly confidential.

5. Gender

☐ Male

☐ Female

6. Age

☐ Under 18

☐ 18–29

☐ 30–40

☐ 41–60

☐ Older than 61

7. Do you have a car available?

☐ Yes

☐ No

8. Could you have used the car today but instead chose the bus?

☐ Yes

☐ No

Thank you for the cooperation!

Please return the completed survey to our staff on-board.

Process Evaluation Form used

Part A. General administrative information

It is important to know who the compiler of this form is. Please fill in the answers in the boxes below.

Reporting period	
Measure leader	
Name	
E-mail	
Compiler of the Process Evaluation Form (Only to be filled in if this is someone other than the Measure Leader)	
Name	
E-mail	

Part B. General content information

B1. In your own words, what are the objectives of the measure?

B2. From your point of view, which groups have been targeted with the measure?

There are predefined answers. Please put an 'X' in the open box before the number. If there are other target groups than the ones mentioned in the table it should be made use of line 10 "other". More than one answer is possible.

	1	Residents
	2	Car drivers/motorists
	3	Public transport users
	4	Cycle/walking groups
	5	Disabled people
	6	Commuters
	7	Visitors (shops/leisure)
	8	Local businesses
	9	General public
	10	Other, please describe

Part C. Content information for this reporting period

C1. Process barriers

Process barriers are events or overlapping conditions that hampers the process to obtain measure objectives/goals. In the checklist below you will find a number of barrier fields and examples of barriers which might have been encountered during the reporting period in trying to reach the objectives as given in question B1.

Barrier fields and examples of possible barriers:

No.	Barrier field	Examples of barriers
1	Political/strategic	Opposition of key actors based on political and/or strategic motives, lack of sustainable development agenda or vision, impacts of a local election, conflict between key (policy) stakeholders due to diverging believes in directions of solution
2	Institutional	Impeding administrative structures, procedures and routines, impeding laws, rules, regulations and their application, hierarchical structure of organisations and programs
3	Cultural	Impeding cultural circumstances and life style patterns
4	Problem related	Complexity of the problem(s) to be solved, lack of shared sense of urgency among key stakeholders to sustainable mobility
5	Involvement, communication	Insufficient involvement or awareness of (policy) key stakeholders, insufficient consultation, involvement or awareness of citizens or users
6	Positional	Relative isolation of the measure, lack of exchange with other measures or cities
7	Planning	Insufficient technical planning and analysis to determine requirements of measure implementation, insufficient economic planning and market analysis to determine requirements for measure implementation, lack of user needs analysis: limited understanding of user requirements
8	Organisational	Failed or insufficient partnership arrangements, lack of leadership, lack of individual motivation or know-how of key measure persons
9	Financial	Too much dependency on public funds (including CIVITAS funding) and subsidies, unwillingness of the business community to contribute financially
10	Technological	Additional technological requirements, technology not available yet, technological problems
11	Spatial	No permission of construction, insufficient space
12	Other	?????????

What are the three most important barriers encountered during the reporting period?

Please fill in the number of the barrier field from the checklist above in the open box according to importance. Please fill in a specification of the barrier in one sentence. This is important to make the barrier more understandable Questions to be answered in this part are: Which impact did the barrier have on the process of the measure and how did it occur? What exactly happened?

Example: If a (institutional) barrier is described just with "Impeding administrative structures, procedures and routines" it is not clear what happened in the city and what negative impact this factor had on the measure. It would be better to additionally write in one sentence a more specific explanation such as "The new complex legislation of procurement for the purchasing of goods and services has caused delays in the process of the public tender necessary for purchasing the automatic control system".

No.	Specification of barrier (max one sentence)		
		1	Most important barrier
		2	Second most important barrier
		3	Third most important barrier

C2. Process drivers

Process drivers are events or overlapping conditions that stimulates the process to obtain measure objectives/goals. In the checklist below you will find a number of driver fields and examples of possible drivers which might have been encountered during the reporting period in trying to reach the objectives as given in question B1.

