

Smart Survey Implementation

Grant Agreement Number: 101119594 (2023-NL-SSI)

Work package 2: Research Methodology

Deliverable M24: Smart advanced stage

Version 1.1 (final deliverable)

Prepared by:

Maren Fritz & Florian Keusch, University of Mannheim, Germany

Johannes Volk, Lasse Häufiglückner & Karen Blanke, DESTATIS, Germany

Claudia De Vitiis, Barbara D'Amen, Fabrizio De Fausti, Francesca Inglese, Barbara Maria Lorè, Angela Pappagallo, Francesco Piccolo, Marco Terribili & Monica Perez, ISTAT, Italy

Theun Pieter van Tienoven, Vrije Universiteit Brussel, Belgium

Patrick Lusyne, Statistics Belgium, Belgium

Danielle McCool, Peter Lugtig & Bella Struminskaya, Utrecht University, the Netherlands

Anne Elevelt, Jelmer de Groot, Maaïke Kompier, Barry Schouten, Jonas Klingwort, Jeldrik Bakker, Deirdre Giesen, Daphne Debije, Janelle van den Heuvel, Jasmine Saitovic & Marina Veldhuizen Statistics Netherlands, the Netherlands

Loreline Court, Fanny Mikol, Simon Moreau, Simon Quantin, Julie Solard, INSEE, France

Nina Berg & Aina Holmøy, SSB, Norway

ESSnet co-ordinator: Remco Paulussen

Workpackage Leader:

Peter Lugtig

Utrecht University

p.lugtig@uu.nl

Index

1 Introduction.....	4
2 Enhancing Recruitment Strategies for Smart Surveys in Official Statistics: Experiments and Insights from Task 2.1	7
2.1 Introduction.....	7
2.2 Standard definitions of participation behavior	7
2.3 Study results	8
2.4 Conclusion	17
3 Machine Learning in smart surveys: insights from task 2.2	19
3.1 Introduction.....	19
3.2 Receipt scanning.....	20
3.3 COICOP classification microservice	26
3.4 GeoService microservice	33
3.5 Transport mode microservice	37
3.6 HETUS activity microservice	46
3.7 Conclusions and take-home messages.....	53
4 Human Computer Interaction and Usability: insights from task 2.3.....	57
4.1 Introduction.....	57
4.2 Test design.....	61
4.3 Results	64
4.4 Receipt Scanning Microservice (RSM)	65
4.5 User interface and user experience	68
4.6 Scanning and creating tentative data.....	72
4.7 Usability attributes	77
4.8 Editing and committing data	78
4.9 Manual entry	82
4.10 Checking expenses overview	86
4.11 Learning curve and support.....	86
4.12 Trust and privacy	88
4.13 GeoService microservice	93
4.14 Tentative geolocation data.....	100
4.15 Geolocation-based entries vs manual entries.....	102
4.16 Trust and privacy	104
4.17 Take home messages: general	108
4.18 Take home messages Receipt Scanning Microservice	109
4.19 Take home messages Geo Service Microservice	111

5 Mode effects and other data integration considerations: Insights from task 2.4	113
5.1 Introduction.....	113
5.2 Country experiments.....	113
5.3 Respondent differences	114
5.4 Outcome differences	118
5.5 Mode effects of specific smart features.....	126
5.6 Conclusions.....	131
6 Conclusion: How does a smart survey methodology look like?	133
6.1 How should you set up a smart survey?	134
6.2 An End-To-End methodology for a smart survey	138
6.3 A continuum of smart survey methodologies	142
6.4 A research agenda	144
6.5 Overall conclusion	147
Appendix A – Country reports around recruitment.....	148
Appendix B – Machine Learning and point-of-interest data.....	387
Appendix C – Country reports on UI/UX tests.....	392
Appendix D – Detailed results from French experiments around mode effects	834
Appendix E – The energy microservice	887

1 Introduction

The goal of the WP2 'Methodology' workpackage of the Smart Survey Implementation project (SSI) is to find out what general methodological elements trusted smart surveys should have so that they can be used in statistical production by European NSIs. Each task focuses on either an 'opportunity' or 'threat' that was identified in the TSS I framework and pilot recommendations for smart surveys. The four subtasks are:

1. The successful **recruitment** of participants for smart surveys.
2. Using **machine learning** to improve Human-Computer Interaction in smart surveys.
3. Usability and **Human-Computer Interaction** in smart surveys.
4. Integrating smart surveys with traditional survey methods by estimating **the mode effect**.

We refer to deliverable M6 (Review stage) for a discussion of learnings from past findings on projects conducted in the context of the European Statistical System and the wider academic context with regards to these four key challenges, and deliverable M14 (smart baseline stage) for an overview of how the central questions from deliverable M14 resulted in a series of smaller and larger field tests that were carried out between 2024 and 2025. In this deliverable, findings from the M6 and M14 deliverables are in places summarized when this is necessary to understand the current deliverable, but details in the design and motivations of our study are here not explained in detail. The goal of the current deliverable is concentrated on how to carry out a smart survey; the methodology of recruiting respondents, how to deal with sensor data, design the app and user Interaction in the app, and how to integrate smart surveys. The goal was to develop an end-to-end methodology for smart surveys. The two use cases that we concentrated on were the Household Budget Study (HBS) and Time Use Survey (TUS), which both are surveys that are part of the European Statistical System (ESS).

Of central importance are several large and small field tests. The large tests aim to answer the question of how respondents can be successfully recruited into smart surveys (task 2.1) and how to integrate smart surveys with traditional surveys (task 2.4). The large field tests were conducted in Norway (HBS), France (TUS & HBS), Belgium (TUS), and Germany (HBS). Norway and France used a smart survey app which was self-developed, and Germany and Belgium used the MOTUS platform as developed by Hbits. All countries used the general population as the target population and drew fresh samples to conduct the field test following a general design, where some key elements of the field tests are shared across the countries. Respondents are recruited using an offline method (e.g., recruitment via interviewers or postal mail). This allows for the comparison of country-level differences in, for example, the success of particular recruitment strategies. Apart from common elements to the fieldwork design, country-specific variations on the design were also used to tailor tests to local circumstances, and to test specific design elements related to recruitment (task 2.1) or mode measurement effects (task 2.4). Deliverable M14 explains these choices in detail.

Chapter 2 discusses the outcomes of these tests for recruitment, and concludes that it is difficult to recruit respondents from fresh-probability samples. The use of interviewers in a tailored invitation strategy can really benefit the success of recruitment for smart surveys. Within the project, additional tests around different ways to use interviewers were scheduled for Italy and the Netherlands, but these tests did not happen within the timeframe of this project. In the case of Italy issues around the conclusion of a DPIA, and in the Netherlands issues in the IT-system at Statistics Netherlands precluded results from these tests to be included in this deliverable. These will be published separately after the conclusion of the project.

Chapter 3 discusses how sensor data that are included as smart elements in a smart survey should be processed. For the HBS pictures of shopping receipts formed the basis of the smart data, whereas for TUS these were geolocation data. In particular, this chapter focuses on how to guarantee that processed sensor data are of sufficient quality to be used in practice. Geolocation data are processed to generate a pre-filled time-diary of travel-episodes and non-travel episodes. In HBS, pictures of receipts are used to extract relevant product lines, read in the products and their prices, and subsequently link these to standard Coicop codes of products. This chapter shows how Machine Learning can be used to process sensor data, but also illustrates some of the difficulties there are on relying on Machine Learning only. In some cases, the quality of processed sensor data is insufficient; a human-in-the-loop may be necessary to further improve the quality of the data.

Chapter 4 directly follows on the chapter on Machine Learning and studies how to integrate process sensor data in a smartphone app, and design the Human-Computer interaction between the respondent and app. At the core of this chapter are a series of small tests conducted in every country throughout the project. The goal of these smaller experiments is to technically test some of the microservices developed in workpackage 3 that process the sensor data, test the Machine Learning standards developed in task 2.2 and integrated in the microservice, and finally to test the Human-Computer Interaction features of smart surveys. This chapter describes the aspects of smart surveys for two microservices that process smart data in detail. It documents aspects of smart surveys that work well and highlights issues respondents face in practice in interacting with steps in the response process. From this chapter follow specific recommendations on how to improve the smart surveys related to TUS and HBS, but also for smart surveys in general.

Chapter 5 investigates how outcome statistics change when moving from a traditional diary study in the context of Household Budget or Time Use, towards a smart survey. One of the main reasons to move to a smart survey is to decrease respondent burden and improve measurement quality. The chapter finds that indeed measurement quality changes when moving to a smart survey. For Time Use, these changes can be relatively large, and consist of changes due to increasing missing data problems, and improved measurement. Suggestions are given to potentially reduce measurement differences between survey modes, for example by tackling the issue of missing data in Time Use diaries, and how to estimate the size of the mode measurement effect in detail using statistical modeling techniques.

This deliverable is concluded by chapter 6. This chapter integrates findings from the different chapters, and establishes a framework for how to design a smart survey methodology. We conclude that there is not one way to design a smart survey methodology. The specific design of a smart survey should be topic- and country-specific. Some of the country-specific considerations for how to design a smart survey may depend on aspects that extend beyond methodological issues, such as legal/ethical considerations (see WP5 of this project), the platform and IT infrastructure used (see WP 3), and most importantly organisational and business processes (See WP4). Still, we believe there is a common and fundamental choice to make on *how smart* a smart survey should be, and that a choice early on the design process can help guide all other choices.

Finally, this deliverable contains a large section of appendices. These appendices are all designed as stand-alone documents that can be read when interested in more detailed results. These include country-specific results from the large field tests (Appendix A), details on what database to use to enrich geolocation data with the purpose of a location visits in the context of TUS (Appendix B), country reports from the usability tests (Appendix C), detailed results from mode-effect studies carried out in France (Appendix D) and a report on a feasibility study to develop a smart survey around energy use (Appendix E).

We hope this deliverable will provide an impetus for further developing smart surveys within official statistics and beyond in the next years. This deliverable aims to show how smart surveys can be conducted successfully. It however also provides lots of specific areas where further research and tests are needed. We hope this deliverable will therefore both serve as a basis, but also as inspiration for the further development of smart surveys in years to come.

Utrecht, 26 April 2025

The authors

2 Enhancing Recruitment Strategies for Smart Surveys in Official Statistics: Experiments and Insights from Task 2.1

2.1 Introduction

This chapter of the deliverable presents the outcomes of Task 1 in Workpackage 2, which focuses on recruitment strategies for smart surveys within the context of official statistics. The primary objectives of Task 1 were twofold: first, to gain insights into participation behavior in smart surveys and assess the viability of smart surveys for official statistical purposes, and second, to identify key factors influencing participation behavior in these surveys.

To achieve these objectives, a series of experiments focusing on the recruitment of participants to smart surveys were conducted across several countries, including Belgium (TUS), Germany, and Norway (both HBS). In addition, a large field test was conducted in France (HBS) that focused on mode comparisons where participants were first interviewed face-to-face and then either assigned to a PAPI diary or could choose between PAPI and an app (see also Task 4 for more details). The design of the individual country field tests are described in deliverable M14 and the individual country reports with results from the different experimental field tests can be found in Appendix A of this report.

The findings of the large field tests highlight that response rates for smart surveys in official statistics are relatively low, and they vary between countries and study designs. Additionally, we identified specific participation characteristics that are associated with lower participation in smart surveys: being of older age, having a nationality other than that of the country in which the study was conducted, having a lower education level, and living in a household with an income below the median. The experiments conducted also reveal that certain design features, such as the use of CATI interviewers in the recruitment process can positively influence response rates, offering valuable insights into the optimization of smart survey methodologies for official statistics.

2.2 Standard definitions of participation behavior

To achieve a common understanding of participation behavior across countries, we build on participation rates as defined by the American Association for Public Opinion Research (AAPOR, 2023). These response rates are well-established indicators in survey methodology and provide a comprehensive overview of respondent engagement, further allowing for comparisons between surveys and across countries. Reporting these AAPOR-defined rates enables a nuanced understanding of participation behavior at different stages of the survey process.

While there are no standard definitions specifically for smart surveys yet, we can apply the logic of web-push surveys where specifically named persons are contacted in one mode (e.g., postal mail) who are then pushed to participate in another mode, here a smartphone app. When calculating response rates in such a hybrid approach, disposition codes related to participant ineligibility or unknown eligibility are determined by considerations related to the sample frame, and among those who are eligible, interview disposition codes are determined by the data collection mode. Since app-based surveys are not specified in the AAPOR Standard Definitions, we use the logic of web surveys here.

2.2.1 Response rate

The AAPOR Response Rate 1 (RR1) is the number of complete interviews (I) divided by the total number of cases, including both complete (I) and partial interviews (P), non-interviews (such as refusals and break-offs (R), non-contacts (NC) and other non-interviews (O)), and cases of unknown eligibility (UH, UR, UO). This rate provides a comprehensive view of overall participation, accounting for all outcomes. The AAPOR RR1 indicates the proportion of the target population that ultimately provided complete

data. Complete interviews (I) are defined in Belgium (note that data were collected in Belgium in two batches with different diary lengths: 7 days in May-June 2024; 2 days in September-December 2024) as cases where activities were recorded on all days of the diary and in Germany as cases where expenditures were recorded on all 14 days of the diary. In Norway and France completes were defined as cases where expenditures were recorded on at least one day.

$$RR1 = \frac{I}{(I + P) + (R + NC + O) + (UH + UR + UO)}$$

2.2.2 App activation rate

Specifically for our studies, we additionally calculate an app activation rate (AAR). That is, the number of all cases in which the app was activated and data collection started divided by all cases. Using the AAPOR Standard Definitions, the number of started cases is the sum of completed interviews (I), partial interviews (P), and the number of break-offs (all cases with code 2.12 among the refusals (R)). The app activation rate informs us about how successful the recruitment was in pushing sample members to start using the app.

$$AAR = \frac{I + P + R_{2.12}}{(I + P) + (R + NC + O) + (UH + UR + UO)}$$

2.2.3 Break-off rate

A second rate that is not specifically defined in the AAPOR definitions but informative for the large field tests is the break-off rate (BOR). We define the break-off rate as one minus the number of completed interviews (I) divided by the sum of completed interviews, partial interviews (P), and break-offs (Rs with code 2.12). This rate informs us about the proportion of people who started in the app but did not complete the entire survey.

$$BOR = 1 - \frac{I}{I + P + R_{2.12}}$$

2.3 Study results

2.3.1 Participation behavior

Table 3.1 summarizes the three rates based on AAPOR classifications for each participating country. Our findings indicate that the participation behavior in the large field tests varied substantially between the four countries. First looking at app activation rates, we see stark differences between countries that employed interviewers in the recruitment process and those who did not. In Norway, 25% of contacted individuals activated the app. On the contrary, in Germany and Belgium, both countries where study invitations were sent via postal mail, around 10% of invited sample members activated the app. Other factors such as the use of a progressive web app, which does not have to be installed on the phone but can be accessed via a browser, in Norway and varying trust in the NSIs might have played a role as well. For France, an app activation rate could not be calculated to lack of access to the raw data.

These differences between countries then also translate to different response rates; France (43.0%) and Norway (22.5%) exhibiting much higher response rates than Belgium (2.3%) and Germany (1.8%). However, it is important to note that the differences in the response rate as well as the break-off rates, might also be attributed to different study lengths and the way completes were defined across countries. For example, Belgium conducted a 7- and a 2-day diary and defined completes as respondents with diary entries on all days. Germany used a 14-day diary and completes were also

defined as respondents with entries on all days (see German field report in the Appendix for different definitions of completes). On the contrary, both Norway and France used a 7-day diary and counted completes as those who provided at least one entry to the diary.

Table 2.1: Participation rates for all participating countries

Country	Gross sample (n)	Completed interviews (I; n)	Break-offs (R _{2.12} ; n)	App activation rate (AAR; %)	Break-off rate (BOR; %)	Response Rate 1 (RR1; %)
BEL	1,269 (7-days)	28 ¹	93	9.5%	23.1%	2.2%
	5,044 (2-days)	118 ¹	420	10.7%	21.9%	2.3%
GER	7,049	126 ¹	152	5.7%	68.3%	1.8%
NOR	1,973	444 ²	66	25.3%	11.2%	22.5%
FRA	2,500	1,076 ³	207 ⁴	- ⁵	16.1%	43.0%

Notes:

¹Cases in Belgium and Germany were counted as completed if activities or expenses were entered on all days of the diary.

²Cases in Norway were counted as completed if expenses were entered on at least one day of the diary.

³Cases in France were counted as complete if expenses were entered in the app or the paper diary on at least one day.

⁴Break-offs in France are cases that completed the initial face-to-face interview but did not complete the diary.

⁵No information on app activation available.

Overall, the findings suggest that the crucial challenge in smart surveys is both getting people to download/activate an app and start with the study but also to continuously provide data until the end of the field period.

2.3.2 Non-participation bias

Next, we assess bias due to nonparticipation in the field tests, based on a comparison of the distribution of variables available on the sampling frame between the gross sample (i.e., sample members invited to participate in the study) and the net sample (i.e., those who completed the study). Note that different variables were available on the sampling frame of the countries, and that no information on the gross sample was available from France. Following Couper et al. (2018), we calculate bias in a variable of interest (y) is the difference in proportions between the gross sample (g) and the net sample (n) as

$$bias(y) = y_n - y_g$$

The standard error (SE) of these differences is calculated as:

$$SE(y_n - y_g) = \frac{n_g - n_n}{n_g} \sqrt{(var(y_n) + var(y_{np}))}$$

Where (np) denotes non-participants.

As shown in Table 3.2, we see a relatively consistent non-participation bias regarding nationality and education across Belgium, Germany, and Norway (note that no frame data was available from France to calculate non-participation bias). In all three countries, people with a nationality different from that country were underrepresented and citizens of the country were overrepresented in the net sample compared to the gross sample. In Belgium, we also see regional differences; people living in Flanders were strongly overrepresented in the net sample compared to people from Wallonia, which might be explained by differences in language across the regions. For Germany and Norway, where no language differences exist across regions, no significant differences were detected. In Belgium and Norway, people with lower educational attainment and those living in households with an income below the median were underrepresented and people with high educational attainment and in households at or above the median income were overrepresented (no education and household income data available on the sampling frame for Germany). In Germany, we see strong differences by age with younger people being overrepresented and older people being underrepresented. People in the oldest age group are also underrepresented in Belgium, while in Norway we only see relatively small bias in age, and people in the age group 45 to 66 years even being overrepresented in the net sample.

Additionally, we found that the education and income bias in Norway could be reduced when CATI interviewers were used for recruitment (see Norway field report in the Appendix for further details). In the group where CATI interviewers were used in the recruitment non-participation dropped from -7.7 to -6.8 for people with low educational attainment, from -8.8 to -0.6 for people with medium education attainment, and from 16.3 to 10.5 for people with high educational attainment. Similarly, CATI interviewers reduced the non-participation bias from -8.2 to -3.9 for people in households below median income and from 5.5 to 4.2 for people in households at or above median income.

Table 2.2: Non-participation bias regarding different sociodemographic variables

		BEL		GER	NOR
		7-days	2-days		
Gender	Male	-5.1 (-12.8, 2.5)	-3.4 (-11.8, 4.9)	-3.2 (-11.8, 5.5)	-3.3 (-7.4, 0.9)
	Female	5.2 (-2.3, 12.7)	3.5 (-4.7, 11.8)	3.2 (-5.4, 11.8)	3.3 (-0.7, 7.2)
Age	18-24	0.6 (-1.7, 3.0)	1.0 (-1.8, 3.8)	20.9 (13.1, 28.7)	-0.1 (-1.8, 1.6)
	25-44	8.0 (0.8, 15.2)	8.6 (0.5, 16.8)	4.9 (-3.6, 13.3)	-2.1 (-6.2, 2.0)
	45-66	0.2 (-7.4, 7.8)	0.6 (-7.7, 8.9)	-20.4 (-27.9, -13.0)	4.1 (0.2, 8.0)
	67+	-8.7 (-13.3 -4.1)	-10.1 (-14.2, -6.0)	-5.3 (-8.0, -2.6)	-1.8 (-4.2, 0.5)
Nationality	Own	20.1 (15.4, 24.9)	18.4 (13.5, 23.4)	13.2 (8.7, 17.7)	5.4 (1.8, 9.0)
	Foreign	-20.1 (-25.4, -14.7)	-18.3 (-23.4, -13.3)	-11.4 (-14.8, -7.9)	-6.2 (-10.0, -2.4)
Education	Low	27.9 (21.5, 34.4)	25.5 (17.5, 33.6)	NA	-7.0 (-8.0, -4.1)
	Medium	-19.6 (-24.7, -14.5)	-18.4 (-23.3, -13.5)	NA	-3.8 (-7.8, 0.3)
	High	-1.2 (-8.4, 6.1)	-0.6 (-8.5, 7.3)	NA	12.6 (9.0, 16.3)

	Un-specified	-7.1 (-9.5, -4.7)	-6.6 (-8.2, -5.0)	NA	-1.8 (-3.6, -0.1)
Region	Flanders	15.1 (8.5, 21.7)	17.5 (10.3, 24.7)	E GER -6.7 (-14.5, 1.1)	Oslo & Viken 1.3 (-2.6, 5.3)
	Wallonia	-15.0 (-22.2, -7.8)	-17.4 (-24.7, -10.1)	W GER 6.7 (-1.1, 14.4)	Inland -0.1 (-2.1, 1.8)
					Agder & SE Norway -1.1 (-3.8, 1.7)
					W NOR -2.2 (-5.8, 1.4)
					Trøndelag 0.4 (-2.0, 2.9)
					N NOR 1.6 (-0.5, 3.7)
Size of community	<5,000	NA	NA	-1.4 (-7.0, 4.3)	NA
	5,000-49,999	NA	NA	3.3 (-4.5, 11.0)	NA
	50,000-499,999	NA	NA	-2.6 (-9.6, 4.4)	NA
	500,000+	NA	NA	0.7 (-7.8, 9.1)	NA
Net household income	Below median	-21.9 (-28.8, -14.9)	-22.0 (-29.1, -15.0)	NA	-5.3 (-9.5, -1.1)
	Equal to or above median	23.8 (17.5, 30.1)	23.9 (16.8, 30.9)	NA	4.5 (0.5, 8.5)

Notes: Bias in percentage points, 95%-confidence intervals in parentheses; Bias where CIs do not overlap with 0 are bolded.

2.3.3 Factors influencing survey participation in smart surveys

As part of the large field tests in Belgium, Norway, and Germany, experiments were conducted with the aim to identify factors influencing participation in smart surveys. The central question guiding this analysis is whether specific design features in the recruitment aimed at increasing participation were effective.

In Belgium, experiments aimed to boost survey participation by using tailored design methods and addressing digital literacy and privacy concerns, including an adapted invitation letter and offering a Paper and Pencil Interview (PAPI) option when reminding non-participants in the TUS. In Germany, efforts focused on communicating the simplification of the data collection in the HBS by utilizing the smartphone's camera for receipt-scanning, thereby reducing participant effort. In Norway, the experimental approach centered on exploring how the use of a trusted and familiar source for accessing the app (the official government website Altinn) influences the willingness to participate in smart surveys within the HBS context, as well as testing the use of interviewers in the recruitment. To assess the effectiveness of these interventions, bivariate tests (Chi-squared tests, t-tests, ANOVA) were employed, examining the influence of each treatment on participation behavior measures in the

respective country. Tables 2.3 through 2.5 provide an overview of the experimental designs, and the hypotheses postulated in deliverable M14, and it summarizes the key findings from each country (note that some of the hypotheses stated in the M14 deliverable could not be tested due to slight deviations between the originally planned and the actually implemented design of the studies).

Table 2.3: Research design, research questions, hypotheses, and results of experiments of large-scale field test in Belgium

Notes: s. ... significant test result ($p < 0.05$); n.s. ... not significant test result ($p > 0.05$).

Country	Design		Research question	Hypotheses	Results		
BEL	<i>Invitation Letter</i>		RQ1: How does the overall design of the invitation letter affect the response rate?	H1: Incorporating user-friendly elements in the design of the invitation will significantly improve response rates in a smart survey. (not confirmed)	<i>Response rates (7-days)</i> 2.4% (traditional) 2.0% (user-friendly) n.s. ($\chi^2(1) = 0.04$, $p > 0.05$)		
	Traditional					User-friendly	
	<i>PAPI follow-up</i>	Yes					318 (25.1%)
	No					315 (24.8%)	319 (25.1%)
	7-day diary (May-June)				<i>Response rates (2-days)</i> 2.3% (traditional) 2.3% (user-friendly) n.s. ($\chi^2(1) < 0.01$, $p > 0.05$)		
	<i>Invitation Letter</i>		RQ2: How does the inclusion of a follow-up PAPI mode affect the response rate in a smart survey? ¹	H2.1: Offering a PAPI mode positively influences the overall response rate in surveys. (not confirmed)	<i>Response rates (7-days)²</i> 2.6% (PAPI) 1.9% (no PAPI) n.s. ($\chi^2(1) = 0.32$, $p > 0.05$)		
	Traditional					User-friendly	
	2,524 (50.0%)	2,520 (50.0%)					
	2-day diary (September-December)						

¹H2.2 could not be tested in Belgium due to a change in the design.

²H2.1 was not tested for the 2-day diary because this experiment was not conducted in the second wave.

Table 2.4: Research design, research questions, hypotheses, and results of experiments of large-scale field test in Germany

Country	Design		Research question	Hypotheses	Results		
GER	Camera function mentioned in invitation		RQ1: How does highlighting the smart features of a survey in the invitation influence its response rate?	H1.1: Stressing the camera feature of the survey in the invitation positively affects the response rate. (not confirmed)	Response rates 1.7% (camera & effort mentioned) 2.1% (camera mentioned, effort not) 1.6% (both not mentioned) n.s. ($\chi^2(2) = 1.89, p > 0.05$)		
	Effort mentioned in invitation	No				2,350 (33.3%)	2,350 (33.3%)
		Yes				2,349 (33.3%)	-
			RQ2: To what extent does highlighting the benefits of the scanning function lead to increased usage of this feature?	H2: Emphasizing both the effort required to complete the survey and the benefits of the scanning feature will lead to higher usage of the scanning function by individuals, compared to merely mentioning the scanning function without stressing the	Proportion of respondents with at least one diary entry who used scanning function at least once 75.3% (camera & effort mentioned) 63.5% (camera mentioned, effort not) 41.8% (both not mentioned) s. ($\chi^2(2) = 22.44, p < 0.01$)		
					Average number of scanned tickets (among respondents who provided at least one entry to diary) 5.1 (camera & effort mentioned) 3.8 (camera mentioned, effort not) 1.9 (both not mentioned)		

effort involved.
(confirmed)

s. ($F(2, 171.73) = 10.47, p < 0.01$)

Percentage of scanned tickets out of all diary entries (among respondents who provided at least one entry to diary)

18.4% (camera & effort mentioned)

15.0% (camera mentioned, effort not)

8.7% (both not mentioned)

s. ($F(2, 181.86) = 8.07, p < 0.01$)

Notes: s. ... significant test result ($p < 0.05$); n.s. ... not significant test result ($p > 0.05$).

Table 2.5: Research design, research questions, hypotheses, and results of experiments of large-scale field test in Norway

Country	Design			Research question	Hypotheses	Results
NOR	<i>Download link</i>			RQ1: Does the type of download site for the app influence response rates?	H1: Higher response rates are associated with downloading the app from a governmental site (Altinn). (not confirmed)	Response rates 21.3% (Altinn) 24.0% (direct link) n.s. ($\chi^2(1) = 1.85, p > 0.05$)
		Altinn	Direct link			
	CATI interviewers used in recruitment	Yes	496 (25.1%)	RQ2: How does the inclusion of a follow-up CATI mode affect the response rate in a smart survey?	H2.1: Using CATI interviewers in the recruitment process positively influences the overall response rate. (confirmed)	Response rates 29.0% (CATI) 16.1% (no CATI) s. ($\chi^2(1) = 46.51, p < 0.01$)
		No	500 (25.3%)			

Notes: s. ... significant test result ($p < 0.05$); n.s. ... not significant test result ($p > 0.05$).

In Belgium, the redesigned invitation letters and the PAPI follow-up option did not have a significant effect on the response rate. In Germany, mentioning the camera feature in the invitation letter did not influence the response rate but led to a significantly higher share of respondents who used the camera function in the diary at least once and to significantly more frequent use of the camera function. In Norway, the use of CATI interviewers in the recruitment significantly increased the response rate. This effect was substantive with a 13 percentage points increase in the CATI interviewer group compared to the group that just received the regular invitation messages via letters and SMS messages.

2.4 Conclusion

Task 1 of Workpackage 2 of the SSI project marks a significant contribution to our understanding of participation behavior in smart surveys and an advancement in refining recruitment strategies for such surveys for official statistics in Europe. By integrating innovative smart features with traditional survey techniques, the experiments conducted in Belgium, Germany, and Norway provide valuable insights into addressing key challenges around willingness to participate in smart surveys and potential threats to data quality due to non-participation bias.

One finding from the large field tests is that it is very difficult to recruit probability samples of the general population to smart surveys without the use of interviewers. In countries where sampled individuals were invited by postal mail to the large field test (i.e., Belgium and Germany), the share of invited individuals who activated the app (6-10%) and eventually completed the diary (2-3%) was low. While these participation rates are relatively low compared to the ones in Norway and France where interviewers were used in the recruitment process, they are not much smaller than the outcomes of studies that invite individuals to complete web surveys via a postal invitation ("push-to-web") and use similar types of conditional incentive. Against this background, one conclusion is that it is not necessarily the smart nature of the survey that deters people from participating but the non-personal invitation mode. Our findings also show that changes in the invitation letters have very little effect on the participation decision. Of course, the findings of the different field tests as part of the SSI project have to be interpreted in light of different study designs (e.g., different field periods, use of progressive web app vs. app that has to be downloaded to the phone) and different country-specific contexts (e.g., regarding trust in NSIs), and we cannot rule out other factors that might have led to the differences in participation behavior across the countries involved in the field test.

The analysis of non-participation bias highlights that certain groups of people are more likely to participate in smart surveys than others. Across all three countries where data about the gross sample was available, we found that people with a nationality other than the country where the test was conducted were significantly underrepresented in the net sample. Nationality in this context might be a proxy for language proficiency, and these findings might be explained by the fact that in all three countries the smart survey was only available in the main national language. We also found bias for other variables that would indicate problems in line with the digital divide. People with lower educational attainment, people who live in households with lower income, and people of older age were less likely to participate. The findings from Norway provide a positive outlook in that using interviewers in the recruitment process was an effective way of reducing non-participation bias due to differences in educational attainment and household income between the participants and the gross sample showing that interviewers might be able to help those who have more difficulties with the technology to participate.

Finally, the experiment in Germany showed that stressing the functionality of the smart feature (i.e., the receipt scanning function) in the study invitation led to a significant and substantive increase of the use of that feature in the diary without reducing the participation rate. This finding shows that advertising the advantages of smart surveys in the invitation can encourage study participants to use that feature, potentially providing higher-quality data at reduced participation burden.

Going forward, the findings of the large field tests provide valuable insights for other NSIs who plan on conducting smart surveys. While the field tests were not able to study the role of interviewers as part of the data collection in more detail, the current findings show that the well-known advantages of interviewer administration in terms of higher participation rates translate from traditional surveys to smart surveys. In countries where interviewers cannot be used, NSIs will need to consider other approaches in the recruitment of participants for smart surveys, including, for example, the use of unconditional incentives and relying on non-probability samples of volunteers. Of course, these approaches come with their own limitations in terms of legal requirements, more complex implementation, and threats to generalizability of the findings.

References

- Bucher, H., Keusch, F., de Vitiis, C., de Fausti, F., Inglese, F., van Tienoven, T. P., Mccool, D., Lugtig, P., & Struminskaya, B. (2023). *Smart Survey Implementation: Workpackage 2: Research Methodology Deliverable M6: Review stage*. Grant Agreement Number: 101119594. European Statistical System.
- Bucher, H., Keusch, F., de Vitiis, C., de Fausti, F., Inglese, F., van Tienoven, T. P., Mccool, D., Lugtig, P., & Struminskaya, B. (2024). *Smart Survey Implementation: Workpackage 2: Research Methodology Deliverable M14: Smart baseline stage*. Grant Agreement Number: 101119594. European Statistical System.
- The American Association for Public Opinion Research. 2023. *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*. 10th edition. AAPOR.
- Couper M.P., Gremel, G., Axinn, W., Guyer, H., Wagner, J., & West, B.T. 2018. New options for national population surveys: The implications of internet and smartphone coverage. *Social Science Research* 73:221-235.

3 Machine Learning in smart surveys: insights from task 2.2

3.1 Introduction

Machine learning (ML) algorithms play an important role in smart surveys to process smart data, although these new types of data provided by sensors (i.e. photo, GPS data, meter data) can be transformed into statistical variables also developing not supervised algorithms. How to deal with smart data is the key question for subtask 2.2 of a smart survey methodology and this chapter.

The SSI project in fact, has developed several microservices (MSs) that implement components to process the smart data collected in a survey, in order to obtain statistical information in the form of prediction or tentative data. Some MS implement ML models, other different types of algorithms. These components are intended to be used independently by the NSI platform.

The level to which automation can be brought to replace the direct acquisition of information or replace manual processes without degrading data quality and/or increasing respondent burden, is the crucial point in the use of algorithms to process smart data (ML or not). In this respect, the user interface plays an important role in facilitating the more or less complex tasks that the respondent is asked to carry out (see Chapter 4).

The problem in the use of ML in a survey then becomes: 1) Build the automation part, 2) Design a mechanism whereby machine alerts human (respondents and/or staff) when it needs input and 3) Design an efficient UI to facilitate human machine interaction (Benedikt et al., 2020).

Key questions were:

1. Under what circumstances results from ML models can be used directly as statistical data, and under what circumstances data should be fed back to respondents?
2. What to do when the quality of the machine learning outcome is too low?
3. When should respondents be asked to provide new input (a picture or open text) because no meaningful information could be extracted?
4. Underlying all the above processes, are the training datasets used in the ML and how to create training data sets.
5. How and when should training datasets be updated or improved?

Generally, ML methods require continuous updating, depending on the applied technique. In the context of SSI, retraining concerns especially Receipt Scanning Microservice (RSM).

This chapter reports of the experience carried out in SSI with the aim to extrapolate the results regarding the processing of sensor data, both using ML and other type of algorithm (not supervised), reviewing the documentation produced in WP3 to describe the microservices and the methodology implemented.

The developed microservices process data collected through a smart survey in two different contexts: photos of receipts and GPS data¹.

¹ The SSI project also dealt with the development of the Energy Microservice, for which a pilot was conducted by CBS but focus of the test was on the feasibility of the data collection and on the behaviour of the respondent, not on the algorithms for processing the data from the meter

In the first context the Receipt Scanning Microservice, including the OCR and COICOP classification microservices has been developed, composed by two components, the first mainly non domain specific and the second specific for the Household Budget Survey (HBS):

- OCR
- COICOP classification.

In the second context, the GeoService Microservice derives from GPS data - geolocation points collect using internal sensors of smart devices - information on stops and tracks using spatial and temporal parameters. Stops and tracks are contextualized using third-party POI databases, such as OpenStreetMap or Google Places. The GeoService Microservice is composed of three components, the first non domain specific and the other two specific for Travel survey and Time Use survey:

- GPS trajectory segmentation (GeoService microservice)
- Transport mode prediction
- Hetus activity prediction

Here we briefly describe the choices made regarding the algorithms implemented, underlining the quality issues regarding the input and the throughput data, extrapolating the requirements and the recommendations to enhance the quality of the output (predicted data). Data quality requirements are then defined considering some general quality dimensions concerning the three levels of the process: input data, throughput and output.

We report also the improvements obtained and the remained open problems and challenges, trying to outline some guidelines for the NSI and the users of these microservices. More in general, one goal of subtask 2.2 is to develop methodological standards around the use of the machine learning models or other type of models in smart surveys.

Furthermore, this chapter summarizes the relevant results of the small tests performed in WP2, with respect to the use of microservices (see Chapter 4). These small tests represent a test of the end-to-end solution but almost exclusively from a usability point of view (respondent journey), not including respondent perceptions on the use and evaluation of (ML) predictions. In any case, we try to bring out from the results some evidences on the actual use of microservices by users and, therefore, on the effect of the implementation of the underlying algorithms on the experience of the respondents.

3.2 Receipt scanning

In this section, we report the methods developed for the smart HBS based on data collected via receipts (photos) in the Receipt Scanning Microservice. The aim is to highlight in a concise and schematic way the most relevant aspects concerning the use of the ML.

The Receipt Scanning Microservice² is supported by two microservices:

- the OCR Microservice (part 1)
- the COICOP Classification Microservice (part 2).

² The contents are taken from SSI deliverable 3.4.

Both microservices are based on a complex pipeline to extract relevant information from receipts such as store names, dates, items purchased and prices and classify products into the corresponding 5-digit Classification of Individual Consumption by Purpose (COICOP).

The focus is on understanding how automation affects the quality of output and when the automated pipeline designed into microservices needs to be supplemented by human intervention (both by the respondent and the NSIs).

The pipelines of the two microservices include:

- for the OCR Microservice, the implementation of ML models both in the OCR algorithm core, and in the pre-OCR and post-OCR steps (see OCR microservice);
- for the COICOP classification microservice, the classification pipeline has been optimized and harmonised based on a mix of machine learning and string matching (see COICOP classification microservice).

3.2.1 OCR MICROSERVICE

Method

The microservice to process scanned receipts is based on Optical Character Recognition (OCR), and automated coding, using different AI/ML models in its pipeline. The OCR Microservice returns all available information that can be used by the platform (and its UI/mobile and web application) and/or accordingly to the COICOP Classification Microservice.

The microservice is developed according to a complex pipeline that includes pre-OCR and post-OCR processing steps. The first phase is functional to improve the quality of the image, as this has an impact on the accuracy of the OCR output; pre-OCR consists in cropping image to receipt (remove background) and receipt orientation.

The second phase, in which the correction of specific receipt information is performed, is functional to improve the accuracy and usability of the OCR output.

The models implemented in the pre-OCR processing can follow two approaches for receipt detection, semantic segmentation and object detection, at user's choice. These models separate receipt from background: respectively, SegFormer segments areas of text from complex backgrounds and YOLOS detects individual words or lines of text. The two models have different characteristics in terms of accuracy and processing time/cost. The SegFormer model has high accuracy due to a multiscale architecture, is efficient (simple and fast decoder) and adapts to different resolutions, but is computationally expensive compared to traditional CNNs, and overloaded for simple tasks (e.g. text and background separation). The YOLOS model is fast (end-to-end processing), good at recognising discrete elements (e.g. words, lines of text) and less sensitive to changes in scale than CNN-based YOLO, but it is less accurate on very dense or distorted text and requires a lot of annotated data for effective training.

For an optimal choice of the two models, it is advisable to use:

- SegFormer if you need to separate text from complex backgrounds.
- YOLOS to quickly detect blocks of text or words.

The choice also depends on the desired trade-off between accuracy and speed.

For the development of the OCR step to extract the text, different OCR models were compared: EasyOCR, Tesseract, PaddleOCR, TrOCR. From these, the optimised and fine-tuned PaddleOCR model was selected to provide the best accuracy for the specific task required. PaddleOCR is an open source OCR framework optimised for real-world scenarios, supporting multiple languages and special characters, and providing pre-trained models for text extraction. The model has significant advantages: it has high accuracy for printed text, works well with standard fonts (used in receipts), even on noisy backgrounds, and supports skewed and distorted text. It is also a light and fast model. The model has some performance limitations if the receipts have special fonts (italicised text or manual annotations may reduce performance), complex layouts, very dense text, tables or non-linear structures. The accuracy of the model also depends on the quality of the image: crumpled, reflective or low-resolution receipts reduce accuracy.

PaddleOCR is an excellent choice due to its speed and accuracy. Optimising PaddleOCR on receipts therefore requires a pre-processing phase, which is currently implemented in the microservice, a fine-tuning phase, which must be performed by training on a customised dataset of receipts, and a post-processing phase, also implemented in the microservice.

Post-processing phase (post-OCR) uses OCR output that includes text and text locations (bounding boxes). That information is used:

- to understand the receipt i.e. trying to give a meaning to the recognized text (by OCR),
- to correct OCR mistakes (to be implemented),
- to produce a final output which contains all receipt details (and some metadata).

In order to understand the receipts, the pipeline applies a fine-tuned model of LiLT (Language-Independent Layout Transformer) that combines text and layout (text position) information to label a text box. A variety of labels were defined within the SSI project to arrive to standardisation, ranging from store address to tax price.

LiLT is a model designed to understand layouts and text even in languages not seen during pre-training. It is particularly useful for improving the structuring of OCR output (e.g. grouping logical fields in receipts/invoices). Fine-tuning LiLT requires: preparation of the input dataset containing the raw OCR text (output from PaddleOCR) and labels (with 'date', 'price', 'shop' fields); configuration of the model with training; customised post-processing.

LiLT is language independent, works well with languages not in the original training set, takes into account the position of text (e.g. column alignment in invoices), improves raw OCR and extracts specific fields. On the other hand, it is computationally expensive and requires annotated data.

In addition, to improve the accuracy and usability of raw OCR, text analysis techniques are applied to extract and correct specific information on the receipt. This post-processing phase includes main steps. Output of receipt understanding and its corrections, contains all found information as well as metadata.

In the following table, the main characteristics of microservice, pipeline and algorithms development are summarised.

Table 3.1: Characteristics, pipeline and algorithms development for OCR

OCR MICROSERVICE (RECEIPT SCANNING MICROSERVICE - PART 1)	
Characteristics	<ul style="list-style-type: none"> • non-domain specific • AI/ML models implemented
OCR Pipeline	
Input data	<p>images (typically only one image)</p> <p>(pdf) e-ticket (which accordingly is processed as an image)</p>
Steps for OCR development:	<ol style="list-style-type: none"> 1. Pre-OCR: receipt detection and rotational correction 2. OCR: text detection and text recognition 3. Post-OCR: machine learning-based and rule-based logic <ul style="list-style-type: none"> • document (receipt) understanding • final post-processing which has as an output a json file.
1.Pre-OCR: pre-processing the received image or pdf	<p>Processing to improve image quality => <i>OCR accuracy</i></p> <p>Pre-OCR step: receipt detection, orientation correction and image cropping.</p> <p>Output of the pre-OCR step => correctly oriented, cropped receipt image</p> <p><i>Image correction is under development</i></p>
Receipt detection approaches	<ul style="list-style-type: none"> • Semantic segmentation (SegFormer model) • Object detection (YOLOS model) <p>The object detection is very accurate but less fine-grained than semantic segmentation. Furthermore, it requires quite some processing to correct the receipt orientation. The orientation of the image is corrected by sequentially applying PaddleOCR to derive text orientation.</p>
i. SegFormer Model for Semantic segmentation	<p>The training process for the SegFormer model, is designed to segment and identify the area of store receipt on photos. It consists of a few steps:</p> <ul style="list-style-type: none"> – Data Preparation – Definition of Parameters – Training Procedure <p>A guide for training this model can be found in the code repository</p>
ii. YOLOS Model for Object detection	<p>The training of the object detection model holds 4 steps:</p>

	<ul style="list-style-type: none"> – Collect images and divide them into 3 subsets: train, test and validation for YOLO model – Rotate all images – Annotate and create the output – Train and retrain <p>A guide for training this model can be found in the code repository</p>
2.OCR (optical character recognition) model	<p>PaddleOCR model to extract text</p> <p>A fine-tuned PaddleOCR model provided the best accuracy for this task.</p>
PaddleOCR Model training	<p>Training on a small, labelled dataset of ~300 receipts</p> <p>Training can be split into two groups: detection model and the recognition model.</p> <ul style="list-style-type: none"> • The detection model detects words in an image and binds them to a bounding box with annotation. • The recognition model recognizes shapes within that bounding box and tries to classify them (by 'assigning' meaning to them e.g. characters). <p>A guide for training this model can be found in the code repository</p>
Output	<p>Output of the OCR step includes text and text locations (bounding boxes)</p>
3.Post-OCR: post-processing OCR	<p>The information of the OCR step is used:</p> <ol style="list-style-type: none"> to understand the receipt i.e. trying to give a meaning to the recognized text (by OCR), to produce a final output which contains all receipt details (and some metadata)
i. Receipt understanding	<p>LiLT Model training</p> <p>The training of the receipt understanding model holds 4 steps:</p> <ul style="list-style-type: none"> • Collect receipts: Images need to be correctly oriented and cropped. Collect receipts in 3 sets (train, test, validation) • Apply OCR to get boxes with text: Tools: e.g. PaddleOCR. • Annotate and compose the dataset. • Train and retrain. <p>A guide for training this model can be found in the code repository</p>

ii. Corrections of OCR mistakes	<p>Post-processing methods to extract and correct specific receipt information using text analysis techniques, enhancing accuracy and usability.</p> <p>The main steps for extracting structured information from raw OCR results include:</p> <ul style="list-style-type: none"> • Receipt Date: Extracted using regular expressions that identify date patterns in the OCR output. • Total Price: Identified by scanning for specific keywords. When keywords are detected, the corresponding numerical price values on the same line are captured. • Products and Prices: Columns of products and corresponding prices are identified by examining consistent alignment and similar text height, enabling accurate extraction of product-price relationships. <p>Each step of the extraction can be customized or expanded according to the project's requirements</p>
iii. Final post-processing	<p>Output of receipt understanding and its corrections, contains all found information as well as metadata.</p>

3.2.2 Small test: Destatis and VUB HBS tests

The experiences of the small tests conducted by Destatis and VUB have shown that the preliminary activity of training the models is very expensive and that a very high number of labelled receipts, and therefore a significant number of resources working on them, are needed.

However, carrying out a training phase as complete and extensive as possible is the key to improving both the quality of the output and the respondent's experience. In fact "Poor quality of the scanning results entail a high correction effort. This causes confusion, dissatisfaction and sometimes resignation among the participants." (see next chapter 4 for details). Moreover, "Many errors in scan result compared to the original ticket, result in a high effort to correct", highlighting the crucial importance of an automatic and effective treatment of the scanning errors to limit as much as possible the involvement of the respondent.

In conclusion, there is a high confidence in technology and accuracy of the scan results.

3.2.3 Conclusions and recommendations

OCR accuracy depends on the Pre-OCR and Post-OCR phases and the training of the OCR model where the size of the labelled receipt dataset can be relevant.

There are two ambition levels: one is to understand and classify every potential type of text box to be found on a receipt. This means the objective is to also derive date of purchase, shop names, etc. The less ambitious strategy is to identify only the products and prices. In essence, it comes down to what

is asked from respondents to provide or supplement and what is being extracted. The two options have very different UI-UX implications.

The main recommendation for an NSI that wants to use RSM, having different language, different types of shops, is to represent the retail market variability as much as possible in the collection of receipts to be labelled and used for training. Besides, it is advisable to use the automatic receipt scanning only for the covered supermarket and use the manual entry for the other types of shops.

NSIs should take into account the aspects that may be country-specific:

- Language (some countries in addition have multiple languages, e.g. Switzerland and Luxembourg)
- Receipt lay-out
- Diversity and balance of receipt layouts
- Punctuation
- Currency
- Range of text box types
- OCR performance

In general, it is not possible to define how many tickets are needed to train the models. Each NSI needs to conduct preliminary analysis and evaluations of the model performance, taking into account also the cost of manual annotation. The first step may be to test the OCR trained in one country how it works on receipts from another country. The performance can give an indication of how much training is needed and therefore of the extent of the annotated receipt data set. The amount of required training data is a function of the diversity of formats. These depend on the range of shop types to be accepted (and included in training data), i.e. supermarkets, garden centers, restaurants, cinemas etc. It is good to specify what kind of shops are accepted. A strategy may be to start with supermarkets and gradually expand.

