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

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.

THE CIVITAS INITIATIVE
IS CO-FINANCED BY THE
EUROPEAN UNION

ISBN 978-3-8309-7881-7
© Waxmann Verlag GmbH, 2013
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

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, electrostatic, magnetic tape, mechanical, photocopying, recording or otherwise without permission in writing from the copyright holder.

Table of Contents

Preface | 9

1 Introduction | 12

1.1 What is evaluation? | 12

1.2 What is the point of evaluation anyway? | 13

1.3 How does evaluation work? | 15
1.3.1 When should I evaluate? | 15
1.3.2 Are there different types of evaluation? | 17

1.4 Whom do I need to get involved? | 18

1.5 Why do we need this book? | 19

2 Impact evaluation | 23

2.1 Objectives of your measure | 25

2.2 Effect analysis | 29

2.3 Indicators | 33

2.4 Evaluation design | 36
2.4.1 Randomised control group designs | 39
2.4.2 Quasi-experimental control group design | 40
2.4.3 Designs without control groups | 41
2.4.4 Choosing the evaluation design | 42

2.5 Data collection | 46
2.5.1 Individual in-depth interview | 48
2.5.2 Questionnaire survey | 49
2.5.3 Focus groups | 52
2.5.4 Observation techniques | 52
2.5.5 Basic guidelines for conducting a survey | 54
2.5.6 Advantages and disadvantages of data collection methods | 56
2.5.7 Sample size and sampling issues | 58

2.6 Data analysis | 61
2.6.1 Data preparation and back-up | 61

2.6.2 Data processing and analysis | 66
2.6.3 Qualitative data analysis | 73

2.7 Discussion of results from data analysis | 77

3 Process evaluation | 79

3.1 Aim and outcome | 79

3.2 Barriers and drivers | 81

3.3 Activities related to barriers and drivers | 84

3.4 Methods for process evaluation | 85
3.4.1 Standardised forms | 85
3.4.2 Learning Histories | 87
3.4.3 Other methods | 92

4 Information reporting and utilisation | 94

4.1 Result presentation | 94
4.1.1 Data presentation | 96
4.1.2 How not to present your data |100

4.2 Up-scaling of results | 103
4.2.1 Introduction | 103
4.2.2 Why and when is up-scaling relevant? | 104
4.2.3 Preconditions for up-scaling | 105
4.2.4 Defining the scale | 106
4.2.5 Considerations for up-scaling | 107

4.3 Transferability | 109
4.3.1 Introduction | 109
4.3.2 Framework for transferability assessment | 110
4.3.3 Conclusions to transferability | 116

5 And now: back to reality | 117

5.1 Some myths about evaluation | 117

5.2 'Things' which can mess up your evaluation | 119

5.3 Strengthening the quality of your evaluation | 121

6 The very last chapter | 124

6.1 Some words to the evaluator | 124

6.2 Some words to politicians | 125

7 Evaluation examples | 127

7.1 Utrecht Road Safety Label | 127
7.1.1 Problem description and measure context | 127
7.1.2 Objectives | 128
7.1.3 Preparation and implementation stages | 128
7.1.4 Cause-and-effect relationships | 131
7.1.5 Indicators and data collection | 132
7.1.6 Business-as-usual scenario | 134
7.1.7 Data analysis and results | 134
7.1.8 Impact result interpretation | 135
7.1.9 Process evaluation | 136
7.1.10 Result exploitation | 137
7.1.11 Appraisal of the evaluation approach | 138
7.1.12 Conclusions | 138
7.1.13 Annex | 139

7.2 Utrecht Cargohopper | 140
7.2.1 Problem description and measure context | 140
7.2.2 Objectives | 141
7.2.3 Preparation and implementation stages | 141
7.2.4 Cause-and-effect relationships | 143
7.2.5 Indicators and data collection | 144
7.2.6 Business-as-usual scenario | 145
7.2.7 Data analysis and results | 146
7.2.8 Impact result interpretation | 147
7.2.9 Process evaluation | 148
7.2.10 Result exploitation | 149
7.2.11 Appraisal of the evaluation approach | 150
7.2.12 Conclusions | 150
7.2.13 Annex | 151