Driver fields and examples of possible drivers:

No.	Driver field	Examples of drivers
1	Political/strategic	Commitment of key actors based on political and/or strategic motives, presence of sustainable development agenda or vision, positive impacts of a local election, coalition between key (policy) stakeholders due to converging (shared) believes in directions of solution
2	Institutional	Facilitating administrative structures, procedures and routines, facilitating laws, rules, regulations and their application, facilitating structure of organizations and programs
3	Cultural	Facilitating cultural circumstances and life style patterns

4	Problem related	Pressure of the problem(s) causes great priority, shared sense of urgency among key stakeholders to sustainable mobility
5	Involvement, communication	Constructive and open involvement of policy key stakeholders, constructive and open consultation and involvement or citizens or users
6	Positional	The measure concerned is part of a (city) program and/or a consequence of the implementation of a sustainable vision, exchange of experiences and lessons learned with other measures or cities
7	Planning	Accurate technical planning and analysis to determine requirements of measure implementation, accurate economic planning and market analysis to determine requirements for measure implementation, thorough user needs analysis and good understanding of user requirements
8	Organisational	Constructive partnership arrangements, strong and clear leadership, highly motivated key measure persons, key measure persons as 'local champions'
9	Financial	Availability of public funds (including CIVITAS funding) and subsidies, willingness of the business community to contribute financially
10	Technological	New potentials offered by technology, new technology available
11	Spatial	Space for physical projects, experimentation zones
12	Other	?????????

What are the three most important drivers encountered during the reporting period?

Please fill in the number of the driver field from the checklist above in the open box according to importance. Please fill in a specification of the driver in one sentence.

This is important to make the driver more understandable. Questions to be answered in this part are: Which impact did the driver have on the process of the measure and How did it occur? What exactly happened?

Example: If a (political) driver is described only with "strong commitment of local authorities", it is not clear to the outside reader which impact on the measure process this driver is causing. It is necessary to write in one sentence which local authority or person is concerned and what has changed concerning the process of the measure due to this commitment. An example is: "The alderman for city development has promoted

the measure in such a way that also business became interested in the measure and this now company XXX is an principal partner"

No.	Specification of the activity (max one sentence)		
		1	Most important activity
		2	Second most important activity
		3	Third most important activity

C3. Activities

Activities are actions taken by one or more measure partners to handle the barriers and/or to make use of the drivers to reach the measure objectives. In the checklist below you will find a number activity fields and examples of possible activities taken during the reporting period to overcome the barriers or to make use of the drivers.

Checklist of activity fields and examples of possible activities:

No.	Activity field	Examples of activities
1	Political/strategic	(Co-)development of vision on sustainable development or sustainable mobility, (Co-)development of a program towards sustainable development or sustainable mobility, discours with key stakeholders (politicians etc) about the sustainability problems to be solved.
2	Institutional	Analysis of and/or proposals to change impeding rules, structures, legislation, organisational structures etc.
3	Cultural	Facilitating cultural circumstances and life style patterns.
4	Problem related	Thoroughly analysing problems towards sustainable mobility to be solved, activities to explain the pressure of the problem, all activities towards sharing the sense of urgency among key stakeholders to sustainable mobility.
5	Involvement, communication	Consultation of target groups by workshop, conference, focus group, expert meeting, face-to-face interviews or questionnaires, telephone interviews or questionnaires or web based questionnaires, public awareness campaign about the sustainability problems to be solved, bringing together key stakeholders to discuss the sustainability problems to be solved (sharing different viewpoints), public awareness campaign about the measure through media activities, involvement of key stakeholders (politicians etc) in the measure.