3.3 COICOP classification microservice

3.3.1 Method

This section describes the COICOP Microservice³, which has the task of assigning a 5-digit COICOP code to each product description that has been extracted from a receipt using the OCR Microservice. Hence, the COICOP Microservice is a classical text to classification task.

NSI-specific techniques will have to be developed by each NSI and training is not part of the SSI. Yet, different methodological considerations will be outlined to inform data scientists and methodologists who want to deploy their own NSI specific model into the OCR Microservice. The SSI project will also provide a generic string-matching approach which can be used without advanced technical adaptations, but should be tested thoroughly.

Regardless of the technique employed, assigning a 5-digit COICOP code to a product description requires a dataset linking product descriptions to COICOP codes, referred to here as the *matching*

³ This paragraph is taken from SSI deliverable 3.4.

data. When a new receipt is scanned and product descriptions are identified, the task is to find the best possible COICOP match for each product description based on the matching data. Consequently, an extensive matching data corpus is a prerequisite for the COICOP Microservice.

One source of matching data is scanner data from price statistics, collected from retail scanners at points of sale and typically used for compiling national price indices. This data is comprehensive and systematically gathered but is often limited to specific retail sectors, such as supermarkets or drugstores. This data usually gets automatically classified to COICOP and therefore serves as a directly usable matching data source.

An alternative approach is to gather matching data manually, for example, by collecting receipts directly from consumers or retailers. Once the data is collected, a crucial step is coding all extracted product descriptions to their corresponding COICOP codes before using the data as matching data.

Lastly, manually curated tag lists created and maintained by subject-matter experts can serve as a matching data source. While such lists may not cover the full range of products found in dynamic receipt data, they often include the most essential products.

The techniques implemented for assigning COICOP code to each product row extracted from a receipt (OCR microservice) are:

- **Automatic string matching** involves directly comparing receipt text to given source materials, offering simplicity and efficiency. This technique is straightforward to implement; however, it struggles with variations in spelling, abbreviations, and typos.

Automatic string matching uses a combination of direct string-matching, similarity-based matching, and substring tokenization to identify potential COICOP matches. The integration of fuzzy matching ensures that even when exact matches are unavailable, a well-fitting COICOP code is identified based on string similarities. By prioritizing shop-specific matching data when available, the process enhances both efficiency and accuracy.

- **Machine learning** uses models trained on the source materials, i.e., the receipt texts and the corresponding COICOP labels. The model learns to recognize more generic patterns and structures with which it can classify previously unseen receipt texts to a certain degree. The downside is the significant initial effort required for setting up a training dataset and a training pipeline as well as the need for substantial computational resources, especially if there is a need to train multiple models.

- **Manual string** searching involves the respondent in the coding process. In this solution, each product row of the receipt is entered into a search algorithm and the respondent selects the best-fitting COICOP category. This method is the most burdensome for the participant but can be highly precise since it leverages the respondent's expertise on their own private purchases. Also, the respondent could change the search string if the receipt text itself is not diagnostic enough.

These techniques are applied in sequential way: if no match is identified through the string matching, the machine learning (ML) is initiated. The application of the ML models is preceded by a text preprocessing phase necessary to transform textual data into representations suitable for ML. The models that can be used in this phase are N-gram and FastText (Destatis has adopted a FastText model for integration). N-gram based preprocessing with two classification modelling techniques - logistic regression and random forest - which yielded similar results in terms of accuracy, are explored.

N-grams start by transforming product descriptions into a series of character strings. N-grams can make the pattern more resistant to small variations or spelling errors. FastText learns word embeddings, which are dense vector representations that capture semantic meanings. In addition, FastText incorporates sub-word information, allowing the model to understand and generalise across different word forms and spelling variations. FastText excels in scenarios where text data is rich in linguistic nuance or domain-specific terminology. It is able to train quickly on large datasets, making it very efficient for large-scale classification tasks. Furthermore, its ability to perform well with limited computational resources makes it an attractive choice for (near) real-time applications.

ML generates a prediction even if the classification is uncertain. To mitigate misclassifications, the confidence score of a prediction is taken into account. Predictions with a confidence score above a predefined threshold are automatically accepted. If the score falls below this threshold, manual string searching is activated.

The following table summarizes the main characteristics of the COICOP classification microservice, the structure and the methods implemented.

Table 3.2: Characteristics, pipeline and algorithms development for COICOP classification

RECEIPT SCANNING MICROSERVICE – COICOP Classification Microservice - part 2	
Coicop microservice - task of assigning a 5-digit COICOP code to each product row that has been extracted from a receipt using the OCR Microservice	
Characteristics	<ul style="list-style-type: none"> • Domain specific (specific techniques will have to be developed by each NSI and training is not part of the SSI) • Sequential Structure of the component <p>(1) automatic string matching,</p> <p>(2) machine learning, and</p> <p>(3) manual string searching</p>
Requirement for the deployment	Matching data: scanner data (limited to retail sectors); annotated receipts; manual tag lists
Steps of the process for sequential use of the techniques - Automatic string matching, Machine learning and Manual string searching	
<ul style="list-style-type: none"> • Step 1a: If specific matching data for a given store is available, attempt automatic string matching on the matching data of this selected store. • Step 1b: Attempt automatic string matching on the full matching data. • Step 2: Prediction made from machine learning model trained on the full matching data to find the highest scoring COICOP match. • Step 3: Allow the user to manually search within the full matching data, with the option to alter the search string. • Step 4: COICOP classification by the NSI during data post-processing. <p>Each step is initiated in sequence and only if the previous one does not yield a satisfactory result for a given product description.</p>	

The criteria for such thresholds are country-specific and require calibration.	
<p>STEP 1</p> <p>String-matching</p>	<p>1a - Store-specific string-matching</p> <p>If specific matching data for a given store is available, attempt automatic string matching on the matching data of this selected store</p> <p>1b - Generic string-matching</p> <p>Attempt automatic string matching on the full matching data</p> <p>Step 1 uses a combination of direct string-matching, similarity-based matching, and substring tokenization to identify potential COICOP matches.</p> <p>If no match can be identified, the machine learning step 2 is initiated</p>
<p>STEP 2</p> <p>Text preprocessing</p> <p>Machine learning</p>	<p>Text preprocessing phase</p> <ul style="list-style-type: none"> • N-gram-based transforms product descriptions into a series of character strings • FaxText learns word embeddings, incorporates subword information, allowing the model to understand and generalize across different word forms and spelling variations. <p>N-gram make the model more resilient to minor variations or spelling errors.</p> <p>The accuracy of the FastText model was quite similar to the n-gram based models, but the processing time and efficiency was much better.</p> <p>Both approaches offer unique benefits. N-gram-based models are straightforward to implement and interpret, but can take up much more computational resources than FastText. The choice between these techniques depends on factors such as data availability, computational constraints, and accuracy with NSI-specific data.</p> <p>Machine learning model trained on the full matching data to find the highest scoring COICOP match. ML:</p> <ul style="list-style-type: none"> • has higher flexibility • allowing for the classification of product descriptions that may not exactly match entries in the existing matching data • learn patterns from data, enabling them to generalize to a certain extent beyond the exact words seen during training – particularly useful when dealing with variations in wording, spelling errors, abbreviations, or newly introduced products <p>Success of ML approach:</p> <ul style="list-style-type: none"> • heavily depends on the quality and quantity of labelled training data (i.e., the matching data) • the choice of modelling and pre-processing techniques

	<p>Performance and accuracy of these approaches will depend heavily on country-specific characteristics of the data and should be evaluated individually.</p> <p>Step 2 generate a prediction, even if the ML model is uncertain about its classification. Incorrect classifications is identified by considering the confidence score of a prediction. Only predictions with a confidence score above a predefined threshold should be automatically accepted. If the score falls below this threshold, the product description should remain unclassified (i.e., as a "no match"). It would then get passed on to Step 3</p>
STEP 3	Allow the user to manually search within the full matching data, with the option to alter the search string.
STEP 4	COICOP classification by the NSI during data post-processing.
<p>Analyses to quantify the success of each sub-step:</p> <ul style="list-style-type: none"> – Annotation task (multiple thousand receipts to create a sufficiently large test corpus) carried out to build a test corpus that can be used as a means to validate and test the developed classification pipeline – Analyses of product inventory and dynamics of the available store data – Analyses of classifier performance 	

3.3.2 Data quality

For the COICOP classification microservice, data quality requirements are defined considering the quality aspects involved in the three levels of the process: input data, throughput and output.

Input data:

- Matching data: Coverage, Timeliness and updating
- OCR output: errors

The *coverage* of the matching data is a critical factor on two levels:

- if the data only includes products within a specific COICOP range, the COICOP Microservice should be restricted to those ranges;
- once a specific retail sector is included, it is crucial to ensure that coverage within this sector is sufficiently comprehensive.

Experience tells us that coverage also applies to chains of shops, even within a specific range of COICOP and retail. Results in every country show big diversity in texts between major shopping chains and a model trained on one chain performs poorly for a different one.

Combining multiple matching data sources can significantly enhance both coverage and comprehensiveness.

Timeliness of the matching data is another critical factor. Maintaining and regularly updating the matching data is essential, as new products are frequently introduced. Regular updates are necessary to ensure that the data remain reliable and relevant. Such updates may also concern unclassified product descriptions to be added to the matching data along with the corresponding COICOP codes.

OCR output quality impacts on the accuracy of COICOP classification, but the effect is not an increase in incorrect classifications. Instead, it resulted in a higher number of product descriptions remaining unmatched to a COICOP code, that results in increased manual efforts (in step 3 or step 4) rather than increased error rates.

OCR-NLP errors can be: a) missed or altered characters in a product text, b) missed or altered numbers in a price, c) spurious product-price, d) missed product-price.

3.3.3 Output data

The output of the string matching provides the unclassified cases that become input to the ML step. Unclassified cases must be identified by an error tolerance threshold.

The output of ML generates a prediction even if the classification is uncertain. Unclassified cases must be identified considering the confidence score of a prediction.

In order to assess the quality of the COICOP classification prediction, an accuracy score is defined to establish if the prediction is satisfactory or needs the manual intervention. Error tolerance thresholds have to be set (by each NSI), depending on how tolerant or conservative the pipeline will be. These decisions should be influenced by the amount of human resources a given NSI has available to correct errors in post-production phase.

To find out how much error can be tolerated, a compromise has to be made taking into account the staff resources and the level of automation that one wants to apply in the system, as all predictions below this threshold are subjected to human manual coding. Setting these thresholds, especially if several classification methods are chained together, is a delicate step. These thresholds could also be changed over time if the model seems too conservative or too lean.

Two different tests have been carried out⁴, the first by CBS and the second one by Destatis.

3.3.4 Internal validity test (CBS)

CBS conducted tests where the scanner data used as matching data was divided into two separate sub-datasets. The first one was used as actual matching data, while the second was used as test data to validate the accuracy of the developed techniques and run comparison benchmark tests. This test approach can be called an internal validity test. Conceptually this test assumes that product descriptions on receipts are identical to production description in scanner data records. Given that deviations between these two data sources are likely, the accuracy scores should be interpreted with caution when it comes to the absolute levels. Using this scanner data, we test how well an ML approach classifies product descriptions. These tests are centred around two research questions:

- How well do ML models classify product descriptions into COICOP-classifications, and does it change over time?
- How well do ML models classify product descriptions from supermarkets not encountered during training?

This test considered several ML algorithms, including some string-matching methods for comparison: Logistic Regression, Logistic Regression SGD (Stochastic Gradient Descent, Naïve Bayes (Multinomial)*, Random Forest Classifier**, Multilayer Perceptron Classifier*, Constant Predictor,

⁴ For details see SSI deliverable 3.4

Exact String-matching Algorithm, FastTextClassifier, DeStatis String-Matching Pipeline, DeStatis String-Matching Pipeline (w/o article-ID matching).

The base performances of the ML algorithms were first tested on previously seen supermarkets and the change over time. The tests show that the best model scored on the first month 89,9% accuracy, with a monthly accuracy drop of 0,5%. Similar performances, however, were tested for Destatis string-matching pipeline, meaning that there is no significant difference between the ML approach and the string-matching approach.

The second aspect tested is how well ML algorithms deal with supermarkets absent in the matching data. The results showed that all algorithms performed poorly, with the best ML model scoring on average only 48,1% accuracy. Although ML algorithms performed better than string-matching pipelines (13,3%), they are still poor at classifying unseen supermarkets. In addition, string matching is more sensitive than ML to OCR errors.

In conclusion, ML models classify product descriptions well, but **only for supermarkets on which they are trained**. They **classify poorly on supermarkets absent from the matching data** set. All models perform worse over time due to changes of product catalogues.

3.3.5 External validity test (Destatis)

Destatis collected a sample of receipts from actual supermarkets and manually classified product descriptions to COICOP. This data was then used with the different COICOP classification techniques to validate and benchmark their accuracy and performance. These tests mimic the actual production use case and can be termed external validity tests. However, they come with the limitation that the sample size was not very large and test results should be interpreted with caution as well. The conducted tests focused on evaluating COICOP accuracy and runtime performance.

This test data consisted of the original OCR product description extraction, the manually corrected product description, a COICOP code, and a store name. This provided the opportunity to test the COICOP Microservice with both **corrected and uncorrected OCR strings**. Comparisons between uncorrected and corrected OCR product descriptions revealed **differences, but these were less severe than initially anticipated** in the project. Nonetheless, investing in algorithms to correct OCR-extracted product descriptions would be worthwhile. Microservice pipeline in sequence decreased from 77% to 72% with uncorrected data. Interestingly, this decline in correct classifications did not lead to an increase in incorrect classifications. Instead, it resulted in a higher number of product descriptions remaining unmatched to a COICOP code. It is unlikely that this result generalizes, but in this case, it would result in increased manual efforts (in step 3 or step 4) rather than increased error rates.

The results highlight in conclusion the **added value of the machine learning component**, as it successfully captures cases that the string-matching approach alone could not classify.

3.3.6 Conclusions and recommendation

The COICOP microservice has not been tested in a large-scale field test within SSI, and so conclusions are mostly based on internal tests.

Recommendations on the construction of the **matching data**:

- In order to enhance both coverage and comprehensiveness of the matching data set, it is advisable combining multiple matching data sources.
- To ensure reliable and relevant matching data, regular updates are necessary; frequently unclassified product descriptions should be added to the matching data along with their corresponding COICOP codes; this requires either an automated data base monitoring or manual quality control measures.
- NSIs should consider the dynamic of product in time when deciding frequency of retraining:
 - The EAN/GTIN code of a product changes, implying that a link to COICOP needs to be re-established. Often the consequence of a change in metrics or look.
 - The EAN/GTIN code is the same but the product text changes.
 - New products are introduced.
 - Products are taken out.
 - Products may have seasonal fluctuations.

Recommendations on output accuracy and thresholds:

- Calibrating thresholds at each sub-step seems critical, as error rates depend heavily on how conservatively each step is applied. These values have a large influence on the number of incorrect COICOP classifications in relation to those predictions where no assignment was made (no matching result). Depending on the resources of a given NSI for quality control and manual post-processing, the decision to set these thresholds more conservatively or not will vary. A system was found where an ML prediction is accepted as good enough only if the prediction score was above 0.55 (with some manual rule-based checks).
- It is advisable to be conservative at the beginning of calibration, to obtain less false positive.
- Since these thresholds will vary by country, universal recommendations cannot be provided. However, it is vital for NSIs to dedicate time and resources to this calibration process.
- Chaining the various string-matching sub-steps with the machine learning approach adds predictive power, making sequential use highly recommended, calibrating better the error tolerance in the trade-off with the resources.

Further improvements could be achieved in the future by involving the respondent, for example:

- asking some info on the receipt or about the household habits;
- considering hierarchical classification: If classification at fifth digit is too uncertain, the respondent may be asked to detail a prediction at higher COICOP level.

Employing location-tracking data is potentially very useful for respondents as a framework to construct diaries concerning both travel and daily activities. Processing location-tracking data to predict the behaviour of respondents requires advanced algorithms exploiting different smart features and external auxiliary information.

3.4 GeoService microservice

Within Workpackage 3 of SSI the GeoService Microservice is developed to process sensor data in order to provide tentative input to the timeline of the diary.

What is in common to Travel surveys and Time use surveys is the first processing step, the segmentation in stop and track, which provides the input variables for the algorithms devoted to predict the survey variables. A fundamental part in this process is the addition of auxiliary information

from a third party, such as map services, which provide information about the context around the GPS points and the Point of Interest (POIs) located nearby the GPS points.

The GPS data itself, and the subsequent algorithms used to segment the tracks and predict travel mode, trip purpose and activities undertaken at a location, are not fail-safe: GPS tracks may be incomplete, with missing data due to people's behaviour or the situation in which the GPS signal may be lost; GPS tracks may be complete, but, for example, the trip detection algorithms may be too sensitive (over-identifying trips) or not sensitive enough (under-identifying trips).

Heterogeneity in GPS sensor quality between different types of smartphones affects the measurements and the statistical information derived from them. This heterogeneity becomes a crucial issue in the context of smart surveys, since an uneven distribution of smartphone brands and models can be observed across countries, as shown in Deliverable M14: Smart baseline stage (Appendix A).

The improving contribution on the predicted variables of contextual data used to add location details may be limited by the quality of the map services themselves, which again may vary from one European country to another.

Data quality refers to the degree to which the requirements of a set of intrinsic data characteristics are met. Geospatial data are generally collected and used for many different purposes; the quality of geospatial data must be evaluated considering the purpose and context in which they are used.

An important quality indicator of geographic data for use in predictive modelling is the completeness of map elements (ISO 2022), in particular points of interest (POIs), as it directly influences the prediction of trip distribution, trip purpose and stops. A convenient approach is to examine the quality of different object classes, such as the road network. The most accurate method of quality assessment is to compare a dataset with its true value. However, estimating the true value is expensive, complex and time consuming. A more appropriate method is to compare the quality of a dataset relative to a benchmark dataset of documented high quality.

This issue has been explored in depth comparing the two main mapping services used in the GPS-based methods, Google Places (GP) and OpenStreetMaps (OSM) through a comparison, which have been carried out randomly selecting points within urban areas of 11 countries: the study compares POI data collected within a 50-meter radius around each point. The study highlights a significant difference between the countries and a significant under-coverage of OSM compared to GP. The results are detailed in Appendix A.

3.4.1 Components of GeoService microservice

GeoService Microservice makes use of GPS data to support Time Use, Travel and mobility surveys. The general idea is to provide to the users/respondents a framework of places (stops) and travels (tracks) and the mode of transportation in order to support them in keeping their timeline up-to date.

The design of the GeoService Microservice holds two important elements: the definition of the stop-track clusters and the prediction of the travel mode upon the track clusters and the HETUS activities upon the stops.

The present section focuses separately on the three components of the microservice:

- Stop-track cluster segmentation;
- Transport mode prediction;
- HETUS classification activity prediction.

3.4.2 GeoService microservice – segmentation

The algorithm for segmentation of GPS data in stop and track is – ATS/OPTICS, adaptive algorithm. The algorithm overcomes the use of fixed spatial and temporal thresholds that indicate when the moving object has been stationary for a significant period of time. The trajectory segmentation follows a multigranular perspective and develops an automatic, user-adaptive, parameter-free solution that flexibly adapts the segmentation criteria to the specific user and geographical areas involved.

Paper “Individual and collective stop-based adaptive trajectory segmentation” from Agnese Bonavita, Riccardo Guidotti and Mirco Nanni, as published in *Geoinformation* (2022) 26:451-477.

ATS/OPTICS is powerful but requires careful tuning and can have problems with noisy, sporadic or very dense data.

In the following table, features algorithms and development of the algorithm are summarised.

Table 3.3: Characteristics, pipeline and algorithms development for Geolocation

GeoService microservice for segmentation – based on geolocations	
Segmentation of GPS data in stop and track	
Characteristics	<ul style="list-style-type: none"> • Non-specific domain • Type of algorithm: ATS/OPTICS, adaptive algorithm
PROS	<ul style="list-style-type: none"> • Handles noisy data and variable speeds well. • Not require the number of clusters to be fixed in advance (unlike k-means). • It is able to overcome the presence of artefacts introduced by GPS errors (e.g. an object's coordinates change even though it is not actually moving).
CONS	<ul style="list-style-type: none"> • Sensitivity to parameters. The choice of ϵ (maximum distance), min_points and min_time greatly influences the results. • Difficulties with low-frequency data. • It does not distinguish between significant and irrelevant stops, which is a more semantic classification.
Design and algorithm development	
Input data	<ul style="list-style-type: none"> • From GPS, parameters: timestamp, longitude, latitude and accuracy. The list of geolocation points needs to be time-ordered • Extra information on stop clusters by connecting to a POI or places API: <ul style="list-style-type: none"> ○ Point-of-interest: find places inside or nearby stop clusters (e.g. OpenStreetMap, Google Places)

Steps of Stop detection algorithm	
<ul style="list-style-type: none"> • Filter GPS points based on accuracy, • Determine which GPS points are significant stop points • Cluster the stop points, and • Post-processing <ul style="list-style-type: none"> ◦ reduce number of clusters (merging) ◦ guarantee stops and tracks alternately 	
Pre-processing Filter GPS points	<ul style="list-style-type: none"> • The algorithm requires an accuracy of 100m, but this value is configurable. • Outliers are eliminated by filtering GPS tracking points according to their accuracy
ATS to determine significant stops	<p>Private location (home, work etc.) is always regarded as a stop.</p> <ul style="list-style-type: none"> • GPS points are a stop point when more than t seconds is spent between the current GPS point and the next GPS point that is more than x meters away. • Algorithm can be tuned by changing the temporal and/or spatial parameters. By default, the implementation used 50m as spatial parameter (as advised in the paper) and 180s for the temporal parameter. • Temporal parameter from the GPS data is derived by means of a Thompson tau statistic
Cluster stop points	In order to get stop clusters, the project decided to use the OPTICS algorithm (derived from DBSCAN), which applies a density-based technique on spatial data.
Split spatial clusters based on time	<p>This step is necessary because OPTICS does not take into account the time aspect of the data.</p> <p>Results of cluster stop points step are further processed to split the spatial clusters based on time as well</p>
Post-processing	<p>To deliver a clean output:</p> <ul style="list-style-type: none"> • the clusters of stops are merged, • then the clusters of tracks between the stops are added.
Output data	<p>Set of stops and tracks:</p> <ul style="list-style-type: none"> • to utilize this output as input for the second part of microservice. • It's beneficial to add points of interest (POIs) to the stop and track clusters (GP, OSM).

Small test: Belgium, Germany and Italy

The results of the three small test of the GeoService carried out by StatBel, Destatis and Istat for TUS diary (see following Chapter 4) show that “The general benefit of the geo-assisted entry depends on

users' entry behaviour and their activities". In fact, if respondent entries the activities at the end of the day the GeoService is useful to recall and reduce burden.

As for the relationship between places and activities, the tests show that the GeoServices is useful when the place changes during the same activity, while it is a hurdle when the activity changes in the same place and the respondents has to entry it manually.

3.5 Transport mode microservice

Method

This section documents the development of the transport mode prediction algorithm specifically designed for integration in the Geoservice microservice. The algorithm is based on a decision tree, using GPS data and infrastructure information from OpenStreetMap (OSM).

The development of the algorithm was based on data collected by Statistics Netherlands. The algorithm was also evaluated on open geo-data that is publicly available in the SSI Git-repository (<https://github.com/essnet-ssi/geoservice-ssi>) and was collected within the scope of the SSI project.

Transport mode prediction currently lacks a universally established algorithm. Existing methods predominantly rely on rule-based approaches, decision trees, or machine-learning techniques. For developing a transport mode classification algorithm within the SSI project, it was decided to base it on a decision tree due to its simplicity and interpretability.

We briefly discuss their advantages and disadvantages and compare rule-based approaches and machine-learning models.

Some advantages of decision-tree models are listed below:

- Interpretability: Decision trees provide a transparent, easy-to-understand decision-making process, making them ideal for explaining predictions.
- Non-linearity: They can model complex relationships between input features without requiring linear assumptions.
- Feature importance: Decision trees naturally rank features based on their importance, helping understand key factors affecting transport mode choices.
- Categorical & numerical data: They can process different data types (e.g., GPS coordinates, timestamps, categorical travel modes) without complex preprocessing.
- Computational efficiency: Training and prediction are relatively fast, making them suitable for real-time transport mode prediction in smart surveys.
- Missing data: Decision trees can handle missing values better than some machine learning models using surrogate splits.

Some of the disadvantages of decision-tree models are listed below:

- Overfitting: Decision trees can overfit the training data, leading to poor generalization unless pruning techniques are applied.
- Sensitivity to noisy data: Small variations in input data can lead to different splits, making the model unstable.

- Limited expressiveness: Decision trees can handle complex patterns but may struggle with highly complex relationships between features compared to deep learning models.
- Bias in splitting criteria: Splitting criteria like the Gini index or Information Gain tend to favor features with more levels, which might lead to biased predictions.

Next, a brief comparison of decision trees with rule-based and machine-learning approaches is given. Decision trees are more flexible and scalable than rule-based approaches but may overfit the data. Rule-based methods are static and rely heavily on expert knowledge, which may not generalize well. Decision trees can be trained automatically, while rule-based approaches are handcrafted.

Table 3.4: Performance comparison of two approaches for transport mode prediction

Theme	Decision Trees	Rule-based approaches
Flexibility	Adapts to patterns in data automatically	Requires manually defined rules
Interpretability	Easy to understand	Easy to understand but harder to maintain
Scalability	Scales well with data size	Becomes complex with increasing rules
Handling new data	Can retrain to adjust	Needs manual updates
Accuracy	Higher with enough data	Limited by predefined rules

While decision trees offer simplicity and interpretability, they may not be as accurate as ensemble methods (like Random Forests) or deep learning models. Next, a comparison between decision trees, Random Forests (machine learning), and neural networks (deep learning) is given.

Table 3.5: Performance comparison of decision tree and ML approaches for transport mode prediction

Theme	Decision Trees	Random Forests	Neural Networks
Interpretability	High	Medium (ensemble of trees)	Low (black box model)
Accuracy	Moderate	High	Very High
Computational costs	Low	Low	High
Overfitting	Pruning needed	Less prone (ensemble effect)	Requires regularization
Training speed	Fast	Slower than single tree	Slowest

Alongside developing the decision-tree model, a rule-based algorithm for transport mode classification was also developed for comparison. The results obtained during the development of the rule-based⁵ algorithm contributed to the advancement of the decision-tree model.

The choice was determined by the evidence that the rule based allowed for multiple classifications, requiring a priority handling, which is not needed when using the decision tree because each track will be assigned one predicted transport mode.

The algorithm was developed on a sample from the Dutch general population collected from 2022 to 2023. Data from 255 participants were used for the development. The dataset contains 4.298 tracks and a total of about 20 million GPS observations.

Table 3.6: Characteristics, pipeline and algorithms development for Transport mode

TRANSPORT MODE PREDICTION MICROSERVICE – GeoService Microservice part 2	
Transport mode microservice - task of to each track that has been defined from GPS data using the GeoService (Segmentation ATS/Optics) Microservice	
Characteristics	<ul style="list-style-type: none"> – Domain specific (Travel survey or Time Use Survey) – Type of algorithm: decision tree model
Requirement for the deployment	<p>Required: Events data, GPS data, and map service (OSM) data associated to a track</p> <p>Optional:</p>
Input data	<p>Event data: Contains one row for every track, with: user_id, start_time and end_time</p> <p>Location data: Contains GPS measurements with: user_id, timestamp, longitude, latitude, altitude, bearing, accuracy.</p> <p>OpenStreetMap data: open data from download.geofabrik.de/Europe.html for the regions the GPS points were observed in.</p> <p>The above three sources were used to create the following features:</p> <p>GPS-based features: Calculated for every event, based on the corresponding location data. Features that are created: for {speed, acceleration, jerk, snap, altitude, accuracy, bearing} the following features are calculated {average, standard deviation, skewness, kurtosis, percentiles {5, 10, 15, 20, 80, 85, 90, 95}, IQR, minimum, maximum}. Proportions of speed within intervals {0-5, 5-15, 10-30, 15-30, 30-50, 45-80, 80-120, 120+} km/h.</p>

⁵ The rule-based algorithm with results is included in the Appendix of the deliverable from WP3 and is not included in this deliverable.

	<p>Track features: length, duration, number of data points, time difference with previous observation {average, median, standard deviation, minimum, maximum}, number of long gaps, day of week, weekend indicator, altitude, proportion altitude below sea level.</p> <p>OpenStreetMap-based features: For various OSM point-based tags {bus station, bus stop, railway, light rail, subway, tram, busway, tram stop, railway halt, railway station, bicycle} a count and normalized count was created. For other OSM route-based tags {bus route, metro route, train route, tram route} the distance of the geolocations route summarized as {minimum, maximum, mean, standard deviation}.</p> <p>Some of these variables are used as model input. The variables used as input depend on the training data and may differ slightly for other training sets. See results for feature importance table.</p>
Train/test Data	<p>CBS data (train): In 2022-2023 a large-scale field test in the Netherlands resulted in 4298 labeled events with sufficient quality of GPS observations. This was used as the basis for training the transport mode prediction model.</p> <p>Open SSI data (test): In 2024, five employees of CBS and students from Utrecht University participated in a small-scale field test in the Netherlands and Germany. This resulted in 137 labeled events with high quality of GPS observations and no errors in the labels. This was used to test the transport mode prediction model.</p> <p>OpenStreetMap data (auxiliary): OpenStreetMap data of The Netherlands and of Germany were used to create the OSM-based features for the training and test data.</p> <p>By partitioning users into separate subsets for training (70%) and testing (30%), ensuring that no user appeared in both sets. Stratification was applied based on each user's dominant mode of transport to maintain a balanced representation of transport modes.</p>
Steps of the algorithm	
Pre-processing	<ul style="list-style-type: none"> • Data selection. Pre-screening of “high quality” users is applied, requiring at least 2000 observations per user for at least one hour. Events are removed if they do not meet the requirements of a duration of at least 10 hours and at least 10 observations. • Preparation. To prepare for feature creation, all events are given an event_id and the GPS geolocations are given corresponding event_id. Geolocations without event_id are removed. • Feature creation of GPS-based and OSM-based features.

Development of the decision tree algorithm	<ul style="list-style-type: none"> Python's sklearn package is used to train a decision tree model within the following parameter boundaries: max_depth: [3, 5, 7], min_samples_split: [10, 25, 50], min_samples_leaf: [5, 10, 15, 20, 25], criterion: ['gini', 'entropy'], 'max_features: [1, 3, 5, 7].
--	---

Results

The classification report for the training and test set shows how the model performs.

Table 3.7 Results on classification of training and test data, feature importance

Evaluation on development set: results	
Classification report for training data	<ul style="list-style-type: none"> Model performs well for high-frequency classes like car and walking, achieving precision, recall, and F1-scores around 0.85-0.87, indicating strong and balanced performance for these categories. Bike and train also show relatively high F1 scores (0.79 and 0.73, respectively), with the train having a high recall (0.83) despite moderate precision. The model struggles with underrepresented classes. Other and tram have 0.00 F1 scores, as the model fails to classify these instances correctly. Bus has low performance, with an F1 score of 0.07, mainly due to extremely low recall, meaning most buses are misclassified as other categories (especially cars). Metro performs better, with precision and recall around 0.60-0.61, but still shows room for improvement. The overall accuracy is relatively high at 81%, the macro average F1-score of 0.49 and balanced accuracy of 0.51 indicate poor performance in less common classes. The weighted average F1-score of 0.78 is boosted by the well-classified dominant classes, masking the severe misclassification of minority classes. <p>The model may be biased towards common classes, struggling to capture the nuances of less common transport modes</p>
Classification report for test data	<ul style="list-style-type: none"> The model performs well for high-frequency classes like car, bike, train, and walking, with relatively high precision, recall, and F1-scores. Car has an F1-score of 0.85, and walking achieves 0.84, reflecting consistent performance compared to the training set, where these classes also had high scores. Train maintains strong recall (0.87) and a high F1-score (0.82), showing that the model reliably identifies most train instances. Similarly, bike achieves an F1-score of 0.76, demonstrating the model's ability to generalize reasonably well to this class. There were various misclassifications, especially for underrepresented classes. For example, the class "Other" was never classified correctly, with instances being mistaken for car,

	<p>bike, or walking -- mirroring the training set where "Other" had an F1-score of 0.00. Bus also performs poorly, with all instances misclassified, mostly as car or bike, leading to a 0.00 F1-score, just like in training data.</p> <ul style="list-style-type: none"> • The model struggles to learn meaningful patterns for rare classes, likely because of class imbalance and overlapping features. The metro and tram classes continue to be problematic. Metro shows a slight improvement over the training set, with a 0.21 F1-score on the test set, but remains low, with many instances misclassified as train. Tram remains entirely misclassified, with an F1-score of 0.00, indicating the model failed to generalize this class from the training set to the test set. • <p>These results indicate that minority classes are poorly represented in the decision boundaries, possibly because the model is biased toward more common classes like car and walking.</p>
Feature Importance	<ul style="list-style-type: none"> • The feature importance shows that the model relies heavily on a few key features, with "bus route mean distance" as the most influential feature, contributing 22,3 to the decision-making process. • This suggests that distance patterns along bus routes play a critical role in distinguishing transport modes. Speed-related features also dominate the model's decisions with metrics like "speed IQR" (17,7%), "speed percentile 10" (5,1%), and "speed standard deviation" (3,1%) collectively contributing a large share of the importance. This heavy reliance on speed variation could explain the model's struggles with modes with overlapping speed ranges (e.g., bus vs. car or metro vs. train). • Interestingly, railway station count (10,4%) is another essential feature, likely helping the model identify train and metro trips. Proportion-based speed features (e.g., proportion 45–80 km/h, 7,9%) also influence predictions, possibly helping differentiate slower modes like walking from faster ones like cycling or driving. Lower-ranked features, like "jerk percentile 85" (0,9%) and "tram route standard distance" (0,75%), contribute minimally. Overall, the model leans heavily on speed and distance metrics.

The feature importance of the top-20 (out of 39) is shown in the following Table.

Table 3.8: Feature importance - top-20

Feature	Importance
---------	------------

Bus route mean distance (OSM proximity)	0.22
Speed IQR	0.18
Railway station normalized count (OSM count)	0.10
Speed proportion 45-80	0.08
Speed percentile10	0.05
Speed proportion 5-15	0.04
Speed standard deviation	0.03
Speed average	0.03
Speed percentile90	0.03
Speed percentile80	0.02
Bus route standard deviation distance (OSM proximity)	0.02
Bus route maximum distance (OSM proximity)	0.02
Speed proportion 15-30	0.02
Speed percentile85	0.02
Acceleration kurtosis	0.02
Accuracy percentile85	0.01
Jerk percentile85	0.01
Speed proportion 80-120	0.01
Tram route standard deviation distance (OSM proximity)	0.01
Speed median	0.01

3.5.1 Test on open data

The decision-tree model was also tested on a dataset exclusively reserved for testing, with no portion used during the development or training phases, ensuring an unbiased evaluation of the algorithm's generalization capabilities.

The dataset was collected in the summer of 2024 within the scope of the SSI project using the most recent version of the CBS smartphone app available at that time. The data was collected to obtain data with high-quality labels without errors for the transport mode. This data was collected by a small group of CBS staff and staff from the University of Utrecht. Furthermore, the data contains tracks within the Netherlands and Germany. Accordingly, this test set will inform how well the algorithm generalizes to a different app version/sensor configuration and data collected in a different country. Data from 5 users with 137 tracks are available. The label “ferry” appears in this dataset but not in the observed labels in the development data. However, this label was kept to study what prediction would

result from this label. This is especially interesting because GPS signals usually get noisy when the smartphone is close to or on the water.

The classification report highlights performance variations across transport modes. Walking and the train achieve the highest F1 scores (0.79 and 0.75, respectively), indicating relatively good performance. The bike also shows moderate recall (0.70) but low precision (0.21), leading to a modest F1-score of 0.32. In contrast, bus and tram categories are never correctly identified, resulting in F1-scores of 0.00. Metro has a low F1-score (0.14) due to poor precision and recall. Ferry could not be correctly classified, because it was not in the observed labels. The confusion matrix showed that the three ferry instances were classified as biking. The overall accuracy is 48%, the balanced accuracy is 32%, and the macro F1-score of 0.22 reflects substantial class imbalances and misclassification issues, particularly for underrepresented classes.

Table 3.9: Test on open data results

Class	Precision	Recall	F1	Support
Bus	0.00	0.00	0.00	5
Ferry	0.00	0.00	0.00	3
Bike	0.21	0.70	0.32	10
Metro	0.11	0.20	0.14	5
Tram	0.00	0.00	0.00	27
Train	0.86	0.67	0.75	9
Walking	0.96	0.67	0.79	78
Accuracy			0.48	137
Macro avg.	0.24	0.25	0.22	137
Weighted avg.	0.62	0.48	0.53	137

3.5.2 Conclusion on test results

This section presents the development and evaluation of a decision-tree-based transport mode classification algorithm for smart surveys.

The model achieved an overall accuracy of 81% (balanced accuracy 51%) on training data and 80% (balanced accuracy 46%) on the test set, with weighted F1-scores of 0.78 and 0.79, respectively. While the model generalizes fairly well, class-specific imbalances and misclassification patterns persist. It effectively identifies common modes like car, bike, and walk but struggles with bus and tram, which are rarely predicted and often misclassified.

When tested on the open geo-data, exclusively reserved for model evaluation, the results indicate that the decision-tree-based model does not generalize well, particularly for less frequent transport modes (accuracy is 48% and balanced accuracy is 32%). While it performs reasonably well for walking and train, its failure to correctly classify bus, tram, and car suggests poor generalization to unseen

data. The low macro F1-score and imbalanced precision-recall values highlight a strong bias toward dominant classes, leading to frequent misclassifications. This result suggests the developed algorithm may overfit to patterns in the training data rather than learning robust, generalizable decision rules.

The feature importance shows that OSM-based distance metrics and GPS speed features heavily influence decisions, likely causing confusion between modes with similar speeds and infrastructure. The confusion matrices reveal systematic misclassifications, such as walk and bike, while other classes are not structurally misclassified as one specific other class.

Around 200 GPS- and OSM-based features were considered, but only 39 were influential in the final decision tree, mainly GPS speed and OSM distance metrics. The model favored distance over OSM count variables, suggesting that counting infrastructure is less effective. Further research is needed to assess OSM data quality and coverage, as low or incomplete counts may explain their limited impact.

The decision tree and rule-based model are simple approaches, but machine learning could also be explored. However, challenges like data imbalance, rare classes, label errors, and missing key features persist. Contextual data, such as vehicle ownership, transit subscriptions, WiFi connections, or smartphone services, could improve accuracy. Yet, no machine learning model alone can fully address these issues.

For a fair comparison of general algorithm performance, a reference public dataset should be utilized, such as the open geo-dataset collected within the scope of this project. Otherwise, there will always be tailored solutions that do not generalize or are not comparable.

3.5.3 Conclusions on the model and the input data

The model results indicate that:

- The model generalizes relatively well, but deeper analysis shows considerable class-specific imbalances and misclassification patterns.
- The model performs well in identifying high-frequency classes
- The feature importance highlights that speed-related features predominantly drive the decision-making process. In contrast, features related to altitude, bearing, and trip duration contribute minimally. This heavy reliance on speed could explain the model's difficulty distinguishing between modes with similar speed ranges. The confusion matrices reveal some systematic misclassifications, for example walk and bike. The other class instances are spread across multiple categories, reflecting this group's lack of distinctive patterns.
- While the current decision tree model provides a strong foundation, there is clear room for improvement in handling minority classes and refining feature selection. The model's accuracy and robustness across transport modes could be enhanced by addressing these areas, leading to a more reliable transportation mode classification system.
- The prediction model could be enhanced by the use of some user information on frequent modes.

The results of the model are strongly influenced by issues concerning GPS and OSM. Regarding GPS signals:

- The quality of GPS varies depending on smart-phone, frequency and sensor configuration.
- GPS signals are only useful to a certain extent by capturing variations in speed, but they have nearly no other GPS-based feature. This is a shortcoming of these smartphone sensors.

Regarding the external auxiliary information:

- OSM did not help to improve the classification quality for the transport modes Walk, Bike, and Car.
- OSM data on transportation and travel infrastructure – including features such as roundabouts, traffic junctions, stop signs, speed cameras, and streetlamps, - did not help improve the classification performance.
- The number of OSM features has to be limited due to computational efficiency

3.6 HETUS activity microservice

The Hetus activity prediction microservice⁶ is a domain specific component, which exploits the GPS data to produce tentative data, supporting respondent in the TUS survey, in the compilation of the daily activity diary.

The last part of the microservice associates HETUS activities distribution (with assigned probabilities or scores) to each stop identified through trajectory segmentation of GPS data.

As pointed out in Deliverable 2.1, while the use of geotracking data has been explored by NSIs for the prediction of travel variables (travel mode and travel purpose, starting with the segmentation of GPS data in stops and trips), the methodological research for the prediction of the daily activities performed at the located places, as classified for the HETUS purpose, is more oriented to commercial purpose than to official statistics needs.

HETUS's daily activity tracking is also based on comprehensive data on existing land use and popular locations around travel destinations (Cheng et al., 2022). In order to extract features from the raw POI data, some issues need to be carefully considered: how to aggregate trivial POIs; how to encode POI data.

The approach outlined so far is therefore a hybrid one in which the statistical model adopted to predict survey variables combines both a dynamic model for extracting features from trajectories and a static model for extracting geographic data from maps. The model aims to overcome the accuracy limitations of ad hoc rule-based approaches, as such deterministic rules tend not to capture the stochastic nature of GPS data.

Method

The component for Hetus classification was developed through a probability model that exploits (following the reasoning in Cheng et al., 2022) several information, such as place categories taxonomy, timing of the stop, country-specific indicators derived from previous TUS survey data and user characteristics. Categories of place from the third party (Google Places GP or OSM) are mapped to the HETUS classification of places.

⁶ The microservice is described in WP3 deliverable 3.4 and in the github repository

The objective is to produce a prediction of the most likely activities carried out by the user during a stop. For the activity taxonomy, we refer to HETUS, "Harmonised European Time Use Survey".

To predict the activity performed by the user during a stop, several features are derived from the spatiotemporal characteristics of the stop, from the types of Points of Interest (POI) near the stop (provided by the Geo Location microservice).

The algorithm exploits mainly the data collected by the TUS surveys, the socio-demographic characteristics of the respondent (age groups and employment status) and the contextual data (map service data). The choice of a probabilistic model derived from the lack of a labelled dataset needed to train a machine learning algorithm, consisting a set of stops with the corresponding activity.

A crucial step is the linking between the textual description (deriving from the map) of each POI found in the stop and the categories of places. The type of classification depends on the map service used. In Google Maps, the name and type of the POI can be used directly to link a TUS place. In Open Street Maps, the name of the place and the values of tags such as Amenity, Shop, Office, Leisure, etc., are used. A linking table between the descriptions of GP and those of OSM is used to classify the POI in the TUS places and then to the HETUS activities. The link between the tag and the TUS place is made through keyword searches and regular expressions.

Each point of interest is classified according to the Italian Time Use Survey classification of the places (TUS place) where the user's activity takes place. The TUS place classification is a specific classification of locations used internally at the statistical institute that provides TUS data. In particular, in the developed prototype, the algorithm is based on data from the Italian TUS survey.

The algorithm works through three main steps. First for each POI in the stop, a score is calculated based on the median distance between the POI and the GPS points associated with the stop, weighted by their accuracy. This score is further weighted by the probability of carrying out any activity in the TUS place associated with the POI and within the specific time slot assigned to the stop. The formula is the following:

The probabilities in the formula are estimated by the TUS data. A shortlist of POIs with the highest POI scores is defined. The selection is made by identifying where the slope of the POI Score curve changes significantly, using the *elbow criterion*.

Secondly, for each POI in the shortlist, the probability of performing an activity according to the HETUS classification is calculated. This probability is decomposed using a Bayesian approach, as outlined in the following formula:

where a is the HETUS activity, x is the tuple (user's condition, user's age class, TUS PLACE of the POI) and t is the duration of the stop, while μ and σ are respectively the mean and standard deviation associated with an activity for a user's condition, user's age class, and TUS PLACE of the POI. In accordance with the chosen approach, the model parameters and the probabilities are estimated using aggregated counts from the TUS survey data.

Finally a rank (ActivityScore) of the HETUS activities is assigned to the stop, based on a final score calculated aggregating the probabilities of the activity weighted by the POI-score associated with the activity for each POI in the shortlist.

Input data and the pipeline are described in table 10.

Table 3.10: Characteristics, pipeline and algorithms development for HETUS activity prediction

HETUS ACTIVITY MICROSERVICE – GeoService Microservice - part 3
--

Hetus activity microservice - task of assigning a 2 or 3-digit HETUS code to each stop that has been defined from GPS data using the GeoService (Segmentation ATS/Optics) Microservice	
Characteristics	<ul style="list-style-type: none"> • Domain specific (Time Use Survey) • Type of algorithm: statistical probability model (not supervised ML)
Requirement for the deployment	<ul style="list-style-type: none"> • Required: GPS data and map service (OSM or GP) data associated to a stop • Optional: TUS country data, respondent information (occupational status and age classes)
Input data	<ul style="list-style-type: none"> • GPS features - Longitude, Latitude, Accuracy, Timestamp, Speed • Stop features - Centroid Longitude and Latitude, Time Start and Time End, Duration, Radius • Map features - POI Longitude and Latitude, Tag (textual description provided by MapService) • Profile of the user – Tus data (occupational status and age classes)
Steps of the algorithm implemented for assigning HETUS code to each stop	
Pre-processing	<ul style="list-style-type: none"> • Data cleaning operations by filtering out low-quality GPS points, specifically those with low accuracy or excessively high speed. • Each stop is assigned a time slot (morning, lunch time, afternoon, dinner time, night, early morning). • The category of POI is assigned based on the textual description in the map service used. • Each type of POI is classified according to the Italian TUS classification of the places (TUS place).
POI-score	<ul style="list-style-type: none"> • For each POI in the stop, a score is calculated based on the median distance between the POI and the GPS points, weighted by their accuracy. This score is further weighted by the probability of carrying out any activity in the place associated with the POI and the specific time slot assigned to the stop • Identification of a shortlist of POIs, applying the elbow criterion to the POI scores
Probability model	<ul style="list-style-type: none"> • For each POI in the shortlist, the probability of performing an activity according to the HETUS classification is calculated. This probability is decomposed using a Bayesian approach • Model parameters are estimated using aggregated counts from the TUS survey data
Rank of HETUS activities	<ul style="list-style-type: none"> • A rank (ActivityScore) of the HETUS activities is assigned to the stop, based on a final score calculated aggregating the probabilities of the activity weighted by the POI-score associated with the activity for each POI in the shortlist

3.6.1 Results⁷

The HETUS Classification algorithm was tested using the available annotated data, obtained from GPS data collected using the app CBS-Odin (made available by CBS for downloading from Google Store), for 4 respondents over 4 days.

The algorithm was initially tested in its complete configuration, utilizing all available input data and features. The algorithm's performance - accuracy and reliability - was thoroughly evaluated under these conditions.

