Conditions, locations, sources, and method of data collection

Ceteri Paribus - under the same conditions - is a central rule in scientific data collection. These conditions are to be answered in the data collection plan of sigmaGuide (col. J). These condition or blocking variables provide a way to differentiate the analyses and compare conditions.

You have defined the locations (col. L) in your *Gemba-Walk*. The sources of data collection are field data - e.g. about the concrete extent of littering as well as data about the trigger of littering among the citizens. The methods of data collection are: field data via on-site survey. The related file in the Sustainability-Project-Files is: **GembaWalk-Data-Sheet**. You can adapt this data collection sheet to your own ideas.

In addition, you need to collect data from citizens via a questionnaire. These data should make the individual influences on littering transparent. I have prepared a questionnaire for you to use to collect this data from the people around you. Collecting this data in this format is mandatory and cannot be customized. But of course I am open to any feedback for improvement.

The related file in the Sustainability Project Files is: **Questionnaire-Data-Sheet_Example**. This sample file contains some dummy data and serves only to give you an overview. The updates of the collected data of all participants will be published regularly in the course unit: 01/03 Downloads. And the questionnaire is linked here:

https://forms.office.com/Pages/ResponsePage.aspx?id=DQSIkWdsW0yxEjajBLZtrQAAAAAAAAAAAN__oL4XjFUQUlDVUNaVzBRQzlWMUVMSFRHOE5NUEpWWS4u

Field-Study

You have already determined the locations for data collection on a map of your *Gemba-Walk* (Project-Story-Book). The following proposal outlines a small quasi-experimental design[29] for the blocking-/ conditions-variables (Data Collection Plan/ Column J). This design offers several options for ad hoc hypotheses (Tab 7-1). The complete table is included in the Sustainability-Project-Files[30]. Of course you can adapt this design to your own ideas in your Data-Collection-Plan – you can extend but you should not reduce it too much. Because if you later want to earn your Lean Six Sigma Black Belt with us, you will build on this design and the data collected.

Condition-/ Blocking Variables - Design

x_G1_GembaWalk	x_G2_ProjPhase	x_G3_date	x_G4_time	x_G5_LocationNumber	x_G6_Location	x_G7_locType	x_G8_locGenCleanliness	x_G9_locGenCrowdedness	x_G10_locDaysSinceLastRemovalService
1	MEASURE	2023-01-01	7:30	1	busstop	Traffic Zone	low	high	0
1	MEASURE	2023-01-01	7:45	2	park	Green City Areas	high	low	0
1	MEASURE	2023-01-01	8:00	3	fast-food-restaurant	Catering Zone	medium	high	0
1	MEASURE	2023-01-01	8:15	4	playground	Kids Zone	?	?	0
1	MEASURE	2023-01-01	8:30	5	super-market	Shopping Zone	?	?	0
2	MEASURE	2023-01-02	7:30	1	busstop	Traffic Zone	?	?	1
2	MEASURE	2023-01-02	7:45	2	park	Green City Areas	?	?	1
2	MEASURE	2023-01-02	8:00	3	fast-food-restaurant	Catering Zone	?	?	1
2	MEASURE	2023-01-02	8:15	4	street crossing	Kids Zone	?	?	1
2	MEASURE	2023-01-02	8:30	5	super-market	Shopping Zone	?	?	1
3	MEASURE	2023-01-03	7:30	1	busstop	Traffic Zone	?	?	2
3	MEASURE	2023-01-03	7:45	2	park	Green City Areas	?	?	2
3	MEASURE	2023-01-03	8:00	3	fast-food-restaurant	Catering Zone	?	?	2
3	MEASURE	2023-01-03	8:15	4	street crossing	Kids Zone	?	?	2
3	MEASURE	2023-01-03	8:30	5	super-market	Shopping Zone	?	?	2
4	MEASURE	2023-01-04	7:30	1	busstop	Traffic Zone	?	?	3
4	MEASURE	2023-01-04	7:45	2	park	Green City Areas	?	?	3
4	MEASURE	2023-01-04	8:00	3	fast-food-restaurant	Catering Zone	?	?	3
4	MEASURE	2023-01-04	8:15	4	street crossing	Kids Zone	?	?	3
4	MEASURE	2023-01-04	8:30	5	super-market	Shopping Zone	?	?	3
5	MEASURE	2023-01-05	7:30	1	busstop	Traffic Zone	?	?	4
5	MEASURE	2023-01-05	7:45	2	park	Green City Areas	?	?	4
5	MEASURE	2023-01-05	8:00	3	fast-food-restaurant	Catering Zone	?	?	4
5	MEASURE	2023-01-05	8:15	4	street crossing	Kids Zone	?	?	4
5	MEASURE	2023-01-05	8:30	5	super-market	Shopping Zone	?	?	4
6	MEASURE	2023-01-06	7:30	1	busstop	Traffic Zone	?	?	5
6	MEASURE	2023-01-06	7:45	2	park	Green City Areas	?	?	5
6	MEASURE	2023-01-06	8:00	3	fast-food-restaurant	Catering Zone	?	?	5
6	MEASURE	2023-01-06	8:15	4	street crossing	Kids Zone	?	?	5
6	MEASURE	2023-01-06	8:30	5	super-market	Shopping Zone	?	?	5
7	MEASURE	2023-01-07	7:30	1	busstop	Traffic Zone	?	?	6
7	MEASURE	2023-01-07	7:45	2	park	Green City Areas	?	?	6
7	MEASURE	2023-01-07	8:00	3	fast-food-restaurant	Catering Zone	?	?	6
7	MEASURE	2023-01-07	8:15	4	street crossing	Kids Zone	?	?	6
7	MEASURE	2023-01-07	8:30	5	super-market	Shopping Zone	?	?	6
8	CONTROL		7:30	1	busstop	Traffic Zone	?	?	0
8	CONTROL		7:45	2	park	Green City Areas	?	?	0

Table 7-1 Plan for collecting stratified samples via condition variables in a quasi-experimental design

This design includes 7 Gemba-Walks before for ANALYSE and 7 Gemba-Walks after improvement for CONTROL. They lead to 5 locations, which are assigned to different location types (aligned with Questionnaire, see below). The degree of the location general cleanliness and crowdedness is in the example subjectively evaluated on a 3-level-rating-scale. You will need to recode the nominal values to integer values (low= 1, medium= 2, high= 3) to use the ordinal scale of this variable in your calculations. The Gemba-Walks take place on 7 consecutive days at the same time each. If your local garbage collection cleans the areas and trash cans 1/ week, and you know the date, then also the elapsed days since the last service are evident.