6	Positional	Put the measure concerned into a running sustainability program (combined with the strategic actions), activities to exchange experiences with other measures / cities (workshop, conference, focus group etc).
7	Planning	Raising or attempting to raise additional 'time budget' for the measure , (re)conduct the economic and technical planning as well as analysis to determine requirements of measure implementation, (re)conduct market analysis to determine requirements for measure implementation, thoroughly analyzing user needs analysis to better understand the user requirements.
8	Organisational	Activities to raise the competences of the measure partners (for example special courses etc), activities to raise the motivation of the measure partners (for example extra measure meetings).
9	Financial	Raising or attempting to raise additional financial budget for the measure, developing a context which is attractive to the business community to contribute financially.
10	Technological	Raising or attempting to raise additional technical resources for the measure (all kind of equipment), all kind of actions to solve technological problems.
11	Spatial	(Attempts) Adjusting the construction permissions, creating experimental and /of investment zones / city parts / corridors
12	Other	

What are the three most important activities taken during the reporting period?

Please fill in the number of the activity field from the checklist above in the open box according to importance. Please bare in mind that there should be a link between the barriers and drivers as mentioned before. Please fill in a specification of the activity in one sentence. This is important to make the activity more understandable. Example: The (political) driver is a strong political commitment in the participation in the campaign to raise awareness for sustainable mobility. The (involvement) activity taken (to make use of the driver) may be described as: "Involvement of committed politicians in the awareness raising campaign activities such as: Conferences, Meetings, Public discussions".

No.	Specification of driver (max one sentence)		
		1	Most important driver
		2	Second most important driver
		3	Third most important driver

Part D. Any other comment

If you have any other comment you can note this below. If there are any ambiguities in the previous parts of the form, it is advisable to make use of this box for explanations. This might be for instance be applicable if there are mentioned several barriers in part C1 but no actions taken by a measure partner to overcome them in part C3. Why have there been no actions taken?

THANK YOU VERY MUCH FOR YOUR COOPERATION

8 Index

9 List of useful evaluation literature

9.1 Evaluation literature

European Commission: EVALSED – The resource for the evaluation of socio-economic development. Bruxelles, 2008. http://ec.europa.eu/regional_policy/sources/docgener/evaluation/evalsed/downloads/guide2008_evalsed.pdf.

Evalsed consists of two main parts:
THE GUIDE and three SOURCEBOOKS. The GUIDE is designed primarily for decision-makers – specifically those who design and manage evaluations in order to enhance decision making on socio-economic development policies. The SOURCEBOOKS (available only online) are of particular interest to practitioners and those wishing to impart or acquire evaluation skills.

European Commission: Evaluating EU activities – A practical guide for the Commission services. Bruxelles, 2004.
http://ec.europa.eu/dgs/secretariat_general/evaluation/docs/eval_activities_en.pdf.

This guide provides guidance on all kinds of evaluations whether ex-ante, interim or ex-post evaluation and whether they concern expenditure programmes or policies but is mainly directed to large-scale, EU-funded activities.

W. K. Kellogg Foundation: Evaluation Handbook – Philosophy and Expectations. Battle Creek (MI), 2004.
http://www.epa.gov/evaluate/pdf/eval-guides/evaluation-handbook.pdf

This handbook provides a framework for thinking about evaluation and outlines a blueprint for designing and conducting evaluations, either independently or with the support of an external evaluator/consultant. It is written primarily for project directors who have direct evaluation responsibilities.

9.2 Evaluation resource web sites

There is a variety of web sites with a comprehensive selection of evaluation resources, including guides, tools, trainings, links to other evaluation web sites, international associations and organisations, internet discussion groups, etc. The following list is a sample and the sites were last accessed in January 2013:

The Evaluation Center: Evaluation Checklists. Western Michigan University, 2011. http://www.wmich.edu/.

Evaluation Portal. 2011. www.evaluation.lars-balzer.name/.

EvaluationWiki.org: A public compendium of user-defined terms and other monitoring and evaluation information. 2011. www.evaluationwiki.org.

IFRC Monitoring and Evaluation web page. 2011. www.ifrc.org/MandE.

National Science Foundation (NSF): User-Friendly Handbook for Mixed Method Evaluations. 2011. http://www.nsf.gov/pubs/1997/nsf97153/start.htm.

OECD/DAC (Organisation for Economic Co-operation and Development/Development Assistance Committee): Evaluation of Development Programs web site. 2011. http://www.oecd.org/dac/evaluationofdevelopmentprogrammes/.