Subsequently, to identify the relevance of the considered features for predicting HETUS activities, the algorithm's performance was further assessed by varying the input data settings. This involved testing different combinations of features to assess their impact on prediction accuracy and overall model performance. This was designed to simulate situations in which NISs do not have available or do not want to use data from a time-use survey, or there are constraints on the use of personal data in the microservice.

From the testing phase, it has emerged clearly that a necessary condition for the algorithm to work well is that there is good-quality auxiliary information. In fact, if the stop identified by the segmentation does not contain useful POIs, the prediction is not obtainable. Furthermore, the use of Google Place as an auxiliary source improves the prediction compared to the use of Open Street Map. Therefore, in fact, the possibility of predicting the activity is based on the accuracy and completeness of the map service.

The accuracy of the algorithm is a function of the quantity and quality of the GPS points of the stop and the quality of the stop segmentation has an impact on the quality of the prediction. The evaluation of activity prediction accuracy reveals several cases, both correct and incorrect prediction, depending on the context of the stop. Below is a detailed breakdown of these cases:

Cases of Correct Activity Prediction:

- Both POI and Activity Correct:
 - The correct POI is identified in the stop and the first activity listed in the shortlist matches the actual activity. This represents an ideal scenario where both the location and activity are accurately predicted.
- Incorrect POI but Correct Activity:
 - Even if the identified POI is incorrect, the activity is still assigned correctly. This occurs when multiple similar POIs are present around the stop, and the algorithm assigns the correct activity despite the POI mismatch.

Cases of Incorrect Activity Prediction:

- Private Residence with Nearby POIs:
 - The stop corresponds to a private house (e.g., a friend's home) that is not the respondent's residence. While some POIs may be present nearby, the activity prediction is incorrect due to the identification of an irrelevant POI. In such cases, the respondent is expected to correct the place during validation.
- Correct POI but Atypical Activity:

⁷ The detailed results are presented in Deliverable 3.4

- The correct POI is identified, but the predicted activity is atypical for that location (e.g., a concert in a public park). This results in an incorrect activity assignment due to the unusual nature of the event.
- Ambiguous POIs with Multiple Activities:
 - The map provides several POIs in the stop, each associated with different activities. The algorithm selects an incorrect activity due to the ambiguity in the available choices.
- POI Without Tags:
 - In cases where the POI lacks sufficient tags or metadata, classification becomes impossible, leading to an inability to predict the activity accurately.

The assessment phase also evaluated different scenarios that concern both the availability of TUS survey data and aspects related to privacy, i.e. the use of the respondent's personal data within the microservice. The performance of the algorithm when using different algorithm settings for Google Places map service was tested:

- Rule based: the algorithm does not use the probabilities based on TUS data but only the context information linked directly to activities, using a table linking places to HETUS activities.
- Probabilistic, no personal data: the algorithm uses the probabilities based on TUS data (only time slot and places) but not the personal user data (age group and employment status).
- Probabilistic, no time slot: the algorithm uses the probabilities based on TUS data with the personal user data but without time slot (distributions that also include frequencies for time slots may not be very meaningful as based on few data).

The use of data referring to the respondent (age group and employment status) does not seem to improve the quality of the prediction, although the analysis is based on a few data. Exploiting the time slot seems instead very relevant: excluding it, in fact, produces the worst results, even if using it requires, in general, the availability of the TUS survey data.

In the following tables a selection from the results is shown, comparing the performance of the algorithm using GP and OSM.

Table 3.11: Number of segmented STOPS

	Google Places	Open Street Map
Stop segmented by ATS/OPTICS	111	111
Stop with duration > 5 minutes	83	83
Stop with POIs useful for activity prediction	71	56
Stops with identified POIs (not labelled home/work)	21	24

Table 3.12: Performance of the prediction with Google Place and Open Street Map

		Google Places	Open Street Map
TOP 1	MATCH	11	7
	NO MATCH	10	17
TOP 2	MATCH	12	8
	NO MATCH	9	16
TOP 3	MATCH	13	8
	NO MATCH	8	16

We can conclude that the most relevant features for the quality of the prediction are the place type and the time slot, while the personal data do not particularly improve the prediction, although the limited set of data used for the evaluation does not allow us to draw unequivocal conclusions on this matter. This indicates that the time slot and the duration of the stop provide in any case a significant contribution to the prediction and its accuracy. Finally, the probabilistic approach seems to perform slightly better than the *rule-based* approach.

Finally, it is difficult to draw general conclusions since the performance of the algorithm has been tested on a few cases, so it might be appropriate in the future to evaluate its use on a larger data set in order to also improve the aspects that do not seem to work.

3.6.2 Country-specific issues

In the context of GPS data processing, country specificity issues are more mitigated than in the context of HBS. In fact, what is linked to the country is above all the map, both in terms of the possibility of using API to Google Places and for the quality and updating of the information contained in the map service used.

The phenomena investigated through the use of GPS data, movements and daily activities, evolve slowly and do not present a problem of constant updating except for the map, which however, if it is GP, is constantly updated by the third party.

In the geotracking case the country specificities have impact on the algorithm specification itself, but only on the quality of the output in relation to the quality of the input (sensor and GPS quality) and of the auxiliary information. The issue of the specificity of the model parameters (for example for the use of indicators from the TUS survey) is perhaps the only one.

3.6.3 Geoservice - quality evaluation

The GPS data itself, and the subsequent algorithms used to segment the tracks and predict travel mode, trip purpose and activities undertaken at a location, are not fail-safe: GPS tracks may be incomplete, with missing data due to people's behaviour or the situation in which the GPS signal may be lost; GPS tracks may be complete, but, for example, the trip detection algorithms may be too sensitive (over-identifying trips) or not sensitive enough (under-identifying trips).

Heterogeneity in GPS sensor quality between different types of smartphones affects the measurements and the statistical information derived from them. This heterogeneity becomes a crucial issue in the context of the ESS, since an uneven distribution of smartphone brands and models can be observed across countries, as shown in Deliverable M14: Smart baseline stage (Annex A).

The contribution on the predicted variables of contextual data used to add location details may be limited by the quality of the map services themselves, which again may vary from one European country to another.

Data quality refers to the degree to which the requirements of a set of intrinsic data characteristics are met. Geospatial data are generally collected and used for many different purposes; the quality of geospatial data must be evaluated considering the purpose and context in which they are used.

An important quality indicator of geographic data for use in predictive modelling is the completeness of map elements (ISO 2022), in particular points of interest (POIs), as it directly influences the prediction of trip distribution, trip purpose and stops. A convenient approach is to examine the quality of different object classes, such as the road network. The most accurate method of quality assessment is to compare a dataset with its true value. However, estimating the true value is expensive, complex and time consuming. A more appropriate method is to compare the quality of a dataset relative to a benchmark dataset of documented high quality.

This issue has been explored more in depth on the mapping services used in the GPS-based methods, Google Places (GP) and OpenStreetMaps (OSM). The analysis has been carried out randomly selecting a sample of geographical points within some urban areas of the seven countries participating in SSI project. The POIs are identified within a 50-meter radius around each sample point in the two maps. The study highlights, first of all, a notable difference in the number of POIs present in the radius considered for GP and OSM and a significant under-coverage of OSM compared to GP. Moreover, the results show a significant difference between the countries considered. The results are detailed in Appendix B.

Regarding the fact that the microservice produces tentative data, it can be added that ideally it can be implemented iteratively. Once respondent control and modification are added, the steps can be iterated. For example, the respondent might indicate that a route was actually a multimodal trip, so the segmentation can be repeated with that knowledge. The same goes for activity. A respondent might react as if a stop had no intention, like a traffic jam. But even without respondent interaction, it could be that the three steps inform each other.

3.7 Conclusions and take-home messages

The SSI project has worked on many algorithms for processing smart data, some of these algorithms belong to the category of machine learning, others are unsupervised non-ML algorithms. Implementing new algorithms for smart surveys requires a time-consuming activity of methodological research, training data acquisition and testing.

The main findings and conclusions are:

- Implementing new algorithms for smart surveys require a time-consuming activity of methodological research, training data acquisition and testing. All the micro services that have

been developed, are based on very complex methods have been implemented and in some cases in sequence following complex and articulated pipelines.

- The most suitable algorithms have been chosen for each situation or different algorithms have been implemented from which the user can choose according to their needs.
- The Trade-off between accuracy and speed or between granularity and computational complexity is a crucial issue.
- All the algorithms require very important and very relevant input and training data, that the user (NSI) has to construct.
- For the services developed it happens that the output produced requires manual intervention, by the respondent for geo or travel, by the NSI for Receipt Scanning.
- The automation of smart data processing is improved significantly but manual intervention remains needed.
- The level of respondent involvement to improve output quality must take into account both the effects on burden and collaboration, as well as the legal and ethical aspects associated with the trade-off between accuracy and minimisation.

Table 3.12 presents a synoptic view of the characteristics and main issues of the implemented methods

Receipt scanning microservice:

- For OCR and COICOP classification complex pipelines are defined in order to enhance the quality of the output.
- OCR accuracy depends on the Pre-OCR and Post-OCR phases and the training of the OCR model where the size of the labelled receipt dataset is crucial
- For an NSI that wants to use RSM, is necessary to represent the retail market variability as much as possible in the collection of receipts to be labelled and used for training.
- For the COICOP classification
- Matching data preparation is the relevant phase (coverage and timeliness are crucial quality dimensions for matching data).
- Chaining the various string-matching sub-steps with the machine learning approach adds predictive power.
- Calibrating thresholds at each sub-step seems critical, vital for NSIs to dedicate time and resources to this calibration process.

Geoservice microservice

- The quality of GPS signal and of the map services (OSM and GP) have a big impact on the prediction (tentative data). The errors and missing data are due to sensor failure or map problems
- The segmentation algorithm has been strongly improved (ATS/Optics allows adapting parameters), but some limits emerged from the field tests, where too long stops do not help the identification of multi-activity in the stops. For transport mode a decision tree algorithm has been chosen for the GSM, preferred to the initial rule-based method
 - The model performs well in identifying high-frequency classes
 - The feature importance highlights that speed-related features predominantly drive the decision-making process
 - Additional user information can improve the prediction

- For the classification of activities, the probability algorithm (based on TUS data and POIs) produces good results if the map contains POIs. The algorithm works better if using Google Places, a greater number of POIs are found and they are generally up to date

Table 3.12: Synoptic view of the algorithms implemented in the SSI microservices

Microservice	Component	INPUT data	Algorithms	Training data needed	Main issues	Field test
Receipt scanning	OCR	Scanned receipt	SegFormer YOLOS PaddleOCR LiLT	Receipts, country specific	Training Quality of the photo OCR quality Involvement of respondent in quality enhancing	Small test: Destatis VUB High expectation
	COICOP classification	OCR output Matching data	String matching N-gram FastText Logistic Random Forest	Matching data, country specific	Country specificity Trade-off between accuracy/resources for manual intervention	-
Geoservice	Segmentation	GPS data	ATS/Optics	NO	Quality of GPS signal	Small test Destatis VUB, ISTAT GeoService not always useful for respondent
	Transport mode prediction	GPS data, OSM	Decision Tree	Survey data (CBS)	GPS configuration Map quality	-
	Hetus activity prediction	Segmentation output (GPS, Stop, OSM/GP), TUS data, User info	Probability model	TUS data	Dependence on the quality of map service (OSM/GP)	-

References:

- Benedikt, L., Joshi, C., Nolan, L., de Wolf, Nick. & Schouten, B., (2020) Optical Character Recognition and Machine Learning Classification of Shopping Receipts - @HBS>An app-assisted approach for the Household Budget Survey
- Bucher, H., Keusch, F., de Vitiis, C., de Fausti, F., Inglese, F., van Tienoven, T. P., Mccool, D., Lugtig, P., & Struminskaya, B. (2023). *Smart Survey Implementation: Workpackage 2: Research Methodology Deliverable M6: Review stage*. Grant Agreement Number: 101119594. European Statistical System.
- Bucher, H., Keusch, F., de Vitiis, C., de Fausti, F., Inglese, F., van Tienoven, T. P., Mccool, D., Lugtig, P., & Struminskaya, B. (2024). *Smart Survey Implementation: Workpackage 2: Research Methodology Deliverable M14: Smart baseline stage*. Grant Agreement Number: 101119594. European Statistical System.

Cheng J., Zhang X., Luo P., Huang J., Huang J. (2022). An unsupervised approach for semantic place annotation of trajectories based on the prior probability. *Information Sciences*, Volume 607, August 2022, Pages 1311-1327. Elsevier ed.

ISO 19131:2022 - Geographic information — Data product specifications - (Edition 2, 2022)

4 Human Computer Interaction and Usability: insights from task 2.3

4.1 Introduction

This document describes the small-scale tests that assess the usability of the microservices developed and implemented in an end-to-end solution as part of the Smart Survey Implementation (SSI) project.

Situating within the SSI project

The SSI project aims to develop, implement and demonstrate the concept of Trusted Smart Surveys, by realising a proof of concept for the complete, end-to-end, data collection process and demonstrating a solution. This solution combines:

1. The involvement and engagement from citizens.
2. The acquisition, processing and combining of data collected from smart devices and other applications.
3. The contribution to trustworthiness and guarantee of strong privacy safeguards.

The SSI project focus on (a) an end-to-end solution for a smart survey framework that is sufficiently mature to be applied in several ESS application, (b) 'going smart' by conceptualizing and operationalizing several micro services, (c) identifying developing, implementing and eventually demonstrating these microservices within the perspective of an end-to-end process for (at least) Time Use Surveys (TUS) and Household Budget Surveys (HBS), and (d) considering smart surveys from a respondent perspective.

WP2.3 is mainly concerned with the latter, albeit some tests may relate to the design and implementation process of the microservices. Full reports of tests conducted within each country can be found in Appendix C.

Aim of task 2.3

Task 2.3 set out to test the usability of the microservices as part of an end-to-end solution on a small scale. The testing basically comes down to the distinction that Cattell (1971)⁸ makes in forms of intelligence. The development of the microservices is based on crystallized intelligence. It is the result of the use of accumulated knowledge that depends on prior learning and increased with experience. In other words, it is what makes the microservices *smart*. This has been demonstrated in the work of WP3. The usability of the microservices depends on fluid intelligence. It is the problem-solving and adaptability that is independent of learning or education when encountering novel situations. In other words, it allows assessing if microservices are *clever*. This report presents this assessment. Are the smart solutions (i.e. microservices) clever enough to complete the tasks set out when given to the uninitiated?

4.1.1 Usability and human computer interaction

As outlined in the SSI project's M6 deliverable, the usability of the solution is the key concern within the Human Computer Interaction (HCI) spectrum. Usability refers to the ease of use and the quality of the user's experience with a platform or application. It has been argued that usability is strongly linked to the three key elements of the solution provided within the SSI project:

⁸ Cattell, R.B. (1971). *Abilities: Their structure, growth and action*. New York: Houghton Mifflin.

Ad. 1. The involvement and engagement from citizens relate to the usability of the application to recruit and retain users. Key usability attributes are engagement, accessibility, clear instructions, time efficient, and the availability of error handling options.

Ad. 2. The acquisition, processing and combining of data collected from smart devices and other applications relates to the usability of the application to complete complex tasks. Key usability attributes are clear and intuitive user interface (UI), clear task flow and guidance, error prevention mechanisms, (in app) training, and (in app) feedback and support.

Ad. 3. The trustworthiness and guarantee of strong privacy safeguards relate to the usability of the application to share personal data. Key usability aspects are trust and credibility, security and privacy, transparent communication, data collection efficiency, and user control of data/information.

4.1.2 Microservices

WP2.3 assessed the usability of two microservices. The first is the Receipt Scanning Microservice (RSM) which consists of the OCR microservice and the COICOP classification microservice. The OCR microservice is non-domain specific and is designed to perform Optical Character Recognition and Document Understanding. In this context, it will be used to scan receipts and extract information on store name, products, (unit) prizes, discounts, and any other relevant information (e.g., VAT). The COICOP classification microservice relates to the statistical domain of the Household Budget Survey (HSB) and is restricted by country and/or language requirements. In this context, it classifies a product or service to a COICOP category. The usability tests of the RSM have been done within the domain of HBS and were limited to the OCR microservice. The RSM is conditional to the training of the models. This training is specific to countries, languages, type of receipts, et cetera. This applies to both OCR and COICOP categorisation (see WP2.2). Note that at the time of testing, the training of these models is still limited and not optimised. Throughout the results it will therefore become clear that training is crucial for good usability.

The second microservice is the GeoService Microservice (GSM). Again, the GSM exists of a non-domain specific Geoservice microservice. It classifies geolocations into stop-track clusters and predicts the travel mode based upon the track clusters. The HETUS classification microservice relates to the statistical domain of the Time Use Survey (TUS) and is restricted by country and/or language requirements. It aims to predict the most likely activity carried out during a stop following the HETUS activity taxonomy and using additional spatiotemporal information, nearby points of interest and other user characteristics. The usability tests of the GSM is limited to the GeoService microservice and has been done within the domain of TUS and Mobility Studies.

4.1.3 Platforms or applications

As cited above, the aim of WP2.3 is to test the usability of platforms or applications that have the developed microservices integrated into an end-to-end solution for Smart Surveys. Currently, the small-scale usability testing, only the data collection platform MOTUS had successfully integrated both microservices into an end-to-end solution. Starting from the central objective within this project to achieve a proof of concept for the complete, end-to-end, data collection process, the discussion of the results will start from the findings of the MOTUS platform.

The Receipt Scanning Microservice (RSM) is supportive of Household Budget Surveys (HBS). An end-to-end solution is understood here as the entire respondent journey of a HBS. That is, downloading and installing an application, logging in and unloading, and keeping expenses in a diary, and closing

the diary and completing the survey. The last two steps are not covered here. The RSM plays a role in entering expenditure. More specifically, the RSM is fed photographed receipts that are then processed, scanned, subjected to optical character recognition (OCR) and document understanding. The RSM then feeds the information back to the host platform or application for subsequent presentation to respondents. The usability test is thus mainly concerned with how the microservice is integrated into an end-to-end solution and how this makes the solution smart.

As mentioned, MOTUS is the only data collection platform that offers a smart solution by integrating the RSM developed within the SSI project into the end-to-end solution. Many of the steps in the end-to-end flow of HBS, however, are generic and independent of RSM. These include downloading, installing, logging in, onboarding, manually entering expenses (as an alternative to RSM), taking a picture of a receipt, et cetera. Since part of the focus of usability testing is on the entire respondent journey, the results of the end-to-end solution integrating RSM for these steps are compared with these of other applications. For example, differences or similarities can be viewed in (the click path leading to) taking a picture, the actions to edit processed data, or simulations of processing time.

GeoService Microservice (GSM) supports Time Use Surveys (TUS) or Mobility Surveys (MS), among others. At the time of carrying out the usability tests of the integration of GSM into an end-to-end solution within the MOTUS platform performed by Destatis and VUB (in collaboration with Statbel), no comparable applications were available among the partners to perform benchmark tests or simulations on. In this case, therefore, the results relate only to the solution provided by the MOTUS platform.

The individual country reports can be found in Appendix C.⁹

4.1.4 Testing the developed microservices in end-to-end solution

The actual testing of the Receipt Scanning Microservice (RSM) and GeoService Microservice (GSM) as developed within WP3 of this SSI project is done only on the MOTUS platform. The MOTUS platform is developed by hbits and consists of a front office and a back office. The back office serves to build a study, collect data, monitor progress, and analyse data through several designated builders. The front office relates to the application with which the users interact via a user interface (UI). The MOTUS application exists as a web version for browsers and as a mobile version for smartphones and tablets with iOS and Android operating systems. Both the back and front office are connected through Application Programming Interfaces (APIs) to the MOTUS core which holds the database required to build a survey and collect data. Adapter APIs allow the MOTUS core to connect to external information such as, in this case, data from RSM and GSM microservices. All processing is done in stand-alone microservices for reasons of optimization, data security and privacy. It is through these Adapter APIs that the RSM and the GSM are connected to and processed by the MOTUS core in order to be presented via the front office API to the users of the MOTUS application. The MOTUS application is used by Destatis (Germany) for testing both the RSM and GSM, by the Vrije Universiteit Brussel (VUB, Belgium) for testing the RSM, by the VUB in cooperation with Statbel (both Belgium) for testing the GSM, and by ISTAT (Italy) for testing the GSM.

⁹ Note that the results from the VUB/Statbel test on GSM in MS with MOTUS have been integrated directly in the final deliverable because of limited time and because of the VUB being the task leader integrating all country findings into one report.

4.1.5 Benchmarking

The benchmarking of the non-domain specific OCR microservice is done by SSB (Norway) and Insee (France). Conceptually, the Norwegian HBS web application is similar to the MOTUS platform with integrated RSM. Here, too, respondents take a photo from a receipt, but the difference is that this image is externally processed by the data extraction platform Veryfi. The extracted data is presented in the web application as a list of expenses. In this list, respondents can find their processed receipt and see the expense items. Based on previous testing, the Norwegian HBS web application limited the editing options to reduce respondent burden, as editing was not frequently done (i.e., new expense items can be added, but existing ones cannot be edited or deleted). The total amount of the expense was taken from the OCR processing of the receipt, if recognized or available. Otherwise, the sum of the item lines was used.

Insee uses an earlier version of the Dutch @HBS application that has been adapted to be integrated in the Insee systems. Conceptually, the French HBS application is designed to limit the respondent burden. It does not include microservice elements. The application allows respondents to manually enter expenses or take a photo of a receipt. In case of the latter, the application only conducts a photo check to see if there are written characters present. This check is minimal since it only checks for recognition of a written character. User guidelines ask respondents to provide a 'readable' image. Respondents are prompted to provide total amount, shop name and date and then validate the photographed receipt. OCR processing and classification is done externally and post-survey.

4.1.6 Simulating the Receipt Scanning Microservice

Finally, CBS (The Netherlands) simulated Receipt Scanning Microservice (RSM) processes using an interactive prototype. Respondents were able to download the @HBS app in the respective stores, onboard, and enter their expenses manually or take a photo of a receipt. To simulate the RSM, the receipts were provided by CBS and data from the receipts were pre-entered in the simulation with different processing time and correctness as follows:

- Short receipt with five seconds processing time and no errors
- Medium receipt with 20 seconds processing time and some errors
- Long receipt with 50 seconds processing time and some errors

Due to the simulation, the personal receipt could only be entered manually. The manual entry nor the simulation of scanned results allowed for classification of expenditures.

Table 4.1 gives an overview of the partners involved in the usability testing, the platforms or applications used, and the microservices tested.

Table 4.1 Overview partners, platforms and testing

Partner	Country	Platform	Testing		Benchmarking		Simulation	
			RSM ^a	GSM ^b	Scan/OCR	Tracking	RSM	GSM
Destatis	Germany	MOTUS	✓	✓				
VUB	Belgium	MOTUS	✓	✓ ^c				
SSB	Norway	PWA			✓			
Insee	France				✓			
Istat	Italy	MOTUS		✓ ^d				
CBS	The Netherlands						✓	

^aReceipt Scanning Microservice^bGeoService Microservice^cIn collaboration with Statbel (Belgium)^dFindings will be added in the Appendix

4.2 Test design

4.2.1 Receipt Scanning Microservice (RSM)

To test the RSM, Think Aloud tests were used. In these experiments, respondents are presented with tasks to perform. While performing these tasks, they tell out loud what they are doing, thinking and expecting. (See also M6 and M14 deliverable.) Table 4.2 lists the tasks that were presented to respondents. This task list was followed by all partners during testing to the extent the tasks were applicable to the application/method used (e.g. the French HBS application did not present the expense items to the respondent to edit; Norwegian HBS web application did present the expense items to the respondent, but the items could not be edited). After completing the tasks, a short interview evaluated the tasks and asked some questions about trust, reliability and privacy and security.

4.2.2 Geo Service Microservice (GSM)

Testing of the GSM was carried out by Destatis, Istat and by VUB/Statbel. Institutions took different approaches. VUB/Statbel focused purely on mapping travel behaviour as a function of mobility studies. Respondents were recruited via email. If they accepted to participate, they were asked a) to complete and return a consent form, b) to watch an instructional video, c) to install the MOTUS application and always allow location sharing, and d) to keep track of their trips in a time diary for two days. In doing so, they were asked to alternate between using the manual entry method and the entry method based on geolocation data. Afterwards, respondents were invited to discuss their experiences with the researcher in individual or duo interviews lasting about 30-45 minutes.

Both Destatis and Istat tested the GSM as part of a time use survey. Additionally, to the manual diary entry creation, the microservice shows time frames for stationary activities and movements as a thought support for filling out time use diaries. Test persons with different socio demographics were recruited and the test was conducted remotely. In a preliminary interview, test persons were instructed to install the application and how to proceed. In the two following days, the test persons used the application including the microservice independently in their everyday lives to fill out a time use diary. For the first day, they were asked to use the geo-assisted entry option with location data suggestions in order to create entries, while for the second day they could choose freely whether to

use the geo-assisted entry or the manual entry. After the independent testing, a second interview of about 45 minutes followed, in which detailed questions on the test persons experiences and usability were discussed.

Table 4.2 Task list usability test - an overview:

Task #	Receipt type	Usability focus	Rationale	Observer focus
Task 1	-	Warm-up/primer	Practice think-aloud	Actively stimulate verbalizations of thoughts
Task 2	-	Install app and log-in	Prepare/access	
Task 3	A1. Short and clear	Choose approach	Start easy	
Task 4	A2. Short and clear	Force inversed approach	Make sure both approaches are tested	
Task 5	B1. Medium in need of editing	Forced scanning approach	Continue with more complex task/receipt, that we know needs editing.	Does the participant notice errors from scanning and does the participant edit errors?
Task 5-bis		Forced editing	A chance to edit errors.	
Task 6	C1. Personal	Choose approach	Continue with personal receipt. Which approach they choose now that we know that they are aware of both approaches.	Are considerations about sensitivity and personal information more prominent when sharing their own receipt?
Task 6-bis		Forced editing	A chance to edit errors.	Do they notice errors from scanning? Do they edit errors?
Optional, depending on time.				
If used 'manual' method for task 6.				
Task 7	C2. Personal	Force scanning	Continue with personal receipt. Which approach they choose now that we know that they are aware of both approaches.	Are considerations about sensitivity and personal information more prominent when sharing their own receipt?
Task 7-bis		Forced editing	A chance to edit errors.	Do they notice errors from scanning? Do they edit errors?
If used 'scanning method for task 6.				
Task 8	B2. Long in need of editing	Forced scanning approach	Continue with more complex task/receipt, that we know needs editing.	Does the participant notice errors from scanning and does the participant edit errors?
Task 8-bis		Forced editing	A chance to edit errors.	
Platform specific				
Task 9	All receipts	Find and check Expense Overview	Navigation and usability	

4.2.3 Test respondents

Since these are small-scale tests, it is impossible to use a representative sample. When recruiting test subjects, an attempt was made to achieve a distribution in terms of age and sex, and if possible, education and operating system. Table 4.3 provides an overview of the test subjects per test and per partner.

Table 4.3 Overview test participants per country

		RSM ^a or equivalent					GSM ^b		
		Destatis	VUB	SSB	Insee	CBS	Destatis	VUB/Statbel	Istat
Number		19	18	11	15	20	14 ^c	15	11
Age range ^f	Youngest age	18	24	18	>25	24	19	24	33
	Oldest age	69	82	76	<65	75	67	74	60
Sex	Women	10	9	4	7	7	7	9	6
	Men	9	9	7	8	13	7	6	5
Education	Low/Student	4	3	/ ^d	/ ^d	5	4	/ ^d	/
	Medium	11	4	/	/	8	5	/	/
	High	4	11	/	/	7	5	/	11
OS	iOS	9	9	/ ^d	7	/ ^d	14	9	/
	Android	10	9	/	8	/	/ ^e	6	11

^aReceipt Scanning Microservice

^bGeoService Microservice

^cDestatis also conducted a focus group with 5 respondents

^dNot asked/Not recorded

^eAt time of testing MOTUS application with GSM was only available in Test Flight for iOS

^fData refers to the age of the youngest/oldest respondent

4.3 Results

The results of the usability tests of the end-to-end data collection solution with Receipt Scanning Microservice (RSM) and Geo Service Microservice (GSM) are discussed below. Recall that only the MOTUS data collection platform offered a complete end-to-end solution integrating both microservices developed within this project. Therefore, and for each microservice separately, the results are firstly discussed from the MOTUS data collection platform and – in the case of RSM – secondly benchmarked with findings from the other applications. Thirdly, results are briefly related to the usability attributes (overview tables can be found at the end of each subsection) and fourthly, some general recommendations are discussed.

Note that qualitative research is based on small samples. As a result, results are not presented in the form of percentages or numbers. Instead, the terms occasionally/few, sometimes/some, and often/many are used. This provides an orientation on the frequency of observed behaviours and evaluations. Furthermore, qualitative research not only shows results, but it also asks about the "why" underlying a result and therefore provides a better understanding of the processes that lead to such a result.

Moreover, everything discussed below occurred for at least some respondents. Opinions or issues of only one respondent are not reflected here unless it is considered highly relevant. In the various sections, we will first show which issues occurred, to what extent they were also found in the benchmark studies and the simulation study, how these findings relate to the various usability attributes and the proposed components of the solution to achieve Trusted Smart Surveys, and finally, what recommendations there are. Note that both the microservices developed within the project and

the data collection on which these microservices were tested are under continuous development. It is possible that findings from the test have already been partially incorporated in subsequent updates.

4.4 Receipt Scanning Microservice (RSM)

4.4.1 Installing and onboarding

For all applications and platforms, the actions related to installing the application, onboarding via welcome screens, logging in and viewing the overview page generally went down well with respondents, especially as these are actions they are familiar with because they are similar to other applications they install and/or use.

All respondents read the invitation letter. The intensity with which they read does vary and partly depends on the type of letter. The VUB used an excerpt from letters with no official letterhead and only a short explanation, QR codes and login details. All respondents read this letter in full. Destatis used official, extended letters, complete with letterhead and printed on both sides. As a result, some respondents missed the back side and went straight to the QR codes.

The latter underlines the finding that almost all respondents are used to scanning QR codes to be linked to, for instance, a website or App Store or Play Store. As major advantages, respondents mention the ease of use (i.e., no need to click and type in the app) and the fact that they can be sure to get straight to where they need to be (i.e., prevents installing the wrong app). Usually, almost all respondents also know which of the QR codes to scan (i.e., the QR code for the App Store or for the Play Store). Only one older respondent (>80 years old) scanned the wrong QR code and therefore experienced difficulties installing the app.

Installing the app from the respective stores also poses no problem for the respondents as this is nowadays a very normal action. Partly because of this, it is noticeable that hardly any respondent really pays much attention to the information about the app available in the stores. Respondents generally indicated that they never look at reviews, that this takes too much time, or that they simply assume that if an app is allowed in a store, it must be reliable because their admission criteria are very strict.

The MOTUS application provides respondents with general onboarding screens about language settings, functionality, privacy and notifications. While most respondents expect to see onboarding screens, they do not read them (thoroughly). Again, a common argument here is that they have already made the choice to install the app by now and thus do not find this information relevant and it is too much text which means it takes too much time to read it. Interestingly, respondents who do read the onboarding screens say they find the information relevant. Also noticeable is that few respondents allow the app to send notifications.

Logging in is easy for almost all respondents. Again, this is because these are actions that people are used to. Few respondents had difficulty copying the credentials from the letter to log in. Only a few respondents started by entering an email address instead of the username as shown in the letter. However, many respondents did ask for the option of displaying the password via an eye icon to check whether it was copied correctly from the letter.

To log in, respondents had to accept the privacy policy and terms and conditions. Although these were clickable links, none of the respondents took the time to read the privacy policy or terms and conditions. Logging in was largely done in two types of actions. On the one hand, respondents tried to

log in without accepting the privacy policy and terms and conditions, both as a conscious effort to avoid having to do so or because they had simply read over it. On the other hand, respondents ticked it without thinking, as it is just a formality anyway and not doing so will only result in the app not allowing them to proceed.

The main screen on the app was clear to many respondents. Due to delays in testing on the MOTUS application by the VUB, some tests were conducted after the preset end date of the registration period. On the main screen, the task for the spending diary was coloured orange as a result. After explanation, one respondent commented that an orange diary gives a better signal of what needs to be done/what needs to be clicked, as it is in line with all the other buttons, such as the confirmation buttons or the plus button to create a new diary entry.

Box 4.1 Click and action path for installing and onboarding

MOTUS

Read invitation letter > open camera > scan QR code > click on link to open app store > download and install app > (dis)allow notification on pop up notification > read info on and set language > click next > read info on functionality > click next > read info on privacy > click next > read info on and (dis)allow notifications > click next > read info on results > click let's go > enter username > enter password > check off general privacy statement > click login (if not checked off general privacy statement: acknowledge pop up notification by clicking on ok)

4.4.2 Benchmark

The above results are further substantiated by findings from the benchmark and simulation tests. Like the VUB, Insee also used an excerpt of an invitation that showed two QR codes for the Appstore and PlayStore and login credentials. Hardly any respondent experienced difficulties installing the French HBS application. Only one respondent searched for the application in the wrong store. Some respondents did encounter difficulties entering the provided passwords, which they said had to do with the illegibility of special characters.

Like Destatis, CBS used materials as they would be used in a real study: a full two-sided recruitment letter and a leaflet for their Dutch @HSB application. Here too, some respondents missed the back side or only scanned the letter and went straight to the QR codes. As a result, most respondents did not understand the response task. CBS also mentioned that a few respondents were not able or willing (because of security considerations) to use a QR code. Those who search for the application in the respective Appstore mentioned that recognizing or matching name and logo is an important verification to install the correct application. Respondents with low digital skills could not install the app without help. When using provided login credentials, the importance of having an option to check the password entered through an eye icon is substantiated by the finding in the Dutch @HSB application where most respondents used the eye icon to verify their password.

The Norwegian HBS application is a web application that is accessed through a weblink, which, too, is a well-known way to access an application. Less clear is the screen that provides the options to continue in the browser or store the app on their smartphone. Most respondents continue in the browsers whereas the preferred option from SSB is to store the app on the smartphone. Respondents

were not provided with login credentials but used their bank-ID to login, which they expect as a requirement to login for public services.

Different from the MOTUS application, the Dutch @HBS application provides respondents with specific onboarding screens that tell them to keep an expense diary for their household. After seeing the onboarding screens, in combination with having read the letter many respondents still do not understand the response task. Respondents think the primary purpose of the application was to provide them with an overview of their expenses. They also do not understand the need to complete multiple tasks (e.g., questionnaire and expense diary), the need to register expenses at the household level, the need to keep the expense diary for two weeks, and what and to what detail expenses need to be entered.

In the French HBS application, respondents went through an automatically launch tutorial which was considered helpful. Also, the Norwegian HBS application has short and sharp onboarding screens asking for consent, explaining what is expected, and give a brief tutorial about how expenses can be registered. Despite this, in the Norwegian HBS application respondents rush through the screens and the few who read them, struggle to connect them to the tasks at hand.

In the Norwegian HBS application respondents are also presented with notifications (e.g., tips regarding their participation) during onboarding, but here too respondents tend to rush through these notifications without reading it, eager to proceed to what they considered their primary task – registering expenses.

Like in the MOTUS application, the home screen of the Dutch @HBS and French HBS and the Norwegian HBS was considered clear and easy to understand, although some respondents with low digital skills in the Dutch @HBS application mentioned that it was unclear how to proceed (i.e. locate the button to add expenses).

4.4.3 Usability attributes

The use of QR codes to download and install the app positively relate to accessibility and time efficiency. Yet accessibility and time efficiency decrease when login credentials are becoming too complex. The absence of an option to show the password relates negatively to error handling options. The option in MOTUS to automatically login after the first time improves accessibility and time efficiency. The excerpt of the invitation letter shows that instructions for downloading, installing and logging in are clear. Long invitation letters provide transparent communication but are in their current form not read by respondents. This is also true for the privacy declaration. Even though it is an obligation to provide this information, it is simply accepted because respondents cannot proceed otherwise. The user interface and task flow of the login, onboarding and home screen is intuitive. Trust and credibility as well as security and privacy is created by the trustworthiness of the issuing/executing institution even before the respondent opens the application.

4.4.4 Recommendations

Installing an application - whether by scanning a QR code or not - and then logging in is such an ingrained action that most respondents get away with it quickly and hardly encounter any problems. This has both its advantages and disadvantages. The biggest advantage is that installing and logging into the app is not a stumbling block to participating in a survey for respondents with a basic level of

digital skills. **To be able to also include respondents with lowest digital skills, in-person help setting up the application is needed.**

While logging in is a very familiar act, it should be noted that credentials should not be too long and complex (e.g. no passwords combining 10 letters, numbers and other characters). The jadedness of respondents when it comes to privacy statements means that accepting them is not a hurdle either.

Of course, this is preceded by the need to recruit respondents (see WP2.1). One of the biggest drawbacks of the above ingrained actions, is that respondents do not read the invitation letter properly - they jump straight to the familiar QR codes - and largely ignore the onboarding screens. Furthermore, even those who did read the letter (and leaflet) in full, often did not understand the response task. This raises the question of whether respondents inform themselves well enough and know what they are signing up for. **So, recruitment is not just about getting as many respondents into the app as possible, but also about finding a way to inform them about what is expected of them** (see also WP2.1). Note that informing does not necessarily mean providing respondents with *more* information beforehand but could take form of *concise* information spread across the process of recruitment, installation, onboarding and login. This is no easy task, given that most respondents simply not read (all) information. Striving to inform respondents as much as possible about what is expected of them and how they can participate conflicts with their onboarding behaviour where they hardly read presented information. **Offering information incrementally in the click path only when it is needed**, with or without the option to turn it off permanently, can ensure that information is better absorbed by the respondent. At the same time, a help button can ensure that respondents can seek assistance when they need it.

Note that there is an important caveat to these findings. The limited focus on (additional) information in the letter, in the App Store or Play Store, and on the application's onboarding screens may have been partly influenced by the test environment. A test environment builds trust and also allows questions to be asked. In line with the protocol, respondents were also assured that the interviewers would help remove the app and all related data after the test. Respondents of the tests conducted by the VUB explicitly stated that they had little to no concerns about privacy and data security precisely because the tests were conducted by a university, served a scientific purpose, and had ethics committee approval. Similarly, respondents of the test conducted by Destatis, SSB, Insee and CBS trust the app because they know that these NSIs published it.

4.5 User interface and user experience

Overall, the user interface (UI) and user experience (UX) of the MOTUS and other application are positively rated as “professional” and “intuitive”. However, “so many people, so many wishes”. There is little point in discussing the UI and UX from front to back. Even more so because the microservices developed within this project and their integration into the data collection platform MOTUS and planned integration in other applications is aimed at a general target audience and that inevitably leads to the situation where one size does not fit all. Yet that does not mean that there are some general improvements possible in the UI and UX of the microservices as they run on the different applications.

4.5.1 Results

Many of the suggestions for improving the UI and UX stem from the complexity of the tasks to be performed. The tests show that it is not only the complexity of the task itself that causes problems, but also the communication in the task execution process. This involves communication in its broadest sense. More specifically: meaning of word usage, consistency of word usage, and colour structure throughout the click path.

The meaning of words used should not be underestimated. Within both the VUB and Destatis test surveys, the meaning of 'expenditure' or 'expenses' was ambiguous. In the MOTUS application, respondents were found to interpret 'expenditure' or 'expenses' as the total of the receipt when entering it manually (i.e., the sum of all products). This led them to try to note in the text field to add a product what type of products they had bought (e.g. groceries, lunch, household products). Such terms are too generic for the underlying classification microservice to make relevant suggestions. A similar situation that occurred for some (older) respondents was that instead of clicking on the plus icon to add expenses, they clicked on 'monthly expenses' (see Figure 4.1). Although this block within the expenditure diary is meant for recording monthly fixed expenses, these respondents appeared to be reasoning from 'last month's groceries' and so they mistakenly ended up in the wrong screen. At this point, the complexity of study and the ignorance of respondents that they are wrong makes it very difficult, confusing and in one case frustrating for the respondent to perform the tasks.

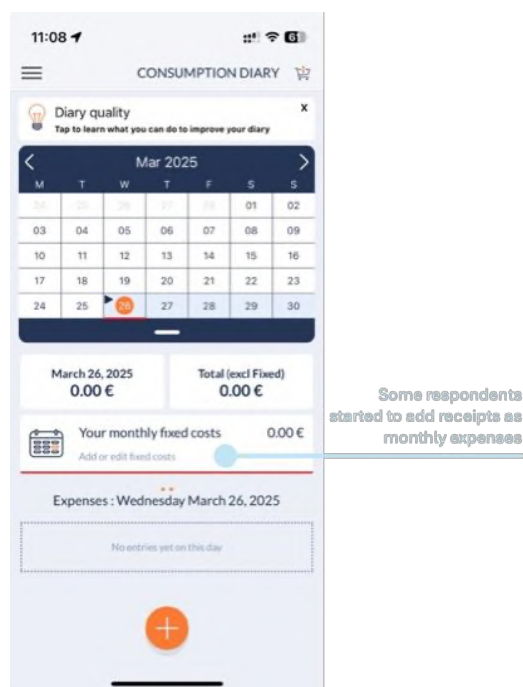


Figure 4.1 Some respondents tapped on monthly expenses

A second example recurred in both the VUB and Destatis test and concerned the options for adding expenditure to the diary. These options in the German version, for example, are "Kassenbon scannen" (EN: scan receipt), "Ausgabe eingeben" (EN: add expense), and "Ich hatte heute keine Ausgaben" (EN: I had no expenses today). The confusion appeared to be in the word "scannen" (EN: scan). Especially in relation to purchases, scanning tends to be interpreted as scanning a barcode. Within the developed microservice, however, scanning refers to taking a picture of the receipt. Nevertheless, it occurred several times that respondents who intended to photograph their receipt still chose the 'add

expenditure' option. This could be avoided by referring to "Kassenbon fotografieren" (EN: photograph receipt) instead of "Kassenbon scannen" (EN: scan receipt).

Confusion also arose from lack of consistency in word usage throughout the execution process of the pending task as well as mismatch of words in, for example, pop-up notifications or other instructions on the one hand and the actual wording in the relevant buttons on the other. An example is that a pop-up notification reminds respondents to 'save' a change while the relevant button in the application says 'update'. These are not directly situations where the respondent gets stuck, but it does make them hesitate for a moment. Within the complexity of the actions to successfully complete the task execution process, this kind of situation can nevertheless lead to confusion.

Finally, such a similar situation occurs with the communication contained in the colours of the buttons. Respondents are quite aware that - in the case of the MOTUS front office - the orange buttons are the buttons to perform actions (e.g. the plus icon is orange, buttons like 'next' and 'save' are orange). One of the respondents from the VUB survey articulated the implementation process as follows: 'You just have to follow the orange trail'. Slight confusion then arose at a pop-up notification where the colour code was reversed, and it was the orange button that returned the respondent and the dark blue button that made the respondent confirm the chosen action. Again, these are not directly things that cause the respondent to get stuck in the survey, but they are things that stand out and potentially cause despair.

When it comes to the data collection platform MOTUS, respondents noted some additional issues that could lead to a better UX. First, it did not appear clear to some (older) respondents that the main screen of the spending diary could be scrolled. Particularly on smartphones with smaller screens, and especially in combination with a collapsed calendar, the newly added expenses often (partially) fell off the screen so that respondents could not find the issue at all or saw them but not the pencil icon to edit the issue. Secondly, and relatedly, a large proportion of respondents found it strange that new issues were not added at the top of the issue list but below it. For many, this was counterintuitive as they are used to this being the case from other apps that use timelines. Thirdly, almost all respondents mentioned that when manually entering an expense or editing an expense created by the microservice, the numpad appears, but then there is no 'done' button that makes the numpad disappear again after entering, say, the amount of a purchase. It took almost all respondents a while to realise that they had to tap the screen somewhere outside the numpad to continue with their input.

4.5.2 Benchmark

In the Dutch @HBS and Norwegian and French HBS applications, CBS, SSB and Insee reported similar language issues with the ambiguity of the meaning of 'expenses'. In terms of UI and UX of the Dutch @HBS application, some respondents mentioned reduced legibility due to font colour, type and size. Other issues raised related to the functionality of the 'back' button, which closed the application instead of moving back within the application.

Additional findings from the Norwegian HBS application reveal that use of wordings is to be considered carefully. Respondents mention that the functional difference between 'save' and 'add item' is not always clear. Additionally, respondents expect these buttons to be at the bottom of the screen and not at the top as is currently the case. A potential bug in the Norwegian HBS application often triggered warning messages for no reason. Observations reveal that these visual elements are disturbing and contradict the appreciated minimalist design of the app. This result can be translated to a more general finding of being aware not to over-inform respondents with visual elements. In similar vein,

during the processing of the receipt in the Norwegian HBS application a progress bar was shown, and it said 'scanning'. For many respondents it was not clear that this meant that in the background relevant information was extracted from the receipt.

In the French HBS application, the choice between manual input and photographing was very clear and intuitive with a camera icon for the latter option and a keyboard icon for the former. Respondents appreciated the simplicity of the design and compare it with banking apps. Here too, respondents questioned the sorting of the list of expenses albeit different from what the respondents using the MOTUS application mentioned. From a committing and/or editing point of view, the latter preferred a chronological list based with last entered receipt on top. However, respondents in the French HBS application rather see their expenses sorted based on *purchase date* and not *entry date*. The Norwegian HBS application resulted in similar findings where respondents mentioned the ambiguity of the wording 'most recent' at the top of the expenses list. It was not clear whether this meant most recent photo taken (i.e. entry date) or most recent expenses (i.e. purchase data).

4.5.3 Usability attributes

Again, user interface and user experience are highly subjective. In general, results show that the usability attributes on engagement, accessibility, and clear instructions are negatively affected by inconsistencies in language and colour palette, unconventional language, buttons in unexpected locations on the screen, and a long click path. Especially for older, less skilled respondents this decreases the usability. The task flow and guidance are clear although respondents need training. In app training is available via the Assistant but this has not been tested. An interviewer that supports and helps respondents is desired. The user interface could be improved by making clear respondents can scroll down and reconsidering the sorting of the expense list.

4.5.4 Recommendations

User interface appreciation (UI) and user experience (UX) are very personal. Nevertheless, there appear to be some universal experiences that can lead to improvements in UI and UX. These experiences fall into simplicity, consistency and expectations based on previous experiences. **In case of simplicity, the click path and expected next steps should be clear.** When the central aim of the application is to register expenses, respondents should be drawn to the click path to execute that task. If multiple tasks are involved, concise communication is important. In addition, some wordings might be replaced by universal signs such as × or ✓. **In the case of consistency, thorough checks are recommended.** Note that these can be language sensitive. These include a) checking whether word usage leads to the right intention (i.e., the desired meaning and any possible interpretation), b) checking whether word usage is consistent between instructions given and the naming of relevant buttons, and c) checking whether the colour palette remains the same throughout the click path. It must be mentioned that contrary to the other applications, wordings in MOTUS are not hard-coded but can be changed at will using xliif files. This recommendation is therefore mainly addressed at the users of MOTUS and not the developers but nonetheless very pertinent. **In the case of expectations, it is recommended whether expectations or anticipated actions can be easily performed.** For example, in the case of a timeline or chronological view, are the most recent entries shown at the top, or if this timeline or chronological view is long, scrolling is visibly possible.

These are not major, insurmountable hurdles, but their importance should nevertheless not be underestimated. For many respondents, it concerns an application that they would not install by themselves but only install in function of a survey. Moreover, it concerns a survey that expects them to perform complex operations for an extended period. Any confusion that can be eliminated in this case improves the user experience and, consequently, the retention and continuation of the survey respondents.