7.3 Tallinn Knitting Bus | 151
7.3.1 Problem description and measure context | 151
7.3.2 Objectives | 152

7.3.3 Preparation and implementation stages | 153
7.3.4 Cause-and-effect relationships | 154
7.3.5 Indicators and data collection | 156
7.3.6 Data analysis and results | 156
7.3.7 Impact result interpretation | 158
7.3.8 Process evaluation | 159
7.3.9 Results exploitation | 160
7.3.10 Appraisal of the evaluation approach | 160
7.3.11 Conclusions | 161
7.3.12 Annex | 161

8 Index | 173

9 List of useful evaluation literature | 175

9.1 Evaluation literature | 175

9.2 Evaluation resource web sites | 175

9.3 Data collection and analysis methods | 177

9.4 Links to helpful software | 177

10 References and picture credits | 179

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.

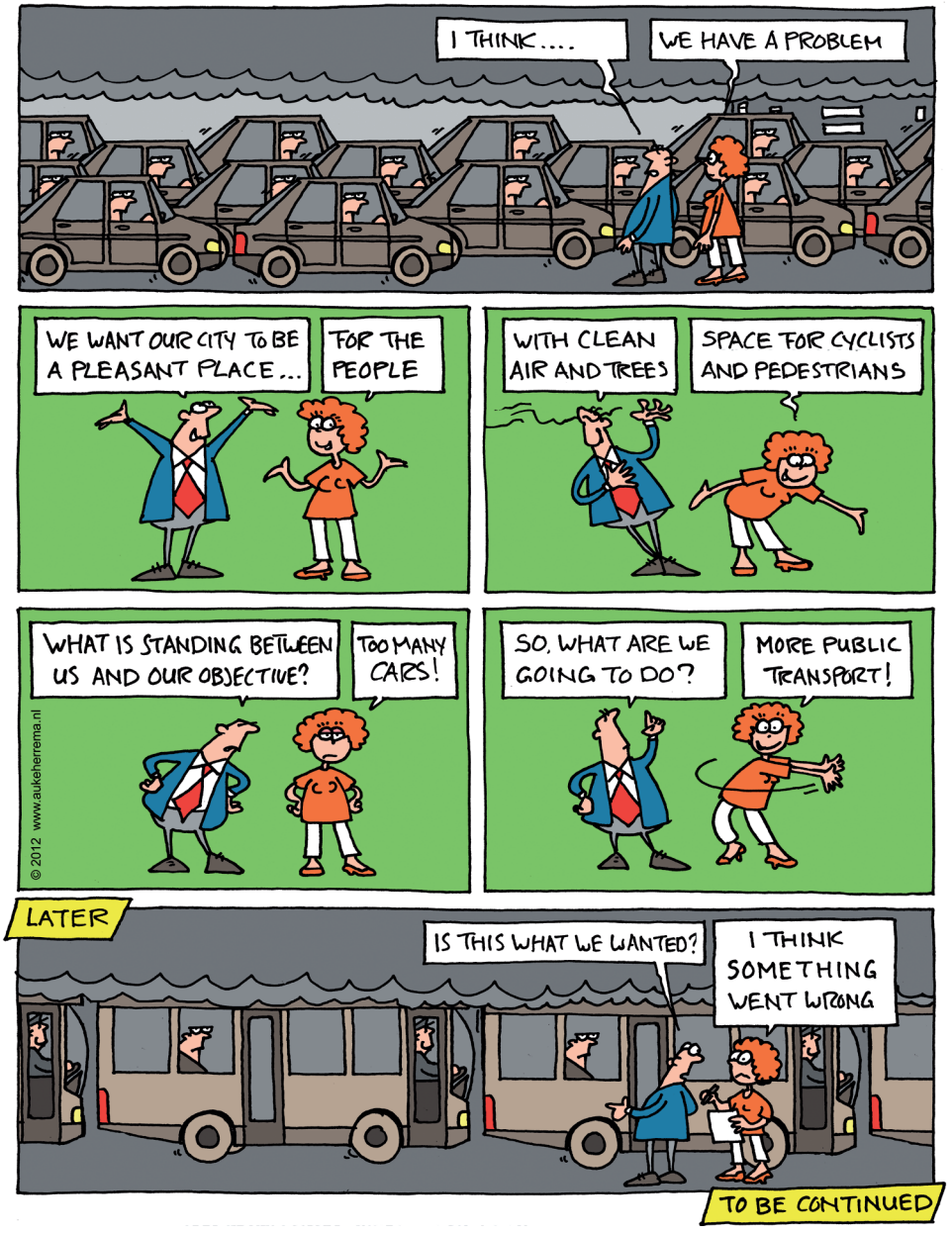
So far there were CIVITAS I (2002-2006) and CIVITAS II (2005-2009). The third programme, CIVITAS PLUS (2008-2013), is about to come to an end and will be followed by CIVITAS PLUS II (2013-2017). In the present programme there are five so-called collaborative projects, namely, ARCHIMEDES, ELAN, MIMOSA, 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.aukeherrema.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?