Resources for Methods in Evaluation and Social Research web site. 2011. http://gsociology.icaap.org/methods/, including a series of user-friendly beginner guides, http://gsociology.icaap.org/methods/BasicguidesHandouts.html.

UNDP Evaluation web site. United Nations Development Programme – Evaluation. 2011. http://www.undp.org/eo/.

UNEG (United Nations Evaluation Group) Evaluation Resources web site. 2011. www.uneval.org/evaluationresource/index.jsp?ret=true.

UNICEF (United Nations International Children's Emergency Fund): Web site & External Links for Evaluation and Lessons Learned. 2011. www.unicef.org/evaluation/index_18077.html.

MaxSumo: Guidance on how to plan, monitor and evaluate mobility projects. 2009. http://www.mobilitymanagement.org/index.phtml?Main_ID=2174&tID1=2359&tid=2359.

PROSPECTS – Procedures for Recommending Optimal Sustainable Planning of European City Transport Systems: Deliverable No. 2 – Evaluation Tools. January 2002. http://www.ivv.tuwien.ac.at/fileadmin/mediapool-verkehrsplanung/Diverse/Forschung/International/PROSPECTS/D2v6web.pdf.

9.3 Data collection and analysis methods

Field, Andy: Discovering Statistics Using SPSS. 3rd edition, Sage Publications, London, 2009.
Gonick, Larry; Smith, Woollcott: Cartoon Guide to Statistics. HarperCollins Publishers, New York, 1993.

Griffith, Arthur: SPSS for Dummies. 2nd edition, Wiley Publishing, Indianapolis, 2010.

Griffiths, Dawn: Head First Statistics. O'Reilly, Sebastopol, 2009.

Hand, David J.: Statistics – A Very Short Introduction. Oxford University Press, Oxford, 2008.

Khan Academy: Free collection of micro lectures on video.
http://www.khanacademy.org/.11) Takahashi, Shin: The Manga Guide to Statistics. No Starch Press, San Francisco, 2009.

Rumsey, Deborah: Statistics II for Dummies. Wiley Publishing, Indianapolis, 2009.

Rumsey, Deborah J.: Statistics for Dummies. 2nd edition, Wiley Publishing, Indianapolis, 2011.

Schmuller, Joseph: Statistical Analysis with Excel for Dummies. 2nd edition, Wiley Publishing, Indianapolis, 2009.

Statistics for the Terrified: Tutorial on statistic (for charge).
http://www.conceptstew.co.uk/PAGES/home.html.

Stat Trek: Free online tutorials on statistics and matrix algebra. http://stattrek.com/.

9.4 Links to helpful software

Statistics for the terrified:
http://www.conceptstew.co.uk/PAGES/home.html.

Statistical Analysis with IBM® SPSS® Statistics:
http://www-01.ibm.com/software/analytics/spss/.

Statistical Analysis with SAS/STAT® Software:
http://www.sas.com/technologies/analytics/statistics/stat/.

Statistical Analysis with Stata® Software:
http://www.stata.com/.

Statistical Analysis with R:
http://www.r-project.org/.

Statistical Analysis with Microsoft Excel:
http://www.microsoft.com.

Qualitative Analysis with Kwalitan:
http://www.kwalitan.nl/engels/index.html.

Qualitative Analysis with Nvivo:
http://www.qsrinternational.com/products_nvivo.aspx.

Qualitative Analysis with Hyperresearch:
http://www.researchware.com/products/hyperresearch/download.html.

Qualitative Analysis with MAXQDA:
http://www.maxqda.de/.

10 References and picture credits

All cartoons: ©Auke Herrema, www.aukeherrema.nl

Chapter 7.1 Utrecht Raod Safety Label: Figure 7-1 ©city of Utrecht

Chapter 7.2 Utrecht Cargohopper: Figure 7-4 ©Geisje Hoetjes

Chapter 7.3 Tallinn Knitting Bus: Figure 7-9 left: ©City of Tallinn, right: ©Eileen O'Connell