4.6 Scanning and creating tentative data

Scanning the receipt is initially perceived by almost all respondents as by far the easiest and most expected way to transfer data from the receipt into the diary. Nevertheless, many respondents do not perceive the handling process as optimal. The main findings are listed below.

4.6.1 Results

First, for most respondents, the first part of the click path is not immediately obvious. After they click on the plus icon, the distinction between ‘scan receipt’ and ‘new expense or return’ is confusing. A ‘new expense’ can also be added by photographing a receipt. Respondents find the click path to get to the option to take a photo too long. Generally, after clicking on the + icon, they simply expect to see a camera icon. This is much more in line with the expectations they have from what they are used to from other apps. In the minds of most respondents, they are doing nothing more than simply taking a picture of a receipt. While the options to upload an existing receipt from the photo library or from files are relevant - especially considering digital receipts - these additional choices appear to distract from the task at hand. In short, respondents expect to be in their camera mode with two clicks to photograph the receipt. Note that in more recent versions of the MOTUS application, this issue has been addressed.

Second, older respondents struggled with the option to crop the photo. They interpret the option to crop as something that should be done to increase the quality of the receipt photo. They therefore spend (too) much time getting the cropping margins around the receipt exactly right, which leads to frustration in several cases. One respondent sees cropping as a positive option because it allows cutting non-relevant information from the receipt. Note that in more recent versions – including the version used in the test by Destatis – of the MOTUS application, cropping of the photographed receipt is no longer offered.

Third, the processing time from the time of uploading and finding the tentative entry in the expenditure diary was almost unanimously considered too long. The data collection platform MOTUS communicates reasonably well during processing, but respondents' expectations of processing speed do not match the actual processing time. This third point is explained further below.

After uploading, a throbber appears on iPhone devices but not on Android devices. Respondents who have the latter type of devices say they immediately think they have done something wrong the moment it takes a while for the application to show the next screen. However, even for iPhone users, the throbber gives only a brief respite. They too think the app ‘crashes’ after only a few seconds and reason that they have done something wrong. Note again that in more recent versions of MOTUS this issue has been addressed.

Once the photo has been uploaded, the screen appears making it clear via animation and text that the receipt is being processed. Respondents are very pleased with the animation, and many see it as a positive signal that ‘something is happening’. The text could possibly be simplified. In the Destatis test study, the text starts with ‘Ihr Ticket wird mit OCR (Optical Character Recognition) bearbeitet.’ (EN: Your receipt will be processed using OCR (Optical Character Recognition)). This text is considered too difficult and too technical. The text also indicates that processing the receipt may take several minutes. This is where the biggest bottleneck is. Respondents benchmark and indicate that they are used to much faster applications. None of the respondents reason that processing a receipt might be complex and therefore take longer. Consequently, respondents indicate that ‘several minutes’ is discouraging. Moreover, the tests at VUB showed that despite most testers reading the text carefully and thus knowing that their receipt could take long(er) to process, they became impatient after 10-20 seconds and assumed that processing had failed.

As a result of the long(er) processing of the photographed receipt, many respondents assumed that they did something wrong (i.e., bad photo quality or incorrect cropping) or that the application crashed at that point. In many cases, respondents restart the task. In some cases, they are interrupted during this operation by the pop-up notification that their previously photographed receipt has been processed and is available in their spending diary. In other cases, respondents go through the entire task before their previously photographed receipt is processed. Mistakenly, they then think that it has now gone faster only to find out later that their receipt is now duplicated in the expense diary. The longer-than-expected or accustomed processing time of the photographed receipt thus leads to incorrect actions in the diary and/or impatience and frustration among respondents.

Fourth, although many respondents found their spending directly in the timeline, there appeared to be some who had difficulty doing so. For this, they indicated several things. For some, their (most recent) expenditures were not visible on a small screen because they fell off the screen. This proved extra confusing when combined with the ignorance that they could scroll or the habit of seeing most recent entries at the top of a timeline. Others did not realise they had to click the ‘show more’ button. And still others could not find their expenses because the total amount of the tentative entry did not match the total amount of the photographed receipt. The latter is discussed further below.

Fifth, the total amount in the tentative data often did not match the total of the photographed receipt. The accuracy varied a lot. In some cases, the total deviated only slightly in others the deviations were very substantial. During testing, several possible causes of this were identified.

4.6.2 Discounts cause problems

In the VUB test study, receipts from the supermarket Delhaize were used. This supermarket runs a programme to encourage healthy(er) eating. Within this programme, a ‘Nutriboost discount’ of 10% is given on healthy products. This discount is displayed directly below the product. In almost all cases, the microservice adopts this correctly. However, the receipt layout shows at the bottom of the product list first the total without the Nutriboost discount, then the total Nutriboost discount, and then the final total from which the Nutriboost discount has been subtracted. At this point, the microservice often adopts the total Nutriboost discount as a product (see Figure 4.2). Since the microservice calculates the total based on the products entered - and not, like the benchmark applications, adopts the total from the receipt - the final total comes out too high.

Poor quality of receipts causes problems

Different receipts were used within the test study. Interestingly, long(er) receipts or receipts with a format different from the standard receipt did not lead to worse results. The quality of the photo taken by most generations of smartphones is good enough for the OCR microservice to process the photo. Even when zooming out to get a long receipt in the photo or accidentally taking the photo in landscape instead of portrait. However, it should be mentioned that taking image from long receipts was considered inconvenient. Respondents had to move a way quite a distance to get the full receipt in the picture. For even longer receipts than the ones tested, this might at some point jeopardize the quality and hamper OCR processing. The quality of the receipt does play a role. Drummed-up receipts where folds run exactly over product prices sometimes cause the microservice to miss a comma or miss a digit, resulting in some products being presented extremely expensive or, on the contrary, extremely cheap in the tentative data (see Figure 4.4).



Example of drummed-up ticket that required substantial editing

Figure 4.4 Drummed-up receipt

4.6.4 Special layout of receipts causes problems

How product data and product prices are displayed on receipts also makes a big difference. The microservice adopts most common receipts well. During testing, at least two types of receipts were identified where the microservice was unable to adopt data correctly. The first case was a receipt where both the unit price and the product price are on the same line. In this case, the microservice copied the unit price. This led to both cheaper and more expensive products (see Figure 4.5). The

second case was a receipt where the product price had three decimal places. This made products very expensive because the microservice interprets this comma as a thousand separator.

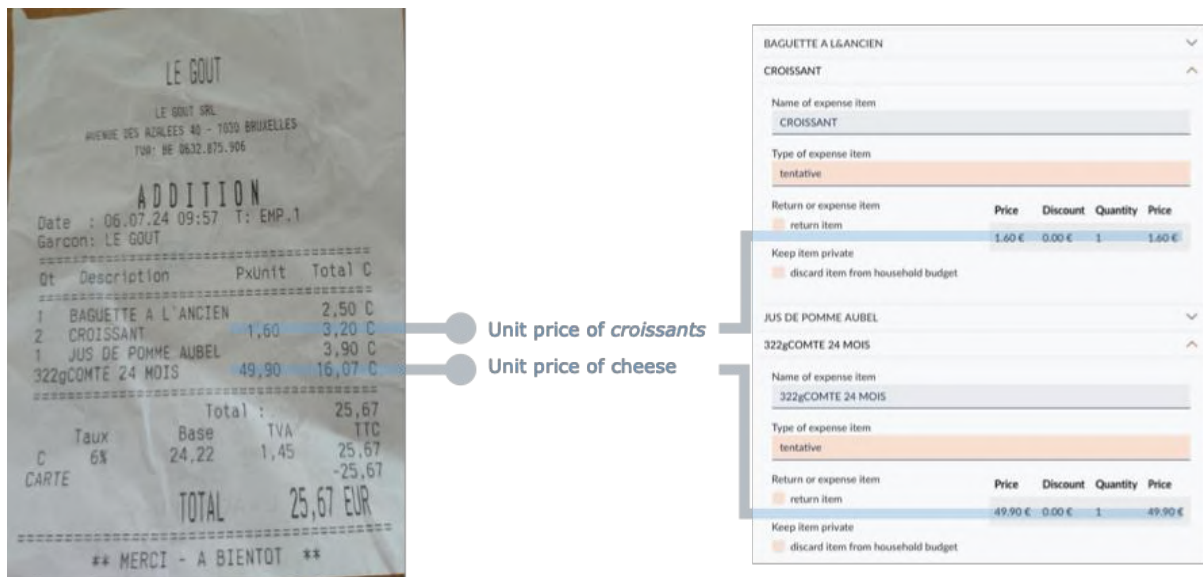


Figure 4.5 Issue with unit price

4.6.5 Additional issues

Some additional issue occurred on an irregular basis and even within the same receipt. These issues related to:

- Wrong comma interpretation: 139 Euro instead of 1.39
- Missing rows: no product, no price
- Missing price: product name but no price
- Missing product name: price but no product name
- Phantasy product: 3 products (without product name) each 80,000 Euro (!)

4.6.6 Benchmark

In the Dutch @HBS, the French HBS, and the Norwegian HBS application respondents also favour the photo method over the manual entry method. Here too, it is considered to be faster, more reliable, provided more detailed data, more time efficient, and it requires less cognitive effort, especially when receipts are long. Interestingly, some respondents in the Norwegian HBS application interpret the positioning and visual presentation of the button for 'take a picture' at the top left of the screen as 'first and best' option, demonstrating that respondents can be drawn into preferred click paths.

The benchmark and simulation tests furthermore resulted in similar findings for the scanning and processing time. Several respondents remark that the Dutch @HBS app asks for audio access along camera access. Contrary to the instructions, respondents do not ensure that all corners of the receipt are visible when taking the picture. Note here that within the RSM, this is not a requirement for OCR to extract information from receipts.

Contrary to findings with the MOTUS application, respondents using the French HBS application faced less issues with the cropping function after having taken a photo. In terms of expectations, respondents were surprised that they were prompted to enter the total amount manually. They expected this to be extracted from the photographed receipt, which further underlines the importance of the RSM design started within this project.

Like the MOTUS application, in the Dutch @HBS application confusion also arose about the word usage of 'scan'. A large proportion of respondents believe they should scan the barcode on the receipt. Other obstacles to scanning include not knowing which part of the receipt to photograph, difficulty photographing long receipts, and doubt about the quality of the photo taken. Similarly, respondents in the Norwegian HBS application mention that 'take a picture' is not necessarily understood as 'scanning'. This goes to show that wordings in applications are language sensitive.

Like the MOTUS application, respondents in the Norwegian HBS app mentioned that processing time was substantial but only a few would put up with it. With the Dutch @HBS application it was also found that long(er) processing time made respondents doubt themselves. Several respondents indicated that they would re-photograph the receipt. The longer CBS simulated processing time, the more inclined respondents were to prefer manual input. Similarly, for the simulated errors in the tentative data in the @HBS application, not all respondents notice these errors. Those who do notice that the receipt is not displayed correctly respond similarly to respondents of the MOTUS application. They either say they will take a new picture, leave it as it is, or try to correct it. In the latter case, though the more errors as results of the processes CBS simulated, the more inclined respondents were again to enter the receipt manually. Some indicated that they would eventually find it easier to enter receipts manually immediately rather than having to make many corrections to the processed receipts.

Finally, younger age groups (18-45 years) in the Norwegian HBS application explicitly mention that keeping paper receipts from their purchases is an issue. Paper receipts are not always provided as standard and young(er) people do not have a habit of collecting them. They mention that being able to upload receipts from loyalty applications or emailed invoices would be useful. The desire to be able to upload digital receipts was also mentioned by several respondents in the Dutch test.

4.7 Usability attributes

Apart from the UI/UX issues (i.e., wordings, click path), scanning or photographing receipts contributes significantly to the usability of the applications. It positively relates to engagement and accessibility and instructions are clear. Respondents expect the scanning option to be time efficient, but their expectations are (too) high, which means they judge the current state of the application not time efficient enough. The user interface is intuitive, and the task flow is clear once respondents have familiarized with it. The option to just photograph the receipt without having to crop or line up corners – note that the test version still had cropping so this is not tested – prevents respondents making errors or doubting their abilities. Although communication about processing the photographed receipts is clear in the sense that it is correct, respondents do not perceive it that way because of their high expectations of the RSM. Data efficiency can be improved by more training data. The current state of the RSM generates (too) many errors. Finally, presenting data processed from the receipt as tentative data can give respondents a sense of control over their data. However, it is not always clear that data are tentative and need to be committed.

4.7.1 Recommendations

Based on the above, several recommendations to improve usability have been identified. First, since respondents prefer to photograph their receipts both out of habit and expectation, it is recommended that this **click path be as short and clear as possible**. Two clicks, graphically represented by a + and a photo camera icon, would be desirable. Second, the quality of the microservice is such that cropping a photo is not necessary. It is therefore advisable to **omit cropping** because respondents interpret it as a necessity (and possibly as something they did wrong). Third, ideally, the processing speed of a photographed receipt would be increased so that it is closer to respondents' expectations. Within the processing process, communication is extremely important. The throbber immediately after uploading is an important sign to respondents that the application is working. So is the notification that the receipt is being processed, although the communication itself could be simpler and possibly include an assurance that the respondent has done everything right (e.g. "Your receipt is being processed. You can close this window."). Note that even with communication, some respondents are (too) quick to conclude that it does not work. **Expectation management and concise communication is crucial**. Fourth, **the importance of training data should not be underestimated**. The better trained the microservice is, the better the results of the tentative data. Training relates to a number of elements and is not limited to language and type of receipts. Training also relates to the alignment of boxes that recognize and isolate text (e.g. this explained the above-mentioned issue with mixing up commas and points), training the 'meaning/destination' of boxes, and training to deal with special elements like the Nutriboost discount mentioned above. Since this all varies significantly both within and between countries, this seems to be something that needs to be done at the country level. An additional stumbling block not tested, but identified by, e.g. respondents in Belgium, is the multilingualism of receipts. Training data are also very important in this area. Note, finally, that the ability to upload digital receipts in the microservice for processing was not tested. This is an interesting avenue to avoid the problem of poor-quality receipts. Some respondents also mention this to increase convenience, for example, if receipts from loyalty programme apps could be uploaded directly. This is certainly a track to explore further both in terms of technical and legal feasibility.

Box 4.2 Click and action path scanning and creating tentative data*

MOTUS

Click + icon > click Use Receipt > select from Take Photo, Choose Photo, Choose File > (dis)allow access to camera > (dis)allow access to photo library > bring receipt in focus of camera > take photo > click Use Photo or Retake > crop image > wait for throbber > read message on processing > wait or click Return to Diary > click on OK in pop up notification that receipt has been processed

* Note that some updates to MOTUS have already made some adjustments, including the omission of cropping the image.

4.8 Editing and committing data

When the photographed receipt has been processed and appeared as tentative data in the timeline in the expenditure diary, respondents should adjust and/or commit it to the expenditure diary.

Committing implies that they send the data to the database for researchers/NSIs to use. The tentative data appears in the timeline as a block with an orange border around it. Below the total is the instruction to check and validate the expenditure. Many respondents missed this information. As a result, and especially when this is the first expense in the diary for respondents, it is not always clear that this data has yet to be committed. A few indicated that at first glance, the fine orange border was too indistinguishable from the committed entries.

The first thing almost all respondents do is check whether the total as shown in the tentative entry matches the total on the receipt. If so, or if 'close enough', most respondents leave the entry as it is. This is problematic for at least three reasons. First, it leaves the data uncommitted, even when the scan result is completely correct. Second, while the total may be correct, details may not be correct, or context may be missing. Third, 'close enough' was usually cited after a few uses with incorrect outcomes. Respondents attributed small discrepancies to rounding errors or glitches in processing the receipt, but did not consider this a substantial problem because it was approximately correct anyway.

When entry was incorrect, almost all respondents clicked on the expense. This unfolded the option to edit or delete it. The icons to do this are in an orange block. Only for one respondent with a small screen, the unfolded piece fell off the screen, leaving the respondent feeling that nothing was happening. At the same time, the pencil icon is universally known to respondents as the option to, in this case, edit the expense. Most respondents said they wanted to edit the expenditure because they like to see that everything is correct or because they believe it is in the interest of the study.

By clicking on edit, respondents get an overview of all products and their price. Most respondents go through this list but often clicked only on some products that are incorrect. Especially for longer receipts with multiple errors they did not open or correct them all. Note that due to the lack of a classification microservice, respondents also received the warning that information about the category of the expense was missing. In general, respondents considered editing the expenses a tedious and time-consuming task. Many stated that they would not be willing to correct every incorrect product for each scanned receipt in their everyday lives. Having to tap on each incorrectly recognized product to correct product name and price, as well as to add the missing COICOP category is too much of a burden for respondents. Several respondents tried to overcome this by looking to adjust only the total or by trying to introduce a separate discount or just make a product more expensive so that the final total was correct.

Respondents who accurately reviewed all products experienced few problems in adjusting the price. However, many respondents did have difficulty entering discounts for three reasons. A more general reason was that respondents mostly think of discounts as a negative amount. They were confirmed in this by the display of discounts on receipts. Product discount is displayed as a positive amount in the data collection platform MOTUS. Respondents often initially thought that this was the problem of the wrong grand total and believed they should put a '-' sign in front of this. A more specific reason occurred mainly in manual entry, but was also sometimes considered as a shortcut in corrections and related to omitting the product discount and entering discount as a separate 'product'. Respondents thought they could get away with this because a receipt often also showed the total discount. A final, very specific reason occurs when the microservice reads discount as a separate product line.

In other words, the product discount is thus not part of the product introduction, but a 'product' in itself, so to speak. In the background, this works when the discount is correctly adopted. The product is then zero euros with the discount as stated on the receipt. However, when the discount is not copied correctly and respondents want to change it, they get the message that the discount cannot be greater

than the product (see Figure 4.6). Since respondents see that discount is included as a separate entry as the result of the processing of the receipt, but since they cannot enter and/or adjust it themselves, this leads to great confusion. Not all respondents realize without assistance that they must apply the discount correctly to the product to which the discount relates. Finally, indicating product categories was also often problematic.

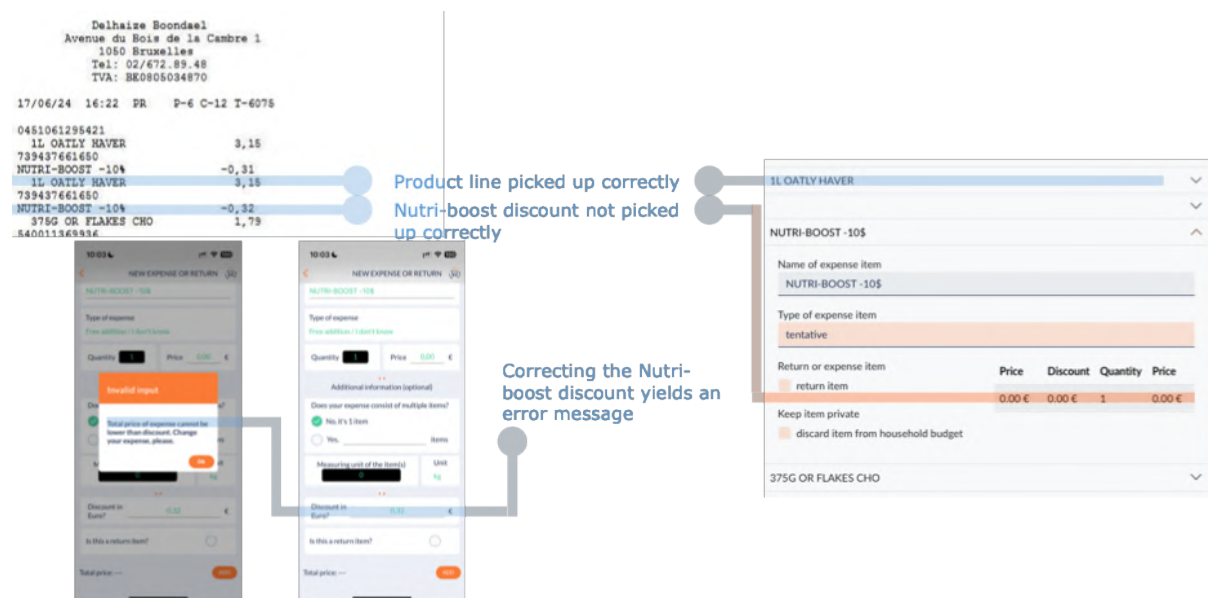


Figure 4.6 Discount included as product but not editable

When respondents checked all entries and made all corrections and clicked update, the tentative data were committed. The committed data are no longer shown in an orange box and the edit and delete options are now shown in a blue block. Also, the grey text that the entry still needs to be validated has now disappeared. Most respondents clearly saw and understood the difference. Yet, especially in Destatis' test study, it was found that visual cues in orange (i.e. the frame around the tentative entry), in light grey (i.e. the text to validate the tentative entry), and in red (i.e. lines at the bottom of the frame where context of the expense is missing) are rarely an impetus to make corrections.

4.8.1 Benchmark

In the Dutch @HBS application, respondents were presented with two different editing screens. On the first screen the total amount is not visible, and expense items can only be added in the second screen. This turns out to be confusing for respondents and using one single editing screen is recommended. Moreover, results show that respondents try different options to delete an expense item: left-swiping, tapping the bin-icon, or setting the amount to zero. This shows the importance of not assuming that all gestures are universal. Also, in the Dutch and Norwegian test it was not clear for all respondents that they should check and if needed edit the data.

In the French HBS application, after photographing their receipt, respondents only needed to provide the total amount, the shop name, the date, some context and validate the receipt. The application only performed a photo check (i.e., the presence of written characters). The OCR processing and

classification happened in the back office. No processed data had to be edited. However, under the expenses tab respondents were able to find an overview of all their entered expenses. Tapping on an expense gives them access to their photographed receipt

The findings with the Norwegian HBS application reveal that few respondents navigate to the product list after their receipt is processed. This underlines the finding that it is not in respondents' nature to further check or edit the details of the receipt. If not prompted, SSB reports that some respondents forget that there are more details collected from the receipt than what the overview screen shows (i.e., date, store name and total amount). Some respondents mention that if checking is important, they expect prompts to remind them to do so.

4.8.2 Usability attributes

Instructions are not always clear to respondents that they need to commit the data. Most respondents do so eventually, but only because the total amount is incorrect. If the total amount were to be correct, they would either leave it or look for a button to confirm and not expect to go to an editing screen. Recall that in the tests all possible context of expense items was requested from respondents. In that form, editing expense items is too burdensome and some context is not well understood or makes less sense in relation to the expense item it is contextualizing. From this perspective, it negatively relates to data collection efficiency and clear task flow and guidance and it detracts from their engagement. Albeit the user interface uses colour change from orange (tentative) to blue (committed), not many respondents actively notice this. Similar, not all respondents notice the coloured warnings that give them feedback about missing information. Although all these elements are designed to give respondents control over their data, not all respondents understand or acknowledge this.

4.8.3 Recommendations

The all-important communication from the application that should be very clear to the respondent is that the data are tentative and thus should be committed. The fact that respondents tend to measure the correctness of the receipt process by whether the total is correct or not makes it clear that respondents do not realise or care that more detail is captured from the receipt than they suspect. Again, communication is important that makes it clear that the details of the receipt are important and therefore need to be checked and corrected where necessary. This communication (both visually and via prompts) is a challenge as this is not clear either in the MOTUS application or in the benchmark (if applicable) and simulation applications. So, a combination of good signal colours and clear, legible text giving the instruction is crucial. For example, 'click on the expense to edit and commit to the diary'. Consider adding a third icon besides the trash icon to delete and the pencil icon to edit to commit the expense immediately. However, it should be noted here that respondents check expenditure against the total which does not necessarily mean that all context is complete or correctly filled in.

Signal colours to fill in missing information are also important on the screen where respondents do the operations. Note here again that the classification microservice was not yet operational and so respondents had to indicate the category of each expenditure anyway. The more missing information respondents had to fill in, the less user-friendly they found the app and the less inclined they were to rate the reliability of the OCR microservice highly. When a lot of information was missing or incorrect,

some respondents indicated that they might as well have entered it by hand. Nevertheless, the will and habit of using the OCR microservice is high and almost all respondents see its potential. This again underlines **the importance of training data on the one hand and the need for a well-developed classification microservice on the other**. The less the respondents have to do, the higher the user-friendliness and willingness to use the OCR microservice.

Box 4.3 Click and action path editing and creating committed data

MOTUS

Click tentative entry > click pencil icon > click on product/service > change information (i.e. expense, category, price, ...) > click on update > *repeat if applicable* > click on save

4.9 Manual entry

4.9.1 Results

For most respondents, the first part of the click path for manual entry is clearer than for receipt entry via the RSM microservice. Clicking on the plus icon and then ‘new expense or return’ makes sense to many. The first part of the manual entry, this is, date, country and shop, is also easy to fill in. However, from the moment they click on ‘add an expense or return’ almost all respondents indicate that it is not clear what is expected of them, and many do not know initially how to proceed on the first page already.

4.9.2 Unclear how to use the search field

Respondents are first presented with the search field in which they need to type a single product and assign it to a category. The most common mismatches of expectations are a) that respondents enter the total price of the receipt in the text field, b) that respondents give a generic description of all products on their receipt (e.g. groceries), c) that respondents describe their product too accurately. In the first two cases, this is due to the fact that respondents reason top down, that is, from the total of the receipt. It is not clear to some that the approach is bottom up where each product has to be entered individually to eventually arrive at the total of the receipt. In the latter case, this is due to the fact that respondents do not know in how much detail they must then describe their product. A respondent within the VUB survey sums this up well: ‘What do I need to enter? Food, fruit, bananas, or bananas from South America?’

4.9.3 Unclear how classification works

A number of respondents do not understand the concept of classification. During testing, the COICOP classification was used. After entering a product, a list of product groups appears from which respondents had to choose one within which the product they bought falls. The logic of COICOP classification does not work for respondents. They indicate that at the lowest level used, there are still too many categories, they do not always understand the names of the categories, and categories are too similar. Moreover, combined with the ignorance with how much detail they have to enter their

products, many entries lead to no suggestions or suggestions of categories unrelated to the product. In many cases, respondents choose the option to describe the product in their own words. Note that this problem also arises when tentative data created by the OCR microservice asks to assign a product category to each identified product.

4.9.4 Unclear what most of the context means

The data collection platform MOTUS allows you to set how much context is requested for each issue. These include the questions about quantity, whether it is multiple items, what the unit of measurement is, and whether a discount was given (see Figure 4.7). In the context of evaluating usability, the test actually asked for very extensive context.

The basic and more familiar context (i.e. quantity, price and discount) yielded only one notable issue. For example, it is not equally clear to some that under 'price' they have to enter the unit price and that it is multiplied by the number. If they enter the number correctly but instead of giving the unit price, they give the total price, their spending will be higher than on the receipt. The extended context in the test study creates more confusion. The question of multiple items is never properly understood. Almost all respondents think it is about the other products on the receipt and thus lose their train of thought that each product must be a separate introduction. A few others, in turn, think that if they bought a box of eggs, for example, they should fill in here how many eggs are in the box. Indicating the unit of measurement and quantity is also confusing. This is mainly due to the fact that this question always appears regardless of the product. When it comes to products charged by weight, respondents can still follow (e.g. 2 kilos of apples). However, when it comes to products that are simply charged by piece, respondents get stuck (e.g. one loaf of bread - even if respondents manage to set the unit of measurement to 'grams', the receipt often does not show the weight of the bread).

4.9.5 Unclear how to proceed with numpad

What did prove to be a stumbling block for almost all respondents is unrelated to the context being queried at issue but has to do with the number pad that pops up when a number has to be entered. The numpad itself has no 'done' or 'return' button. Some of the respondents do not know what to do after entering the price, for example, especially since the numpad also falls over the 'save' button at the bottom of the page. When these respondents realise that they have to tap on the entry screen above the numpad, they do not perceive it as user-friendly. Incidentally, this appears to be a more common annoyance. Even respondents who know that this is how they make the numpad disappear do not find this user-friendly. Many let it be known that this applies not only to this application but also to others. Finally, a few are surprised that they are asked whether credit card was used to pay, but not whether cash was paid.

Household budget diary

Tijddagboek BTUS24

COICOP CLASSIFICATION

Select the classification you want to use

OPTIONAL FIELDS

Entry overview

☐ Hide country question

☐ Hide where did you make this expense question

☐ Hide professional expense question

☐ Hide paid with credit card question

Expense details screen

☐ Hide quantity question

☐ Hide consist of multiple items question

☐ Hide measuring unit of the expense question

☐ Hide discount question

Figure 4.7 Optional fields in the household budget diary in MOTUS

4.9.5 Benchmark

The Dutch @HBS application did not yet have a search field linked to a database for shops or for products and services implemented, nor an underlying classification structure for entering the expenses. Respondents entered the date, the shop, the product, and the price. Here respondents ran into the problem that not all receipts contained recognizable product names or respondents started typing in the exact wordings from the receipt. Additionally, findings from both the Dutch @HBS and French HBS application underline the top-down reasoning of respondents. Here too respondents used more generic terms like ‘groceries’ and thus were unaware of the level of detail required. The test setup of the Dutch @HBS application without the search databases revealed that respondents anticipated this to exist. Many expected autofill options or not to see that the store they entered last time, is still unknown to the application. This is underlined by the French HBS where such a database exists, and respondents managed to find their products and/or services once they understood that the required level of detail was the expense items.

Similar to the MOTUS application respondents are easily confused by the additional context. First, both the Dutch and the French HBS application asked whether it concerns a foreign expense or an online expense using tick boxes. Several respondents think they must choose one option. Second, in the Dutch @HBS application, when it came to product details, respondents were confused about what type of information was required for the quantity field (e.g., number of carrots or kilograms of carrots). Third, in the Dutch @HBS application, respondents indicated that it was not always clear how to report discounts, deposits, and refunds. Several strategies emerged. Some just ignored it, others subtracted the discount from the original price by heart or using the phone’s calculator, yet others used a minus sign, and some even tried typing discount as a formula in the price input. Two respondents reasoned that deposits and stamp savings should not be entered because they are not expenses.

Finally, the majority of the respondents in the Norwegian application mention that manual registration is the least desired option, too burdensome, and too prone to make errors. Knowing the level of detail required from the scanning option (many tried this option first), many consider it undoable to reproduce that level of detail in a manual registration. A few others mention that they feel more in control when entering data manually, which is positive from a usability point of view, but it actually implies that it gives them more opportunities to manipulate the data.

4.9.6 Usability attributes

The click path presents a clear task flow and guidance. In app help is available via the Assistant although this has not been tested in the app. Instructions on the detail level of the entry (i.e., entering each expense item separately) are not clear enough. The user interface is generally well appreciated except for the use of the numpad. Additionally, the more context asked about expense items, the lower the perceived data collection efficiency. The use of the COICOP classification for respondents to categorize their expenses reduces data collection efficiency as well, because they are not familiar with it. Respondents experience more control over their data entries which also contributes to security and privacy, albeit from reasoning that they can leave out sensitive expenses. However, manual entry, in particular of longer receipts, is considered too much effort and too time consuming, negatively affecting engagement and time efficiency.

4.9.7 Recommendations

For manual input, there are three main recommendations. First, it is necessary **to communicate clearly to the respondent whether the relevant entry screen refers to the ‘expense’ or the ‘expense item’**. There is enough space visually to make this clear to respondents and this helps avoid consequential errors. Additionally, it would be useful to somehow make it clear to respondents in what detail they should describe their expense item.

Second, and related to the previous one, **if categorisation is important, then it should make sense to respondents**. The COICOP classification does not. Another option is a microservice that does the classification of products in the background. Ideally, it is trained with a list of all product names, ~abbreviations, and ~codes and runs this classification in the background without showing it to respondents, or at least only in case verification is needed. The less the respondents have to do for this, the better.

Third, it is important to **consider how much context is desirable with the expense items in a study**. From a scientific and analytical point of view, more context obviously leads to more reliable and valid data. However, from the users' point of view, this extra context can be confusing and implies more clicks. On this point in particular, respondents indicated that it does take a lot of work to enter receipts manually. A large proportion note that for weekly shopping, it is no mean feat to fill this in accurately and they often wonder aloud whether they would commit to the real household budget surveys.

Box 4.4 Click and action path for manual entry

MOTUS

Click + icon > Click on ‘new expense or return’ > Click on ‘add expense or return’ > Provide details (date, country, shop) > Type product name and select category > Provide details about expense (quantity, price, multiple items, measuring unit, discount, return item) > Click on add > *repeat if multiple products* > Provide final details (professional expense, paid by credit card) > Click on save

4.10 Checking expenses overview

4.10.1 Results

Because of time constraints and other choices for optional tasks, the VUB did not test the task of looking up the expenses overview. Destatis did offer this task to respondents. For many respondents, the overview of their expenses is an important element in the application. They appreciate it as a reward for efforts made, they see it as a clear overview of what they have entered, and they are positive about the personalised feedback of their spending. Some even see it as an opportunity for self-monitoring. A few respondents have difficulty finding the button to see an overview of their spending. This button is located at the bottom of the timeline of entered receipts and relates to the previously mentioned lack of clarity that one can scroll down in the spending diary overview.

4.10.2 Benchmark

The Dutch @HBS and the Norwegian HBS application did not test the task of looking up the expenses overview in a visual presentation such as a pie chart presenting totals per expense categories or other categorizations. The French HBS application provided charts of expenses by date or store. All participants were able to access this overview and appreciated it. In the Dutch test several respondents commented spontaneously that they expect the app to provide them with more insight in their spendings.

4.10.3 Usability attributes

Respondents see the expense overview as a reward for their efforts and even a way to self-monitor their expenses. This positively relates to their engagement and is appreciated as in app feedback and support. The user interface is intuitive. Some respondents ask for it to be more interactive and more informative.

4.10.4 Recommendations

The overview is appreciated by respondents. Recommendations for improvement are very limited and mainly relate to the UI/UX. Possibly the spending summary could be housed under a clear button instead of in a block similar to the receipts entered. Furthermore, the header 'confirm' to leave the overview looks strange and would be better called 'close'.

4.11 Learning curve and support

4.11.1 Results

The design of the test study was to throw respondents into the deep end. Although they had been made clear what the purpose of the application and OCR microservice was, they had not been told how it worked. They were simply given the task of entering expenses from the receipt into the expense diary. A number of actions are already so ingrained in the daily use of ICT that it was very easy to perform even for those who are unfamiliar with smartphones and applications. These include scanning

a QR code, installing an application, logging in, giving permission to use e.g. the camera, taking photos of a receipt, and the meaning of universal icons (e.g., bin, photo camera, pencil, plus sign).

However, when it comes to actions that are new or where they are less familiar, it becomes more difficult for some of the respondents. These do not include typing text in a search field, ticking a checkbox, or typing an amount. It is mainly about what content is expected and about how they navigate through the application. Positively, all but a few, least digitally skilled respondents eventually figure out how to do this through trial and error - and thus without interviewer intervention. On the negative side, it appears that this learning curve is very steep. Repeatedly, respondents indicate that if it were not for the test study, they would have already given up. Apart from some of the recommendations made above, this has to do not only with the user-friendliness of the microservice and application, but mainly with the complexity of the task and the actions involved.

4.11.2 Benchmark

The Dutch @HBS test underlines the steep learning curve. Findings show that respondents became increasingly more efficient, especially when manually entering expenses. Even respondents that needed step by step instructions for their first entry, were able to make the next entry on their own. Most of the mistakes made in the receipt scanning (e.g. scanning the barcode or not scanning the whole receipt) were quickly solved by trial and error, indicating that photographing receipts is a common action.

The French HBS test confirms this too. Despite the automatically launched tutorials, some respondents still need to be guided in terms of what was expected next from them. Many respondents used trial and error to complete the task. More digitally skilled users picked it up quite quickly, but the digitally less skilled mentioned the complexity and inability to complete the task undermined their confidence to use the application.

The Norwegian HBS test did not focus explicitly on the learning curve, but it did confirm that respondents learned manoeuvring quickly. However, it reveals that older age groups generally struggle more with technology, including how to use the app and how the relevant information is compressed on small screens. Middle age groups are the most willing and able to meet the survey requirements.

4.11.3 Usability attributes

The steep learning curve negatively affects engagement, accessibility and time efficiency. This can be mitigated with clear instructions and training, support and follow up, both in-app and through the help of an interviewer. The absence hereof – or it not being tested in case of the in-app Assistant – may negatively affects usability. However, too many instructions can mask usability too.

4.11.4 Recommendations

The main lesson to be drawn from the above is that while smart features make performing the tasks easier (i.e. scanning is faster than manual input), they do not necessarily make respondents understand the task better. Moreover, it should be kept in mind that while such features are 'smart' because they automate some actions and perhaps sometimes even automate more than necessary,

respondents are not always 'clever' enough to use the smart features or follow the logic behind them. From this follows the recommendation that **online and smart surveys do not mean that interviewers no longer play a role.**

On the one hand, the role of the interviewer remains unchanged. Human contact still plays an important role in convincing respondents to participate in a study (see WP2.1). On the other hand, the role or task of the interviewer does change compared to traditional face-to-face or telephone interviews. For doing complex(er) diary studies such as HBS or TUS that are characterised by less obvious or less common actions and thus have a steep learning curve, it is important that the interviewer helps and explains to the respondent how to perform tasks with the application. While non-personal aids (e.g. instructional video, the Assistant in MOTUS) will suffice for a large proportion of respondents, there is also a proportion of respondents who want to be able to ask questions or be confirmed in performing the task as expected. The interviewer will thus partly become a personal helpdesk or case manager. The tests show that once respondents get the hang of it - with or without help and explanation - they can easily repeat the tasks. An alternative is to work with a group of experienced respondents who are, for example, part of a panel or that enters through a deliberate enrolment procedure and thus are well informed in advance about what is expected of them. Another alternative is – as explored in France – to still rely on mixed-method approach to complex diary studies.

4.12 Trust and privacy

4.12.1 Results

There is no mistrust. All respondents have trust in the application. This confidence is mainly formed by the professional design and the absence of advertisements. In Germany, the fact that Destatis is the application's publicist inspires trust. In Belgium, the fact that it is the Vrije Universiteit Brussel that conducts the survey also inspires trust.

As a result, respondents have no problem entering their expenses in the application. Only a few said they omitted sensitive expenses (e.g. spending in a pharmacy). It also made no difference to enter a receipt from the researchers or their personal receipt they were asked to bring. Most respondents are aware that their receipt is stored on a server and the data is further processed by an NSI or by researchers. Also, as a result of trust in the application but especially in the implementers of the study, respondents are not concerned about their privacy and data security. Of course, test participants are typically more positive (i.e., more trusting) than respondents in a probability sample or those that are hard to recruit. Finally, respondents believe that using the OCR microservice leads to more reliable data.

4.12.2 Benchmark

Similar to the results above, both CBS, Insee and SSB are considered trusted institutes, and no respondents spontaneously raise concerns about their privacy. The supported public administration-ID login in the Norwegian HBS application definitely adds to this. Respondents in the French and Norwegian HBS applications mentioned that the conservative design and simple use of the application adds to its seriousness and trustworthiness.

As a result, they trust the application and only a few mentioned that they would not report certain expenses (e.g. medical expenses or mortgage, or jewellery). One respondent in the Dutch @HBS application mentioned that since the expenses are to be kept at the household level, (s)he would not want to share expenses with her/his parents in the household. Some respondents in the Norwegian HBS test did mention that they might try to avoid being (wrongly) perceived or stereotyped, for example, when buying products like alcohol, candy, or energy drinks.

In the French and Norwegian HBS applications respondents perceive scanning as more accurate than manual entry, although the Norwegian HBS test reveals that respondents might put too much confidence in technology and the system's ability to assure the quality and accuracy of the data. Most perceive checking and editing the results from the scanned receipts not as their responsibility whereas they do so in case of manual entry. In the Dutch @HBS test several respondents indicate they think scanning yields a better data quality. The reasons given for this are that the data entered will be more complete and that it will require less effort. Some think that the photo enables the statistical office to check and correct the scanned data. Others, however, say that they cannot judge this or that the input method does not matter. Respondents in the French HBS application consider the scanning method more reliable because it reduces errors. Of course, in the setup of the French HBS application respondents were not shown any data processed from the receipt photo and so they might be less aware that this step could generate errors compared to respondents who used an application with simulated or linked OCR processing.

Accuracy seems of little concern when the total amount is low because it is thought to have little impact on statistics. Concerns about spelling accuracy (e.g. product names or shop names) are almost non-existent. Respondents do not see this as an issue and thus do not see the need to correct it.

4.12.3 Usability attributes

The UI/UX of the application contributes up to a certain point to trust and credibility and to security and privacy. However, it is mainly the issuing or executing institution's trustworthiness and transparent communication that positively affects trust, security and privacy. Also, the availability of RSM is seen as professional and an efficient way to collect data and thus contributes to the credibility. The fact that respondents are given control over the data and information they share, is not perceived as adding to trust, security and privacy.

4.12.4 Recommendations

The main finding to take away from these results is that not only the professionalism of the application but the trust in the publicist of the application and executor of the study create trust. If that trust is there, privacy and data security are not likely to be a stumbling block to participating or not. Trust in NSIs can be further enhanced by:

- Having a well thought out recruitment and communication plan.
- Having several modes of contacts.
- Provide information about surveys' authenticity on institutes' websites.
- Use interviewers (i.e. human contact).

In countries where trust in governments and/or NSIs is lower, it might also be an option to collaborate with university partners.

Note that data on groceries is generally not perceived as sensitive information. Entering other expenses or fixed monthly costs might alter respondents' opinion on trust, privacy and data security.

Table 4.4 Usability Attributes: Overview of how findings on current state of RSM as Integrated in an end-to-end solution relate

The solution that realizes and demonstrates proof of concept for the complete, end-to-end, data collection process contains:	Identified key usability attributes	Installing and onboarding	User interface and use experience	Scanning and creating tentative data	Editing and committing data
1. The involvement and engagement from citizens	Engagement		▽	▲	▽
	Accessibility	▲	▽	▲	
	Clear instructions	►	▽	△	▽
	Time efficient	►		△	▼
	Error handling options	▼			
2. The acquisition, processing and combining of data collected from smart devices and other applications	Intuitive user interface	▲	▽	△	►
	Clear task flow and guidance	▲	▲	▲	►
	Error prevention mechanisms			◄	
	(In app) training		◄		
	(In app) feedback and support		▼		►
3. The contribution to trustworthiness and guarantee of strong privacy safeguards	Trust and credibility	▲			
	Security and privacy	▲			
	Transparent communication	►		△	
	Data collection efficiency			△	▽
	User control over data/information				►

▲ = Good practice (positively affecting usability)

△ = Can be improved (partially positively affecting usability)

► = Present but not perceived by respondents

◄ = Present but not tested

▽ = Need to be improved (partially negatively affecting usability)

▼ = Required (negatively affecting usability)

Table 4.4 *Continued*

The solution that realizes and demonstrates proof of concept for the complete, end-to-end, data collection process contains:	Identified key usability attributes	Manual entry	Checking expense overview	Learning curve and support	Trust and privacy support
1. The involvement and engagement from citizens	Engagement	▼	▲	▽	
	Accessibility			▽	
	Clear instructions	▽		▽	
	Time efficient	▼		▽	
	Error handling options				
2. The acquisition, processing and combining of data collected from smart devices and other applications	Intuitive user interface		▲		
	Clear task flow and guidance	▲			
	Error prevention mechanisms				
	(In app) training	◀		▽	
	(In app) feedback and support		▲	▼	
3. The contribution to trustworthiness and guarantee of strong privacy safeguards	Trust and credibility				▲
	Security and privacy	▲			▲
	Transparent communication				▲
	Data collection efficiency	▽			△
	User control over data/information	▲			▶

▲ = Good practice (positively affecting usability)
 △ = Can be improved (partially positively affecting usability)
 ▶ = Present but not perceived by respondents
 ◀ = Present but not tested
 ▽ = Need to be improved (partially negatively affecting usability)
 ▼ = Required (negatively affecting usability)

4.13 GeoService microservice

4.13.1 User interface and user experience

The display of geolocation data in the MOTUS application is a new user interface and experience (UI and UX). This section discusses the general UI and UX of displaying geolocation data.

4.13.2 Results

Almost all respondents recognize the map marker and link that to where the geo location data should be. They also immediately notice that there are two types of diaries. Many indicate the difference based on colour: the 'white diary' and the 'green diary' (see Figure 4.8). In the Istat test, some respondents remarked that the colours setting of the text in the white diary is too light and not easily readable.

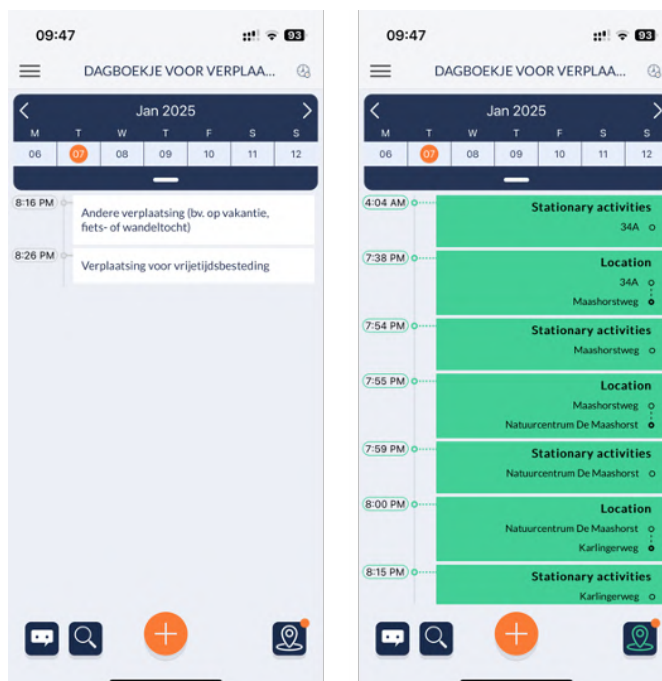


Figure 4.8 The 'white' and the 'green' diary in MOTUS

However, the difference between both diaries is not clear to most respondents. First, respondents assume that both diaries are linked. For example, in the VUB/Statbel test, some respondents expected that when they make an entry based on geolocation data and correct the end time or mode (i.e., if the mode is different from what is suggested) this correction is also visible in the geolocation data. Similar, in the Destatis and Istat test, respondents expect that activities can be added to the geolocation data list.

This ambiguity is further fuelled by the fact that the plus icon (to manually add a new entry) also remains below the overview of the geolocation data. This while the geolocation data actually allows adding an activity from the tentative stop-track records but this option is only visible in the detailed view (i.e., when tapping on one of the green geolocation blocks; see Figure 4.9). Moreover, if using

the plus-icon from within the geolocation diary, the manual entry is made in the 'white' diary, but the respondent is led back to the 'green' diary and does not see the entry made. This is also true for geolocation-based entries (see further).