Do you already have ideas what comparisons and thus hypotheses you can derive from this design?

[29] https://en.wikipedia.org/wiki/Quasi-experiment

[30] File: GembaWalk Data-Sheet.xlsx in the Sustainability-Project-Files.zip

Influences (x)

The variables from Table 7-1 are labelled as influences (x) because they determine the conditions for your data collection. All direct influences (x) - as well as the problems (Y) - result from your data collection plan (Tab. 7-2).

x_G11_TrashCanVisibility	x_G12_TrashCanDistance	x_G13_TrashCanFillStatus	x_G14_TrashCanUsability	x_G15_TrashCanFeatures
10	1	0	0	standard
4	1	1	1	ashtray tops
1	1	2	0	pet waste bags
		3		standard
		4		standard

Table 7-2 Variables representing the negative influences (x) of the Data-Collection-Plan

The influences to be measured in this example are: the visibility of the trash can (ordinal: 0 - 10), e.g. evaluated from the bench of a bus stop, its distance (cardinal: meter), its filling level (empty= 0 - overfilled= 4), its usability (ok= 0, defect= 1). If you like, then you can additionally discriminate the features of the respective trash can (e.g.: standard; ashtray tops; pet waste bags).

Problems (Y)

Problems extracted from the Data Collection Plan are the littered area (square meters), the (estimated) number of pieces of litter in that area (pieces), and the (estimated) percentages for the litter types: Food, paper, metal, plastic, glass, chewing gum and cigarette butts (Tab. 7-3). The numbers behind the names of the seven litter-type variables - e.g. Y_G25_Food_**1** - represent their respective harmfulness, e.g. Food= 1, Paper= 2 – Cigarette Butts= 7. These litter-types and their evaluated harmfulness correspond to the litter-types and respective harmfulness in the questionnaire (see below), to allow cross references.

Y_G16_LitterArea	Y_G17_LitterPieces	Y_G18_Food_1_perc	Y_G19_Paper_2_perc	Y_G24_CigButts_7_perc	Y_perc_Check
2	100	5%	20%	30%	100%
5	200	10%	25%	20%	100%
3	300	20%	5%	15%	100%
					0%
					0%

Table 7-3 Section for Variables representing the problems (Y) of the Data-Collection-Plan

To balance the effort and accuracy of data collection, in this example first the overall number of litter pieces are counted/ estimated first in the area of each Gemba-Walk location (Tab 7-3). Then, the percentage of the different litter-types is estimated. The percentage check (right) verifies that your estimated proportions add up to 100%. The estimated number of all litter pieces at a location and the proportions of the different litter-types then allow the number of pieces of litter-types to be calculated (Tab 7-4).

325_Food_1_pcs	Y_G26_Paper_2_pcs	Y_G27_Metal_3_pcs	Y_G28_ChewGum_4_pcs	Y_G29_Glas_5_pcs	Y_G30_Plastic_6_pcs	Y_G31_CigButts_7_pcs
5	20	15	0	10	20	30
20	50	10	0	40	40	40
60	15	30	15	45	90	45
0	0	0	0	0	0	0
0	0	0	0	0	0	0

Table 7-4 Number of estimated pieces of litter-types derived from total number and percentages.

The following table gives an overview to all suggested variables for your field-data collection, their meaning, variable names, scales, and their values/ range.

Meaning	Variable Names	Scale-Level	Values/ Range
Serial number of Gemba Walk	x_G1_GembaWalk	cardinal	1 - N
DMAIC phase of data collection	x_G2_ProjPhase	nominal	MEASURE (= before), CONTROL (=after)
Date of Gemba Walk	x_G3_date	Date/ Time	
Time of Gemba Walk	x_G4_time	Date/ Time	
Location of Data Collection	x_G5_LocationNumber	nominal	e.g.: bus stop; park; fast-food-restaurant; playground, super-market
Location Type of Data Collection	x_G7_locType	nominal	e.g.: traffic zone; green/ water area; catering zone; kids zone; shopping zone
Location cleanliness	x_G8_locGenCleanliness	ordinal	e.g. high= 3; medium= 2; low= 1
Location crowdedness	x_G9_locGenCrowdedness	ordinal	e.g. high= 3; medium= 2; low= 1
Visibility of Trash Can from Location	x_G11_TrashCanVisibility	ordinal	0 - 10
Distance of Trash Can from Location	x_G12_TrashCanDistance	cardinal	meter
Time-Interval since last Trash Can Emptying	x_G10_locDaysSinceLastRemovalServic	cardinal	days
Fill Status of Trash Can	x_G13_TrashCanFillStatus	ordinal	e.g. overfilled= 4, full= 3; medium= 2; low= 1; empty= 0
Usability of Trash Can	x_G14_TrashCanUsability	nominal	e.g.: 0= ok; defect= 1
Configuration/ Features of Trash Can	x_G15_TrashCanFeatures	nominal	e.g.: standard; ashtray tops; pet waste bags
Littered Area under observation	Y_G16_LitterArea	cardinal	square-meter
Number of Trash Pieces within Littered Area	Y_G17_LitterPieces	cardinal	number of pieces
Percentual proportion of Food within Littered Area	Y_G18_Food_1_perc	cardinal	percentual proportion
Percentual proportion of Paper within Littered Area	Y_G19_Paper_2_perc	cardinal	percentual proportion
Percentual proportion of Metal within Littered Area	Y_G20_Metal_3_perc	cardinal	percentual proportion
Percentual proportion of Chewing Gum within Littered Area	Y_G21_ChewGum_4_perc	cardinal	percentual proportion
Percentual proportion of Glas within Littered Area	Y_G22_Glas_5_perc	cardinal	percentual proportion
Percentual proportion of Plastic within Littered Area	Y_G23_Plastic_6_perc	cardinal	percentual proportion
Percentual proportion of Cigarette Butts within Littered Area	Y_G24_CigButts_7_perc	cardinal	percentual proportion
Calculated pieces of Food within Littered Area	Y_G25_Food_1_pcs	cardinal	calculated pieces
Calculated pieces of Paper within Littered Area	Y_G26_Paper_2_pcs	cardinal	calculated pieces
Calculated pieces of Metal within Littered Area	Y_G27_Metal_3_pcs	cardinal	calculated pieces
Calculated pieces of Chewing Gum within Littered Area	Y_G28_ChewGum_4_pcs	cardinal	calculated pieces
Calculated pieces of Glas within Littered Area	Y_G29_Glas_5_pcs	cardinal	calculated pieces
Calculated pieces of Plastic within Littered Area	Y_G30_Plastic_6_pcs	cardinal	calculated pieces
Calculated pieces of Cigarette Butts within Littered Area	Y_G31_CigButts_7_pcs	cardinal	calculated pieces