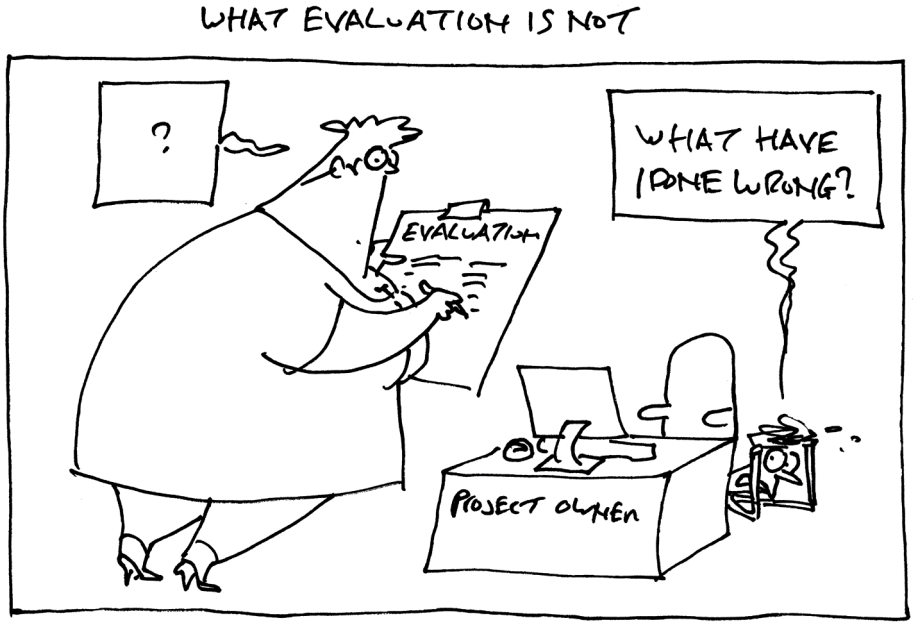
Why is it important to conduct an evaluation if a project is running well and everything seems fine? Surprisingly, this question often pops up at city-level evaluation. Ironically, the answer is the same that drives us to learn from less successful pilot measures. Evaluation is a natural thing and every 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.