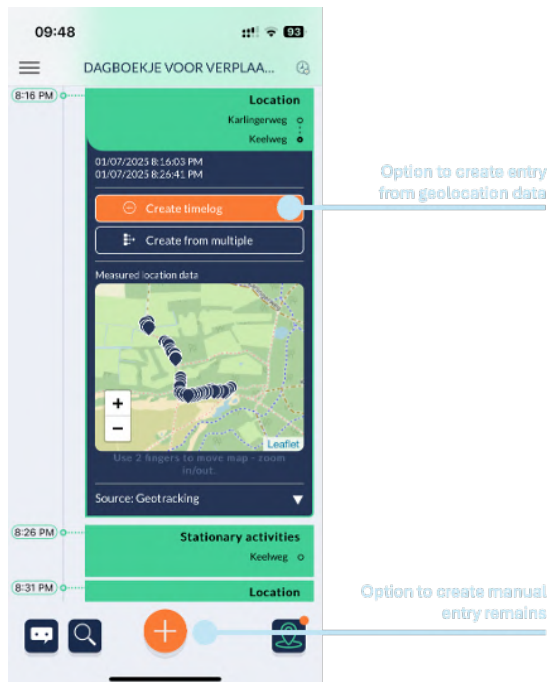


Figure 4.9 The plus-icon remains in the geolocation diary

Although almost all respondents recognize the map marker icon and tap it to go to the geolocation diary, not many notice the colour difference and in the Istat test some respondents said they initially thought that clicking this button would activate geolocation instead of showing the list of movements. The map marker icon – like the Assistant icon and the magnifying glass icon – is white when not used and turns green when used (see Figure 4.10). Missing this colour change results in respondents not knowing how to leave the geolocation diary to return to the actual diary. In addition, some Istat testers highlighted that the text that appears when the icons are clicked is not clear enough and users get confused when trying to understand what it means.

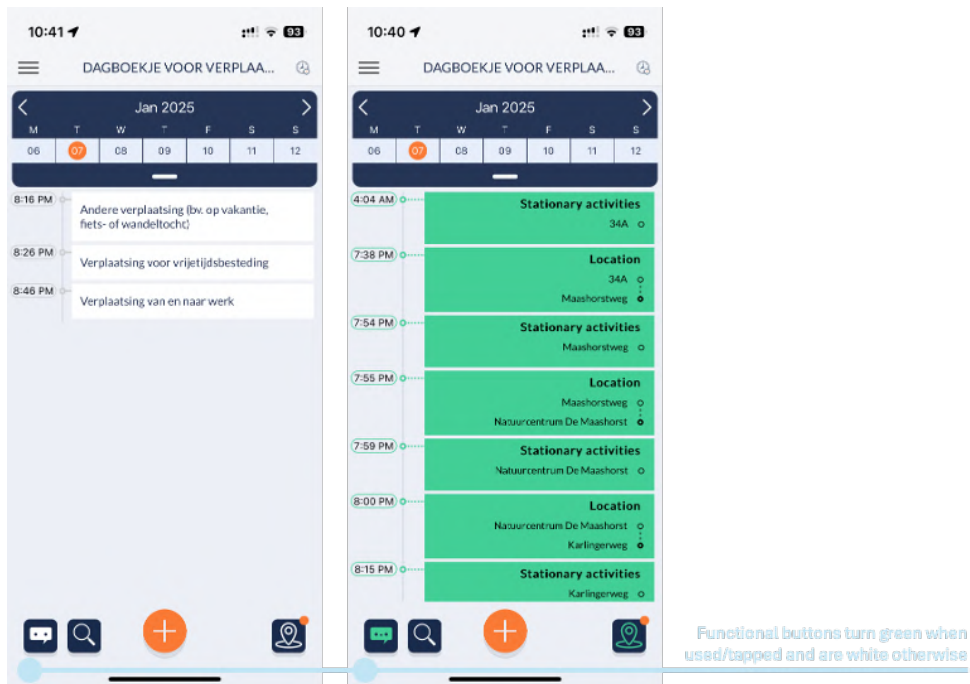


Figure 4.10 Colour changes in functional buttons

All respondents are invited by the green blocks to tap on them and reveal more details. The appreciation of the details are mixed. Some respondents consider them irrelevant or mention that the small map in combination with the high number of markers make it fiddly to use. Especially in case of longer trajectories (see Figure 4.11). Others are surprised by the level of detail and appreciate the geolocation markers on the map to see where they have been.

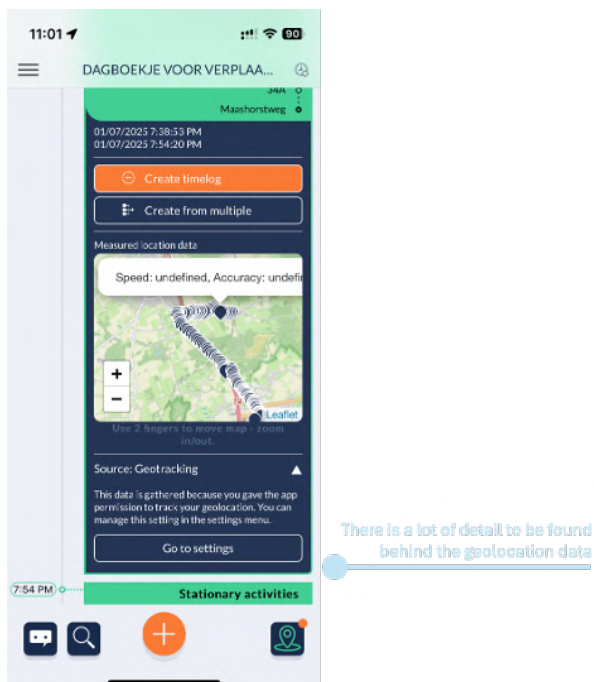


Figure 4.11 Detail behind the geolocation data

Given that all respondents (eventually) click on a green block means that all respondents also (eventually) find the orange button to create an entry based on geolocation data. In terms of the click path to make a geolocation-based entry, some respondents in the Destatis test mention that not having to open the details first, would improve usability. This was not mentioned in the VUB/Statbel and Istat tests. Conversely, respondents in the VUB/Statbel test find it strange that they must both save their entry - after completing the context questions - and then also confirm to insert their entry in their diary (see Figure 4.12). For most, this is an unnecessary extra step. This observation is not made in the Destatis test. Note that very few respondents eventually found out why this step is necessary (see further). Respondents were not able to directly see their entry because after confirming insertion of a geolocation-based entry, respondents are guided back to the geolocation diary. A few looked here for their entry and – in combination with not knowing how to get back to the real diary – could not find their entry and thought they did something wrong. According to the Istat test, most of the respondents found it easy and intuitive the use of the geolocation service to create diary entries, whilst only a few respondents found the orange button integrated into the geo location data somewhat confusing. In the Istat test, some respondents highlighted difficulties related to the diary layout. In particular, the representation of activities that continue over midnight. Some suggested it would visually be easier to understand if activities are split up in the diary such that each diary starts with an activity at midnight.

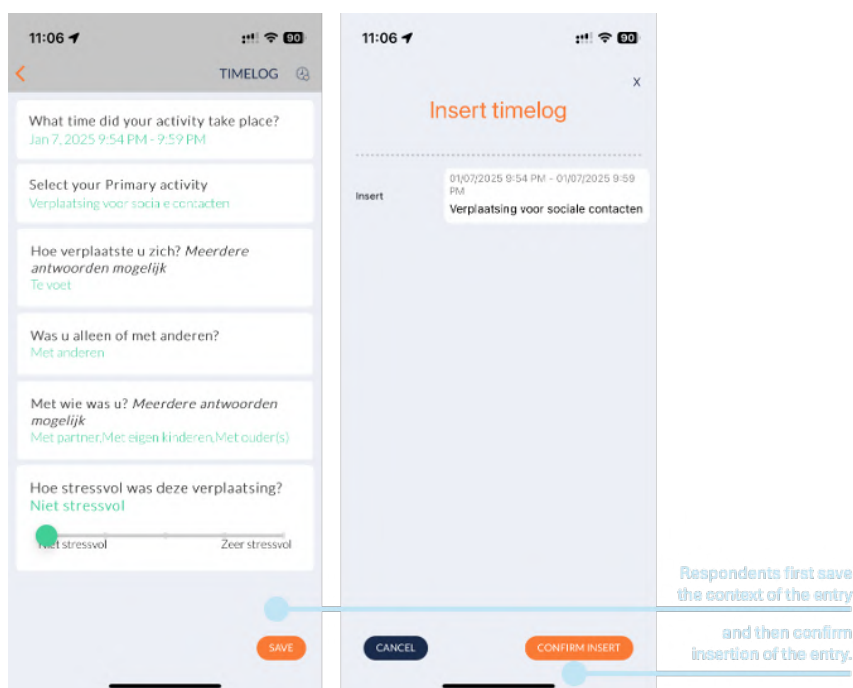


Figure 4.12 Saving the context of the entry and confirming the insertion of the entry

The list of geolocation data is in most cases very accurate and detailed, although some respondents reported missing start times, gaps in tracking, or unfamiliar default locations. This accuracy meant that

even the smallest stop-track changes are picked up. The example in Figure 4.13 shows a nature walk during which a respondent stops several times to, for example, watch birds. Another example often cited is while travelling by train or metro. Stops at different stopping places interrupt the travel route, making a long train journey display a long list of displacements and stationary activities in the geolocation diary. Visually, this makes the list of geolocation data for several respondents overwhelming.

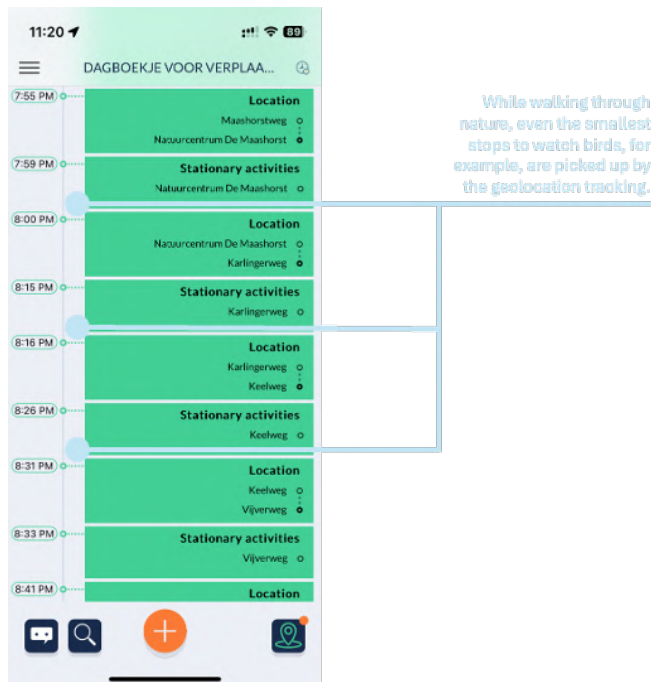


Figure 4.13 Very accurate geolocation tracking detects even the smallest stops

Although only a few indicated that they do not use geolocation data to add entries to the diary for this reason, the majority did not realise (immediately) that there is an option to include stops and tracks together. This option is called 'create from multiple' and is shown under the orange 'create timelog' in the detail overview. When a respondent clicks this, checkboxes become visible under the start times of the different geolocation blocks. This way, multiple, contiguous blocks can be selected (see Figure 4.14). At first, respondents sigh that they have to select each block, which, in the case of the long train journey example, is a lot of work. Through trial and error, some respondents find out that they can select only the last block, and the intermediate blocks are automatically selected as well. Once everything is clear, respondents indicate that this is a useful feature. Similarly, in the Istat test, only one respondent spontaneously recognised the availability of the "create from multiple" option. However, the respondent reported that it was not immediately obvious how to merge or divide periods, because the check boxes are considered too small and appeared in a peripheral position in the field of view. In fact, the checkbox for the first block is in the top left-hand corner and the others may not even be visible because the map takes up most of the screen.

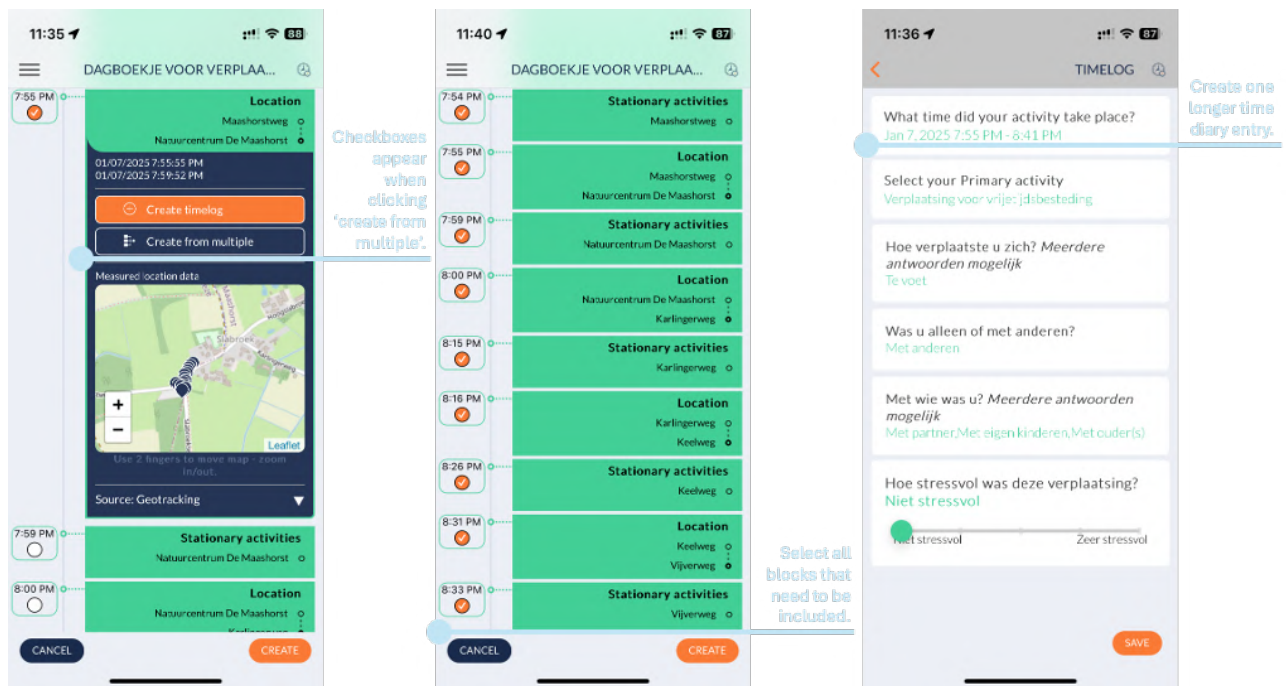


Figure 4.14 Create time diary entry from multiple geolocation blocks

In the VUB/Statbel test, respondents were asked to use the time diary as a mobility diary and only register their displacements. When using the geolocation data to make entries in the diary, almost all respondents mentioned that they missed a visual clue to know which displacements they already committed to the real diary and which ones they still had to commit. Many suggested a fading or lighter green colour. This was suggested by some Istat respondent as well. Moreover, some suggested to use two different colours to differentiate stationary and moving activities. In addition, one respondent suggested a visual status changes (e.g., different colour) when activities were merged or separated. At this point, it became clear to some respondents why MOTUS used the additional step to confirm the insertion of the activity. This was the only screen where they were informed that they were either recommitting an entry or creating an entry that overlapped with existing entries (see Figure 4.15). Not knowing which geolocation data they committed already and only knowing at the end of answering all context questions that there is a potential conflict in their real diary is not considered very user-friendly.

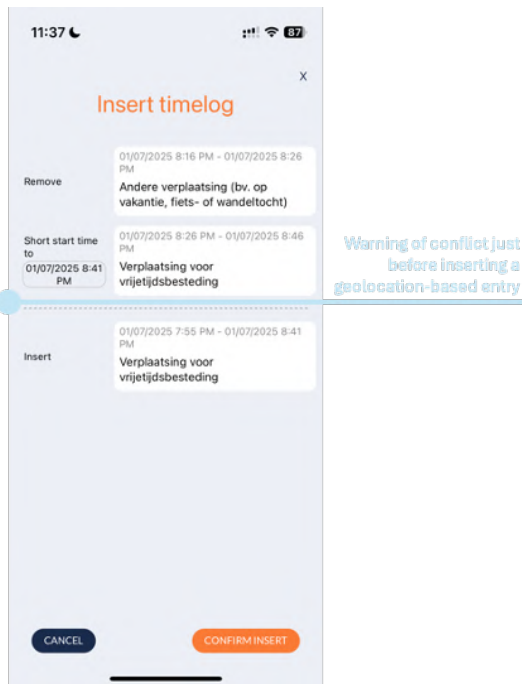


Figure 4.15 Warning about potential conflict between diary entries

Finally, in the VUB/Statbel test with focus on mobility only, a few respondents mentioned that prompts that notified them of a new displacement that needs to be added to their actual diary would be helpful.¹⁰

4.13.4 Relation to usability

User interface and user experience are highly subjective and personal. In general, results show that the usability attributes on engagement, accessibility, and clear instructions are negatively affected by ambiguities. The usability attributes on intuitive user interface and clear task flow and guidance are affected in a similar way by unexpected or counterintuitive click paths. It must be said that this relates to how the GeoService Microservice (GSM) data are eventually presented in the MOTUS application and not how the GSM itself performs.

4.13.4 Recommendations

Several usability issues might be solved with **communication and managing expectations from respondents**. These are usability issues that once known to respondents are easily solved. Focus should thereby lie on:

- Explaining the difference between the two diaries.
- Explaining the tentative (supportive) nature of geolocation data in the green diary and the need to commit data into the white diary.
- Explaining the on/off, forth/back toggling and corresponding colour change of the functional icons.

¹⁰ Interview observation: such a suggestion might be very useful in a mobility study but is not compatible with the current way of tracking. To stay with the example of the train journey, this would imply that respondents would get prompts at every train stop.

Some other suggestions to improve usability relate to **the interface and respondents' intuitive behaviour and expectations**. The most relevant suggestions are listed below:

- Do not display the plus-icon in the geolocation data overview.
- Instead add the plus-icon and the multiple-icon in the green geolocation block.
- Always redirect back to the white diary regardless of whether a manual or geo-based entry is made.
- Change the colour of the geolocation data used to create diary entries.
- Avoid light green fonts on white screens and, in general, reconsider colour contrast for easier reading.
- Eliminate the confirmation step in geo-based entries but notify about clashes as part of the time settings of the entry (i.e. before selecting main activity category).

4.14 Tentative geolocation data

4.14.1 Results

As already mentioned, most of the respondents were surprised by the accuracy of the geolocation tracking data and for most of the respondents it worked well. However, some issues occurred. The most frequently mentioned problem was the late appearance of geolocation data. Many respondents noted that there was a delay in generating the geo-data. Some respondents experienced that the geolocation data were not available until the next day, and for a few, not at all.¹¹ Some other respondents noted that their smartphone battery drained noticeably faster.¹² And finally, there were several respondents for whom the times of the geolocation data did not match their actual behaviour. Mostly it concerned the start time, suggesting that the geolocation tracking mode of the application had a delayed response (e.g., to become active from sleep mode). This prompted some of the respondents to try and correct the geolocation data. Delayed appearance of the geolocation data and high smartphone battery consumption have also been mentioned also by the Istat testers.

One of the observations of Destatis' test was that each geolocation data block simply stated: "start location" and "end location". Respondents found this not very useful information and would rather see more detail here. In the updated version used during the VUB/Statbel test, street names or names of locations were already displayed here. Respondents rated this as useful and clear. This also applied to the suggestion of the means of transport used. However, here respondents noted that the infrastructure does not always allow them to estimate the means of transport correctly. For example, when a tram line lies between two carriageways (see Figure 4.16).

The most often mentioned issue with geolocation tracking, however, related to the stops of public transport modes splitting a trajectory in multiple stops and tracks as mentioned above. Knowing the detail of the geolocation tracking, respondents wondered whether the microservice could not simply realise that one continues along the same train, metro or tram track and thus cut out the stops.

¹¹ These were respondents with a Samsung Galaxy a50 device and an iPhone 15 Pro and those who used Google Pixel 7a and Samsung Galaxy a34

¹² These were respondents with a Samsung Galaxy s22 devices.

When it comes to using geolocation tracking, some respondents mentioned that they can imagine situations in which they do not want to be tracked or – when simultaneously mentioning battery life – just want to be able to turn geolocation tracking off. They question why they cannot do this from within the application but must go to settings instead.



Figure 4.16 Geolocation tracking where the tramline runs in between the R22 ringway in Brussels

4.14.2 Relation to usability

Tentative geolocation data relate positive to engagement. Generally, respondents are pleasantly surprised by the detail. Accessibility to geolocation data would benefit even more by in app access to permission settings. Depending on the study, geolocation data are a time efficient way to record transport activities or support reconstruction one's day. However, delays in data availability and inaccuracies hamper time efficiency. Feedback on geolocation data is good but would benefit more from more detail. However, at the same time, more detail jeopardizes security and privacy.

4.14.3 Recommendations

The GeoService microservice is still under development. In future development, it is recommended to be further improved in the following terms:

- Eliminating delays on geolocation data availability and improving time accuracy.
- Additional learning algorithms could possibly improve mode or transport suggestion and/or merge trajectories with (many) short stops into one block.
- Provide the ability to turn geolocation tracking on/off in the app, like the time tracker for ongoing activities in the time diary.

4.15 Geolocation-based entries vs manual entries

4.15.1 Results

As it turned out, the type of survey makes a big difference in the use of geolocation-based entries. Recall that Destatis and Istat set up a time use survey in which respondents had to keep their activities for 48 hours whereas the VUB/Statbel set up a mobility survey in which respondents only had to keep track of their displacements. In case of the latter, almost all respondents would only use geolocation-based entries. In this case, respondents really use the geolocation data as tentative data that they only need to commit to the actual diary and not necessarily as anchor points or reminders. The way they commit the data varies. Some do it immediately after changing location, others only at the end of the day. A few even ask for prompts to commit the tentative data. When asked about manual input, most (working) respondents indicated that for weekday geolocation data are not actually needed because their schedule follows a routine to a large extent. On weekends, this is different and that is when manual input is considered less convenient. In addition, some respondents indicated that geolocation data prevents them from having to use the time setting. In particular, iOS users denounce the arrows and miss the scroll wheel.

It is a different story in the time use survey for two main reasons. Firstly, since respondents in a TUS are asked to record all their activities for at least 24 hours, many respondents indicated that immediately after stating their activity, they already entered it in the time diary. In this case, they find the support of geolocation data unnecessary. Related to the above, the delay in availability of geolocation data often rendered it impossible to use these data when recording activities close to real-time. Not even if it involves changing location. Secondly, if there is no change of location, geolocation-based tentative data provides a long single suggestion for an activity (e.g. time spent at home). However, a TUS requires detailed time use activities (e.g. ironing, cooking food, washing dishes, taking care of children, watching TV). The operations required to first commit the entire location block to the actual diary and then divide it into shorter activities via the split function in the diary is too complex and cumbersome, according to respondents (see Figure 4.17). In both cases (i.e. real time entries and no change of location for a long time), respondents prefer manual entry. At some point there might even be a risk of suggesting that stationary blocks might be conserved a single entry carrying in it the risk of under-report activities. In a large-scale survey, this could lead to a lack of data on the description of daily activities.

However, there are also respondents who see the geolocation data not necessarily as tentative data but as a reminder. These are respondents who enter their activities in bulk into the time diary in the evening. They indicate that the geolocation data act as anchor points that help them reconstruct their day. So, in this case, it is not necessarily about committing tentative geolocation-based entries, but rather supporting information to complete their time diary.

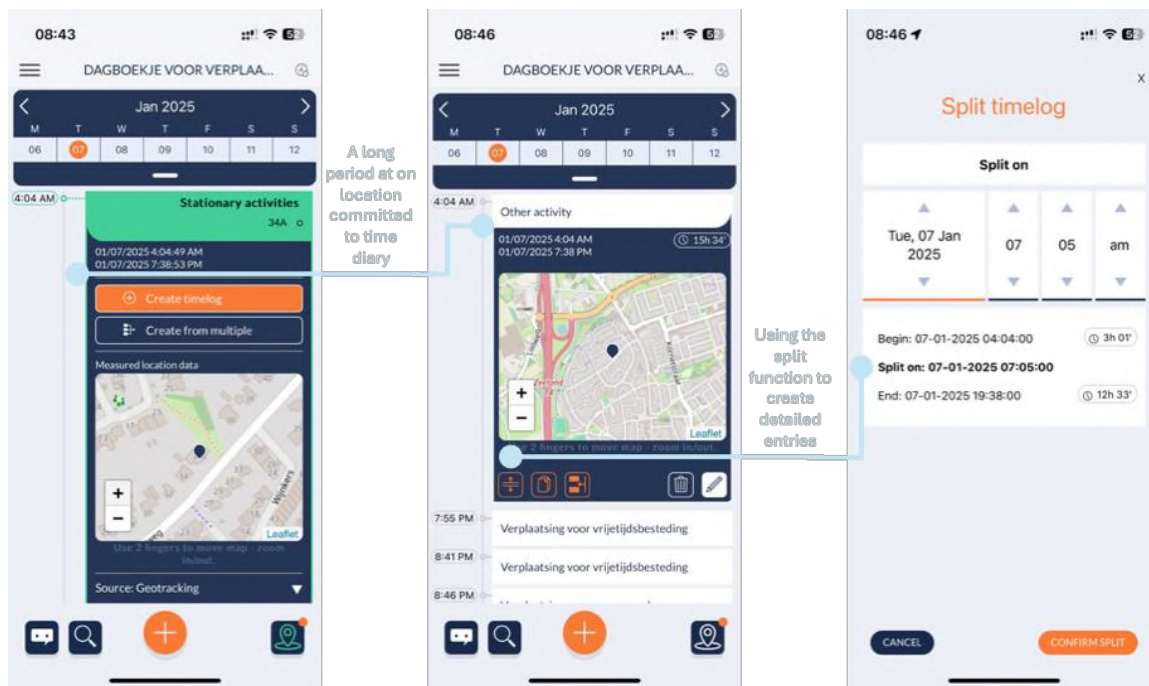


Figure 4.17. Committing and splitting of long stationary suggestion to diary

Box 4.4 Click and action path geolocation-based entry

MOTUS

Click map icon > click “time” > select start and end time > click “confirm” > click “main activity” > select main activity from category list > click “confirm” > click, answer and confirm context (i.e. secondary activity, location, travel mode, presence of others, ...) > click “save”

Box 4.5 Click and action path manual entry

MOTUS

Click + icon > click geolocation block > click “create time log” > click “main activity” > select main activity from category list > click “confirm” > click, answer and confirm context (i.e. secondary activity, location, travel mode, presence of others, ...) > click “save”

4.15.2 Relation to usability

The relation of geo-based versus manual entries to usability attributes again relates to the study objective. Geo-based entries are positively affecting engagement and time and data collection efficiency in mobility studies but less so in time use surveys. The geolocation data are presented as tentative data that need to be committed to the (time) diary. Although this is designed to give respondents control over their data, not all respondents understand or acknowledge this. As mentioned in the section on the Household Budget Survey the click path for manual entries is more intuitive than for geo-based entries affecting the usability attribute of intuitive user interface.

4.15.3 Recommendations

The main recommendation here is **to recognise that the function of geolocation data varies by survey type**. This leads to different uses of geolocation data and therefore different expectations in terms of usability. For example, it seems that in mobility surveys, geolocation-based data is quite desirable to quickly and easily add displacements to a travel diary. In terms of usability, then, the easier it is to go from tentative geolocation data to committed diary data, the better.

However, in the case of time-use research, geolocation data seems to have more of a supporting or recalling role that helps respondents reconstruct their day in the time diary. In this case, it is less often about committing tentative data – especially in the case of extended time with no change of location – but rather anchor points that make respondents remember when they were where during the day. In terms of usability, this is precisely where having a clear overview of the day is important.

4.16 Trust and privacy

4.16.1 Results

Similar to the tests of the Receipt Scanning Microservice (RSM), respondents in this test also indicated that much of the trust lies with trust in the performer of the survey and/or the publicist of the application. The design of the application and the flawless functioning of the app also contributes to this.

Many respondents indicated that they see little problem in sharing their location as long as it is clear to them what it is for. For example, Destatis asked respondents when they generally allow geolocation tracking, and the answers showed that this was mostly when they saw very clearly its usefulness or added value. Think of using maps to navigate, location sharing to plan routes with public transport, or to order a taxi or Uber, or to share locations chat apps. From that reasoning and combined with the reliability of the implementing institutions, respondents see the usefulness of turning on geolocation within the test survey. However, it does appear that still relatively many respondents assume that the researchers get the geolocation data anyway and a few even think that the researchers use the geolocation data to check the diary.

Concerns also emerged from the Istat test. Some respondents reported that they usually do not allow the geolocation tracking or that they allow it tracking under certain conditions. These conditions too related to the clarity and accuracy of the information about the use of geolocation data, whether the geolocation data actually help completing the survey or improving the user experience, and that the survey using geolocation data is a certified statistical survey and privacy is respected. Moreover, respondents require guarantees that data will not be shared with third parties and that the location information will remain on their mobile phone and transparency about the use of this data, the security measures taken, the limitation of storage, and the possibility to revoke consent at any time. Related to this, some respondents remarked that they remained logged in even if they close the app (e.g., swipe the app upwards on iPhone). This too raised some privacy concerns and made them ask for a clearer option in the app to log out.

When respondents were asked about suggestions based on geolocation tracking, again the majority said they had no problem with this. As mentioned above, respondents in Destatis' test liked to see

more detail of the start and end points, and respondents in VUB/Statbel's test liked to see an underlying algorithm processing the data to avoid splitting trajectories of public transport into multiple stops and tracks. Respondents in Destatis' test also expressed interest in more detail in their displacement data, including travel mode, travel distance, travel time, and travel speed.

Finally, in the VUB/Statbel test, respondents were presented with a hypothetical situation in which the stationary geolocation data would be further enriched with information about location (e.g. football club) and matching suggestion of activity (e.g. playing sports) to further enhance the ease of use of completing a diary. Responses to this were threefold. Some respondents found this useful and would possibly use it as a function of their participation in the study. A large proportion of respondents still felt this was too intrusive on their privacy. A sports club might still be possible but certainly work and home location was considered a no-go by many. This echoes the comment made by some respondents during Destatis' test that they would like their home location not to be marked with precision. Finally, there were also some respondents who were less confident in the suggestions of activities (e.g. is one at the football club as a sportsman, as a spectator, to bring their child, or on a social club night). This is in line with the finding from Destatis' test in which large blocks without a change of location tend to contain many different, detailed activities.

4.16.2 Relation to usability

Similar to the conclusion on the Receipt Scanning Microservice (RSM), the UI/UX of the application contributes up to a certain point to trust and credibility and to security and privacy. However, it is mainly the issuing or executing institution's trustworthiness and transparent communication that positively affects trust, security and privacy. Allowing the GeoService Microservice (GSM) to track respondents is almost secondary to this. Several respondents do start to have concerns about their privacy when presented a hypothetical situation wherein the GSM becomes controlling. The fact that respondents are given control over the data and information they share, is not perceived as adding to trust, security and privacy.

4.16.3 Recommendations

Again, as with RSM, **the use of GSM must be accompanied by clear communication: why and how are geolocation data used**. Respondents are generally willing to turn on their geolocation tracking, especially because they are used to doing this in other apps. However, a condition for this is that they see the benefit of it. This may be fairly easy to make clear for a travel survey, but that it may be more difficult in a time-use survey.

An **option to toggle geolocation tracking on/off in the application** itself can facilitate such communication. From the RSM tests we found that respondents do not or hardly read onboarding screens. At the same time, that is the moment when the application asks to (always) allow geolocation tracking. If the benefit or necessity of this is not clear at that moment, there is a chance that permission will not be granted. (Note that for the tests, respondents were explicitly asked to enable geolocation tracking.) Once in the application, it appears difficult for less technically savvy respondents to find where in the settings they can still enable location services for the MOTUS application.

From a privacy and security perspective, is recommended to provide **a simpler (more accessible) way to log out** from the application. The question remains of course whether the geolocation remains

active when the respondent logs out. Moreover, this may lead to other user issues such as logging in again, forgetting - and therefore having to reset - the password, etc.

Table 4.5 Usability Attributes: Overview of how findings on current state of GSM as Integrated in an end-to-end solution relate

The solution that realizes and demonstrates proof of concept for the complete, end-to-end, data collection process contains:	Identified key usability attributes	User interface and use experience	Tentative geolocation data	Geo-based entries vs manual entries	Trust and privacy
1. The involvement and engagement from citizens	Engagement	▽	▲		
	Accessibility	▽	△		
	Clear instructions	▽			
	Time efficient		△	△	
	Error handling options				
2. The acquisition, processing and combining of data collected from smart devices and other applications	Intuitive user interface	▽		▽	
	Clear task flow and guidance	▽			
	Error prevention mechanisms				
	(In app) training				
	(In app) feedback and support		△		
3. The contribution to trustworthiness and guarantee of strong privacy safeguards	Trust and credibility				▲
	Security and privacy		(▽*)		▲ (▽*)
	Transparent communication				
	Data collection efficiency				
	User control over data/information			►	►

▲ = Good practice (positively affecting usability)

△ = Can be improved (partially positively affecting usability)

► = Present but not perceived by respondents

◄ = Present but not tested

▽ = Need to be improved (partially negatively affecting usability)

▼ = Required (negatively affecting usability)

*In the hypothetical case where geolocation data are becoming controlling of their (time) diary, for example, because geolocation data are used to suggest activities or identify activities that conflict with geolocations.

4.17 Take home messages: general

On complexity

Conducting Household Budget Surveys (HBS) and Time Use Surveys (TUS) presents significant challenges due to their complexity and the burden they place on respondents. These surveys might benefit from a shift away from random population sampling to relying on panels. However, engaging respondents effectively remains a challenge, particularly when the tasks assigned to them are intricate and time-consuming.

A major difficulty lies in helping respondents understand the purpose of the survey and their specific responsibilities. In case of HBS, participants need clear guidance on how to categorize expenses correctly in the app, how detailed their entries should be, and the importance of including all expenses for all household members. Additionally, when moving towards smart features such as a Receipt Scanning Microservice (RSM), a critical task—collecting all physical receipts—is often overlooked, as many receipts are lost, no longer used, or exist only in electronic form. Without these, survey data may be incomplete or inaccurate. In case of TUS, participants too need clear guidance on how detailed their entries should be. When moving towards smart features such as a GeoService Microservice (GSM) for which respondents need to give explicit permission to be tracked, the purpose of doing so, needs to be very clear.

Another key insight is that respondents do not share the same perspective as researchers when it comes to data collection (i.e., they do not know its origin and purpose). Traditional paper-based surveys were seen as highly burdensome, but replacing them with digital tools does not automatically resolve the issue. For example, many respondents do not see the need for complex classification systems such as COICOP or HETUS and expect survey apps to be as intuitive and easy to use as any other modern application. If these expectations are not met or new, clear expectations are not set, engagement and retainment may be affected.

On interviewers

Additionally, online smart surveys do not replace the need for interviewer support. While smart features can streamline some aspects of data collection, respondents still require human assistance for clarification, motivation, and troubleshooting. Neglecting this aspect could result in lower data quality and higher dropout rates.

On smart versus clever

Indeed, developments such as a Receipt Scanning Microservices (RSM) and GeoService Microservice (GSM) provide smart solutions to improve data collection. However, their ‘smartness’ must be aligned with the ‘cleverness’ of respondents to use them. In other words, their application must be carefully managed to avoid making HBS and TUS even more complex.

An important example here, is the part of smart features within a privacy-by-design approach that includes respondents’ control over their data. Smart features like the microservices designed within this project present smart data as tentative data that only become committed data (i.e., data to be used by researchers/NSIs) after the respondents commit the data to the dairy. This is not clearly perceived by respondents and thus not acted upon, potentially leaving tentative data uncommitted.

On communication

Effective communication is crucial throughout the entire survey process. It is essential to inform respondents when and why they are being contacted and provide multiple touchpoints for engagement. A combination of letters, text messages, homepage updates, and interviewers should be used to ensure that respondents recognize official communication and prepare for participation. In a digital landscape where people tend to ignore unknown contacts, building familiarity through multiple modes of outreach helps legitimize the request and can significantly improve response rates.

On implementation

Furthermore, even when a microservice is available for a specific task, each country that adopts it must go through additional implementation steps. This includes integrating the microservice into a national survey app, training the algorithm with national input, and adapting it to local characteristics such as language and receipt structures. These steps require time, resources, and technical expertise, making cross-country implementation more complex than it may initially seem.

On trust and engagement

Trust in survey apps plays a pivotal role in participation. Respondents are more likely to engage with apps that are backed by National Statistical Institutes (NSIs) or universities, have a modern, intuitive and user-friendly design, and are free from advertisements. Future research could explore how the requirement to download a study-specific app influences usability perceptions, engagement levels, and overall trust in the survey process.

Encouragingly, participants are open to granting access to smart functions—such as geo-tracking and camera use—if they understand the rationale and perceive a clear benefit. This reinforces the importance of transparent and effective communication about why these features are necessary and how they contribute to the survey's success.

Finally, processing speed and result quality are critical for maintaining engagement. Microservices used in survey applications must deliver results quickly and accurately—respondents do not want to wait for slow processing, nor do they want to correct errors manually. If they encounter delays or inaccuracies, they may experience confusion, dissatisfaction, and ultimately decide to drop out of the survey. Ensuring seamless, high-quality performance is therefore essential for retaining respondents and collecting reliable data.

4.18 Take home messages Receipt Scanning Microservice

On using RSM

Using the Receipt Scanning Microservice (RSM) has emerged as the preferred method for data collection in a Household Budget Survey (HBS), as it significantly reduces respondent burden compared to manual entry. However, expectations for scanning accuracy and efficiency are extremely high. Respondents compare these microservices to well-established commercial scanning apps, such as those used by insurance companies, and expect a flawless, near-instant experience. Any deviation

from this standard—such as slow processing, incorrect recognition, or frequent need for corrections—can result in frustration and survey dropouts.

Despite these challenges, RSM remains the best path forward. To minimize manual entry, RSM should be combined with additional data streams, such as barcode scanning, links with loyalty cards, online bank statements, and digital invoices.

On training

Training is essential for improving receipt scanning accuracy, but training an effective RSM requires a large and diverse dataset. It is crucial to collect a wide range of receipts from different shops and supermarkets to ensure that the algorithm performs well across various formats. Additionally, sufficient manpower is needed for manually labelling and categorizing receipts, a step that remains crucial for machine learning development. However, when the survey is repeated, the process becomes more efficient, enabling reuse and continuous improvements, and reducing the need for manual intervention over time. Without comprehensive training data, scanning accuracy will suffer, leading to increased respondent effort and dissatisfaction. Additionally, training should be maintained and updated to accommodate new products, new stores, changes in receipts, and so on.

On manual entry

A critical takeaway is that offering *only* manual data entry is not a viable option. Respondents strongly dislike entering receipts manually, especially when they are long. Manual operations are preferably avoided, which shows from respondents' reluctance to spend time correcting errors in scan results. At this stage, respondents only consider manual entry when the effort required to fix scanning mistakes is higher than simply inputting data manually. That said, manual entry should undoubtedly always be part of a HBS, but a well-functioning RSM is crucial for engagement and retainment in HBS.

On classifications

Another major challenge is the discrepancy between how respondents and statisticians classify expenses. While statisticians require detailed and structured classification, respondents do not naturally think in terms of categories like COICOP and often do not understand why such classifications are necessary. Therefore, clear communication is essential. Respondents need to be informed upfront about how much detail is required in their entries (e.g., should they report simply “food,” specify “fruit,” or provide full details such as “organic fair-trade banana from Peru”?). On the other hand, this must be handled delicately, as emphasizing the complexity of the task too much may discourage participation. Wherever possible, classification should be automated through a dedicated microservice, using machine learning to reduce the burden on respondents while maintaining data integrity. And even at this point the discussion might be focussed on how much of this classification is fed back to respondents to check or is kept in the back-office for manual in-house checking. Classifying expenses on a detailed level cannot be left only to the respondents.

On UI and UX

Consistency in labelling and naming conventions is crucial to ensure usability. All function names should be unambiguous, uniform, and aligned with standard user expectations. For example, a function labelled “scan receipt” should do exactly that—avoiding vague or misleading terminology that could cause confusion. Similarly, users expect conventional labels for familiar actions, such as 'save entry' for storing data, which helps them feel confident in their interactions with the app.

The way scan results are displayed also plays a key role in usability. The latest scanned receipt should always appear at the top of the list, ensuring that respondents can immediately see their most recent entry without having to search for it. If scrolling is necessary (e.g., due to small screen sizes), visual cues should guide users, helping them navigate their scanned receipts effortlessly. Overlooking scan results due to poor visibility can lead to confusion, errors, and reduced engagement.

On data collection

Ultimately, balancing data quality with respondent burden is essential. While increasing context and detail improves data reliability, it also raises the question: how much can we reasonably ask of respondents before participation becomes too demanding? Finding this balance will be critical in ensuring high-quality data collection while maintaining a positive user experience.

4.19 Take home messages Geo Service Microservice

On using GSM

Like the Receipt Scanning Microservice (RSM) the GeoService Microservice (GSM) offers a powerful tool for improving data collection in time use surveys, helping to reduce recall bias and respondent burden. However, geodata can serve different purposes, and its role in the survey affects its usability. Sometimes, geo-tracked locations serve as tentative data –suggestions that respondents can review and commit to their diary. Other times, they function as anchor points, helping to reconstruct the day and infer activities that need to be logged. Recognizing these different roles is essential to improving usability and data accuracy.

Moreover, there is a balance between a supportive and controlling perception of a smart feature like GSM. While geolocation data can enhance engagement and retainment in a time use or mobility survey, respondents must not feel monitored or controlled. Clear communication is key. Respondents should be informed about what geodata is collected, how it is used, and why it benefits the survey and their participation in it. While not all respondents will be interested in these details, transparency fosters trust and willingness to participate.

On geo-based versus manual entry

Geolocation based entries are the way to go in mobility surveys. To maximize GSM’s value in time-use or mobility diaries, the data it provides must be accurate, useful and directly transferable to diary

entries. Current delays in availability, small inaccuracies in starting and ending times, and the issue of long stop-track lists when using public transport could be improved. Additional derived context such as the prediction of transport mode (e.g., walking, driving, public transport) is considered to improve usability.

Geolocation based entries in time-use surveys are less straightforward. Large blocks of being stationary in one location does not provide much extra help. The problem here is that in situations in which a respondent frequently visits a specific location and GSM could suggest likely activities performed there, this could be considered too invasive for some respondents. So, enriching these suggestions with activity-based insights would not necessarily make GSM a more valuable tool for respondents. Obviously, a GSM is less straightforward in its use as a smart feature than RSM.

Manual entry will always remain an important option. Some respondents prefer to manually input activities rather than rely on automated suggestions, and GSM is only beneficial when respondents change locations. If a person spends most of their day in a single location but engages in different activities, GSM does not significantly reduce their workload. Additionally, the benefit of GSM depends on how respondents enter activities in a time diary. Those who log activities at the end of the day may find geo-based suggestions particularly useful as memory aids, while those who enter activities in real time may not rely on GSM as much.

On trust and privacy

It would be beneficial if users have an easy, obvious way to deactivate GSM, preferably from within the application. Location tracking should be an opt-in feature rather than a requirement, and respondents should feel in control of their participation. If deactivation is complicated or unclear, some participants may opt out of the survey entirely rather than engage with features they do not understand or trust.

On UI/UX

The user interface (UI) must clearly distinguish between the timeline with geolocation data and the 'real' diary entries to reduce confusion. Similarly, it must be clear to respondents whether they are creating a manual or geo-based entry. Furthermore, respondents should also easily recognize which geolocation data they already committed, especially when working of a long list of stop-track entries. All this is especially important to ensure that respondents who want to use geo-based suggestions actually do so, rather than defaulting to manual entry out of habit or uncertainty

5 Mode effects and other data integration considerations: Insights from task 2.4

5.1 Introduction

This chapter presents the outcomes of Task 2.4, focusing on the evaluation of mode effects introduced by smart surveys. The primary aim is to understand how these effects may impact data quality and comparability over time, and suggest mechanisms to compensate and adjust for them over time. The experiments embedded within the field tests, as discussed in Chapter 2, were designed to address these concerns and provide insights into how National Statistical Institutes (NSIs) can best transition towards Smart Surveys.

Mode effects are a crucial consideration because they directly influence the quality and comparability of data. If data differ significantly due to mode effects, it becomes challenging to compare them over time, which is essential for longitudinal analysis. The innovative features that may be utilized in Smart Surveys are expected to introduce larger mode effects with respect to traditional survey modes than previous mixed mode implementations have delivered. This is due to two main factors: the potentially different characteristics of respondents who complete smart surveys (representational differences), and the nature of the data collection itself (measurement differences).

The introduction of smart surveys potentially risks losing older participants who may be less inclined to use apps, but does offer some benefits. These include capturing data from younger individuals and hard-to-reach subpopulations by decreasing time, effort, and language burdens.

The structure of this chapter follows from low-difference to high-difference scenarios. We begin by examining qualitative differences in responses across various surveys, including France's Time Use Survey (TUS), and Household Budget Survey (HBS), Belgium's TUS, and Italy's TUS. Following this, we compare outcome differences between paper and web respondents on the basis of crossover and matched pair analysis for France's TUS. Finally, we explore potential differences arising from low-smart features such as buttons for user interaction, as well as from high-smart features, comparing an app-based mobility study using geolocations to a traditional diary. More details on the analysis on the size and causes of mode effects resulting from the field tests in France can be found in Appendix D.

5.2 Country experiments

As described in Deliverables M6 (Bucher et al., 2023) and M14 (Bucher et al., 2024) large field tests with embedded experiments intended to address the outstanding mode effects questions were originally planned for France, Belgium, The Netherlands, and Italy. Circumstances prohibited the timely resolution of the field tests in The Netherlands and Italy, and low response rates from Belgium led to an early stoppage of the experimental condition that investigated how paper could be used as a follow-up mode. We refer readers to the earlier deliverables for a theoretical explanation on why these experiments were run (M6), and how they were designed (M14). Below, we will summarize the design of experiments conducted in France and Belgium that inform the results for most of remainder of this chapter.

Time Use Survey in France

The methodology employed for this study involved an address-based, household-based probability sample of 2,100 drawn from the population register. Invitations to participate were sent via postal mail. The study utilized a sequential mixed-mode crossover design, where respondents were randomly assigned to complete either the web-based (CAWI) or paper-and-pencil (PAPI) Time Use Survey (TUS) diaries first. This randomized order assignment assured that each respondent experienced both modes. Additionally, respondents were randomly assigned a specific day of the week (e.g., “Wednesday” or “Saturday”) and were asked to complete the diary for that particular day. They reported on the same day for both modes, but separated by one week's time. This approach was designed to permit a comprehensive comparison of both selection and measurement effects across the two modes.

Household Budget Survey in France

The Household Budget Survey (HBS) in France employed an address-based, household-based probability sample of 2,500 drawn from the population register, with invitations sent via postal mail. Households were randomly assigned to either a PAPI-only condition or a choice condition (between PAPI and app-based). The survey protocol consisted of two face-to-face interviews, each lasting approximately one hour, with the objective of reconstructing the entire budget of a household. Respondents were required to complete both questionnaires and a spending diary for a week between the two visits.

Time Use Survey in Belgium

In Belgium, the TUS design sampled individuals from 6,313 households, proportionally selected from residences in Flanders and Wallonia¹³ to participate in a 7-day app-based diary. Invitations were sent via letter, and the original experimental design randomly assigned non-responders to one of two conditions: a PAPI alternative enclosed with the follow-up letter, or no PAPI alternative with the same follow-up letter. This approach allowed for a comparison of response rates and preferences between the app-based and paper-based diaries. More details on the experimental design and its implementation can be found in Chapter 2 of this deliverable.