Table 7-5 Meaning, names, scales and values/ range of the variables for collecting field data

After each Gemba-Walk with data collection, please think about how you can make it more reliable and valid - perhaps also more efficient. These would be steps towards continuous improvement - as part of Lean.

Questionnaire

Access

A questionnaire serves as second data source. It is publicly available at this address:
https://forms.office.com/Pages/ResponsePage.aspx?id=DQSIkWdsW0yxEjajBLZtrQAAAAAAAAAAAAN__oL4XjFUQUlDVUNaVzBRQzlWMUVMSFRHOE5NUEpWWS4u

and additionally via this QR-Code:

Figure 7-6 QR-Code for Survey on Environmental Littering

To get familiar with the questionnaire, it's best to try it out yourself. The average processing time is 6 minutes. Then you can and should share this link and code via eMail or social media to collect answers. Because the data collection with this questionnaire is mandatory. The related file in the Sustainability Project Files is: Questionnaire-Data-Sheet_Example.xlsx. This sample file contains some dummy data and serves only to give you an overview. The updates of the collected data of all participants will be published regularly in the course unit: 01/03 downloads.

Questionnaire Construction

The questionnaire consists of 5 parts:

a. Littering Risk (Y)
b. Knowledge about Litter Harmfulness (x)
c. Motivation to (avoid) Litter (x)
d. Support by the City (x)
e. Information about the respondent (x) + assignment key

a. Littering Risk (Y)

a. Littering Risk														
Y_Q1	May I ask you how often you eat, drink or possibly smoke in public outdoor places?	0 never	1	2	3	4	5	6	7	8	9	10 always		
Y_Q2	Would you tell me, how often you throw away garbage outdoors, like bottles, cups, bags, or cigarette butts?	0 never	1	2	3	4	5	6	7	8	9	10 always		
Y_Q3	Do you throw trash out the window when you are on the road, e.g. from the car, bus or train?	0 never	1	2	3	4	5	6	7	8	9	10 always		
Y_Q4	What would you basically drop outdoors? Please rank these waste types from: top = most likely to bottom = least likely.	Food: Fruit peels, leftovers		Paper: Sheets, boxes, bags, cartons		Metal: Cans, caps, foil		Chewing Gum:		Glas: Bottles, jars		Plastic: Bottles, boxes, bags, cups, wrappers		Cigarette Butts:

Table 7-7 Questions and response options about the frequency and severity of littering - in the data combined in a Littering-Risk-Indicator (see below).

The questions Y_Q1 and Y_Q3 (Tab 7-7) refer to the frequency of littering and use an 11-point rating scale for the answer. Question Y_Q4 refers to the severity of littering. It uses a 7-level ranking scale for given litter types. That means 7 predefined littering types are to be ranked by respondents in order of their individual littering willingness. A risk indicator is derived from the answers to these 4 questions (see below). All specific responses, as well as the Littering-Risk-Indicator, serve as problems (Y) in this questionnaire, representing an estimation of actual littering behaviour.

b. Knowledge about Litter Harmfulness (x)

Knowledge about the harmfulness of the predefined litter-types is considered as an influence (x) on the litter risk (Y): the more is known about the harmfulness of the litter types, the lower the risk. This item also uses the same 7-point ranking scale (Tab. 7-7) for the given litter-types (Tab 7-8).

The harmfulness of the litter-types (x_Q5) cannot be evaluated objectively. In principle, the harmful effect of cigarette butts on groundwater has been proven. The short-, medium-, and long-term effects of plastic litter on the world's oceans is also undisputed. However, in my opinion, broken glass on a children's playground poses a greater immediate danger than a plastic cup. My evaluation of food scraps as harmless waste up to the very harmful cigarette butts - even on a children's playground - is therefore to be seen as a compromise (Tab 7-8, x_Q5). If necessary, you can adjust the harmfulness rating to your needs.

b. Knowledge about Litter Harmfulness								
x_Q5	How would you rank the environmental harmfulness of these waste types? Please rank them from: top = least harmful to bottom = most harmful.	Food: Fruit peels, leftovers	Paper: Sheets, boxes, bags, cartons	Metal: Cans, caps, foil	Chewing Gum:	Glas: Bottles, jars	Plastic: Bottles, boxes, bags, cups, wrappers	Cigarette Butts:
	Suggestion: 1= least harmful to 7= most harmful	**1**	**2**	**3**	**4**	**5**	**6**	**7**
x_Q6	How big are the harmful effects of litter in these outdoor places?	Shopping Zone: In front of a supermarket or shopping center.	Catering Zone: Street cafe, street restaurant, in front of a fast	Traffic Zone: Side walk, street, bus stop, market place, event	Green City Areas: Parks and other green areas in the city	Kids Zone: Playground, schoolyard, sports field	Forest, Fields: Forest, fields, along highways	Water: Creek, river, lake, shore
	Suggestion: 1= least harmful to 7= most harmful	**1**	**2**	**3**	**4**	**5**	**6**	**7**

Table 7-8 Questions and answer options about the knowledge on littering harmfulness

Knowledge about the harmful effects of litter in different locations (x_Q6) is considered as another influence (x) on the litter risk (Y): the more is known about the harmful effects of littering in a certain location, the lower the risk. This item uses the same 7-point ranking scale - but here for the predefined litter locations (Tab 7-8, x_Q6).