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

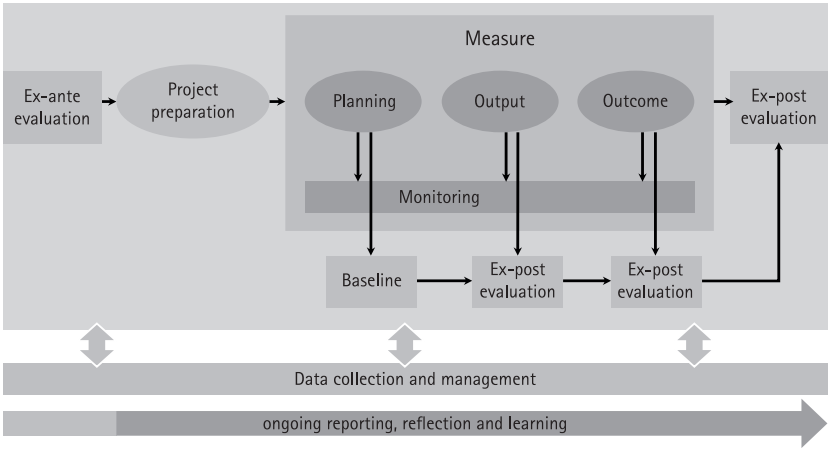
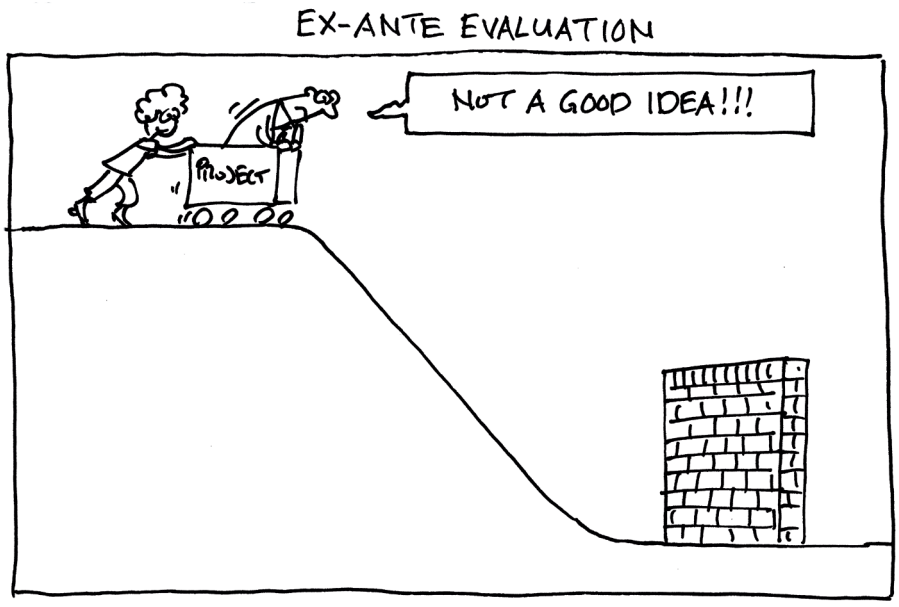


Figure 1–1: Arrangement of evaluation.

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



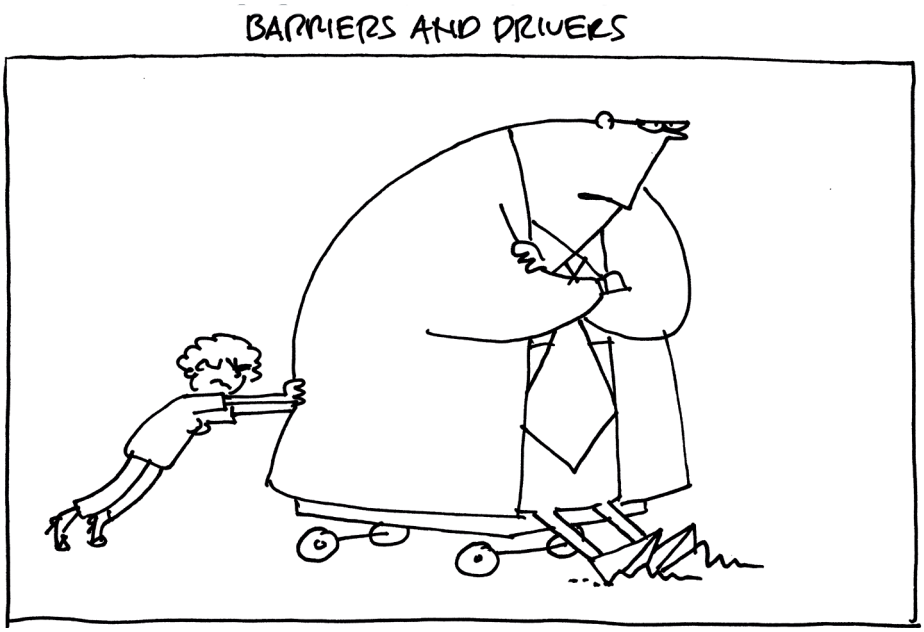
inform on the ongoing project/programme implementation. Often, the data you will need to provide for this can also be useful for the evaluation or vice versa. The final (ex-post) evaluation occurs after the measure/project is completed to assess how well the project/programme achieved its intended objectives. 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).