5.3 Respondent differences

Sampling for all three experiments was performed at the household level. In France, households were approached by an interviewer with a request to participate in the survey, while in Belgium, sampled households were sent a letter by post. Not all households who agreed to participate in the experiment ultimately did participate in the French design, leading to two distinct response opportunities. We differentiate these as response rates (RR) and dropout. Response rate at the household level is split by PAPI (any paper response), and Web (any web response.) Because of the design of the experiment, households can be both PAPI and Web respondents, and indeed over half of responding households completed both modes.

¹³ The Brussels area was not included so that materials could be delivered in one language only.

For Belgium, the experiment with offering PAPI in the first reminder was discontinued after the initial batches due to the low overall response rate and increased expense of sending the paper versions in the reminders. In total, 596 households were sent paper versions of the survey enclosed with the reminder. Of this group, 11 returned completed PAPI surveys, and 7 initiated the app/web-based diary, but did not complete it. Due to the limited sample size, the planned comparison between the two groups was not possible, and response rates are presented with the combined App/PAPI response. Two different response rates are considered: Response Rate (RR1) reflecting households completing two diary days, and App Activation Rate (AAR) reflecting households with either full or partial completion.

Table 5.1 *Per-mode response rates at the household (HH) level across studies with mode effect experimental designs*

		France TUS		France HBS		Belgium TUS	
Demographic Category		PAPI ¹⁴	Web ⁸	PAPI ⁸	Mode Choice ⁸	App/PAP I ¹⁵	App/PAP I ¹⁶
HH Type	Single men	18.6%	16.6%	33.3%	34.1%	1.3%	6.7%
	Single women	25.5%	19.1%	46.2%	37.2%	2.0%	9.5%
	Couple w/o child	24.6%	21.2%	50.3%	50.3%	2.4%	11.9%
	Couple w/ child	26.1%	22.9%	55.3%	48.3%	3.2%	12.3%
	Single parent	19.5%	15.7%	48.3%	48.4%	1.8%	9.8%
	Other	13.5%	13.5%	37.0%	50.0%	2.2%	8.9%
Area	Urban Priority	11.0%	8.7%	35.5%	37.0%	-	-
Home	House	27.8%	22.6%	55.5%	52.6%	-	-
	Owned	28.9%	23.1%	56.4%	52.3%	2.8%	12.2%
	Social Housing	16.0%	14.2%	36.3%	40.7%	-	-
HH income	Decile < 5	-	-	43.9%	42.9%	1.2%	7.8%
HH income source	1+ salary	23.9%	20.3%	49.8%	44.7%	2.9%	11.7%
	1+ pension	26.6%	21.2%	50.8%	52.4%	1.4%	9.5%
	1+ unemployed	22.1%	19.4%	48.2%	44.3%	2.0%	9.5%
Total		23.0%	19.5%	47.1%	44.3%	2.3%	10.5%

Table 5.1 shows a breakout of response rates across household characteristics, across modes where available. In the France TUS, although both PAPI and Web response rates at the household level show evidence of selection bias across various household characteristics, these differences are

¹⁴ RRc – Response rate reflecting household return of at least one diary of the given type

¹⁵ RR1

¹⁶ App Download Rate, as detailed in Chapter 3

generally non-significant between modes. While Belgium's TUS has no between-mode comparison, the patterns in overall non-response trends within household characteristics are broadly similar.

France's HBS field test involved an experiment comparing PAPI-only to mode choice (PAPI or web) at a household level. Overall response rates were lower in the mode choice condition, but these differences were more distinct in some categories, such as single woman households. More information on individual response rate and attrition in the French HBS test can be found in the country report in the appendix.

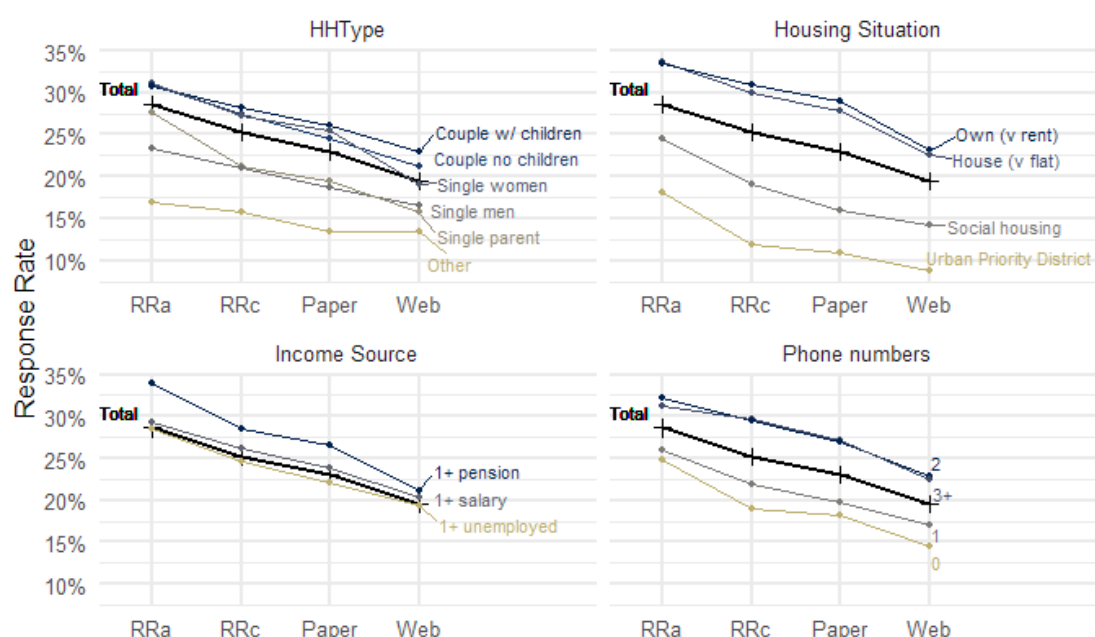


Figure 5.1 Response rate drop-off in France TUS experiment by key household demographic categories. The thicker black line represents the overall average. Lines above therefore demonstrate an amount of overrepresentation within a category, while lines below show underrepresentation.

The French TUS experiment involved a crossover analysis in which sampled households were invited to participate in both the PAPI and the Web mode sequentially. Not all households that agreed to participate returned both diaries for all participating members of the household. Figure 5.1 presents participant response in the form of progressive dropout from the agreement to participate following the initial interview (RRa), return of at least one diary (RRc), return of at least one PAPI diary (Paper), return of at least one Web diary (Web). The dropout pattern remains consistent across household variables, despite initial nonresponse differences. Response rates to PAPI mode are higher than response rates for Web across all household level demographic characteristics.

Although there is little evidence at the household level of serious selection differences between modes, residual endogeneity remains a concern at the individual level. 1045 participants from 601

households returned at least one survey between the two modes, with 57.1% responding to both diaries, 16.7% responding to neither diary, and 26.1% responding to only one of the two diaries. Table 5.2 shows per-mode response by individual characteristics. Overall response differed across individual characteristics: older participants were more likely to return both diaries than younger participants, more educated participants were more likely to return both diaries, and retirees and employed persons were more likely to return both diaries than students. Here too, although response favored the PAPI mode overall, no significant differences were observed across age, education, or employment type between the two modes.

Table 5.2 *France TUS response¹⁷ by personal characteristics, percentage conditional on categories*

Demographic Category		No diaries (N = 179)	Only paper (N = 211)	Only web (N = 74)	Both diaries (N = 581)
Age	< 15	17 (32.1%)	4 (7.5%)	3 (5.7%)	29 (54.7%)
	15-25	23 (24.7%)	15 (16.1%)	2 (2.2%)	53 (57%)
	26-35	23 (16.4%)	23 (16.4%)	11 (7.9%)	83 (59.3%)
	36-45	26 (12.4%)	42 (20%)	7 (3.3%)	135 (64.3%)
	46-55	40 (19.2%)	42 (20.2%)	13 (6.3%)	113 (54.3%)
	56-70	33 (13%)	48 (18.9%)	15 (5.9%)	158 (62.2%)
Diploma	< BAC	85 (25.7%)	70 (21.1%)	13 (3.9%)	163 (49.2%)
	BAC	31 (15.5%)	43 (21.5%)	14 (7%)	112 (56%)
	> BAC	49 (11.4%)	75 (17.4%)	24 (5.6%)	283 (65.7%)
Employment status	Employed	98 (15.7%)	127 (20.4%)	29 (4.6%)	370 (59.3%)
	Unemployed	10 (20.8%)	9 (18.8%)	3 (6.3%)	26 (54.2%)
	Retiree	18 (12.6%)	31 (21.7%)	9 (6.3%)	85 (59.4%)
	Student	28 (27.2%)	14 (13.6%)	4 (3.9%)	57 (55.3%)
	Others	13 (27.1%)	8 (16.7%)	6 (12.5%)	21 (43.8%)

Within the French TUS, the order of modes was randomly assigned by household as either Web-Paper or Paper-Web and unknown to participants at the first interview. Despite random assignment, any potential order effect could bias estimates of mode measurement differences. As shown in Table 5.3, while 77% of Paper-Web respondents completed the questionnaire in the first period, only 54% completed both methods, reflecting a drop of 23 percentage points. Conversely, while 64% of Web-Paper respondents answered the questionnaire in the first period, 57% answered using both methods, i.e. 7 percentage points less. This raises the question of a possible endogenous selection. For example, respondents with a large number of different activities on the same day (and therefore a large number of time slots to be recorded) may have decided out of weariness not to participate in the second survey, especially if the paper diaries were deemed time-consuming to record. A logistic regression on

¹⁷ This response rate is at the individual level, conditional on initial household agreement at the first interview.

the data from Table 5.3 testing for an interaction effect between mode and order finds only a significant effect of mode ($B = -0.626$, $SE = 0.138$, $z = -4.524$, $p < .001$, $OR = .535$).

Table 5.3 *France TUS-specific non-response/dropout according to the sequential nature*

	Paper-Web (N = 516)	Web-Paper (N = 529)	Total (N = 1,045)
No diaries	84 (16.3%)	95 (18.0%)	179 (17.1%)
Paper only	117 (22.7%)	94 (17.8%)	211 (20.2%)
Web only	37 (7.2%)	37 (7.0%)	74 (7.1%)
Both diaries	278 (53.9%)	303 (57.3%)	581 (56%)

To understand mode preferences, especially as they may relate to selection biases where response mode is driven by user choice, respondents to the French TUS were classified into two groups: paper-preference and web-preference. On the basis of their feedback to certain questions related to the difficulty, enjoyment, and time taken across both modes as part of the face-to-face visit conducted after completing both diaries. Classification to paper/digital preference was derived from a multiple correspondence analysis (MCA) and resulted in a division of 55% with web preference, 40% with paper preference¹⁸. Table 5.4 illustrates the differences in mode preference across various respondent characteristics. For example, under-35s represented a larger share of the web-mode preferers (33.7%) compared to paper (18.3%), which may pose difficulties when considering the overall lower response rate among young persons, in combination with the lower response rate for the digital version. Similarly, males made up a higher proportion of web-preferers (50.3% versus 42.4%), which may present difficulties when considering mode selection effects, as men were also underrepresented in overall response. A complete writeup of this analysis is available in the appendix.

Table 5.4 *Differences across respondent characteristics in user-reported mode preference, by relative percentage share.*

	Paper preference	Web preference
Under-35s	18.3%	33.7%
Males	42.2%	50.3%
Graduates with bacculaureate or higher	66.1%	73.7%
People in employment or studying	68.3%	78.2%
Household type "Couple with children"	43.5%	50.3%
Household monthly disposable income > €3000	52.2%	62.5%

5.4 Outcome differences

Although differences in selection between modes must be considered during data integration, this aspect of the process is unlikely to differ substantially with respect to Smart Survey modes. This is unlikely to be true for mode measurement differences (Bucher et al., 2023). Consequently, as

¹⁸ 5% were excluded from grouping due to high rates of non-response.

discussed in the M14 Deliverable (Bucher et al., 2024), selection of an appropriate data integration process should depend largely on the magnitude and nature of the outcome differences.

This section aims to clarify the existence of pure mode differences in measurement and establish the need for careful data integration. By analyzing data from various experiments, we demonstrate that, ignoring any potential differences in selection bias, the measured variables may be quite different, both with respect to measures of quantity as well as precision. Researchers must therefore make decisions on how to integrate their data carefully: simply combining data from different modes is likely to lead to inaccuracies.

The experiments performed in the French TUS, in which respondents completed both modes offer some clear insights into the patterns that may be expected with respect to differences in measurement of time use behaviors between modes. While there was insufficient data to compare aggregate differences to PAPI and app-based response within the Belgian TUS, the experiment provided the opportunity to investigate outcome differences between countries.

Deliverable M14 (Bucher et al., 2024) proposed investigation of outcome differences along two lines: mode differences with respect to quantity, and with respect to precision. Quantity differences reflect on potential missing data, based either on measured levels of missingness itself, or on variables that may be used as a proxy for establishing likely missing data (e.g., the length of the day recorded in TUS, the total number of entries over a period of time in HBS). Measures of precision are expected to reflect on the reduction of inaccuracies in the data that are recorded and may involve comparison of start and end times or task specificity.

Data Quantity

It is expected that the addition of certain smart features may improve both visible and invisible missing data. For example, an app or web program may be programmed to disallow leaving certain fields blank, such as the end time or activity performed while registering time use, or the name of a product category during registration of an expenditure. There may also be mode differences that lead to increased missingness with the introduction of a smart mode, for example, it may be easier to scan a PAPI diary to determine whether all entries have been completed.

The French TUS crossover experiment in which respondents completed both modes separated by a week presented an opportunity to look at these differences in a very concrete way. Due to the large amount of non-response following the first interview, in combination with the potential of endogenous selectivity biases, the simple analysis of the experiment with respect to the crossover design was judged to be insufficient to produce unbiased results of true measurement effects. For example, respondents with many different activities on the same day (and therefore many time slots to be recorded) may have decided out of weariness not to participate in the second survey, especially if the paper diaries were deemed time-consuming to record. For this reason, the results were also investigated following a matching procedure in which respondents were compared against similar respondents on the basis of the first completed diaries. Following this, it was possible to establish how sensitive each comparison was to potential differences in selection, in accordance with the method

proposed in Rosenbaum (2002). The maximum sensitivity at which the result would remain significantly different is denoted by Γ and noted in the table footnotes for measures for which this was calculated. A full description of the analyses and results is available in the appendix.

Table 5.5 *Mode differences in data quantity estimates from the French TUS experiment*

	Crossover Constant Effect (Paper - Web)	Matching Mean Effect (Paper - Web)
Total diary duration	-53 minutes [-73;-34]*	-67 minutes [-102;-32]*
Time period covered	-27 minutes [-44; -11]*	-44 minutes [-73; -14]*
Frequency of diary truncation ¹⁹	-	15.1% [14.9%; 15.3%]* ^a
Frequency < 2 sleeping periods ²⁰	-	14.1% [14.0%; 14.1%]* ^b
Frequency 0 leisure time declared	-	17.7% [17.5%; 17.8%]* ^c
Total travel time declared	-9 minutes [-3; -15]*	-15 minutes [-25; -5]*

Note: * $p < .05$; Sensitivity analysis results: ^a Measurement effect significant while $\Gamma \lesssim 1.9$, ^b Measurement effect significant while $\Gamma \lesssim 2$, ^c Measurement effect significant while $\Gamma \lesssim 2.8$

Table 5.5 presents the results of the mode measurement effects analysis on the basis of the French TUS experiment, supported by these two methods of comparative analysis. Contrary to earlier expectations, the differences across a variety of data quantity measures clearly demonstrate that there is a larger proportion of missing data as measured by the smart mode. In both the Crossover and Matching analysis, the total duration, as calculated by the difference between the latest period and the earliest period recorded, was approximately one hour less in the web version than in the paper version. The covered time period, calculated as the sum of all time periods logged on a given day, was also significantly lower in the smart survey, although the difference was less than that of the total diary duration, estimated at a difference of 27 minutes on the basis of the Crossover analysis, and 44 minutes on the basis of the Matching analysis.

The frequency of diary truncation can additionally be used to establish differential missing data patterns between the two modes. The French TUS made use of single-day diaries running between 4am and 4am the next day. A diary is truncated if it does not have activities that end at the 4am that completes the period. Table 5.5 shows that the web diary with smart features was more likely to have truncated ends than the PAPI version.

Taken in conjunction with the greater difference in total diary duration, this may signal that two different missing data processes are at work here. While web diaries are more likely to have missing

¹⁹ Difference of marginal proportions, McNemar's $\chi^2(1, N = 311) = 20.152, p < .001$.

²⁰ Difference of marginal proportions, McNemar's $\chi^2(1, N = 311) = 19.26, p < .001$.

data occurring at the beginning and end of the measurement period, when this truncation is excluded, web diaries are less likely to have periods of missing data occurring within the recorded duration.

Web respondents are also more likely to record only one sleeping period when two would be expected for most persons reporting a complete diary, as well as fewer than three meal periods. While fewer sleeping periods would be a likely consequence of a measurement difference due to the increased frequency of diary truncation, this is less demonstrably the case with the difference in reported leisure periods.

Data Precision

One of the expected benefits of the addition of smart features was to improve the accuracy of entered data. In a Time Use Study (TUS), this might include more precise start and stop times for activities, or a larger number of total recorded activities. For a Household Budget Survey (HBS), this might be improvements to the specificity of provided labels. The experiments were designed to assess these differences through the comparison between modes, such as with the crossover experiment in the French TUS, or comparisons between the other experimental arms in the Belgian TUS and French HBS field tests.

Figure 5.2 (left) shows the distribution of sleep duration times as recorded in matched web/paper pairs as part of the French TUS. In 32% of the pairs, the difference declared by the two respondents do not exceed one hour. However, the median difference is 20 minutes, with respondents in the web condition declaring less sleep than respondents in the paper condition. Despite a significant difference, there is considerable variance in the differences between the matched pairs. The lion's share of the effect is due to a higher proportion of (very) large negative differences, meaning that when people report very low sleep durations, they are more frequently web respondents. Over 25% of people sleeping less than five hours, as calculated within the web respondents, is unlikely. More likely is that the differences is indicative of a higher proportion of missing data occurring within the web response.

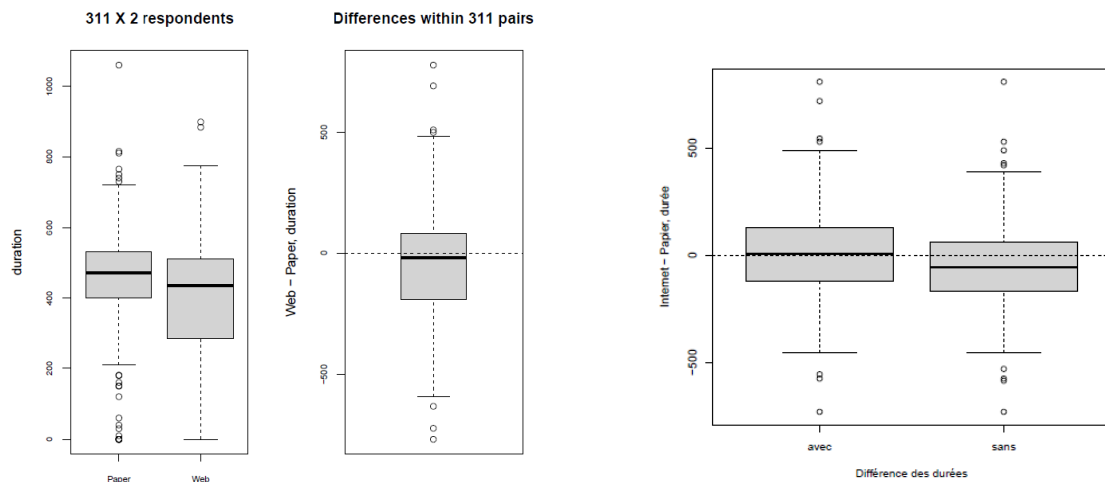


Figure 5.2 Differences in sleeping times (left) between Web and Paper within 311 matched pairs of Paper and Web respondents in the French TUS field test, and sleeping times (right) with or without correcting personal care codes to reflect leisure time.

Data Quantity measures can often be used to signal issues with Data Precision. For example, Figure 5.2 (right) demonstrates a duration difference between the smart and non-smart mode in recorded leisure time, with and without correction for a suspected issue, which is discussed in more detail in section [Specific Smart Features](#). 25% of app-based respondents declared no leisure time in their diary, compared with 7% of paper-based respondents, with a correspondingly large difference in average duration of leisure activities between the two modes. However, after accounting for the input difference between the two modes, the duration difference becomes non-significant.

Table 5.6 *Estimates of the measurement effect of internet data collection over different activity durations*

	Crossover-based constant effect Minutes [95% CI]	Matching-based mean effect Minutes [95% CI]
Personal and physiological time	-1 [-20; 18]	-5 [-36; 25]
Working and studying	-8 [-35; -5]*	-22 [-50; 0]
Housework and recreation	-6 [-16; 3]	6 [-11; 23]
Care for others	4 [0; 8]	4 [-4; 11]
Sociability	25 [17; 33]*	27 [14; 40]*
Leisure	-51 [-64; -38]*	-45 [-67; -24]*
Travelling	-20 [-27; -14]*	-32 [-43; -20]* ^a
Sleep duration	-56 [-72; -40]*	-53 [-79; -26]* ^b
Meal duration	-15 [-22; -9]*	-16 [-28; -4]* ^c

Note: * $p < .05$; Sensitivity analysis results: ^a Measurement effect significant while $\Gamma \lesssim 1.7$,

^b Measurement effect significant while $\Gamma \lesssim 1.4$, ^c Measurement effect significant while $\Gamma \lesssim 1.2$

Table 5.6 details the duration differences across various activity classifications within the Time Use Survey. As described in the section on Data Quality, the duration of multiple different activities were compared on the basis of two different analyses: the original crossover design, and after matching. Analysis of the crossover experiment produced additive and constant measurement effects of data collection via the web mode. This can be compared to the average measurement effect, in which the difference is obtained from matched pairs selected from respondents to the first collection period allocated either to the web mode or the paper mode.

Investigation of data precision presupposes that the mechanisms for differences in data *quantity* are well-understood. In the French TUS, which offered the most concrete method to assess these differences, the differences in missing data between the two modes were larger than expected, limiting the comparison of data precision.

The durations in Table 5.6 are uncorrected for the total difference in recorded time, which means that on average, there is less time to assign across activities in the smart mode than PAPI. It is therefore not surprising that on average across most categories, there is either no difference, as with the activities personal and physiological time, working and studying, housework and recreation, and care, or a negative difference, as with the categories leisure (around 50 fewer minutes), travelling (around 25 fewer minutes), sleep duration (around 55 fewer minutes), and meal duration (around 16 fewer minutes). However, one interesting finding from the French TUS experiment is that web respondents were more likely to report time spent socializing (around 25 minutes on average.) One potential explanation for this difference – differences in how people code their own response – is elaborated on in section Smart features.

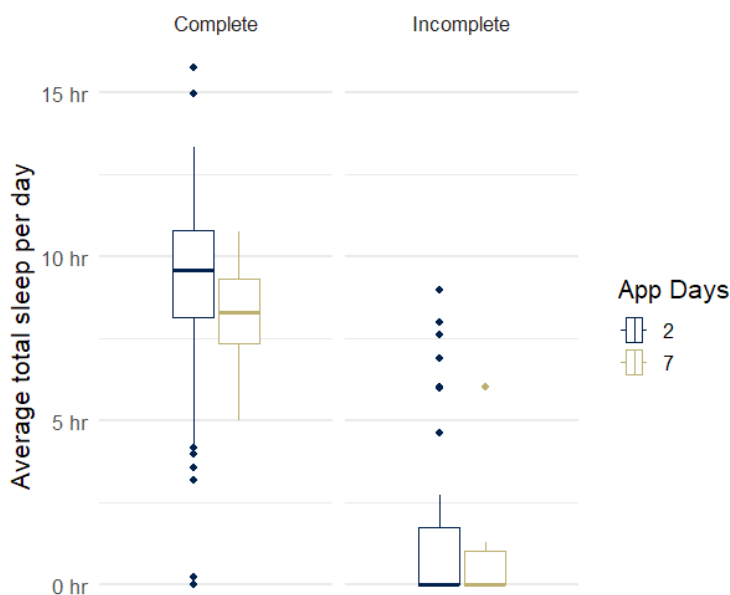


Figure 5.3 Average sum total sleep time per day from complete and incomplete respondents in the Belgian TUS field test. Later batches assigned 2-day diaries report slightly longer durations, and have a higher proportion of very-long and very-short durations.

We can compare the results for sleep duration and number of sleeping periods from the French TUS to the Belgian TUS. While the French TUS involved a one-day diary, the Belgian TUS asked respondents to report either on seven days (in the first two batches), or on two days (in the later three batches.) In total, 248 respondents to the Belgian TUS recorded at least one activity. Of these 248, 134 (54%) of the diaries are marked as complete. 184 respondents (74%), reported at least one sleep activity, including 130 complete diaries (97%) and 54 incomplete diaries (47%). Figure 5.3 shows the difference in average total sleep duration between the two-day and seven-day diaries, and across diary completion status. Respondents completing the diary in the app reported on average 9 hours of total sleep time per day when reporting on two days, and 8.3 hours per day when reporting on seven days. The seven-day average is close to the PAPI average from the French TUS. The Belgian study by incorporating more days is better able to distinguish between incomplete diaries, which in this case produces estimates more consistent with expectations.

Differences between app-measured and diary

To supplement the experiments that were planned but not executed within the Italian field test, data collected as part of the 2022 Travel App field test (Schouten et al., 2024). An experiment within this field test compared a PAPI mode with a high-smart survey conducted simultaneously in one subset of respondents. This subset oversampled respondents who were deemed more likely to experience a high burden as part of a self-completion travel diary, either because they had a high frequency of multi-modal trips during their prior participation in a panel study on mobility, or because they were employed in jobs that required frequent travel, such as street sweepers or driving instructors.

Participants were requested to install an app that would capture their geolocations on their own device, and to annotate the recorded data to provide information on transportation mode and travel reason. They were additionally asked to record their activities on a specific day using a self-completion web diary.

In total, only 65 of 124 respondents who provided at least some diary data across both modes recorded sufficient data for comparison. The largest factor was due to problems with missing data in the smart survey; on average the app missed recording 6.2 hours per day. Table 5.7 shows a comparison of various travel metrics derived from each source as well as the difference between the two, both when the traditional self-completion diary was restricted (SCD_r) and not restricted (SCD) to the jointly measured time period.

Table 5.7 Comparison of selected travel metrics recorded on the same day on two different modes.

	PAPI	App	SCD – App (Raw)	SCD _r – App (Raw)	SCD – App (%)	SCD _r – App (%)
Distance	70.9km	60.4km	10.5km	2.6km	7.1%	-.3%
Travel time	148 min	55 min	57 min	38 min	53.3%	45.4%
Trips	5.0	5.4	-.42	-.77	7.0%	-2.5%

Total duration	23.9 hr	17.7 hr	6.2 hr	.6 hrs	26.1%	.4%
Balanced Accuracy ²¹	-	-	-	-	58.0%	71.0%

Overall, app-reported diaries were shorter both in total duration as well as in travel duration. The large difference in total duration is due to missing data, much of which occurs during the night time hours as a planned OS functionality to restrict processing and data connections during these hours. This sort of missingness is generally ignorable, but because there may be other non-ignorable reasons for the missing data, distinguishing between ignorable and non-ignorable missing data is not possible. Additionally, many app-based diaries recorded only a few hours of data, missing entire round trips that were recorded in the diary. This accounts for the majority of the difference in distance estimated between the two modes, as restricting the diary to the app-measured period reduces the percentage difference to less than one percent. However, this accounts only for a portion of the difference in travel duration between the two modes. One likely reason for this difference is that algorithms for establishing movement events may arrive at different conclusions than a person reporting on his or her own behavior, although both may be correct. The section Geolocation Services describes this mode measurement effect in greater detail.

As noted in the section on the GeoService microservice, trip estimation algorithms are complex and may often be either under- or over-sensitive. The average number of trips measured in the app is larger than the number of trips from the diary, even before the time period restriction occurs. After restriction, the overestimation increases by a small amount. One reason for this is likely to be that most cases of trip overestimation occur within a small subset of users for whom positional errors measured by the device cause a number of erroneous short trips to be recorded.

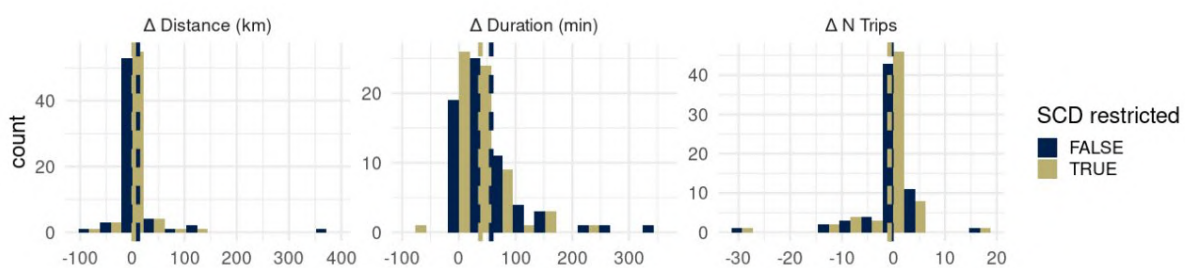


Figure 5.4 Graphical comparison of the distribution of differences across distance, duration, and number of trips between the smart survey mode and the self-completion mode (SCD). The SCD may be restricted to the time period covered in the app to estimate differences in the case of no missing data. Vertical lines show the average difference within each comparison.

²¹ Reflects the accuracy of the established stop/track event classification as the sum of the agreeing classification duration (both stop or both track) duration divided by the total elapsed duration.

Looking at the distribution of the differences in Figure 5.4 shows that for both distance and number of trips, most sets of diaries record quite small deviations, but measurement differences in duration are more systemic. In total, there are only four cases in which the app-measured duration is longer than the self-reported travel duration, and in only one of these does it occur in the unrestricted diary, in which case it is due to the same positional error issues that lead to the increase of number of trips.

The use case for geolocation data in a Time Use Survey is likely to be very different from that of a Travel Survey, as noted in Chapter 4. While a smart Travel Survey may be very interested in the total distance traveled or the segmentation of trips into trip legs to better describe behavior with respect to individual modes, the central features for a TUS would involve 1) the measurement of a central location during a stop, in order to prompt respondents' memory and 2) careful and accurate duration measurement to accurately identify times spent traveling and times spent at a location. While these data demonstrate that the smart features deployed in a travel app are already equipped to provide the first, the second is still quite far out. This was also noted in the section on Tentative geolocation data, in which respondents noted that they experienced problems with start and stop times due to the delayed response of the app to identify that movement had begun.

Although the sample size for this experiment is very small, directly comparing these travel statistics across the same persons and days demonstrates that the mode measurement effect can be quite large in high-smart studies and the statistics may not permit simple integration. In particular, we present this as an example of a study conducted across two modes where integrating the data in a single data set cannot be accomplished without direct consideration of how individual estimates may differ from each other due to the process that generated it.

5.5 Mode effects of specific smart features

Integrating Smart Surveys with Traditional surveys introduces smart features that can potentially enhance data collection and the user experience. They also have the potential to introduce measurement differences between modes. This section describes smart features of a low-smart application as well as a high-smart application. As a low-smart feature, we discuss potential mode measurement effects due to the input mechanisms and navigation features in a smart diary.

TUS Input mechanism selection and navigation

This section contains some highlights from analyses performed at Insee. The full document is included as an appendix. Some key findings are presented here and used to explore potential reasons for the differences in measurement between the modes. While tasks such as data entry may feel like they should be relatively uniform across modes, it's important to keep in mind that smart features such as search fields and buttons for navigation are introduced intentionally to change the user experience, and should therefore also be expected to introduce complexity.

Respondents to the TUS Diary had three options for entering their chosen activity: hierarchical menu navigation through visual buttons, entry into a search field which attempted to find matches for the entered terms in the list of available activities, and lastly, an open text entry field. 76% of all activities

were selected using categorical buttons, 19% of activities were selected on the basis of the search field, and 6% of activities were entered directly via the open field.

In 21.5% of cases where the open field was used, respondents had previously attempted to input text into the search field, providing an opportunity to determine which search terms lacked representation in the activity codes. Terms such as “prepare” or “meal” were frequently entered by respondents without generating appropriate suggestions, indicating a need to enrich the suggestion database. Although this subset of respondents eventually registered their activity by using the open field, it is possible that others who encountered similar difficulties did not do so, and instead left the activity blank. This may account for some portion of the missing meals recorded by web respondents.

19.9% of activity entries entered with an input mechanism other than the buttons had previously attempted to input their activity by navigating the hierarchical menu system. This provides a way to investigate where the button as an input mechanism may have led to measurement differences between the two modes. In 54% of cases where buttons were clicked, but the activity entered for the time slot was eventually selected via another method, respondents clicked the "Personal care" button but did not find their activity in the subsequent panels. This is likely indicative of other misclassifications with some respondents recording leisure activities under "Personal care." Although there was no significant difference found between web and PAPI modes within the category of personal and physiological care, this comparison does not take into account the higher proportion of missing data in the web mode.

The design of the hierarchical button panel is suspected to play a primary role in influencing respondents' decisions to click on specific buttons. For example, the button labeled “leisure” was positioned at the end of the eight categorical buttons displayed. This placement required respondents to scroll down to access it, which was less intuitive depending on the screen type and the respondents' computer skills. Consequently, many respondents used alternative methods to record leisure activities.

Finally, respondents faced challenges in locating specific personal care activities, such as "personal hygiene," “having a meal,” or “breakfast.” These are positioned in the second or third level of hierarchy under the personal care category, which may have caused them to be selected less frequently than otherwise.

Building on the insights from the input selection mechanism analysis, a model considering the ergonomic characteristics of each button and its relative frequency to its corresponding activity in the PAPI version. The hypothesis was that two primary factors influenced the selection of a given activity in the web diaries: the frequency of the activity in the general population, and the ease of access of the button itself.

In their analysis, Insee modeled the probability of an individual selecting an activity during their day as a function of these variables. Ergonomic characteristics of the selection buttons were depth of the button in the interface hierarchy (rank), and whether the button is at the bottom of the list in panels with more than six buttons. An additional covariate was included to examine whether the button label

contained the word “Other” to establish whether a catch-all category might influence selection. Lastly, in addition to other personal characteristics, a per-user variable was included to indicate the propensity towards button usage for activity selection (above or below the median propensity of the sample.)

Table 5.8 *Results from the analysis of button ergonomics in the relative frequency of activity selection in the French TUS web application.*

	Logistic model	Firth's method
Intercept	0.34	0.35
Log-odds weighted frequency of activity	0.79***	0.79***
Female	0.29***	0.29***
Age 35+	0.17***	0.17***
Single parent	-0.22***	-0.22***
Monthly income > €3000	0.09**	0.09**
House (vs apartment)	-0.08*	-0.08*
Button-user	-0.11***	-0.11***
Second level rank	-0.84***	-0.85***
Third level rank	-0.94***	-0.95***
Fourth level rank	-1.05***	-1.06***
“Other”	-0.26**	-0.26**
Low on list	-0.36***	-0.36***

*Note: * $p < .05$, ** $p < .01$, *** $p < .001$;*

The results of this model are presented in Table 5.8. As expected, the positioning variables for hierarchical depth (second, third, and fourth level rank), and depth on the page (low on list) are associated with being less frequently selected than would otherwise be expected based on the PAPI frequency. When the activity “other personal care activities” was included in the sample, the coefficient for buttons containing the word “other” was highly significant and positive. However, once observations corresponding to this activity were removed, the coefficient became significantly negative. This indicates that, with the exception of “other personal care”, web respondents are less likely than paper respondents to select buttons labeled other. Rather than being used as a catch all, web respondents tend to select a similar category, which may be imperfect.

One interesting finding is that respondents who made extensive use of the buttons in the web app reported, on average, fewer distinct activities compared to other respondents. Indeed, while the quick-access buttons in the web app were designed to facilitate data entry, they might create a sort of “cognitive shortcut”: users who heavily rely on these buttons tend to stick to a limited set of familiar activities, compared to those who more frequently use the search bar or free-text field – tools that lead to a greater diversity in activity selection, as users initiate the description in their own words. This suggests that the convenience of buttons might paradoxically reduce the diversity of reported activities.

These analyses led to the following set of recommendations:

- Prioritize the accessibility of frequently used buttons, particularly on smaller screens, or for users with limited technical proficiency.
- Personal care button clearly state that it does not include leisure activities, and leisure activities such as “watching tv” should be given greater visibility within the application's button interface to align more closely with respondent's daily activities.
- Respondent difficulty in accessing various deeper-ranked personal care items should be addressed by either modifying existing button labels or introducing additional, more specific buttons.
- Specific and descriptive button labels can improve clarify and reduce ambiguity in respondents’ choices.

Geolocation services

The 2022 Travel App field described in the Outcome Differences section also provides us with data we can use to investigate the role of a specific high-smart feature in inducing mode measurement effects.

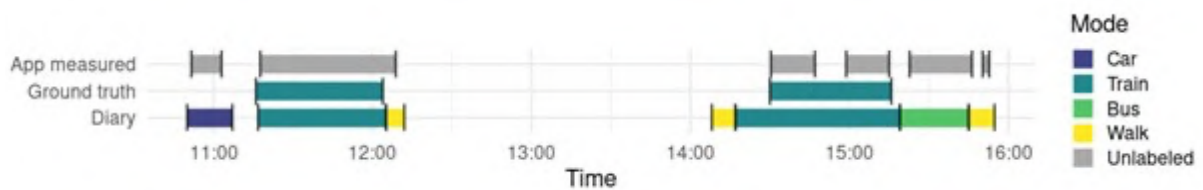


Figure 5.5 Comparison of a single user's travel behavior over one day as recorded in the app (App measured), and in the self-completion diary (Diary.) The actual historical departure times for the train portions of the journey are provided as ground truth. The app-measured trips better reflect the true begin and end time of the second train journey.

Figure 5.5 compares mobility event data for a single user across two modes. Discrepancies exist between the annotated traditional diary and the algorithmically-determined set of starts and stops based on locations measured within the app. While the algorithm identifies roughly the same number of trip legs as the diary (6 in the app, versus 7 from the diary), they do not align perfectly. Both the diary and app record an initial trip beginning around 10:50, labeled as a car trip in the diary. However, the app-measured trip is shorter, starting later and ending sooner.



Figure 5.6 Timestamped geolocations for a user overlaid on a map. Under the assumption that the location was accurately recorded, this demonstrates that the user was already in the train station at 11:04, while their self reported diary entry at this moment reported that the user was still in the car segment of their journey.

Comparing the app's timestamped locations, the person was recorded inside the train station at 11:04:45. Although it might seem that this definitively establishes the end time of the car trip as closer to the app-reported 11:04 rather than the diary-reported 11:07, the recorded locations can be incorrect. Wi-Fi-based triangulation in places with strong, dispersed signals, such as a train station, tends to place people indoors incorrectly. In this case, the app likely estimated the end of the car trip correctly, whereas the diary entry overestimated it. This leads to diary estimates of total trip length tending to overestimate the time relative to the ground truth.

The diary next records a train leg followed by a walking leg, but the algorithm merges these two because the walking distance was short, around 700 meters. Stop detection algorithms often assimilate locations at the end of a trip into the following stop, leading to an underestimation of distance traveled by a few hundred meters relative to the ground truth.

On the return trip, the diary records a walked leg followed by a train leg. The app fails to measure a new track for the walking portion and only begins measuring the train journey about 10 minutes after the respondent indicated starting the train leg. Although the algorithm did not label the walking portion as a track, the app-measured locations confirm it. However, there is a significant discrepancy between the starting times of the train journey.

Generally, it is challenging to establish which source of truth is more accurate. However, the national train system in the Netherlands provides actual arrival and departure times accurate to the minute, allowing for the establishment of a ground truth for train journeys. In this case, the train journey began

at 14:30, aligning with the start time established by the location-based stop detection algorithm. A similar instance occurs in the Train -> Bus leg switch, where the bus ride begins immediately after the train ride. Although no third source of historical data exists here, it is likely that most people wait for the bus to depart, as measured in the app.

The consequence is that the total time spent traveling in both the train and the bus may be overestimated relative to the actual time in transit. Waiting times for public transportation are often agglomerated into the following trip leg. For some purposes, this may be desirable, as it provides a more comprehensive estimate of total commute time. However, for other purposes, such as estimating waiting times, this lack of precision could be a significant drawback.

5.6 Conclusions

- **Mode Effects Impact Data Quality and Comparability:**
 - Mode effects are due to differences between smart and non smart surveys, and this will impact data quality and comparability over time. These effects arise from both representational differences (e.g., respondent characteristics) and measurement differences (e.g., differences between the two methods of data collection).
 - The benefits from smart surveys are often still under development, and it may take some time before smart features resolve into improved measurement. In the meantime, there are risks for losing older participants, as well as a small reduction in response rates.
- **Response Rate and Dropout Patterns:**
 - Response rates and dropout patterns vary across modes and demographic characteristics, but there is limited evidence that there are interactions between the two. For instance, PAPI modes generally had higher response rates than web modes in the French TUS experiment.
 - Younger participants, males, and those with higher education levels showed a preference for web-based surveys, highlighting the potential for selection biases when mode choice is offered to users in the presence of high levels of mode measurement differences.
- **Data Quantity and Precision:**
 - Smart surveys, despite expectations, showed a higher proportion of missing data compared to traditional PAPI modes. For example, web diaries in the French TUS had more truncated ends and fewer recorded meal periods.
 - While smart features like geolocation can improve data precision, they also introduce challenges such as underestimation of travel distance and travel times, usually due to an increase in missing data. On the other hand, geolocation accurately measures multi-modal trips and repeated start-stop intervals, demonstrating the potential for high-smart features to reduce the burden significantly for certain segments of the population.
- **Specific Smart Features and Their Impacts:**

- Input mechanisms in smart surveys, such as hierarchical menus and search fields, influence how respondents report activities. For instance, the placement and labeling of buttons affected the reporting of leisure and personal care activities.
- Geolocation services, while providing more accurate travel data, can lead to discrepancies in trip leg identification and waiting time estimation.
- When comparing between modes, it is not always clear which should be taken as the ground truth. To assume that the paper version is always correct would counter one of the primary reasons for applying smart surveys to a problem.
- **Recommendations for Improving Smart Surveys:**
 - Prioritize accessibility and clarity in the design of input mechanisms to reduce ambiguity and improve data accuracy.
 - Enhance the visibility of frequently used buttons and ensure descriptive labeling to align with respondents' daily activities.
 - Address respondent difficulties in accessing deeper-ranked activities by modifying button labels or introducing more specific buttons.

References

- Bucher, H., Keusch, F., de Vitiis, C., de Fausti, F., Inglese, F., van Tienoven, T. P., Mccool, D., Lugtig, P., & Struminskaya, B. (2023). *Smart Survey Implementation: Workpackage 2: Research Methodology Deliverable M6: Review stage*. Grant Agreement Number: 101119594. European Statistical System.
- Bucher, H., Keusch, F., de Vitiis, C., de Fausti, F., Inglese, F., van Tienoven, T. P., Mccool, D., Lugtig, P., & Struminskaya, B. (2024). *Smart Survey Implementation: Workpackage 2: Research Methodology Deliverable M14: Smart baseline stage*. Grant Agreement Number: 101119594. European Statistical System.
- Rosenbaum, P. R. (2002). Attributing effects to treatment in matched observational studies. *Journal of the American statistical Association*, 97(457), 183-192.
- Schouten, B., Remmerswaal, D., Elevelt, A., de Groot, J., Klingwort, J., Schijvenaars, T., Schulte, M., & Vollebregt, M. (2024). *A smart Travel Survey: Results of a push-to-smart field experiment in the Netherlands*. Centraal Bureau voor de Statistiek. <https://www.cbs.nl/en-gb/background/2024/37/a-smart-travel-survey-results-of-a-push-to-smart-field-experiment-in-the-netherlands>

6 Conclusion: How does a smart survey methodology look like?

In recent years smart surveys have emerged as a promising data collection method, bridging the gap between traditional survey techniques and modern technological advancements. The key characteristic of smart surveys is that they intelligently combine the use of asking questions (i.e., surveys through self-report) with smart features collected via sensors on smartphones, wearables, and other devices. The goals of smart surveys are to improve data quality, reduce participant burden, as well as provide more timely and more granular data. In this chapter, we describe the solutions to the methodological and technical challenges that arise when implementing smart surveys in the field in the context of European Official Statistics data collection, based on findings from the Smart Survey Implementation (SSI) project that were presented in chapters 1-5 of this deliverable.

The SSI project focused on developing an end-to-end methodology for two use cases: household spending and time use. A small feasibility test was conducted around energy use (see Appendix E). National Statistical Institutes (NSIs) across Europe periodically produce statistics on these topics, and have traditionally relied on diary studies, where respondents keep track of all their spending over a specific period (often a week), or keep account of how they spend their time split into 10-minute time slots (often for a week- and weekend-day).

This workpackage 2 focused on how to develop a methodology for doing Smart Surveys. In the context of household spending, respondents were invited to use an app on a mobile device to take pictures of a shopping receipt, from which products were extracted. In the context of time use, geo-tracking was used to pre-populate a time diary in a smartphone app with episodes of travel and non-travel to facilitate respondents in delineating the start and end times of time activities.

Other workpackages within the larger project focus on how to build the technical infrastructure to collect smart survey data (WP3), how to organise the business process (WP4), and how to ensure that smart surveys are conducted compliant to legal and ethical standards (WP5). This chapter is focused on the methodology of doing a smart survey; on *how* to do a smart survey. The earlier chapters in this deliverable document in detail the results of small and large tests carried out, and the consequences that using smart surveys has for the output statistics. In the remainder of this concluding chapter we want to zoom out and discuss several overarching methodological issues that have been the focus of our work and discussions throughout the Smart Survey Implementation project. There is not one way to do a smart survey; we will argue that in order to conduct a smart survey successfully, one needs to take a ‘tailored design’ perspective (Dillman, 2014) in all stages of developing and carrying

out a smart survey and develop a smart survey that is both topic-, situation- and (country)-specific.

6.1 How should you set up a smart survey?

In contrast to traditional surveys, smart surveys have a more complex data collection and storage processes that are related to the technical components of collecting data using sensors, apps, and wearables. There are several steps for implementing the app and sensor-based data collection.

First, one needs to design a technical infrastructure that will enable data collection from participants. In the foreground is programming of the app and functions that it performs (which we call here microservices). Ideally, the app needs to be integrated into the NSI's systems and daily operations. Of equal importance is considering legal issues associated with the special data collection since, for example, there are higher risks associated with collecting sensor data and in this case, a Data Privacy Impact Assessment (DPIA) needs to be set up for each smart survey that researchers plan to use. An example of potential privacy threats is other apps piggybacking on the geolocation collected by the app-based survey, which can be identified and avoided during the implementation.

Second, the technical infrastructure should be populated by participants who are allowing passive data collection such as geolocation tracking or active data collection such as is the case of a respondent scanning receipts and annotating them. The two crucial questions are how to recruit and retain participants, that is, what factors motivate participation and how can continuous participation be ensured?