This subjective assessment is based on the observation that litter is most often removed in shopping and catering zones because the owners of the locations have an interest in the positive external image of their businesses. Traffic zones, green urban areas, and public facilities are on the city's removal plan and are typically cleaned up on a regular weekly basis. Forests, fields, and water areas probably have the lowest clean-up frequency and litter can spread directly to groundwater and oceans here. If necessary, you can adjust the harmfulness rating to your specific needs.

A Littering-Knowledge-Indicator - overall representing the influence of knowledge on littering - is derived from these 2 items (see below).

c. Motivation to (avoid) Litter

The 10 questions on motivation (x_Q7 - x_Q16) in table 7-9 cover intrinsic beliefs (Q7 - Q10) and extrinsic incentives to (not) litter (Q11 - Q16). Answers can be given on a 11-point rating scale.

c. Motivation to (avoid) Litter												
x_Q7	How far do you carry a piece of trash to a trash can?	would not carry	1 m	5 m	10 m	20 m	50 m	100 m	200 m	500 m	1 km	carry any distance
		1	2	3	4	5	6	7	8	8	9	10
x_Q8	How often do you pick up strange litter that is on your path and take it to the nearest trash can?	0 never	1	2	3	4	5	6	7	8	9	10 always
x_Q9	How would you characterize the condition of your own apartment?	0 messy	1	2	3	4	5	6	7	8	9	10 tidy
x_Q10	Do you separate waste at home, e.g. paper, plastic, glass, organic waste, chemicals and residual waste, in different containers?	0 never	1	2	3	4	5	6	7	8	9	10 complete
x_Q11	When I am pressed for time, I may litter.	0 disagree	1	2	3	4	5	6	7	8	9	10 agree
x_Q12	Sometimes I put garbage in a place where it cannot be directly seen.	0 disagree	1	2	3	4	5	6	7	8	9	10 agree
x_Q13	In places where there is already garbage lying around, it is not so bad to add something to it.	0 disagree	1	2	3	4	5	6	7	8	9	10 agree
x_Q14	When I'm out with others, I'm more likely to also litter than when I'm out alone.	0 disagree	1	2	3	4	5	6	7	8	9	10 agree
x_Q15	I don't mind the garbage on the ground. And a bit more or less doesn't matter.	0 disagree	1	2	3	4	5	6	7	8	9	10 agree
x_Q16	Throwing garbage away is not a problem. The garbage removal service is paid to clean it up.	0 disagree	1	2	3	4	5	6	7	8	9	10 agree

Table 7-9 Questions and answer options about the motivation to (not) litter

A Littering-Motivation-Indicator is derived from these 10 items (see below).

d. Support by the City

The 3 questions on the support by the city to remove litter - x_Q17 - x_Q19 (Tab 7-10) can again be answered on an 11-point rating scale. The questions cover the distribution density of public trash cans, the overall cleanliness of the city, and proactive measures to prevent littering.

d. Support by the City												
x_Q17	How well is your town equipped with trash cans?	0 very bad	1	2	3	4	5	6	7	8	9	10 very good
x_Q18	How clean are the streets and public places of your city in your opinion?	0 very clean	1	2	3	4	5	6	7	8	9	10 very dirty
x_Q19	My city actively promotes litter reduction, e.g. by signs and public campaigns.	0 disagree	1	2	3	4	5	6	7	8	9	10 agree

Table 7-10 Questions and answer options about the city in addressing littering.

A city-Littering-Support-Indicator is derived from these 3 items (see below).

e. Information about the Interviewee

The questions x_Q20 - x_Q24 (Tab 7-11) ask for sociodemographic information about the respondent: Place of residence - country and city, the gender, age, and education.

e. Information about the Respondent											
x_Q20	Country										
x_Q21	City										
x_Q22	Gender	Woman	Man	Non-binary	Prefer not to say						
x_Q23	Age	< 15 1	15 - 20 2	21 - 30 3	31 - 40 4	41 - 50 5	51 - 60 6	61 - 70 7	71 - 80 8	> 80 9	Prefer not to say 0
x_Q24	Education	Primary 1	Secondary 2	Vocational Training 3	Bachelor 4	Master 5	PHD 6	Prefer not to say 0			
x_Q25	Feedback										
x_Q26	keyword										

Table 7-11 Questions and answer options about socio-demographic information about the respondent, feedback, and a unique keyword that you provide for this survey

Question 25 offers the respondent the opportunity to give feedback on the questionnaire, e.g. on possible improvements.

Attention: **Question 26** has a special function. Here the respondent should enter a **keyword** that you must define and publish with the invitation to participate. What is the purpose of this keyword? The answers of your respondents and the answers of all other participants of this course are stored in one central file. This file with all collected data will be published weekly (course unit 01/ 03, Downloads). Using the submitted keyword you will be able to filter the answers of your respondents in this file. And of course you can compare the answers of your respondents with the answers of other participants of this course - if you like. Therefore, please make sure that the reference to the keyword is visible in your invitation and that its meaning and importance for the respondents is clear. Furthermore, make sure that you use a unique keyword – please avoid obvious keywords such as TUM, Lean, Six Sigma, Green Belt, or similar trivial terms.

Questionnaire Data

The collected data of all participants will be published regularly in the course unit: 01/03 Downloads in the file: Questionnaire-Data-Sheet. It contains these tables:

- **Quest Data-Export**: This table contains all collected data of the survey so far. Copy this data including the variable names into the first row of a Minitab worksheet
- **Quest Data-Legend**: This table contains the questions from the questionnaire, the associated names, scale levels and the value ranges of these variables. Additionally, the table contains derived standardized summarizing indices (Risk, Knowledge, Motivation, City-Support).
- **Relation Questionnaire-GembaWalk**: This table sketches a design on how to relate the questionnaire data to the *Gemba-Walk* data by rearranging them. This is optional but you are invited to try this out.
- **Quest Basis**: This table contains the setup with some basic data on the questionnaire and serves as a reference for the other tables. Please leave this unchanged.

This information may seem complex and confusing at first glance. Therefore, I will summarize the important information for you and explain some details.

Data sheet: Quest Data-Export

Preparing data for subsequent analysis is an important step, and often a laborious, annoying, and tedious one. Personally, I was repeatedly surprised by the discrepancy between the data requested from a company's archives and the format in which it then was provided. In many cases, aggregated data was sent instead of the required raw data, with percentages, or the parameters of subsamples. Unfortunately, you can't always tell from the variable names what they stand for. Thus, all important information on the data in the table: **Quest Data-Export** is described in the table **Quest Data-Legend**.

The Quest Data-Legend (Tab 7-12a-g) contains 4 columns:

1. Questions of the questionnaire
2. Variable names
3. Scale-Level of the variables
4. Value Range of the variables representing the answer options

Administration Data:

The first four variables contain the record ID of respondents, the individual day of processing, its processing start and end-time.

Question/ Meaning	Variable Names	Scale-Level	Values/ Range
Serial record number	ID	cardinal	1 - N
Day of processing	Date	Date/ Time	
Start time of processing	Start-Time	Date/ Time	
End time of processing	End-Time	Date/ Time	

Table 7-12a Quest Data-Legend - Administration data

Littering Risk:

A questionnaire collects individual attitudes - not individual behaviour. However, with this questionnaire we can estimate the littering risk of a person - however, without being able to determine the confidence interval of this estimate. Risks are typically derived by the probability and the severity of an event and serve as the problem (Y) in this questionnaire.

Questions Y_Q1 - 3 are related to the probability of littering. The answer options to these questions are 11-level ordinal scales (Tab 7-12b).

Question Y_Q4 is related to the severity of littering. Here, the 7 litter-types (food, paper, ...) are to be ranked by willingness to litter them outdoors, from 1= least likely to 7= most likely. This one question in the questionnaire results in 2 x 7 variables in the data sheet. The variable: Y_Q4_sevPubThrown_1 in the first block contains the answer with the estimation of the highest likelihood to litter. According to our predefinition this should be: *Food*. But the respondent could of course also have answered: *Cigarette butts* - the most harmful litter type. Thus, with these 7 variables, the individually assessed ranking order of littering willingness is assigned to the predefined ranking order. Ideally, for all respondents, the variable: Y_Q4_sevPubThrown_1 would only contain only *Food*, the variable: Y_Q4_sevPubThrown_2 only *Paper*, etc.

The scale level is nominal, and the value range corresponds to the 7 litter-types.

May I ask you how often you eat, drink or possibly smoke in public outdoor places?	Y_Q1_freqPubConsum	ordinal	0 - 10
Would you tell me, how often you throw away garbage outdoors, like bottles, cups, bags, or cigarette butts?	Y_Q2_freqOutdoorLit	ordinal	0 - 10
Do you throw trash out the window when you are on the road, e.g. from the car, bus or train?	Y_Q3_freqTravelLit	ordinal	0 - 10
What would you basically drop outdoors? Please rank these waste types from: top = most likely to bottom = least likely.	Y_Q4_sevPubLitThrown_1	nominal	Food; Paper; Metal; Chewing gum; Glas; Plastic; Cigarette butts
	Y_Q4_sevPubLitThrown_2	nominal	Food; Paper; Metal; Chewing gum; Glas; Plastic; Cigarette butts
	Y_Q4_sevPubLitThrown_3	nominal	Food; Paper; Metal; Chewing gum; Glas; Plastic; Cigarette butts
	Y_Q4_sevPubLitThrown_4	nominal	Food; Paper; Metal; Chewing gum; Glas; Plastic; Cigarette butts
	Y_Q4_sevPubLitThrown_5	nominal	Food; Paper; Metal; Chewing gum; Glas; Plastic; Cigarette butts
	Y_Q4_sevPubLitThrown_6	nominal	Food; Paper; Metal; Chewing gum; Glas; Plastic; Cigarette butts
	Y_Q4_sevPubLitThrown_7	nominal	Food; Paper; Metal; Chewing gum; Glas; Plastic; Cigarette butts
	Y_Q4_sevPubLitThrown_ord_1	ordinal	1 - 7
	Y_Q4_sevPubLitThrown_ord_2	ordinal	1 - 7
	Y_Q4_sevPubLitThrown_ord_3	ordinal	1 - 7
	Y_Q4_sevPubLitThrown_ord_4	ordinal	1 - 7
	Y_Q4_sevPubLitThrown_ord_5	ordinal	1 - 7
	Y_Q4_sevPubLitThrown_ord_6	ordinal	1 - 7
	Y_Q4_sevPubLitThrown_ord_7	ordinal	1 - 7
Combined and standardized Littering Probability	Y_litProb_std	ordinal	0 - 100%
Combined and standardized Littering Severity	Y_QlitSev_std	ordinal	0 - 100%
Combined and standardized Littering Risk	Y_litRisk_Ind	ordinal	0 - 100%

Table 7-12b Quest Data-Legend – Littering Risk data

In the second variable block (Tab 7-12b) - Y_Q4_sevPubLitThrown_ord_1 - 7, the nominal entries are recoded into the corresponding ordinal values: *Food*= 1, *Paper*= 2, ..., *Cigarette butts*=7. The variable: Y_Q4_sevPubThrown_ord_1 would thus contain the correct value: 1 - if the answer was *Food* – but of course also contain the other, deviant values from respondents. This recoding allows the predefined ranking order to be numerically compared with the individually assessed ranking order and an overall Litter-Risk-Index can be derived.