Chapter 2 of our study shows that recruitment into smart surveys is quite challenging across countries. In Belgium recruitment was tested for an app-based time use survey and Germany and Norway tested recruitment designs with the household budget survey. Completion rates are comparatively low ranging from about 2 percent in Germany and Belgium (defined as all activities on all days, i.e. either 7 or 2 days for Belgium and recording the expenditures on 14 days for Germany) to around 23 percent in Norway (recording expenditures on at least one day). The app activation rates which indicate successful download of the app and start of the data collection were around 6 percent in Germany, 10 percent in Belgium and 25 percent in Norway. The challenge in smart surveys is thus both getting people to download an app and start with the study but also to continuously provide data until the end of the field period. Participant characteristics of being male, having a nationality other than that of the country,

and having a lower education level are associated with lower participation probabilities. These characteristics are associated with nonparticipation biases as well, which are further aggravated by the biases in age with older people less likely to participate. The design choices for recruitment strategies such as using interviewers (CATI in Norway) increases the success of the recruitment compared to mailed invitations (in Germany and Belgium).

Third, data capture and processing needs to be set up. For smart surveys, machine learning (ML) algorithms play an important part. Using machine learning algorithms involves building the automation part of a smart survey, designing the mechanism when the machine (app) alerts humans (participants or researchers) when it needs input, and designing an efficient user interface (UI) to facilitate human-machine interaction (Benedikt et al. 2000). Key questions related to machine learning in smart surveys are about the conditions under which results from ML can be directly used as outcomes or when they need to be fed back to the participant, what should be done if the quality of ML outcomes is too low, when should participants be asked to provide new input because no meaningful information can be extracted (e.g., from a scanned receipt in an app-based household budget survey), and how to create and update training datasets for ML algorithms.

Chapter 3 discusses the use of supervised and unsupervised algorithms focused on two contexts: processing photos of receipts and geolocation data. The Receipt Scanning Microservice (RSM) developed in this project (see WP3) involved Optical Character Recognition (OCR) and COICOP (Classification of Individual Consumption by Purpose) classification. The GeoService Microservice focused on stop and track recognition using spatial and temporal parameters as well as transport mode prediction in the context of an app-based travel survey and activity prediction for a time use survey. The two components of the RSM, OCR and COICOP classification, are based on the complex pipeline to extract information from the receipts such as store names, purchased items and their prices, and classify products into the corresponding COICOP category. The latter is based on a mix of machine learning and string matching (for details, see chapter 3). There are different choices available to the developers on which models to use, and the choice depends, for example, on the trade-off between accuracy and speed. One main finding of the workpackage is that NSIs need to invest in training data to know whether the algorithm works in general and whether it works under different circumstances (e.g., scanning receipts in different countries and from different stores). Significant resources are required to produce training data such as collecting various receipts and labeling them (manual annotation), but the accuracy depends on the factors such as the size of the training dataset. One can also decide to make a choice in terms of extracted information: the minimum is to identify products and prices while the

maximum is to identify every type of text on the receipt, including date of purchase, store name, and the like. The latter strategy will require more input from participants and the choice between these strategies will have implications for the UI/UX of the app. Another choice is to reserve the smartness of the app (i.e., automatic recognition) only for the main stores (e.g., supermarkets) since the amount of the required training data is dependent on the variability of formats. For stores that are not part of a chain, respondents have to resort to manual entry.

The GeoService microservice was developed to process sensor data in order to provide input for the timeline of the time use diary. First, the segmentation into stops and tracks needs to be done. Second, third-party information is added, such as map services (Google Places and OpenStreetMaps) to give context to a visited place (e.g. is it a shop, office building or home). The microservice does three things: (1) defines the stop-track clusters, (2) predicts the travel mode, and (3) predicts time use activities at stops according to the Harmonized European Time Use Survey (HETUS). Detection of tracks can be prone to error due to e.g., missing data. There is an issue of heterogeneity in geolocation sensor quality among different types of smartphones which in smart surveys becomes crucial since smartphone brands vary across countries. Moreover, the quality of the map services can also vary among countries. Small tests in Belgium, Italy, and Germany show that the GeoService reduces burden when respondents enter their activities into the Time Use Diary at the end of the day, but if the activity changes within the same place, respondents still need to enter it manually, and this can cause errors. Input from respondents can be required to improve data quality, for example, transport modes with similar speed can get misclassified such as walking and biking; bus and tram can also be often misclassified, and difficulties arise when being near water (e.g., traveling by ferry). Furthermore, good quality auxiliary information, i.e., the map service, is needed for the algorithm to predict activities well, as our small test showed.

Fourth, smart surveys might require specific involvement from participants. For example, participants should annotate the data they provide, check the output of the machine learning models, or commit the information provided for a particular time period such as specify whether the information is complete, or if there is no information to be provided (e.g., no purchases made on a given day to be recorded in the app-based budget diary) to confirm that this is the case. These tasks can place an additional burden on the participants especially if the usability of the app is not high. In this project, usability testing of smart surveys was performed in several countries with different apps.

Chapter 4 explains that rather than smartness of the application, which focuses on sensor data and how these should be processed by trained algorithms, for usability testing the concept of cleverness was central. Cleverness is defined as the ability of the app to work in

novel situations, that are not encountered (yet) in training data. Clever applications thus allow participants with lower digital skills and experience to complete the tasks set out in the smart surveys. In small-scale tests, the ease of use and the quality of the user's experience with the application was tested. Three aspects central to usability testing were 1) engagement of the participants, 2) ability of the application to perform complex tasks, as well as 3) trustworthiness. Usability testing was done for two microservices: the Receipt Scanning Microservice (Household Budget Survey) and the Geolocation Microservice (Mobility and Time Use domains).

The key finding from chapter 4 is that smart surveys should aim to be clever rather than smart can further be specified into a set of HCI 'requirements': a smart survey should be engaging, easily accessible, come with clear instructions, not take much time (be time efficient), able to handle errors (e.g., when a picture is uploaded that is not a receipt), have an intuitive user interface, task flow and guidance, error prevention methods, and provide feedback and in-app support. Other usability aspects not directly related to the app, but more to the methodology are trust and credibility, security and privacy, transparent communication, data collection efficiency, and user control of data/information.

In practice, one also has to take into account that respondents may not use the app as is intended by researchers. We find that some respondents do not fully read the invitation letter, and that some respondents - especially older ones - have problems using a QR-code as a means to locate the smart survey app that needs to be downloaded. The process of downloading the app itself seemed to work relatively well across applications, but respondents do not always read/watch the onboarding screens that explain how the app works. Explaining what the app does and how it works is something that has to be explained within the app, but also - perhaps at a higher level - in invitation materials, to ensure that respondents know why they need the app, and what they need to do within the app. There are lots of practical UI/UX best-practices that help respondents get through the process of installing and starting the app, such as allowing respondents to check their password when logging in via an 'eye' symbol next to the password that should be included in app design.

Trust and credibility of the smart survey was non-problematic among test persons. The fact that the survey is conducted by the NSI or a university automatically communicates trust, and it is perhaps for this reason that respondents do not commonly read the privacy policies of the app and data collections.

Another finding related to UI/UX design has to do with the steps that the respondent needs to go through after app installation. Respondents with low digital skills not only had trouble installing the app, but also in figuring out what was expected of them once the app was installed on their phone. Support and feedback needs to be available within the app, but in some cases also perhaps via an interviewer that helps the respondent navigate the app and explain what is needed. Information on within-app navigation can perhaps be provided incrementally; so that respondents only receive an explanation on how a feature of the app works (e.g., taking pictures, annotating a receipt), when this becomes relevant to the respondent, with the option of leaving these explanations 'on' or 'off'.

Fifth, data from smart surveys need to be integrated into the data streams of official statistics. Apart from technical requirements for the back-office of NSIs (see WP4), there is a need to combine outputs from smart surveys with old, traditional diary studies. One reason for this is that for both HBS and TUS long time series on output statistics are produced to inform policy-makers and researchers.

Chapter 5 finds that measurement differences exist between smart- and non-smart surveys. These stem partly from differences in the amount of missing data that we see between the modes, but also due to differences in how people select answers in the app (e.g., using semi-automated decision trees on time use vs. free-text entry). For the French time use app, we see that respondents in the smart app spend less time on sleep, leisure, meal and travel time, but more time on social activities, compared to the traditional time use diary. The underreports in sleep and leisure time can be at least partly explained by differences in missing data rates, but for other activities it is likely that respondents simply report different time use due to the use of a smart app vs. traditional diary itself.

One way to bridge these changes in time series when moving from a traditional diary to a smart survey is to calibrate survey estimates. However, if a smart survey is part of a broader mixed-mode design, in which a smart survey is offered to some respondents and a traditional survey is offered to other respondents, there is also a need to bridge gaps in the time series within the same wave of data collection. The M6 deliverable from this workpackage provides guidance on how data can be integrated in such situations.

6.2 An End-To-End methodology for a smart survey

The tests conducted in this deliverable have shown some successes and some difficulties in doing smart surveys in practice. What can we take away from these findings? What would a methodology for smart surveys look like? Where should someone start when trying to move from a traditional survey to a smart survey?

The smart survey app makes a logical starting point of a smart survey methodology. In theory an app can be seen as a blank canvas on which any kind of study can be developed. In practice, some of the decisions of the type of software used for programming the app, and the integration with existing IT systems within NSIs are confined due to technical choices in IT architecture (see WP 3). An app can, for example, be set up as a digital-device only ‘native’ app for Android and iOS operating systems (for tablets or smartphones), or a progressive web-app, which relies on an internet browser. Similarly, a choice can be made for a study-specific app (as is the case with the CBS HBS-app), or for a platform that can be used across multiple studies (such as MOTUS).

Within the SSI project, a choice was made to focus on using a native smartphone app and develop joint ‘microservices’; these are IT components that perform a specific task and can be integrated in different types of app platforms. The microservices focus on processing smart data: they allow sensor data to be processed outside of the app used on the phone, and processed data to be fed back to the app. In this project, several microservices were developed to process geolocation data into episodes of travel and non-travel, to recognize the text and relevant lines on a shopping receipt, and to link product codes of purchases to COICOP product categories. WP3 documents how these microservices work.

Although the microservices central to this project allow data to be sent from a respondent’s phone and back, it is not always necessary to feed data back to respondents. In some cases, feeding data back to respondents is helpful so that they can further annotate or enrich the data, but in other cases it is not. For example, in a smart time use survey feeding geo-data back to respondents in the form of travel and non-travel episodes is needed because the geo-data alone do not give us a sufficient amount of information about how people spend their time. To illustrate, knowing that a respondent was in the same location between 17:00 on a particular day and 08:00 of the next day, does not tell us exactly what kinds of activities the respondent did when, and for how long. The respondent is still needed to perform a crucial task of entering activities, and the smart data are only useful as a reminder to the respondent of the time when he or she came home (or another location), and left home (or another location) the next day.

On the contrary, for shopping receipts, the role of the respondents is to check and commit the processed data. The respondent can help to correct mistakes made in the OCR, such as correcting product names, or check and correct the prices of products. Here, the respondent is the human-in-the-loop. Although in the time use study, respondents can technically also serve as a human-in-the-loop to check the exact geolocations (and times), in practice respondents cannot be expected to correct or improve geolocation data. The entire rationale for collecting geolocation data is to help respondents; we assume that the geolocation data are much better at identifying periods of travel and non-travel than a respondent would be able to from memory alone.

The usability tests that we conducted in this project confirmed that respondents are looking for the smart data to simplify their response task. Some respondents do not feel like carefully checking the text data extracted from receipts, let alone helping in linking product names to COICOP categories. Respondents often assume that because we use a smart element, there is no need for them to check the data carefully. If respondents are not keen on giving feedback, asking them for feedback risks making the app being perceived as too complicated. There is a fine balance between involving and engaging the respondent to give context to the data that is essential (in time use), and involving the respondent too much in terms of the user experience of the app.

This also means that for some steps in the processing of smart data, a choice can be made to deliberately not involve the respondent and do further processing of data within the NSI after data collection. For example, when products on shopping receipts need to be matched to one of the COICOP categories, there are perhaps advantages to doing this step in the back-office of the NSI. There is then more time to, for example, match product lines to other data sources, such as scanner data from supermarkets, or to tailor the processing and matching of shopping receipts to local circumstances. One reason why such tailoring of algorithms may be necessary during or after data collection, is that shops may change their product descriptions shortly before or during fieldwork, that new shops may open for which no training data is available yet, or that new products enter the market that have not yet been assigned an appropriate COICOP code.

Apart from usability considerations for the respondents, the availability of appropriate training data is essential for ensuring that a smart survey works. Ideally, smart data are processed without any errors so that respondents are not needed at all to correct or improve data. This requires a data processing pipeline that minimizes such errors. For example, in the process of using OCR to extract product lines from shopping receipts, at least in theory it

should be possible to read the product descriptions without many errors. Only when receipts are missing, or the quality of the pictures is too low, should a respondent be asked to enter products manually. While training the OCR this is largely a one-off task that requires a diverse set of shopping receipts to be trained on, other tasks require more specific and continuous training data. Before OCR can be used on product lines, one earlier step is to identify the relevant lines on a shopping receipt. Lines that include, for example, the shop's address, or card payment details are not relevant and should be skipped. Learning how receipts are structured so that this processing task is accurately done requires a relatively large training set that needs to be constantly updated due to the addition of new shops, or restructuring of shopping receipts. Instead of using one training set, the training needs to be country-specific too, as even some larger shops that are active in Europe (e.g. Lidl, Aldi) may use different receipt lay-outs across countries. Not having a good and up-to-date training set risks too many errors being fed back to respondents for correction. In absence of good training sets, a smart survey should generally rely less on smart elements, and more on the respondent. For example, a smart HBS survey can explicitly be limited to the scanning of shopping receipts from supermarket chains only. This limits the number of store chains that need to be included in the training set of shopping receipts, and will probably result in better quality of the processed data extracted from the receipts. Respondents will potentially not have to do many corrections on the processed data, making the app more usable. The large disadvantage of restricting the HBS to shopping receipts of supermarkets only is that this needs to be communicated to respondents; in such a situation respondents need to be asked to scan some receipts, but to enter other receipts manually. Communicating what specific receipts from shops can and should be scanned is potentially complicated. For respondents who do a lot of shopping at shops not listed, the response task of having to enter a lot of receipts manually is more difficult, and at least some of the advantages of smart surveys of lowering respondent burden are lost. One other solution to this problem is to do more processing of receipts after data collection, and not involve the respondent in the task of checking and committing data at all.

The main characteristic of the end-to-end methodology is that within a smart survey, the design of the app, processing of smart data, and the respondent task are all linked to each other. Decisions taken on how to process and feed back smart data to the respondent will affect the usability of the app, which in turn may impact whether respondents are willing and able to do the task. In thinking about a methodology for a smart survey it is essential to design and think about all the steps in the data collection together. These should not only work within the business process model of the NSI (see WP4), but also work for the respondent. How to design a methodology for a smart survey depends on several contextual factors, like legal and ethical considerations (WP5), capacities at the NSI (WP4), and the technical

architecture of the app (WP3). The methodology of a smart survey does not follow one particular recipe: the specific design depends largely on the specific outcome statistic that we want to produce, and the role of the respondent. In the next section, we provide some guidance on how in the design phase, one can think about the methodology.

6.3 A continuum of smart survey methodologies

One of the core outcomes of this workpackage on methodology is that there is not one recipe for conducting a smart survey. There are many decisions to be taken on how to design a smart survey, from recruitment methods to the role of algorithms and the respondent as a human-in-the-loop. Choices in the design of all these aspects lead to an almost infinite number of possible ways in how a smart survey can be designed. In this section we argue that a lot of these combinations do not make intuitive sense, and that a successful methodology for a smart survey starts with an informed decision on how smart a survey should be. There is an underlying continuum going from ‘only a little smart’ to ‘very smart’, and a choice on that continuum will help to guide decisions on the specific design.



Figure 6.1: A smart survey design continuum

We can use the studies conducted on the Household Budget Survey in this project as a good example of design choice options to illustrate various design positions on the smart survey design continuum.

A version of the HBS that would be only a little smart could be a version where we ask respondents to use an app and take pictures of their shopping receipts only. Respondents are

not asked to complete many survey questions, nor to annotate the receipts, or correct and commit data. The respondent will just be asked for the very simple task of taking pictures. Both the evaluation of the quality of the picture, the extraction of relevant product lines, the OCR of the text on the receipt, and matching the products of COICOP categories happens during post-processing of receipts at the NSI. The respondent does not act as a human-in-the-loop and the response task is relatively simple. The advantage of this approach is that a relatively simple app would be needed and that respondents only need instructions on how to take pictures. Big disadvantage of this approach is that it is essential that the pictures taken are of high quality, and that there are few possibilities to intervene in case pictures are not of sufficient quality. On top of this, significant resources are required to post-process receipts after data collection, and there is little room for enriching the receipt data with contextual questions.

On the other extreme of the spectrum we would find a very smart survey. In this situation, post-processing at the NSI is kept at a minimum, and the respondent has an active role in producing smart statistics. Checking the quality of the picture, product extraction, OCR and COICOP category matching is all done on the digital device itself, or via a microservice. This approach has the big advantage that the respondent can be used to improve the quality of the data at every processing step, and that it is possible to get richly contextualised receipt data. Big disadvantages are that large resources are needed in training the algorithms needed to process the data and these all have to be done *before* data collection, that the app needs to be quite sophisticated to allow the respondent to interact with the smart data, and that respondents need to be able to understand their task, be willing and able to do what is expected of them.

There are also intermediate solutions: for example, extending a 'little-smart' app to be a bit smarter could use the smart elements to evaluate whether the quality of a picture of a receipt is 'good enough', and then prompt the respondent to retake the picture (with some instructions such as 'retake in a better lit space or at a time with more light') until the quality is evaluated to be good enough. Starting from the 'very smart' design option at the other extreme, a choice can be made to make the app a bit less smart by taking out the phase of matching a product to a COICOP category. In both intermediate solutions, the design choice will impact the need for training data, the design and usability of the app, and ultimately also affect the outcome statistics.

In this project, we tested several levels of smartness. For example, the app that was fielded in Norway was designed as 'little smart' and did not process the content of the receipts on

the phone or via a microservice, but rather moved this to a postprocessing phase together with matching to COICOP categories. In Germany, the app was somewhat smarter: COICOP matching was also not done after data collection at the NSI, but respondents were asked to check and commit the total price of their receipt. These differences in design impact how the app looks like, how instructions to respondents should be designed, and even how the app technically works: the choice not to use the respondent as human-in-the-loop in Norway meant that the app could be designed as a progressive web app.

The continuum as we presented here is illustrated here for the Household Budget Study, but would also apply to other use cases for smart surveys, such as the Time Use Survey or household energy consumption. The specific choices made are topic-specific, and ultimately depend a lot on the specific outcome statistic of interest. Even within the use cases of the Household Budget Study or Time Use Survey, there is a choice to be made at the start of the study on how many receipts would be needed (overall and per respondent), how detailed COICOP classifications would need to be (at what digit level), or at what time intervals one would want to record time use. And apart from the output statistics, there are dimensions of costs, and practical issues (e.g., WP4: do we have all the building blocks ready to do a smart survey?) that will determine where on the smart continuum an NSI will settle in designing the smart survey. What we argue is that it is important to have this discussion early on, and use the smart design continuum as a tool for guiding later decisions in developing an NSI- and study-specific smart survey methodology.

6.4 A research agenda

Although the project Smart Survey Implementation did implement smart surveys in several variations across countries, there are still a lot of things we do not know. We tested several designs in both large and small tests, within different app platforms, and in different countries. The tests focused on specific problems such as how to recruit respondents successfully, how to process smart data with machine learning models, what kind of measurement differences we find between smart- and non-smart surveys, and how to design the Human-Computer Interaction such that the app is usable for respondents. These tests have given us a lot of concrete and usable findings that help us understand how parts of a smart methodology should look like. In our view, the research agenda for the next few years is how to combine different design elements; where to place smart surveys on the smart survey continuum, and how to make an informed decision on this. Below, we outline some issues that in our view deserve more attention in the next years:

Issue 1: How smart should your smart survey be?

- How do we train our algorithms for different output statistics? E.g. In this project, we focused on getting output statistics on household spending using COICOP level-5 statistics. What is the quality of output statistics for a smart survey at this level, and how does quality improve, if we decide to compute output statistics at a lower level, such as COICOP 4?
- How do we balance the respondent as human-in-the-loop with post processing in-house at the NSI across different stages of a smart survey? Experiments are needed here to compare the quality of both approaches, for example, in the phase of matching product lines to COICOP categories.
- Do we limit 'smart' to a particular set of use cases only? For example, do we ask respondents to only scan receipts of particular kinds of shops for which our training models have shown that quality is acceptable (e.g., only supermarket receipts), and ask them to enter other shopping manually? Or do we aim for a generic service that would allow scanning of receipts of all shops?

Issue 2: How to recruit?

- How can we integrate interviewers in the invitation protocol? Our tests revealed that not involving interviewers at all makes smart surveys difficult in terms of response and continuing participation. But how to use them exactly is an issue that needs further research. Should interviewers follow-up on nonrespondents mainly? Should they (also) be used to improve data quality, by, for example, explaining the app to respondents, or checking for missing data in the final stage of data collection? Within SSI, we have not been able to properly test what is the best way for interviewers to help respondents, due to scheduling issues with the large tests in Italy and the Netherlands.
- Recruitment requires a tailored design. A tailored design implies, for example, that messaging in the invitation letters, but also in the app is topic-specific. When we invite people for a smart survey it needs to be obvious why the study is conducted via an app and why sensor data are needed. For example, in the HBS respondents appear to understand why taking pictures of receipts is natural. For geolocations, this may also be true when a travel survey is conducted, but it is not necessarily true for the time use survey. Explaining why geodata is collected appears complicated to the participants, and better communication is needed. This is especially true if input from participants is needed for correction of the time use activities suggested on the basis

of the geolocation when the geolocation service has issues. Better communication does not only pertain to invitation materials, but also on how the app is designed. Some respondents will not read instruction letters in detail or look at onboarding screens within the app. The whole process of installing the app, registering data, and interaction with smart data needs to be 'logical' and easy. This is easier said than done, however; our findings from chapter 4 on Human-Computer Interaction provide a clear research agenda for specific issues to address.

- For some subgroups of respondents, smart surveys are something 'obvious', for others, it is 'too difficult'. We need to bridge this gap if we are to implement smart surveys within Official Statistics. Although the perception survey of workpackage 1 showed that there are large country differences in how prepared respondents are for smart surveys, we find that nonresponse biases are remarkably consistent across countries, with older age groups, migrants, and lower educated being particularly difficult groups to reach via smart surveys. In our fieldwork design, we need to better cater towards difficult-to-reach groups. Our experiences from Norway showed that CATI follow-ups by interviewers helped to reduce age biases, and this may be one way forward.
- There appears to be a need for testing variations of mixed-mode settings, where smart surveys are offered within a mix of survey modes. This mix would primarily consist of self-administered survey modes (app, web and perhaps paper-and-pencil), with some role of the interviewer (see earlier points). How exactly such a mode-mix should look like from the perspective of achieving good response rates, and small nonresponse bias among important demographic subgroups is a topic for further research.

Issue 3: We need a mixed-mode or data integration framework

- Chapter 5 showed that we find measurement differences between smart- and non-smart surveys. These differences are to some degree expected and even desirable (e.g., more granular data collected with the app), but they do provide a challenge in combining data from smart and non-smart surveys. One way forward is to try and decrease the differences by changing our fieldwork design. For example, in France, interviewers are used to visit respondents, check the diaries and to fill in gaps in the diaries. Such edit checks can perhaps be pre-programmed in the app in such a way that respondents are reminded of missing episodes. Further research here is needed to minimize missing data overall, and decrease differences in measurement between smart and non-smart surveys, especially if in the future both modes are used alongside each other (e.g., to reach specific subgroups in the population).

- A change in methodology by moving from a non-smart to a smart survey will result in a break in time series. This change will be study and country specific, and therefore, countries will in the future need to conduct specific studies to estimate the extent of this break in time series. The best way to do this is to do a parallel run, if possible, using a cross-over design as was done in France. Calibration will always be needed to bridge potential breaks in the time series. The M6 deliverable provided a framework for how to do this calibration.

6.5 Overall conclusion

- The design of the smart survey methodology starts with the output. Smart elements should be closely aligned with the goal of study (important for recruitment, legal).
- Relying on extensive annotation, or asking for elaborate feedback from the respondent is not the way to go due to high burden and technical issues. A smart survey needs to be clever, and also work in situations where the respondent is not 'perfect'. One way to start to design a smart survey is to make an informed choice for where you want to be on the smart survey continuum. This often requires research and testing of different solutions, so that empirically, we can evaluate effects on quality and costs.
- Developing smart surveys is a complex task whose success relies on many interconnected parts. For Official Statistics institutions across countries it necessarily will be iterative, in this regard, the Smart Survey Continuum model is beneficial and can be used by the countries not included into our SSI project, as it allows to adjust to the readiness of the population for app-based data collection, technological readiness, as well as the resources available at the country's NSI.

References

- Benedikt, L., Joshi, C., Nolan, L., Henstra-Hill, R., Shaw, L., & Hook, S. (2020). Human-in-the-loop AI in government: a case study. Proceedings of the 25th International Conference on Intelligent User Interfaces (IUI '20). Association for Computing Machinery, New York, NY, USA, 488–497. <https://doi.org/10.1145/3377325.3377489>
- Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. John Wiley & Sons.

Appendix A: Enhancing Recruitment Strategies for Smart Surveys in Official Statistics

This Appendix contains the following country reports on the large field tests of the Household Budget Survey (HBS) and Time Use Survey (TUS).

1. StatBel – Field Report Belgium
2. University of Mannheim & Destatis – Field Report Germany
3. Destatis – Recommendations field test
4. SSB – Field Report Norway
5. Insee – Field test protocol France

Field Report Belgium

Patrick Lusyne

In the context of the Smart Survey Implementation (SSI) project, Belgium organized a large-scale field test using the MOTUS platform to explore the potential of smart surveys for official statistics. The Belgian test focused on a modular, app-based Time Use Survey (TUS), and was designed to experiment with strategies for improving participation (see Section 2.3.3 *Factors influencing survey participation in smart surveys*).

The sampling design followed a stratified random sample on the Belgian Population Register with a gross sample size of **N = 6,313**. The Brussels Capital Region was excluded for operational reasons, mainly due to its bilingual status. Households were selected, and within each household, one individual aged between 18 and 85 was randomly chosen to participate.

Fieldwork took place between **May and December 2024**, and was carried out in two main waves: the first wave in **May–June** (N = 1,269) and the second in **September–December** (N = 5,044). Due to disappointing response rates during the first wave, several adjustments were made to the field protocol as the study progressed. The most important changes were:

- Reducing the diary period from **7 days to 2 days**,
- Limiting the **PAPI (Paper-and-Pencil Interview) experiment** to the first wave only.

These and other smaller adjustments had significant operational implications, particularly in terms of managing the survey process within the back-end systems (i.e., the MOTUS app and Statistics Belgium's data warehouse environment). The fact that these changes were ultimately implemented without major technical issues is, from an operational point of view, one of the key takeaways of the project for Statistics Belgium.

Table 1. Distribution of Sociodemographic Characteristics in Gross Sample (n; percent in parentheses)

		BEL		
		7-days	2-days	All
Gender	Male	656 (51.7)	2,569 (50.9)	3,225 (51.1)
	Female	613 (48.3)	2,475 (49.1)	3,088 (48.9)
Age	18-24	27 (2.1)	100 (2.0)	127 (2.0)
	25-44	421 (33.2)	1,734 (34.4)	2155 (34.1)
	45-66	606 (47.8)	2,401 (47.6)	3007 (47.6)
	67+	215 (17.0)	809 (16.0)	1024 (16.2)
Nationality	Own	883 (69.6)	3,629 (72.0)	4,512 (71.5)
	Foreign	386 (30.4)	1,415 (28.1)	1,801 (28.5)
Region	Flanders	730 (57.6)	2,892 (57.4)	3,622 (57.4)
	Wallonia	538 (42.4)	2,148 (42.6)	2,686 (42.6)

List of Appendices

Appendix A: Invitation and reminder letters

Appendix B: Questionnaires

Appendix A: Invitation and reminder letters

Traditional invitation

{{voornaam_contact}} {{naam_contact}}
{{TX_ADRESS}}
{{CD_ZIP_RES}} {{TX_city}}

Uw berichten

Uw kenmerk
{{nr_ID}}

Ons kenmerk

Bijlagen

Betreft: Tijdsbestedingsonderzoek 2024

Geachte {{mevrouw/heer}} {{naam_contact}},

U bent samen met zo'n achtduizend andere personen geselecteerd voor een onderzoek over de tijdsbesteding van de Belgen. Ik nodig u in naam van Statbel, het Belgische statistiekbureau, uit om deel te nemen aan deze studie.

De kwaliteit van ons onderzoek steunt op uw medewerking. De deelname aan dit onderzoek is vrijblijvend. Een vergoeding van **15 euro** is voorzien voor de tijd en de energie die u hieraan wilt besteden.

De vragen gaan over uw dagelijks leven, met thema's zoals betaald en huishoudelijk werk, onderwijs, vrije tijd, mediagebruik, sociale contacten en verplaatsingen. Uw antwoorden zullen worden gebruikt om het sociaal beleid van België en Europa mee vorm te geven.

De studie verloopt online en bestaat uit drie onderdelen.

1. U vult een vragenlijst in over uw achtergrond, uw opleiding, uw werk, uw tijdsbesteding en uw mening daarover.
2. U houdt uw dagelijkse activiteiten bij in een online dagboekje gedurende 1 week.
3. U beantwoordt nog enkele vragen over de week waarin u uw dagboekje bijhield en uw deelname aan vrijetijds- en sportactiviteiten.

U kunt de vragenlijsten invullen en het dagboekje bijhouden via een webapplicatie op uw computer, laptop of tablet of via een app op uw smartphone of tablet.

Hiervoor werkt Statbel samen met de onderzoeksgroep TOR van de Vrije Universiteit Brussel. Zij ontwierpen het software platform MOTUS dat voor deze studie wordt gebruikt.

Om mee te doen scant u de QR-code hieronder of surft u naar www.motusresearch.io/nl en vul uw gebruikersnaam en wachtwoord in.

Uw gebruikersnaam: {{respondent_username}}
Uw wachtwoord: {{respondent_password}}



Ik kan u verzekeren dat de door u gegeven antwoorden en gegevens strikt vertrouwelijk worden behandeld, conform de statistiekwetgeving (o.a. privacy). Voor bijkomende informatie over de bescherming van uw privacy kunt u terecht op onze website <https://statbel.fgov.be/Privacy>.

Wij danken u reeds voor uw bereidheid tot medewerking aan dit onderzoek.

Hoogachtend,

Philippe Mauroy

Directeur-generaal a.i.

Voor meer informatie kunt u altijd mailen of gratis bellen naar onze hulplijn: TUS@economie.fgov.be of 0800 120 33. Houdt u dan wel uw identificatienummer bij de hand.



User-friendly invitation

{{voornaam_contact}} {{naam_contact}}
{{TX_ADRESS}}
{{CD_ZIP_RES}} {{TX_city}}

Geachte {{voornaam_contact}} {{naam_contact}},

Helpt u ons een beeld te krijgen hoe Belgen hun tijd besteden? Als geen ander weet u hoe u uw tijd besteedt, en welke taken u uitvoert op een dag. Maar u bent vast ook benieuwd naar hoe anderen hun tijd indelen. Hoe lang slapen de Belgen? Werkt de gemiddelde Belg van 9 tot 5? Hoe zit het met de verdeling van huishoudelijke taken binnen een gezin? Lezen jongeren minder boeken dan hun ouders? Bestaat de actieve gepensioneerde of zijn er meerdere verhalen? ...

Statbel, het Belgische statistiekbureau, voert onderzoek naar die tijdsbesteding om zo een objectief zicht te krijgen op hoe u en andere Belgen de dag invullen. Statbel werkt hiervoor samen met de onderzoeksgroep TOR van de Vrije Universiteit Brussel.

Waarom is uw deelname zo belangrijk?

Uw deelname betekent dat we een nauwkeuriger beeld kunnen schetsen van de uiteenlopende manieren waarop Belgen hun tijd gebruiken, waardoor de verschillende overheden een doeltreffender beleid kunnen uitstippelen. Veel sociale en gezondheidsvoorzieningen worden bijvoorbeeld in grote mate ondersteund door onbetaald werk. Uw antwoorden helpen om dit werk te meten en te waarderen, zodat er rekening mee kan worden gehouden bij het nemen van beslissingen.

We belonen deelnemers met een vergoeding!

De studie bestaat uit het beantwoorden van twee vragenlijsten en het bijhouden van uw dagelijkse activiteiten in een dagboekje gedurende 1 week. Voor de deelname aan het volledige onderzoek, ontvangt u een belastingvrije vergoeding van 15 euro.

Meedoen is eenvoudig

U kunt de vragenlijsten invullen en het dagboekje bijhouden via een webtoepassing op uw computer, laptop of tablet of via een app op uw smartphone of tablet.

Surf naar **www.motusresearch.io/nl** of scan de QR-code hiernaast met uw smartphone of tablet. Vul uw gebruikersnaam en wachtwoord in.

- Uw gebruikersnaam: **{{respondent_username}}**
- Uw wachtwoord: **{{respondent_password}}**



We respecteren uw privacy en we verwerken uw gegevens vertrouwelijk en anoniem. Lees onze privacyverklaring op <https://statbel.fgov.be/nl/privacy>. Het is ons alleen om gemiddelden en statistieken te doen.



Heeft u nog vragen?

Mail naar TUS@economie.fgov.be

Bel naar het gratis nummer 0800 120 33 van de FOD Economie.

Hartelijk dank voor uw medewerking!

Met vriendelijke groeten,

Philippe Mauroy

Directeur-generaal a.i. Statbel

Traditional reminder with PAPI

{{voornaam_contact}} {{naam_contact}}
{{TX_ADRESS}}
{{CD_ZIP_RES}} {{TX_city}}

Uw berichten

Uw kenmerk
{{nr_ID}}

Ons kenmerk

Bijlagen

Betreft: Tijdsbestedingsonderzoek 2024

Geachte {{mevrouw/heer}} {{naam_contact}},

We nodigden u een tijdje geleden uit om deel te nemen aan een studie over tijdsbesteding. Als u de vragenlijsten en het dagboekje al invulde, bedanken wij u daarvoor. Kwam u er nog niet toe, dan kan u dit nog steeds doen.

De kwaliteit van ons onderzoek steunt op uw medewerking. Een vergoeding van **15 euro** is voorzien voor de tijd en de energie die u hieraan wilt besteden.

U kunt de vragenlijsten invullen en het dagboekje bijhouden via een webapplicatie op uw computer, laptop of tablet of via een app op uw smartphone of tablet.

Hiervoor werkt Statbel samen met de onderzoeksgroep TOR van de Vrije Universiteit Brussel. Zij ontwierpen het software platform MOTUS dat voor deze studie wordt gebruikt.

Om mee te doen scant u de QR-code hieronder of surft u naar **www.motusresearch.io/nl** en vul uw gebruikersnaam en wachtwoord in.

Uw gebruikersnaam: {{respondent_username}}
Uw wachtwoord: {{respondent_password}}



Verkiest u op papier deel te nemen? Vul het meegestuurd dagboekje en de aanvullende vragenlijst in en stuur deze gratis terug in de voorgedrukte, geadresseerde en gefrankeerde envelop.

Ik kan u verzekeren dat de door u gegeven antwoorden en gegevens strikt vertrouwelijk worden behandeld, conform de statistiekwetgeving (o.a. privacy). Voor bijkomende informatie over de bescherming van uw privacy kunt u terecht op onze website <https://statbel.fgov.be/Privacy>.

Wij danken u reeds voor uw bereidheid tot medewerking aan dit onderzoek.

Hoogachtend,

Philippe Mauroy

Directeur-generaal a.i.

Voor meer informatie kunt u altijd mailen of gratis bellen naar onze hulplijn: TUS@economie.fgov.be of 0800 120 33. Houdt u dan wel uw identificatienummer bij de hand.



User-friendly reminder with PAPI

{{voornaam_contact}} {{naam_contact}}
{{TX_ADRESS}}
{{CD_ZIP_RES}} {{TX_city}}

Geachte {{voornaam_contact}} {{naam_contact}},

We nodigden u een tijdje geleden uit om deel te nemen aan een studie over tijdsbesteding.
Als u de vragenlijsten en het dagboekje al invulde, bedanken wij u daarvoor.
Kwam u er nog niet toe, dan kan u dit nog steeds doen.

We belonen deelnemers met een vergoeding!

De studie bestaat uit het beantwoorden van twee vragenlijsten en het bijhouden van uw dagelijkse activiteiten in een dagboekje gedurende 1 week. Voor de deelname aan het volledige onderzoek, ontvangt u een belastingvrije vergoeding van 15 euro.

Meedoen is eenvoudig

U kunt de vragenlijsten invullen en het dagboekje bijhouden via een webtoepassing op uw computer, laptop of tablet of via een app op uw smartphone of tablet.

Surf naar **www.motusresearch.io/nl** of scan de QR-code hiernaast met uw smartphone of tablet. Vul uw gebruikersnaam en wachtwoord in.

- Uw gebruikersnaam: **{{respondent_username}}**
- Uw wachtwoord: **{{respondent_password}}**



Verkiest u op papier deel te nemen? Vul het dagboek en de vragenlijst in en stuur deze gratis terug in de voorgedrukte, geadresseerde en gefrankeerde envelop.

We respecteren uw privacy en we verwerken uw gegevens vertrouwelijk en anoniem. Lees onze privacyverklaring op <https://statbel.fgov.be/nl/privacy>. Het is ons alleen om gemiddelden en statistieken te doen.



Heeft u nog vragen?

Mail naar TUS@economie.fgov.be

Bel naar het gratis nummer 0800 120 33 van de FOD Economie.

Hartelijk dank voor uw medewerking!

Met vriendelijke groeten,

Philippe Mauroy
Directeur-generaal a.i. Statbel

Traditional reminder without PAPI

{{voornaam_contact}} {{naam_contact}}
{{TX_ADRESS}}
{{CD_ZIP_RES}} {{TX_city}}

Uw berichten

Uw kenmerk
{{nr_ID}}

Ons kenmerk

Bijlagen

Betreft: Tijdsbestedingsonderzoek 2024

Geachte {{mevrouw/heer}} {{naam_contact}},

We nodigden u een tijdje geleden uit om deel te nemen aan een studie over tijdsbesteding. Als u de vragenlijsten en het dagboekje al invulde, bedanken wij u daarvoor. Kwam u er nog niet toe, dan kan u dit nog steeds doen.

De kwaliteit van ons onderzoek steunt op uw medewerking. Een vergoeding van **15 euro** is voorzien voor de tijd en de energie die u hieraan wilt besteden.

De studie verloopt online en bestaat uit drie onderdelen.

1. U vult een vragenlijst in over uw achtergrond, uw opleiding, uw werk, uw tijdsbesteding en uw mening daarover.
2. U houdt uw dagelijkse activiteiten bij in een online dagboekje gedurende 1 week.
3. U beantwoordt nog enkele vragen over de week waarin u uw dagboekje bijhield en uw deelname aan vrijetijds- en sportactiviteiten.

U kunt de vragenlijsten invullen en het dagboekje bijhouden via een webapplicatie op uw computer, laptop of tablet of via een app op uw smartphone of tablet.

Hiervoor werkt Statbel samen met de onderzoeksgroep TOR van de Vrije Universiteit Brussel. Zij ontwierpen het software platform MOTUS dat voor deze studie wordt gebruikt.

Om mee te doen scant u de QR-code hieronder of surft u naar www.motusresearch.io/nl en vul uw gebruikersnaam en wachtwoord in.

Uw gebruikersnaam: {{respondent_username}}
Uw wachtwoord: {{respondent_password}}



Ik kan u verzekeren dat de door u gegeven antwoorden en gegevens strikt vertrouwelijk worden behandeld, conform de statistiekwetgeving (o.a. privacy). Voor bijkomende informatie over de bescherming van uw privacy kunt u terecht op onze website <https://statbel.fgov.be/Privacy>.

Wij danken u reeds voor uw bereidheid tot medewerking aan dit onderzoek.

Hoogachtend,

Philippe Mauroy

Directeur-generaal a.i.

Voor meer informatie kunt u altijd mailen of gratis bellen naar onze hulplijn: TUS@economie.fgov.be of 0800 120 33. Houdt u dan wel uw identificatienummer bij de hand.



User-friendly reminder without PAPI

{{voornaam_contact}} {{naam_contact}}
{{TX_ADRESS}}
{{CD_ZIP_RES}} {{TX_city}}

Geachte {{voornaam_contact}} {{naam_contact}},

We nodigden u een tijdje geleden uit om deel te nemen aan een studie over tijdsbesteding.
Als u de vragenlijsten en het dagboekje al invulde, bedanken wij u daarvoor.
Kwam u er nog niet toe, dan kan u dit nog steeds doen.

We belonen deelnemers met een vergoeding!

De studie bestaat uit het beantwoorden van twee vragenlijsten en het bijhouden van uw dagelijkse activiteiten in een dagboekje gedurende 1 week. Voor de deelname aan het volledige onderzoek, ontvangt u een belastingvrije vergoeding van 15 euro.

Meedoen is eenvoudig

U kunt de vragenlijsten invullen en het dagboekje bijhouden via een webtoepassing op uw computer, laptop of tablet of via een app op uw smartphone of tablet.

Surf naar **www.motusresearch.io/nl** of scan de QR-code hiernaast met uw smartphone of tablet. Vul uw gebruikersnaam en wachtwoord in.

- Uw gebruikersnaam: **{{respondent_username}}**
- Uw wachtwoord: **{{respondent_password}}**



We respecteren uw privacy en we verwerken uw gegevens vertrouwelijk en anoniem. Lees onze privacyverklaring op <https://statbel.fgov.be/nl/privacy>. Het is ons alleen om gemiddelden en statistieken te doen.



Heeft u nog vragen?

Mail naar TUS@economie.fgov.be

Bel naar het gratis nummer 0800 120 33 van de FOD Economie.

Hartelijk dank voor uw medewerking!

Met vriendelijke groeten,

Philippe Mauroy
Directeur-generaal a.i. Statbel

Appendix B:
CAWI - Personal information

Variable name	HETUS-variable	N
Comm_0		Ter bescherming van uw privacy beschikt MOTUS niet over dezelfde (persoons)gegevens als Statbel. Om het goede verloop van het onderzoek te bewaren, vragen we hieronder uw voornaam, familienaam, geslacht en geboortedatum in te vullen.
Firstname_ref		Wat is uw voornaam?
		Open veld (max 300 karakters)
		verplicht veld
Lastname_ref		Wat is uw familienaam?
		Open veld (max 300 karakters)
		verplicht veld
Gender_ref	INC1	Wat is uw geslacht?
		Man
		Vrouw
		X
		Zeg ik liever niet
		verplicht veld
Birthdate_ref	INC2/INC2A/INC2B	Wat is uw geboortedatum?
		Gebruik het icoon om een datum te kiezen of typ dd/mm/jjjj
		dd/mm/jjjj
		range: [01/01/1900-01/01/2006]
		verplicht veld
Email_ref		Gelieve een geldig e-mailadres op te geven. Dit adres zal worden gebruikt voor al de communicaties met betrekking tot dit onderzoek.
		Open veld (max 300 karakters)
		verplicht veld

F

Pour protéger votre vie privée, MOTUS n'a pas accès aux mêmes données (personnelles) que Statbel. Afin d'assurer le bon déroulement de l'enquête, nous vous demandons de remplir ci-dessous vos prénom, nom de famille, sexe et date de naissance.

Quel est votre prénom ?

Champ libre (max 300 caractères)

champ obligatoire

Quel est votre nom de famille ?

Champ libre (max 300 caractères)

champ obligatoire

Quel est votre sexe ?

Homme

Femme

X

Je préfère ne pas le dire

champ obligatoire

Quelle est votre date de naissance ?

Utilisez l'icône pour choisir une date ou entrez jj/mm/aaaa

jj/mm/aaaa

range: [01/01/1900-01/01/2006]

champ obligatoire

Veillez indiquer une adresse de courriel valide. Cette adresse sera utilisée pour toutes les communications relatives à cette étude.