Litter-Risk-Index:

1. The littering probability is standardized to the range: 0 - 100%. Y_Q1 (the frequency of outdoor consumption) serves as a moderator variable for Y_Q2 (general littering) and Y_Q3 (throwing things out of the window while driving).
   Y_litProb_std= Y_Q1 * (Y_Q2 + Y_Q3) / 2
2. The littering severity is calculated via the ratio of: respondents sum of deviations from the predefined target values to the max. possible deviations from the predefined target values.
   Y_litSev_std= Sum($x_{response\_i} – T_{predefined\_i}$) / Sum(max. deviations)

3. The litter-risk-index is the product from litter probability times severity
   Y_litRiskInd= Y_litProb_std * Y_litSev_std

Knowledge about Litter Harmfulness

The knowledge of a person about the harmfulness of the 7 predefined litter types, and their harmful effects in the 7 predefined locations presumably determines the risk of littering and, therefore, the actual behaviour. Thus, knowledge is assumed to be an influence (x) on the littering problem (Y).

Question x_Q5 is related to the knowledge about the harmfulness of the 7 litter-types (Tab 7-12c). Unlike Y_Q4 the 7 litter-types (food, paper, ...) are to be ranked according to their estimated harmfulness, from 1= least to 7= most harmful. Also this one question of the questionnaire results in 2 x 7 variables in the data sheet. The variable: x_Q5_knowPubLitHarm_1 in the first block contains the answer with the estimation of the least harmful litter-type. According to our predefinition this again should be: *Food*. But the respondent could of course also have answered: *Cigarette butts* - the most harmful litter type. Thus, with these 7 variables, the individually assessed ranking order of harmfulness is assigned to the predefined ranking order. The scale level is nominal, and the value range corresponds to the 7 litter-types.

In the second variable block (Tab 7-12c) - x_Q5_ knowPubLitHarm_ord_1 - 7, the nominal entries are recoded into the corresponding ordinal values: Food= 1, Paper= 2, ..., Cigarette butts=7. The variable: x_Q5_ knowPubLitHarm_ord_1 would thus contain the correct value: 1 - if the answer was *Food* – but of course also contain the other, deviant values from respondents. This recoding allows the predefined ranking order to be numerically compared with the individually assessed ranking order and an overall Litter-Knowledge-Index can be derived (see below).

Question x_Q6 is also about knowledge, but here about the harmful effects of litter in predefined locations on the environment. These 7 locations (shopping-zone, catering-zone, ..., water) are to be ranked, from 1= least to 7= greatest harmful effects. Again, this single question results in 2 x 7 variables in the data sheet. The variable: x_Q6_knowPubLitLoc_1 in the first block contains the location with the estimated least harmful effects of litter on the environment. According to our predefinition this should be: *Shopping-Zone*. But the respondent could of course also have answered: *Water Area* - the location with the most harmful effects on the environment. Thus, with these 7 variables, the individually assessed ranking order of harmfulness of the effects in different locations is assigned to the predefined ranking order. The scale level is nominal, and the value range corresponds to the 7 location-types.

How would you rank the environmental harmfulness of these waste types? Please rank them from: top = least harmful to bottom = most harmful.	x_Q5_knowPubLitHarm_1	nominal	Food; Paper; Metal; Chewing gum; Glas; Plastic; Cigarette butts
	x_Q5_knowPubLitHarm_2	nominal	Food; Paper; Metal; Chewing gum; Glas; Plastic; Cigarette butts
	x_Q5_knowPubLitHarm_3	nominal	Food; Paper; Metal; Chewing gum; Glas; Plastic; Cigarette butts
	x_Q5_knowPubLitHarm_4	nominal	Food; Paper; Metal; Chewing gum; Glas; Plastic; Cigarette butts
	x_Q5_knowPubLitHarm_5	nominal	Food; Paper; Metal; Chewing gum; Glas; Plastic; Cigarette butts
	x_Q5_knowPubLitHarm_6	nominal	Food; Paper; Metal; Chewing gum; Glas; Plastic; Cigarette butts
	x_Q5_knowPubLitHarm_7	nominal	Food; Paper; Metal; Chewing gum; Glas; Plastic; Cigarette butts
	x_Q5_sevPubLitThrown_ord_1	ordinal	1 - 7
	x_Q5_sevPubLitThrown_ord_2	ordinal	1 - 7
	x_Q5_sevPubLitThrown_ord_3	ordinal	1 - 7
	x_Q5_sevPubLitThrown_ord_4	ordinal	1 - 7
	x_Q5_sevPubLitThrown_ord_5	ordinal	1 - 7
	x_Q5_sevPubLitThrown_ord_6	ordinal	1 - 7
	x_Q5_sevPubLitThrown_ord_7	ordinal	1 - 7
How big are the harmful effects of litter in these outdoor places? - Please rank these locations from: top = least harmful to bottom = most harmful.	x_Q6_knowPubLitLoc_1	nominal	Shopping Zone; Catering Zone; Traffic Zone; Green City Areas; Kids Zone; Forest, Fields; Water
	x_Q6_knowPubLitLoc_2	nominal	Shopping Zone; Catering Zone; Traffic Zone; Green City Areas; Kids Zone; Forest, Fields; Water
	x_Q6_knowPubLitLoc_3	nominal	Shopping Zone; Catering Zone; Traffic Zone; Green City Areas; Kids Zone; Forest, Fields; Water
	x_Q6_knowPubLitLoc_4	nominal	Shopping Zone; Catering Zone; Traffic Zone; Green City Areas; Kids Zone; Forest, Fields; Water
	x_Q6_knowPubLitLoc_5	nominal	Shopping Zone; Catering Zone; Traffic Zone; Green City Areas; Kids Zone; Forest, Fields; Water
	x_Q6_knowPubLitLoc_6	nominal	Shopping Zone; Catering Zone; Traffic Zone; Green City Areas; Kids Zone; Forest, Fields; Water
	x_Q6_knowPubLitLoc_7	nominal	Shopping Zone: Catering Zone: Traffic Zone: Green City Areas: Kids Zone: Forest. Fields: Water
	x_Q6_knowPubLitLoc_ord_1	ordinal	1 - 7
	x_Q6_knowPubLitLoc_ord_2	ordinal	1 - 7
	x_Q6_knowPubLitLoc_ord_3	ordinal	1 - 7
	x_Q6_knowPubLitLoc_ord_4	ordinal	1 - 7
	x_Q6_knowPubLitLoc_ord_5	ordinal	1 - 7
	x_Q6_knowPubLitLoc_ord_6	ordinal	1 - 7
	x_Q6_knowPubLitLoc_ord_7	ordinal	1 - 7
Combined and standardized Litter-Types Harmfulness Knowledge	x_litHarmKnow_std	ordinal	0 - 100%
Combined and standardized Litter-Locations Harmfulness Knowledge	x_locHarmKnow_std	ordinal	0 - 100%
Combined and standardized Littering Harmfulness Knowledge	x_harmKnow_Ind	ordinal	0 - 100%

Table 7-12c Quest Data-Legend – Littering Knowledge data

In the second variable block (Tab 7-12c) - x_Q6_ knowPubLitLoc_ord_1 - 7, the nominal entries are also recoded into the corresponding ordinal values: Shopping Zone= 1, Catering Zone= 2, ..., Water Area=7. This recoding allows the predefined ranking order to be numerically compared with the individually assessed ranking order and an overall Litter-Knowledge-Index can be derived.

Litter-Knowledge-Index:

1. The litter-type harmfulness knowledge is calculated via the ratio of: respondents sum of deviations from the predefined target values to the max. possible deviations from the predefined litter-type target values. This leads to a standardized value range from: 0 – 100%
   $x_{litHarmKnow\_std} = Sum(x_{response\_i} - T_{predefined\_i}) / Sum(max.\ deviations)$
2. The location harmfulness knowledge is calculated via the ratio of: respondents sum of deviations from the predefined target values to the max. possible deviations from the predefined location target values. This leads to a standardized value range from: 0 – 100%
   $x_{locHarmKnow\_std} = SUM(x_{response\_i} - T_{predefined\_i}) / Sum(max.\ deviations)$

3. The knowledge-risk-index is the average of litter-type harmfulness knowledge and location harmfulness knowledge
   $x_{harmKnowIndInd} = Mean(x_{litHarmKnow\_std}; x_{locHarmKnow\_std})$

Motivation to (avoid) Litter

The motivation of a person is - besides knowledge - another influence on the risk of littering and thus on the concrete behaviour. Psychology distinguishes between two types of motivation: intrinsic and extrinsic. Intrinsic motivation is based on values and derived beliefs that are relatively stable over time. Extrinsic motivation is based on external incentives to behave in a specific situation in a certain way. The two types of motivation are considered in the questionnaire.

The questions **x_Q7 - 10** are aimed at the enduring intrinsic motivation of the respondent (not) to litter, and the questions **x_Q11 - 16** at the situational extrinsic incentives (Tab. 7-12d).

How far do you carry a piece of trash to a trash can?	x_Q7_motLitCarryDist	ordinal	0 - 10
How often do you pick up strange litter that is on your path and take it to the nearest trash can?	x_Q8_motPickOtherLitter	ordinal	0 - 10
How would you characterize the condition of your own apartment?	Y_Q9_motHomeCond	ordinal	0 - 10
Do you separate waste at home, e.g. paper, plastic, glass, organic waste, chemicals and residual waste, in different containers?	x_Q10_motHomeWasteSep	ordinal	0 - 10
When I am pressed for time, I may litter.	x_Q11_motTimePressure	ordinal	0 - 10
Sometimes I put garbage in a place where it cannot be directly seen.	x_Q12_motHideWaste	ordinal	0 - 10
In places where there is already garbage lying around, it is not so bad to add something to it.	x_Q13_motBrokenWindow	ordinal	0 - 10
When I'm out with others, I'm more likely to also litter than when I'm out alone.	x_Q14_motWithOthers	ordinal	0 - 10
I don't mind the garbage on the ground. And a bit more or less doesn't matter.	x_Q15_motPersContrNegleg	ordinal	0 - 10
Throwing garbage away is not a problem. The garbage removal service is paid to clean it up.	x_Q16_motRemovServResp	ordinal	0 - 10
Combined and standardized Intrinsic Motivation	x_motivIntrisic_std	ordinal	0 - 100%
Combined and standardized Extrinsic Motivation	x_motivExtrinsic_std	ordinal	0 - 100%
Combined and standardized Motivation to (not) litter	x_motivBalance_Ind	ordinal	- 100% - 100%

Table 7-12d Quest Data-Legend – Littering Motivation data

The answer options to these 10 questions are 11-level ordinal scales. This allows to directly an overall Litter-Motivation-Index.

Litter-Motivation-Index:

1. The intrinsic motivation is calculated as the average of the answers to: x_Q7 – x_Q10
   x_motivIntrinsic_std= Mean(x_Q7; … , x_Q10)
2. The extrinisic motivation is calculated as the average of the answers to: x_Q11 – x_Q16
   x_motivExtrinsic_std= Mean(x_Q11; … , x_Q16)

3. The balanced motivation index is calculated as the difference: intrinsic motivation - extrinisic motivation. The value range is between: 100% (max intrinsic motivation to not litter and min extrinsic incentives to litter) and: -100% (min intrinsic motivation to litter and max extrinsic incentives to litter)
   x_motivBalance_Ind= x_motivIntrinsic_std - x_motivExtrinsic_std

Support by the City

Questions x_Q17 - 19 relate to perceptions of how well the hometown is managing litter reactively and proactively. The answer options to these questions are 11-level ordinal scales. This allows to derive an overall Litter-City-Support-Index (Tab 7-12e).

How well is your town equipped with trash cans?	x_Q17_cityTrashCanAmount	ordinal	0 - 10
How clean are the streets and public places of your city in your opinion?	x_Q18_cityOverallClean	ordinal	0 - 10
My city actively promotes litter reduction, such as signs on trash cans and public campaigns.	x_Q19_cityProactiveMeasures	ordinal	0 - 10
Combined and standardized Support by the City	x_citySupport_Ind	ordinal	0 - 100%

Table 7-12e Quest Data-Legend – Support by the City

City-Support-Index:

1. The overall city-support is calculated as the average of the answers to: x_Q17 – x_Q19
   x_citySupportInd_Ind= Mean(x_Q17; x_Q18, x_Q19)

Information about the respondent + assignment key

Questions **x_Q20 - x_Q24** (Tab 7-12f) ask for sociodemographic information about the respondent: Place of residence - country and city, the gender, age, and education. The answers can be given as free text and via nominal categories. The nominal entries to: age (Q23) and education (Q24) are recoded into the corresponding ordinal values. This allows to quantitatively relate these ordinally recoded variables to other numerically scaled variables. Please note, that the value: 0 is in both variables related to: *missing value* and correspondingly recoded to: *blank*.

Country	x_Q20_intCountry	nominal	all countries possible
City	x_Q21_intCity	nominal	all cities possible
Gender	x_Q22_intGender	nominal	Woman; Man; Non-binary; Prefer not to say
Age	x_Q23_intAge	nominal	< 15; 15 - 20; 21 - 30; 31 - 40; 41 - 50; 51 - 60; 61 - 70; 71 - 80; > 80; Prefer not to say
	x_Q23_intAge_ord	ordinal	1; 2; 3; 4; 5; 6; 7; 8; 9; 0
Education	x_Q24_intEducation	nominal	Primary; Secondary; Vocational Training; Bachelor; Master; PHD; Prefer not to say
	x_Q24_intEducation_ord	ordinal	1; 2; 3; 4; 5; 6; 0
Feedback	x_Q25_intFeedback	nominal	all entries possible
Keyword for/of the respective investigator to filter a subsample (subgroup) with invited respondents	x_Q26_intkeyword	nominal	published keyword of the inviting investigator

Table 7-12f Quest Data-Legend – Sociodemographic information about the respondent

The **keyword** in x_Q26 serves to filter/ subgroup your respondents from the respondents of other participants. Please evaluate the correct spelling of this keyword by your respondents.

Please note:

1. Responses to risk of littering (Y) and motivation to litter (x) are susceptible to response tendencies toward social desirability. This subjective whitewashing is human because we typically want to portray ourselves as better than we are. Responses to knowledge about the harmfulness of Litter and its harmful effects in different locations tend to be uninfluenced by response tendencies.
2. The standardization of the value ranges 0 – 10 to percentage values from 0 – 100% are basically not allowed for ordinal scales, as the intervals between scale values are unknown. The application of linear transformations in the risk- knowledge- and motivation-indices, e.g. calculation of the mean, are for the same reason not defined for ordinal-scales. Although this is common practice, you should thus put your interpretations under reserve.

Summary

Important influences and relationships from the questionnaire as well as important influences from the field study are summarized in Fig. 7-24.

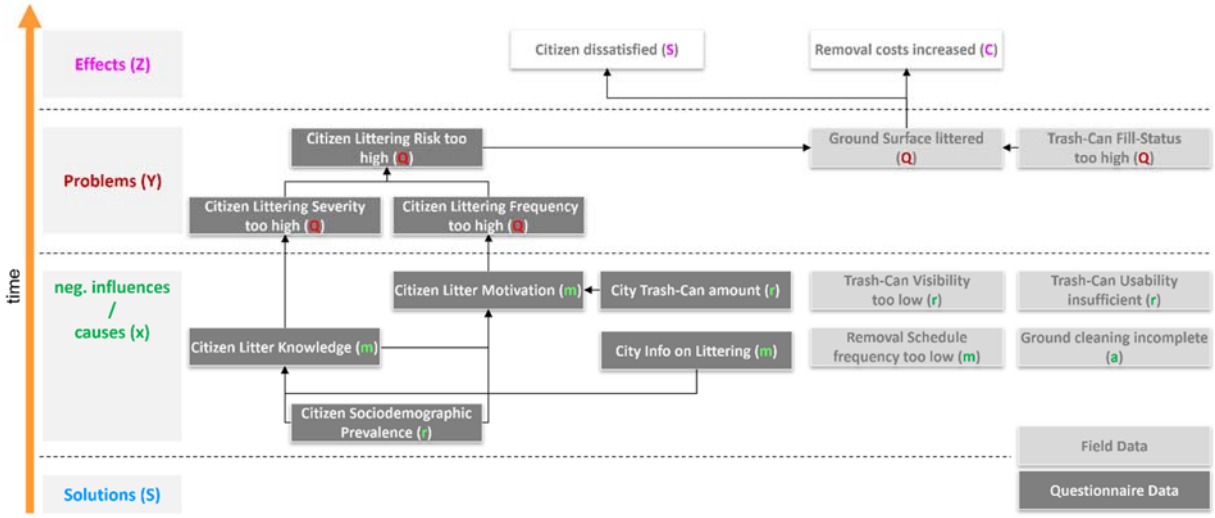

Figure 7-24 Assumed relationships between influences covered by the questionnaire

Please note: The arrows between the problems indicate, that there also can be relationships between the problems (Y). In these cases, also in - Business-Projects – they indicate the relationship from an independent variable to a dependent variable, as we typically indicate by x-Y relationships.