Champ libre (max 300 caractères)

champ obligatoire

E	Code	Logic
<p>To protect your privacy, MOTUS does not have access to the same (personal) data as Statbel. In order to keep the survey running smoothly, we ask you to fill in your first name, last name, gender and date of birth below.</p> <p>What is your first name?</p> <p>Open field (max 300 characters) mandatory field</p>		Firstname_ref
<p>What is your last name?</p> <p>Open field (max 300 characters) mandatory field</p>		Surname_ref
<p>What is your gender?</p> <p>Man Woman X I prefer not say mandatory field</p>		Gender_ref
<p>What is your birth date?</p> <p>Use the icon to choose a date or type dd/mm/yyyy dd/mm/yyyy range: [01/01/1900-01/01/2006] mandatory field</p>	1 2 3 98	Birthdate_ref Birthdate_ref Birthdate_ref Birthdate_ref
<p>Please provide a valid e-mail address. This address will be used for all communications regarding this study.</p> <p>Open field (max 300 characters) mandatory field</p>		End

CAWI - pre-questionnaire

Variable name	HETUS-variable	N	F	E	Code	Logic	Remarks
Comm_0		In deze vragenlijst stellen we een aantal vragen over uw achtergrond, uw opleiding, uw werk, uw tijdsbesteding en uw mening daarover. Het invullen van deze vragenlijst duurt gemiddeld minder dan 10 minuten.	Dans ce questionnaire, nous posons des questions sur vos antécédents, votre formation, votre travail, votre emploi du temps et vos opinions à ce sujet. En moyenne, il faut moins de 10 minutes pour le compléter.	In this questionnaire, we ask a number of questions about your background, your education, your work, your use of time and your opinions about it. On average, it takes less than 10 minutes to complete.			
Comm_1		De volgende vragen gaan over uw huishouden.	Les questions suivantes concernent votre ménage.	The following questions concern your household.		NumAdult	
NumAdult		Hoeveel volwassenen van 18 jaar of ouder maken deel uit van uw huishouden, uzelf inbegrepen?	Combien d'adultes âgés de 18 ans ou plus font partie de votre ménage, y compris vous-même ?	How many adults aged 18 or older are part of your household, including yourself?			
		Aantal (1-20)	Nombre (1-20)	Number (1-20)		NumChild_1	
NumChild_1		Geen antwoord	Pas de réponse	No answer	98		
		Hoeveel kinderen van 6 tot en met 17 jaar maken deel uit van uw huishouden?	Combien d'enfants âgés de 6 à 17 ans font partie de votre ménage ?	How many children aged 6 to 17 are part of your household?			
		Noteer '0' indien geen gezinsleden van 6 tot en met 17 jaar.	Veillez noter '0' si aucun membre du ménage n'est âgé de 6 à 17 ans.	Please note '0' if no household members aged 6 to 17.			
		Aantal (0-19)	Nombre (0-19)	Number (0-19)		NumChild_2	
NumChild_2		Geen antwoord	Pas de réponse	No answer	98		
		Hoeveel kinderen jonger of gelijk aan 5 jaar maken deel uit van uw huishouden?	Combien d'enfants âgés de 5 ans ou moins font partie de votre ménage ?	How many children aged 5 or younger are part of your household?			
		Noteer '0' indien geen gezinsleden van 5 jaar of jonger.	Veillez noter '0' si aucun membre de la famille n'est âgé de 5 ans ou moins.	Please note '0' if no household members aged 5 or younger.			
		Aantal (0-19)	Nombre (0-19)	Number (0-19)		TypeHH	
TypeHH		Geen antwoord	Pas de réponse	No answer	98		
		In welk type huishouden woont u?	Dans quel type de ménage vivez-vous ?	What type of household do you live in?			
		Eenpersoonshuishouden	Ménage isolé (vivant seul)	Single person household	1	Comm_2	
		Koppel zonder inwonende kinderen (+ eventueel inwonende andere personen)	Couple sans enfant résident (+ toute autre personne vivant dans le ménage)	Couple without resident children (+ any other persons living in the household)	2	PosHH_1	
		Koppel met inwonend(e) kind(eren) (+ eventueel inwonende andere personen)	Couple avec enfant(s) résident(s) (+ toute autre personne vivant dans le ménage)	Couple with resident child(ren) (+ any other persons living in the household)	3	PosHH_2	
		Eenoudergezin (+ eventueel inwonende andere personen)	Ménage monoparental (+ toute autre personne vivant dans le ménage)	Single parent household (+ any other persons living in the household)	4	PosHH_3	
		Ander type huishouden (samenwonende broers/zussen, samenwonende vrienden, ...)	Autre type de ménage (frères/sœurs vivant ensemble, amis vivant ensemble, ...)	Other type of household (brothers/sisters living together, friends living together, ...)	5	Comm_2	
PosHH_1		Geen antwoord	Pas de réponse	No answer	98	Comm_2	
		Wat is uw positie in het huishouden? U bent ...	Quelle est votre position dans le ménage ? Vous êtes ...	What is your position in the household? You are ...			
		Partner (in koppel zonder inwonende kinderen)	Partenaire (d'un couple sans enfant résident)	Partner (of a couple without resident children)	1	Comm_2	
		Inwonende andere persoon	Autre personne vivant dans le ménage	Other person living in the household	2	Comm_2	
PosHH_2		Geen antwoord	Pas de réponse	No answer	98		
		Wat is uw positie in het huishouden? U bent ...	Quelle est votre position dans le ménage ? Vous êtes ...	What is your position in the household? You are ...			
		Partner (in koppel met inwonende kinderen)	Partenaire (d'un couple avec enfant(s) résident(s))	Partner (of a couple with resident child(ren))	1	Comm_2	
		Kind	Enfant	Child	2	Comm_2	
		Inwonende andere persoon	Autre personne vivant dans le ménage	Other person living in the household	3	Comm_2	
PosHH_3		Geen antwoord	Pas de réponse	No answer	98		
		Wat is uw positie in het huishouden? U bent ...	Quelle est votre position dans le ménage ? Vous êtes ...	What is your position in the household? You are ...			
		Alleenstaande ouder	Parent isolé	Single parent	1	Comm_2	
		Kind	Enfant	Child	2	Comm_2	
		Inwonende andere persoon	Autre personne vivant dans le ménage	Other person living in the household	3	Comm_2	
		Geen antwoord	Pas de réponse	No answer	98		
Comm_2		De volgende vragen gaan over uw werksituatie.	Les questions suivantes concernent votre situation professionnelle.	The following questions are about your work situation.		Activity1	
Activity1	IND1	Hebt u in de week die vorige zondag eindigde betaalde arbeid verricht, ook al was dit maar één uur?	Pensons à la semaine se terminant dimanche dernier, avez-vous effectué un travail rémunéré, même s'il ne s'agit que d'une heure ?	Did you do any paid work, even for one hour only, during the week that ended last Sunday?			
		Dit kan zijn als werknemer of in uw eigen bedrijf of familiebedrijf of op de boerderij van uw gezin.	Cela peut être comme salarié ou dans votre propre entreprise, dans une entreprise familiale ou dans une exploitation agricole de votre ménage.	This can be as an employee or in your own company or in a family business or on a farm of your household.			
		Ja	Oui	Yes	1	Comm_3	

Activity2	IND1	Neen	Non	No	2	Activity2
		Geen antwoord	Pas de réponse	No answer	98	
		Was u tijdelijk afwezig wegens vakantie, ziekte of omwille van een andere reden?	Avez-vous été temporairement absent à cause de vacances, d'une maladie ou encore pour d'autres raisons ?	Were you temporarily absent due to holiday, illness or for any other reason?		
Activity3	IND2	Ja	Oui	Yes	1	Activity3
		Neen	Non	No	2	Activity4
		Geen antwoord	Pas de réponse	No answer	98	
		Wat is de reden dat u die week helemaal niet werkte?	Pour quelle raison n'avez-vous pas du tout travaillé cette semaine-là ?	Why were you absent for the whole week?		
		Eigen ziekte, letsel of tijdelijke arbeidsongeschiktheid	Maladie, blessure ou incapacité de travail temporaire	Own illness, injury or temporary disability	1	Comm_3
Comm_3		Vakantie	Vacances	Holiday	2	Comm_3
		Moederschaps-, vaderschaps- of ouderschapsverlof	Congé de maternité, de paternité ou parental	Maternity, paternity or parental leave	3	Comm_3
		Verlof wegens studie	Congé éducation	Leave of absence for studies	4	Comm_3
		Arbeidsconflict	Conflit au travail	Labour dispute	5	Comm_3
		Technische of tijdelijke werkloosheid	Chômage technique ou temporaire	Technical or temporary unemployment	6	Comm_3
		Andere reden	Autre raison	Other reason	7	Comm_3
		Geen antwoord	Pas de réponse	No answer	98	
		De volgende vragen hebben betrekking op uw werk. Als u meerdere jobs of meerdere ondernemingen hebt, is het de bedoeling dat u alleen over die job of onderneming vertelt waaraan u het grootste deel van uw werktijd besteedt. Indien voor alle jobs evenveel uren wordt gewerkt, dan wordt de job die het hoogste loon oplevert beschouwd als uw hoofdjob.	Les questions suivantes portent sur votre emploi. Si vous avez plusieurs emplois ou plusieurs affaires, veuillez répondre pour l'emploi (ou l'affaire) auquel vous consacrez le plus de temps. Si vous travaillez le même nombre d'heures pour tous vos emplois, veuillez considérer celui duquel vous tirez votre plus gros revenu comme votre emploi principal.	The following questions refer to your job or business. If you are engaged in two or more jobs or businesses, please consider only the one in which you spend most of your working hours. If you work the same amount of hours for each of the jobs, it will be the job which earned you the highest wage.		Job4
Job4	IND6_1	Tot welke beroepscategorie behoort u?	A quelle catégorie professionnelle appartenez-vous ?	What is your professional status?		
		Deze vraag beantwoorden voor de job (of onderneming) waaraan u het grootste deel van uw werktijd besteedt als u meerdere jobs (of meerdere ondernemingen) hebt.	Veuillez répondre pour l'emploi (ou l'affaire) auquel vous consacrez le plus de temps si vous avez plusieurs emplois (ou plusieurs affaires).	Please answer this question for the job (or business) you spend most of your working time on if you have several jobs (or several businesses).		
SrtSector_1		Private sector - arbeider	Secteur privé, ouvrier(ère)	Private sector - manual (laborer)	1	SrtSector_3
		Private sector - bediende	Secteur privé, employé(e)	Private sector - non-manual (employee)	2	SrtSector_3
		Openbare sector - statutair ambtenaar	Secteur public, fonctionnaire statutaire	Public sector - fixed appointment (statutory)	3	SrtSector_3
		Openbare sector - contractueel	Secteur public, contractuel(le)	Public sector - contract	4	SrtSector_3
		Vrij beroep / zelfstandige zonder personeel	Profession libérale / indépendant sans personnel	Self-employed without employees	5	SrtSector_1
		Vrij beroep / zelfstandige met personeel	Profession libérale / indépendant avec personnel	Self-employed with employees	6	SrtSector_1
		Helper / meewerkend familielid (zonder vergoeding)	Conjoint(e) aidant(e) (non rémunéré(e))	Family worker (unpaid)	7	SrtSector_2
		Geen antwoord	Pas de réponse	No answer	98	
SrtSector_1		ALS Job4 = 5/6 DAN: Onder welke sector valt uw bedrijf/praktijk?	SI Job4 = 5/6 ALORS : Quel est le secteur d'activité de votre entreprise/activité ?	IF Job4 = 5/6 THEN: Within which sector does your company/practice fall?		
SrtSector_2		ALS Job4 = 7 DAN: Onder welke sector valt het bedrijf/praktijk van uw partner of familielid?	SI Job4 = 7 ALORS : Quel est le secteur d'activité de l'entreprise/activité de votre partenaire ou famille ?	IF Job4 = 7 THEN: Within which sector does the company/practice of your partner or relative fall?		
SrtSector_3		ALS Job4 = 1/2/3/4 DAN: In welke sector of in wat soort bedrijf bent u werkzaam?	SI Job4 = 1/2/3/4 ALORS : Dans quel secteur ou type d'entreprise travaillez-vous ?	IF Job4 = 1/2/3/4 THEN: In which sector or in what type of company are you working?		
		U kunt kiezen uit de volgende sectoren	Vous pouvez choisir parmi les secteurs suivants	Please select one of the following sectors		
		Landbouw, bosbouw, visserij	Agriculture, sylviculture, pêche	Agriculture, forestry, fisheries	1	SrtLand
		Delfstoffen, energie, water, afval	Minéraux, énergie, eau, déchets	Minerals, energy, water, waste	2	SrtEnergie
		Bouw	Construction	Construction	3	SrtBouw
		Productie, installatie, reparatie	Production, installation, réparation	Production, installation, repairs	4	SrtProd
		Transport, vervoer, zorgdiensten	Transports, transport de marchandises, services de livraison	Transport, transportation, delivery services	5	SrtTransp
		Handel (bijv. winkel, groothandel, autohandel)	Commerce (par exemple, magasins, commerce de gros, commerce automobile)	Trade (e.g. shop, wholesaler, car dealership)	6	SrtHandel

	Horeca, recreatie, sport	Horeca, création, sports	Accommodation and food services, recreation, sports	7	SrtHoreca
	Zorg	Soins	Healthcare	8	SrtZorg
	Onderwijs	Enseignement	Education	9	SrtOnderw
	Overheid	Gouvernement	Government	10	ActOverh
	IT, telecommunicatie, media	Informatique, télécommunications, médias	IT, telecommunications, media	11	SrtIT
	Financiële dienstverlening, vastgoed	Services financiers, immobilier	Financial services, real estate	12	SrtFinan
	Andere	Autre	Other	13	SrtAnders
	Ik weet het niet	Je ne sais pas	I don't know	97	SrtAnders
	Geen antwoord	Pas de réponse	No answer	98	
SrtLand	Gaat het dan om:	S'agit-il de :	Does it involve:		
	Akkerbouw / tuinbouw	Agriculture / horticulture	Arable farming / horticulture	1	ActLand
	Veeteelt / fokkerij	Production animale / élevage d'animaux	Livestock farming / breeding	2	ActLand
	Gemengd bedrijf (combinatie van akker- of tuinbouw met veeteelt)	Culture et élevage associés (production combinée de produits végétaux et de produits animaux)	Mixed farm (combination of arable farming or horticulture with livestock farming)	3	ActLand
	Dienstverlening in de landbouw	Activités de service dans l'agriculture	Agricultural services	4	ActLand
	Bosbouw	Sylviculture	Forestry	5	ActLand
	Visserij	Pêche	Fisheries	6	ActLand
	Ik weet het niet	Je ne sais pas	I don't know	97	ActLand
	Geen antwoord	Pas de réponse	No answer	98	
ActLand	Waarin is dit bedrijf gespecialiseerd?	Quelle est la spécialité de cette entreprise ?	What is this company's specialisation?		
	Probeer dit zo precies mogelijk te omschrijven door een gewas, diersoort of productiewijze op te geven. Bijv. teelt van aardbeien onder glas, houden van melkkoeien, kweken van vis- en schaaldieren, uitdunnen en in stand houden van bossen.	Essayez de le décrire le plus précisément possible en précisant une culture, une espèce animale ou un mode de production. Par exemple, la culture de fraises sous serre, l'élevage de vaches laitières, l'élevage de poissons et de crustacés, l'éclaircissement et l'entretien des forêts.	Please try to describe this as precisely as possible by indicating a crop, type of animal or manner of production. For example: growing strawberries under glass; keeping dairy cows; farming fish and shellfish; thinning and maintaining forests.		
SrtEnergie	Open veld (max 300 karakters)	Champ libre (max 300 caractères)	Open field (max 300 characters)		Job2
	Gaat het dan om:	S'agit-il de :	Does it involve:		
	Winning van aardolie of gas	Extraction de pétrole ou de gaz	Extraction of petroleum or gas	1	ActEnergie
	Winning van zand, grind of andere delfstoffen	Extraction de sable, de gravier ou d'autres minéraux	Extraction of sand, gravel or other minerals	2	ActEnergie
	Winning en distributie van water	Extraction et distribution d'eau	Extraction and distribution of water	3	ActEnergie
	Productie, distributie of verkoop van energie (elektriciteit, gas, wind, zon)	Production, distribution ou vente d'énergie (électricité, gaz, énergie éolienne, énergie solaire)	Production, distribution or sale of energy (electricity, gas, wind, solar)	4	ActEnergie
	Rioolbeheer, waterzuivering	Gestion des égouts, traitement de l'eau	Sewage treatment, water purification	5	ActEnergie
	Inzamelen, verwerken of recyclen van afval (sloop)	Collecte, traitement ou recyclage des déchets (démolition)	Collection, processing or recycling of waste (demolition)	6	ActEnergie
	Sanering (bijv. bodemzuivering, asbest opruimen)	Dépollution (par exemple, dépollution des sols, désamiantage)	Decontamination (e.g. soil purification, asbestos removal)	7	ActEnergie
	Andere	Autre	Other	8	ActEnergie
	Ik weet het niet	Je ne sais pas	I don't know	97	ActEnergie
	Geen antwoord	Pas de réponse	No answer	98	
ActEnergie	Waarin is dit bedrijf gespecialiseerd?	Quelle est la spécialité de cette entreprise ?	What is this company's specialisation?		
	Probeer dit zo precies mogelijk te omschrijven. Bijv. proefboren voor de winning van aardgas, productie van drinkwater, beheer van het hoofdtransportnet voor elektriciteit, zuiveren van afvalwater, snipperen van papier voor recycling, zuiveren van vervuilde bodems.	Essayez de le décrire le plus précisément possible. Par exemple : forages d'essai pour l'extraction de gaz naturel, production d'eau potable, gestion du réseau principal de transport d'électricité, traitement des eaux usées, déchetage de papier pour le recyclage, traitement des sols pollués.	Please try to describe this as precisely as possible. For example: test drilling for the extraction of natural gas, production of drinking water, management of the main electricity transmission network, wastewater treatment, shredding paper for recycling, purification of polluted soils.		
SrtBouw	Open veld (max 300 karakters)	Champ libre (max 300 caractères)	Open field (max 300 characters)		Job2
	Gaat het dan om:	S'agit-il de :	Does it involve:		
	Grond-, water- of wegenbouw	Génie civil	Land, water or road construction	1	ActBouw
	Installatie (bijv. elektriciteit, verwarming, riolering)	Installation (par exemple, installation électrique, installation de chauffage, installation d'égouts)	Installation (e.g. electricity, heating, sewage systems)	2	ActBouw
	Afwerking (bijv. stukadoren, schilderen, glas zetten, vloeren leggen)	Finition (par exemple, travaux de plâtrerie, travaux de peinture, vitrerie, pose de carrelages de sols)	Finishing (e.g. plastering, painting, glazing, laying floors)	3	ActBouw

	Andere gespecialiseerde bouw (bijv. dakdekken, metselen, heien)	Autres travaux de construction spécialisés (par exemple, couverture, maçonnerie, battage de pieux)	Other specialised construction (e.g. roofing, masonry, pile driving)	4	ActBouw
	Aannemersbedrijf of algemeen klusbedrijf Architectuur (bijv. architecten- of ingenieursbureau)	Entreprise de construction ou d'homme à tout faire Architecture (par exemple, cabinet d'architectes ou d'ingénieurs)	Contracting company or general contractor Architecture (e.g. architectural or engineering firm)	5 6	ActBouw ActBouw
	Andere	Autre	Other	7	ActBouw
	Ik weet het niet	Je ne sais pas	I don't know	97	ActBouw
	Geen antwoord	Pas de réponse	No answer	98	
ActBouw	Wat zijn de belangrijkste activiteiten van dit bedrijf?	Quelles sont les principales activités de cette entreprise ?	What are the most important activities of this company?		
	Probeer dit zo precies mogelijk te omschrijven. Bijv. bestraten, installatie van cv-ketels, dakdekken, projectontwikkeling, ontwerpen van woningen voor particulieren, ontwerp verkeerskundige projecten.	Essayez de les décrire le plus précisément possible. Par exemple : pavage, installation de chaudières de chauffage central, couverture, promotion immobilière, conception de maisons pour particuliers, conception de projets de circulation.	Please try to describe this as precisely as possible. For example: paving, installation of central heating boilers, roofing, property development, design of private dwellings, design of traffic projects.		
SrtProd	Open veld (max 300 karakters) Gaat het dan om een: Fabriek of productiebedrijf Installatiebedrijf Garage / reparatiebedrijf voor auto's, motorfietsen, aanhangwagens Reparatiebedrijf voor consumentenartikelen (bijv. computer, fiets, kleding) Ander reparatie- of onderhoudsbedrijf Ik weet het niet Geen antwoord	Champ libre (max 300 caractères) S'agit-il d'un(e) : Usine ou entreprise de fabrication Entreprise d'installation Garage/atelier de réparation pour voitures, motos, remorques Atelier de réparation de biens de consommation (par exemple, ordinateur, bicyclette, vêtements) Autre entreprise de réparation ou d'entretien Je ne sais pas Pas de réponse	Open field (max 300 characters) Is it a(n): Factory or production company Installation company Garage / repair company for cars, motorcycles, trailers Repair company for consumer articles (e.g. computers, bicycles, clothing) Other repair or maintenance company I don't know No answer	1 2 3 4 5 97 98	Job2 ActProd_1 ActProd_3 ActProd_2 ActProd_3 ActProd_3 ActProd_1
ActProd_1	ALS SrtProd=1 DAN: Wat voor (half)producten worden gemaakt? Probeer dit zo precies mogelijk te omschrijven. Bijv. metalen bouwmaterialen, brood, elektrische huishoudelijke apparaten, kinderkleding.	SI SrtProd=1 ALORS : Quels types de (semi-)produits sont fabriqués ? Essayez de le décrire aussi précisément que possible. Par exemple, des matériaux de construction métalliques, du pain, des appareils électroménagers, des vêtements pour enfants.	IF SrtProd=1: What type of products (or semi-products) are made? Please try to describe this as precisely as possible. For example: metal building materials, bread, electric household appliances, children's clothing.		
ActProd_2	ALS SrtProd=3 DAN: Welke voertuigen (of onderdelen) worden voornamelijk gerepareerd? Probeer dit zo precies mogelijk te omschrijven. Bijv. personenauto's, vrachtwagens, motorfietsen, caravans, carrosserieherstel, bandenservicebedrijf.	SI SrtProd=3 ALORS : Quels véhicules (ou pièces) sont principalement réparés ? Essayez de le décrire aussi précisément que possible. Par exemple : voitures, camions, motos, caravanes, réparation de carrosserie, entreprise d'entretien de pneus.	IF SrtProd=3: What types of vehicles (or parts) are mainly repaired? Please try to describe this as precisely as possible. For example: passenger cars, lorries, motorcycles, campers, chassis repairs, tire service.		
ActProd_3	ALS SrtProd=2/4/5 DAN: Wat wordt geïnstalleerd of gerepareerd? Probeer dit zo precies mogelijk te omschrijven. Bijv. kleding, computers voor consumenten, bewakingsapparatuur voor bedrijven.	SI SrtProd=2/4/5 ALORS : Qu'est-ce qui est installé ou réparé ? Essayez de le décrire aussi précisément que possible. Par exemple, vêtements, ordinateurs grand public, matériel de surveillance pour les entreprises.	IF SrtProd=2/4/5: What is installed or repaired? Please try to describe this as precisely as possible. For example: clothing, computers for consumers, surveillance equipment for businesses.		
SrtTransp	Open veld (max 300 karakters) Gaat het dan om: Transport / vervoer Post-, pakket- of bezorgdiensten Ik weet het niet Geen antwoord	Champ libre (max 300 caractères) S'agit-il de : Transport Services postaux, de colis ou de livraison Je ne sais pas Pas de réponse	Open field (max 300 characters) Does it involve: Transport / transportation Postal, package or delivery services I don't know No answer	1 2 97 98	Job2 TypTransp TypTransp
TypTransp	Hoe vindt het transport of vervoer plaats? Is dat via: Weg Lucht Binnenvaart Zeevaart Spoor Pijpleiding	Comment se déroule le transport ? S'agit-il d'un transport par : Route Voie aérienne Navigation intérieure Navigation maritime Rail Canalisation	How is the transport or transportation performed? Is this by: Road Air Inland water navigation Sea and coastal navigation Rail Pipeline	1 2 3 4 5 6	PersGoed PersGoed PersGoed PersGoed PersGoed PersGoed

PersGoed	Andere (bijv. dienstverlening, opslag, weging)	Autre (par exemple, service, stockage, pesage)	Other (e.g. service, storage, weighing)	7	PersGoed
	Ik weet het niet	Je ne sais pas	I don't know	97	
	Geen antwoord	Pas de réponse	No answer	98	
	Gaat het om personen- of goederenvervoer?	S'agit-il d'un transport de passagers ou de marchandises ?	Does it involve passenger transportation or freight transport?		
ActTransp	Personen	Personnes	Passenger transportation	1	ActTransp
	Goederen	Marchandises	Freight transport	2	ActTransp
	Ik weet het niet	Je ne sais pas	I don't know	97	
	Geen antwoord	Pas de réponse	No answer	98	
	Wat zijn de belangrijkste activiteiten van dit bedrijf?	Quelles sont les principales activités de cette entreprise ?	What are the most important activities of this company?		
	Probeer dit zo precies mogelijk te omschrijven. Bijv. postbezorging, bezorgen van afhaalmaaltijden per fiets, autoparkeerterrein, laden en lossen van ladingen uit zeeschepen, verhuisdiensten.	Essayez de les décrire le plus précisément possible. Par exemple : distribution de courrier, livraison de repas à emporter à vélo, parking, chargement et déchargement de cargaisons de navires, services de déménagement.	Please try to describe this as precisely as possible. For example: postal delivery, takeaway meal delivery by bicycle, car parking, loading and unloading of freight from ocean-going vessels, relocation services.		
SrtHandel	Open veld (max 300 karakters)	Champ libre (max 300 caractères)	Open field (max 300 characters)		Job2
	Gaat het dan om:	S'agit-il de :	Does it involve:		
	Detailhandel (bijv. winkel, (super)markt, benzinstation)	Commerce de détail (par exemple, magasin, (super)marché, station-service)	Retail trade (e.g. shop, market, supermarket, petrol station)	1	ActHandel_1
	Groothandel	Commerce de gros	Wholesale	2	ActHandel_2
ActHandel_1	Handelsbemiddeling	Intermédiaires du commerce	Trade intermediation	3	ActHandel_3
	Ik weet het niet	Je ne sais pas	I don't know	97	ActHandel_2
	Geen antwoord	Pas de réponse	No answer	98	
	ALS SrtHandel=1 DAN: Om wat voor soort detailhandel gaat het precies en wat wordt verkocht?	SI SrtHandel=1 ALORS : De quel type de commerce de détail s'agit-il exactement et qu'est-ce qui est vendu ?	IF SrtHandel=1: What type of retail trade is it exactly, and what is sold?		
	Probeer dit zo precies mogelijk te omschrijven. Bijv. dameskledingwinkel, webwinkel in sieraden, bakker, groente- en fruit marktkraam, supermarkt, benzinstation, autodealer in tweedehandsauto's en lichte bestelwagens, verkoop vrachtwagens aan particulieren.	Essayez de le décrire aussi précisément que possible. Par exemple : magasin de vêtements pour femmes, bijouterie en ligne, boulangerie, étalage de fruits et légumes, supermarché, station-service, concessionnaire de voitures et de véhicules utilitaires légers d'occasion, vente de camions aux particuliers.	Please try to describe this as precisely as possible. For example: women's clothing store, webshop for jewellery, bakery, fruit or vegetable market stall, supermarket, petrol station, dealer in used cars and second hand vans, retail sale of trucks.		
ActHandel_2	ALS SrtHandel=2 DAN: Wat voor producten worden verkocht?	SI SrtHandel=2 ALORS : Quel type de produits sont vendus ?	IF SrtHandel=2: What type of products are sold?		
	Probeer dit zo precies mogelijk te omschrijven. Bijv. bouwmaterialen, levensmiddelen, witgoed, personenwagens en lichte bestelwagens, vrachtwagens, landbouwwerktuigen.	Essayez de le décrire aussi précisément que possible. Par exemple : matériaux de construction, denrées alimentaires, produits blancs, voitures particulières et véhicules utilitaires légers, camions, matériel agricole.	Please try to describe this as precisely as possible. For example: building materials, foodstuffs, white goods, passenger cars and light vans, trucks, agricultural implements.		
ActHandel_3	ALS SrtHandel=3 DAN: Om wat voor soort handelsbemiddeling gaat het precies?	SI SrtHandel=3 ALORS : De quel type d'intermédiation commerciale s'agit-il exactement ?	IF SrtHandel=3: What type of trade intermediation is it exactly?		
	Probeer dit zo precies mogelijk te omschrijven. Bijv. als tussenpersoon optreden tussen kopers en verkopers van brandstoffen, bemiddelen bij handelstransacties met betrekking tot bouwmaterialen, machines, huishoudelijke artikelen, voedingsmiddelen.	Essayez de le décrire aussi précisément que possible. Par exemple : agir en tant qu'intermédiaire entre les acheteurs et les vendeurs de combustibles, médiateur dans les transactions commerciales concernant des matériaux de construction, des machines, des articles ménagers, des denrées alimentaires.	Please try to describe this as precisely as possible. For example: acting as an intermediary between buyers and sellers of fuel, mediating commercial transactions involving building materials, machinery, household goods, food products.		
SrtHoreca	Open veld (max 300 karakters)	Champ libre (max 300 caractères)	Open field (max 300 characters)		Job2
	Gaat het dan om een:	S'agit-il d'un :	Is it a(n):		
	Eet- en/of drinkgelegheid (café, restaurant, snackbar)	Établissement de restauration (café, restaurant, snack-bar)	Eating and/or drinking establishment (cafe, restaurant, snack bar)	1	ActHoreca_1
	Cateringbedrijf / kantine	Traiteur / cantine	Catering company / canteen	2	ActHoreca_1
	Hotel, B&B, gastenverblijf	Hôtel, chambre d'hôtes, pension de famille	Hotel, Bed-and-Breakfast, guest house	3	ActHoreca_2
	Camping, bungalowpark, vakantiepark	Camping, parc de bungalows, parc de vacances	Campground, bungalow park, holiday park	4	ActHoreca_2
	Verhuurbedrijf van vakantiewoningen	Rental company de maisons de vacances	Rental company for vacation homes	5	ActHoreca_2
	Sport	Sports	Sports	6	ActHoreca_3

ActHoreca_1	Attractie- of pretpark	Parc d'attractions ou d'aventures	Theme park or amusement park	7	ActHoreca_4 ActHoreca_4 ActHoreca_4 ActHoreca_5 ActHoreca_5
	Bioscoop	Cinéma	Cinema	8	
	Gokhal, casino of loterij	Salle de jeux, casino ou loterie	Gambling hall, casino or lottery	9	
	Andere	Autre	Other	10	
	Ik weet het niet	Je ne sais pas	I don't know	97	
	Geen antwoord	Pas de réponse	No answer	98	
	ALS SrtHoreca=1/2 DAN: Om wat voor soort bedrijf gaat het dan precies? Probeer dit zo precies mogelijk te omschrijven. Bijv. restaurant, café, fastfoodzaak, ijssalon, bedrijfskantine, discotheek.	SI SrtHoreca=1/2 ALORS : De quel type d'entreprise s'agit-il exactement ? Essayez de le décrire le plus précisément possible. Par exemple, un restaurant, un café, un fast-food, un glacier, une cantine d'entreprise, une discothèque.	IF SrtHoreca=1/2: What type of company is it exactly? Please try to describe this as precisely as possible. For example: restaurant, bistro, fast-food outlet, ice-cream parlour, company canteen, discotheque.		
ActHoreca_2	ALS SrtHoreca=3/4/5 DAN: Om wat voor soort bedrijf gaat het dan precies? Probeer dit zo precies mogelijk te omschrijven. Bijv. hotel, pension voor buitenlandse werknemers, jeugdherberg, gîte, chaletpark, B&B.	SI SrtHoreca=3/4/5 ALORS : De quel type d'entreprise s'agit-il exactement ? Essayez de le décrire le plus précisément possible. Par exemple, un hôtel, un foyer pour travailleurs saisonniers, une auberge de jeunesse, un gîte, un parc de chalets, une chambre d'hôtes.	IF SrtHoreca=3/4/5: What type of company is it exactly? Please try to describe this as precisely as possible. For example: hotel, boarding house for foreign workers, youth hostel, gîte, chalet park, B&B.		
ActHoreca_3	ALS SrtHoreca=6 DAN: Om wat voor sport(voorziening) gaat het of wat zijn de belangrijkste activiteiten? Probeer dit zo precies mogelijk te omschrijven. Bijv. zwembad, fitnesscentrum, verzorgen van vistochten, tennisbond, professioneel schaker, voetbalclub, verhuur van fietsen.	SI SrtHoreca=6 ALORS : De quel type de sport (installation) s'agit-il ou quelles sont ses principales activités ? Essayez de le décrire le plus précisément possible. Par exemple : piscine, centre de fitness, organisation d'excursions de pêche, association de tennis, joueur d'échecs professionnel, club de foot, location de vélos.	IF SrtHoreca=6: What type of sport (or sports facility) is it, or what are the most important activities? Please try to describe this as precisely as possible. For example: swimming pool, fitness centre, fishing trips, tennis club, professional chess player, football club, bike rental.		
ActHoreca_4	ALS SrtHoreca=7/8/9 DAN: Om wat voor soort bedrijf gaat het dan precies? Probeer dit zo precies mogelijk te omschrijven. Bijv. uitbaten van gokautomaten, kermisattracties, (water)pretparken, binnenspeeltuinen, bioscoopzalen.	SI SrtHoreca=7/8/9 ALORS : De quel type d'entreprise s'agit-il exactement ? Essayez de le décrire le plus précisément possible. Par exemple : l'exploitation de machines à sous, d'attractions foraines, de parc d'attractions (aquatiques), d'aire de jeux intérieure, de salles de cinéma.	IF SrtHoreca=7/8/9: What type of company is it exactly? Please try to describe this as precisely as possible. For example: operating slot machines, fairground attractions, (water) amusement parks, indoor playgrounds, cinema halls.		
ActHoreca_5	ALS SrtHoreca=10/97 DAN: Om wat voor soort bedrijf gaat het dan precies? Probeer dit zo precies mogelijk te omschrijven. Bijv. uitbaten van van snooker- en biljartenzalen, escape rooms. Open veld (max 300 karakters) Gaat het dan om een: Ziekenhuis Medische praktijk (bijv. huisarts, specialist, tandarts)	SI SrtHoreca=10 ALORS : De quel type d'entreprise s'agit-il exactement ? Essayez de le décrire le plus précisément possible. Par exemple : l'exploitation de salles de billard et de snooker, de salles d'évasion. Champ libre (max 300 caractères) S'agit-il d'un(e) : Hôpital Cabinet médical (par exemple, médecin généraliste, spécialiste, dentiste)	IF SrtHoreca=10: What type of company is it exactly? Please try to describe this as precisely as possible. For example: operation of snooker and billiards halls, escape rooms. Open field (max 300 characters) Is it a(n): Hospital Medical practice (e.g. general practitioner, specialist, dentist)		
SrtZorg	Para-medische praktijk (bijv. verloskundige, psycholoog, fysiotherapeut)	Cabinet paramédical (par exemple, obstétricien, psychologue, kinésithérapeute)	Paramedical practice (e.g. midwife, psychologist, physiotherapist)	1	Job2 ActZorg_1 ActZorg_1 ActZorg_1 ActZorg_2 ActZorg_2 ActZorg_2 ActZorg_2 ActZorg_2 ActZorg_2
	Begeleid wonen, verzorgingshuis, verpleeghuis	Maison de retraite, maison de soins, maison de repos	Assisted living, care home, nursing home	2	
	Jeugdzorg	Aide à la jeunesse	Youth care	3	
	Maatschappelijk werk	Travail social	Social services	4	
	Thuiszorg	Soins à domicile	Home care	5	
	Kinderopvang	Garde d'enfants	Childcare	6	
	Andere	Autre	Other	7	
	Ik weet het niet	Je ne sais pas	I don't know	8	
	Geen antwoord	Pas de réponse	No answer	9	
	ALS SrtZorg=1/2/3 DAN: Om wat voor soort gezondheids- of zorginstelling gaat het precies?	SI SrtZorg=1/2/3 ALORS : De quel type d'établissement de santé ou de soins s'agit-il exactement ?	IF SrtZorg=1/2/3: What type of health or care institution is it exactly?	97	
ActZorg_1				98	

	Probeer dit zo precies mogelijk te omschrijven. Bijv. huisarts, academisch ziekenhuis, psychiatrische kliniek, verloskundige praktijk, sport fysiotherapie.	Essayez de le décrire aussi précisément que possible. Par exemple, médecin généraliste, hôpital universitaire, clinique psychiatrique, cabinet d'obstétrique, physiothérapie sportive.	Please try to describe this as precisely as possible. For example: general practitioner, university hospital, psychiatric clinic, obstetric practice, sports physiotherapy.		
ActZorg_2	ALS SrtZorg=4/5/6/7/8/9/97 DAN: Om wat voor soort instelling gaat het precies en wie zijn de cliënten?	SI SrtZorg=4/5/6/7/8/9 ALORS : De quel type d'institution s'agit-il exactement et qui sont les clients ?	IF SrtZorg=4/5/6/7/8/9: What type of institution is it exactly, and who are the clients?		
	Probeer dit zo precies mogelijk te omschrijven. Bijv. dagopvang voor mensen met een verstandelijke beperking, zorgboerderij voor jongeren met autisme, opvanghuis voor slachtoffers van geweld, buurthuis, slachtofferhulplijn.	Essayez de décrire cela aussi précisément que possible. Par exemple : centre de jour pour personnes handicapées mentales, ferme d'accueil pour jeunes autistes, centre d'hébergement pour victimes de violence, centre communautaire, service d'assistance téléphonique aux victimes.	Please try to describe this as precisely as possible. For example: day care centre for people with mental disabilities, care farm for young people with autism, shelter for victims of violence, community centre, victim helpline.		
Zorg_Overn	Open veld (max 300 karakters) Kunnen cliënten overnachten in de instelling waar u werkt? Ja Neen (bijv. alleen dagbesteding, ambulante zorg etc.)	Champ libre (max 300 caractères) Les clients peuvent-ils passer la nuit dans l'établissement où vous travaillez ? Oui Non (par exemple, uniquement les soins de jour, les soins ambulatoires, etc.)	Open field (max 300 characters) Can clients spend the night at the institution where you work? Yes No (e.g. only day activities, outpatient care)		Zorg_Overn
	Ik weet het niet Geen antwoord Gaat het dan om: Kleuteronderwijs (inclusief buitengewoon onderwijs)	Je ne sais pas Pas de réponse S'agit-il de : Enseignement maternel (y compris l'enseignement spécial)	I don't know No answer Does it involve: Preschool education (including special education)	1 2	Job2 Job2
SrtOnderw	Basisonderwijs (inclusief buitengewoon onderwijs)	Enseignement primaire (y compris l'enseignement spécial)	Primary education (including special education)	1	TypOnderw
	Secundair onderwijs Hoger onderwijs Ander soort onderwijs (bijv. sport, muziek, taal, autorijles)	Enseignement secondaire Enseignement supérieur Autre type d'éducation (par exemple, sport, musique, langue, leçons de conduite)	Secondary education Higher education Other type of education (e.g. sports, music, language, driving lessons)	3 4 5	TypSec BestuurHoger ActOnderw
	Andere (bijv. dienstverlening voor het onderwijs) Ik weet het niet Geen antwoord Wat zijn de belangrijkste activiteiten van deze onderwijsinstelling?	Autre (par exemple services pour l'éducation) Je ne sais pas Pas de réponse Quelles sont les principales activités de cet établissement d'enseignement ?	Other (e.g. education-related services) I don't know No answer What are the most important activities of this educational institution?	6 97 98	ActOnderw ActOnderw
ActOnderw	Probeer dit zo precies mogelijk te omschrijven. Bijv. muziekles voor kinderen, autorijles, onderwijsinspectie hoger onderwijs.	Essayez de le décrire le plus précisément possible. Par exemple, cours de musique pour enfants, cours de conduite automobile, inspection de l'enseignement supérieur.	Please try to describe this as precisely as possible. For example: music lessons for children, driving lessons, educational inspection for higher education.		
TypSec	Open veld (max 300 karakters) Om welke onderwijsvorm gaat het dan precies?	Champ libre (max 300 caractères) De quel type d'enseignement secondaire s'agit-il exactement ?	Open field (max 300 characters) What type of secondary education is it exactly?		Job2
	Meerdere antwoorden mogelijk Algemeen secundair onderwijs (ASO) Kunst secundair onderwijs (KSO) Technisch secundair onderwijs (TSO) Beroepssecundair onderwijs (BSO) Ik weet het niet Geen antwoord	Plusieurs réponses possibles Enseignement secondaire général Enseignement secondaire artistique Enseignement secondaire technique Enseignement secondaire professionnel Je ne sais pas Pas de réponse	Multiple answers possible General education Arts education Technical education Vocational education I don't know No answer	1 2 3 4 97 98	TypPostSec TypPostSec TypPostSec TypPostSec
TypPostSec	Gaat het om een onderwijsinstelling waar uitsluitend post-secundair niet-hoger onderwijs (ondernemersopleiding, zevende jaar, 4° graad BSO) wordt aangeboden?	S'agit-il d'un établissement d'enseignement offrant exclusivement un enseignement post-secondaire non supérieur (formation de chef d'entreprise, 7ème année, 4ème degré de l'enseignement secondaire professionnel) ?	Is it an educational institution exclusively offering post-secondary non-higher education (entrepreneurial training, 7th year, 4th degree of vocational secondary education)?		
	Ja Neen Ik weet het niet Geen antwoord	Oui Non Je ne sais pas Pas de réponse	Yes No I don't know No answer	1 2 97 98	TypOnderw TypOnderw

TypOnderw	Gaat het om een onderwijsinstelling waar buitengewoon onderwijs wordt aangeboden?	S'agit-il d'un établissement d'enseignement offrant l'enseignement spécial ?	Is it an educational institution where special education is offered?		
BestuurBuiteng	Ja	Oui	Yes	1	BestuurBuiteng BestuurGewoon BestuurGewoon
	Neen	Non	No	2	
	Ik weet het niet	Je ne sais pas	I don't know	97	
	Tot welk onderwijsnet behoort de onderwijsinstelling?	À quel réseau d'enseignement appartient cet établissement d'enseignement ?	To which educational network does the educational institution belong?		
BestuurGewoon	Officieel gesubsidieerd	Organisé par les pouvoirs publics	Official subsidised	1	Job2 Job2 Job2 Job2
	Vrij gesubsidieerd	Libre subventionné	Private subsidised	2	
	Andere	Autre	Other	3	
	Ik weet het niet	Je ne sais pas	I don't know	97	
	Tot welk onderwijsnet behoort de onderwijsinstelling?	À quel réseau d'enseignement appartient cet établissement d'enseignement ?	To which educational network does the educational institution belong?		
	Ingericht door de Gemeenschappen	Organisé par les Communautés	Organised by the Communities	1	Job2 Job2 Job2 Job2 Job2 Job2
BestuurHoger	Provinciaal gesubsidieerd	Subventionné organisé par les provinces	Subsidised by the province	2	
	Gemeentelijk gesubsidieerd	Subventionné organisé par les communes	Subsidised by the commune	3	
	Vrij gesubsidieerd	Libre subventionné	Private subsidised	4	
	Andere	Autre	Other	5	
	Ik weet het niet	Je ne sais pas	I don't know	97	
	Tot welk onderwijsnet behoort de onderwijsinstelling?	À quel réseau d'enseignement appartient cet établissement d'enseignement ?	To which educational network does the educational institution belong?		
ActOverh	Officieel gesubsidieerd	Organisé par les pouvoirs publics	Official subsidised	1	Job2 Job2 Job2 Job2
	Vrij gesubsidieerd	Libre subventionné	Private subsidised	2	
	Andere	Autre	Other	3	
	Ik weet het niet	Je ne sais pas	I don't know	97	
SrtIT	Om wat voor soort overheidsinstelling gaat het?	De quel type d'organisme public s'agit-il ?	What type of governmental institution is it?		
	Bijv. politie, brandweer, gemeente, provincie, FOD Binnenlandse zaken, Defensie, Rekenhof.	Par exemple, la police, les pompiers, la municipalité, la province, SPF Intérieur, la Défense, la Cour des comptes.	For example: police, fire services, municipality, provincial government, FPS Home Affairs, Defence, Court of Audit.		
	Open veld (max 300 karakters)	Champ libre (max 300 caractères)	Open field (max 300 characters)		Job2
	Gaat het dan om een:	S'agit-il d'une :	Is it a(n):		
ActiT	IT-bedrijf (bijv. ontwikkelen software, advies, netwerkbeheer)	Entreprise informatique (par exemple, développement de logiciels, consultance, gestion de réseaux)	IT company (e.g. software development, consultancy, network management)	1	ActiT
	Telecommunicatie bedrijf (bijv. telefoon- of internetaanbieder)	Entreprise de télécommunications (par exemple, fournisseur de services téléphoniques ou d'internet)	Telecommunications company (e.g. telephone or internet provider)	2	ActiT
	Uitgeverij (bijv. van boeken, kranten, computerspellen)	Maison d'édition (par exemple, de livres, de journaux, de jeux informatiques)	Publisher (e.g. books, newspapers, computer games)	3	ActiT
	Radio- of tv-omroep	Radiodiffusion ou télédiffusion	Radio or television broadcaster	4	ActiT
ActiT	Ander media bedrijf (bijv. productie van film- of tv-programma's)	Autre société de médias (par exemple, production de films ou de programmes télévisés)	Other media company (e.g. production of film or television programmes)	5	ActiT
	Andere	Autre	Other	6	ActiT
	Ik weet het niet	Je ne sais pas	I don't know	97	ActiT
	Geen antwoord	Pas de réponse	No answer	98	
SrtFinan	Wat zijn de belangrijkste activiteiten van dit bedrijf?	Quelles sont les principales activités de cette entreprise ?	What are the most important activities of this company?		
	Probeer dit zo precies mogelijk te omschrijven. Bijv. ontwikkelen van games, webhosting, verkoop van telefoonabonnementen, uitgeven van kranten.	Essayez de les décrire le plus précisément possible. Par exemple, développement de jeux, hébergement web, vente d'abonnements téléphoniques, publication de journaux.	Please try to describe this as precisely as possible. For example: game development, webhosting, telephone subscription sales, newspaper publishing.		
	Open veld (max 300 karakters)	Champ libre (max 300 caractères)	Open field (max 300 characters)		Job2
	Gaat het dan om een:	S'agit-il d'un(e) :	Is it a(n):		
SrtFinan	Bank of andere kredietverstrekker	Banque ou autre prêteur	Bank or other credit provider	1	ActFinan
	Beleggingsinstelling of financiële holding	Établissement d'investissement ou compagnie financière holding	Investment institution or financial holding company	2	ActFinan
	Verzekeraar of pensioenfonds	Assureur ou fonds de pension	Insurer or pension fund	3	ActFinan
	Administratie- of accountantskantoor of belastingconsulent	Cabinet d'administration ou de comptabilité ou conseiller fiscal	Administration or accounting firm; tax consultancy	4	ActFinan

ActFinan	Ander adviesbureau of tussenpersoon (bijv. verzekeringen, vermogen, aandelen, hypotheken)	Autre société de conseil ou intermédiaire (par exemple, assurance, capital, actions, prêts hypothécaires)	Other consultancy agency or intermediary (e.g. insurance, assets, shares, mortgages)	5	ActFinan
	Immobiënkantoor (bijv. makelaar, kamerverhuur, syndicus)	Agence immobilière (par exemple, agent immobilier, location de chambres, syndic)	Real estate agency (e.g. broker, room rental agency, syndic)	6	ActFinan
	Andere	Autre	Other	7	ActFinan
	Ik weet het niet	Je ne sais pas	I don't know	97	ActFinan
	Geen antwoord	Pas de réponse	No answer	98	
	Om wat voor soort bedrijf gaat het precies of wat zijn de belangrijkste activiteiten? Probeer dit zo precies mogelijk te omschrijven. Bijv. algemene bank, pensioenfonds, hypotheekadviseur, sociaal verhuurkantoor.	De quel type d'entreprise s'agit-il exactement ou quelles sont ses principales activités ? Essayez de le décrire aussi précisément que possible. Par exemple, banque générale, fonds de pension, courtier en prêts hypothécaires, agence immobilière sociale.	What type of company is it, or what are the most important activities? Please try to describe this as precisely as possible. For example: general bank, pension fund, mortgage consultancy, housing corporation for social rental housing.		
SrtAnders	Open veld (max 300 karakters)	Champ libre (max 300 caractères)	Open field (max 300 characters)		Job2
	Gaat het dan misschien om dienstverlening op het gebied van:	S'agit-il peut-être de services dans le domaine du :	Does it perhaps involve services relating to:		
	Arbeid (bijv. uitzendbureau of arbeidsbemiddeling)	Travail (par exemple, agence d'itérim ou services en matière d'emploi)	Labour (e.g. temporary employment agency or job placement services)	1	ActAnders
	Recht (bijv. advocatenbureau, notaris of deurwaarder)	Droit (par exemple, cabinet d'avocats, notaire ou huissier)	Law (e.g. law firm, notary or bailiff)	2	ActAnders
	Politieke-, levensbeschouwelijke of belangenorganisaties (bijv. politieke partij, kerk, goed doel)	Organisations politiques, organisations philosophiques ou groupes d'intérêt (par exemple, parti politique, église, association caritative)	Political, ideological or special-interest organisations (e.g. political party, church, charity)	3	ActAnders
	Management of organisatie (bijv. adviesbureau)	Gestion ou organisation (par exemple, société de conseil)	Business and other management consultancy activities	4	ActAnders
	Verhuur (bijv. auto's, kleding, sportartikelen, machines)	Location (par exemple, voitures, vêtements, équipements sportifs, machines)	Rental or leasing (e.g. cars, clothing, sporting goods, machinery)	5	ActAnders
	Reizen (bijv. reisbureau, touroperator)	Voyages (par exemple, agence de voyage, tour-opérateur)	Travel (e.g. travel agency, tour operator)	6	ActAnders
	Dierenhulp (bijv. dierenarts, dierenambulance)	Assistance aux animaux (par exemple, vétérinaire, ambulance pour animaux)	Animal assistance (e.g. veterinarian, animal ambulance)	7	ActAnders
	Bibliotheken, dierentuinen en musea	Bibliothèques, zoos et musées	Library, zoo or museum	8	ActAnders
	Grafische vormgeving of industrieel ontwerp	Graphisme ou design industriel	Graphic design or industrial design	9	ActAnders
	Kunst	Art	Art	10	ActAnders
	Reclame of marktonderzoek	Publicité ou études de marché	Advertising or market research	11	ActAnders
	Keuringen of kwaliteitscontrole	Inspection ou contrôle de la qualité	Inspections or quality control	12	ActAnders
	Ander onderzoek (laboratorium, technisch, wetenschappelijk)	Autres recherches (laboratoires, techniques, scientifiques)	Other research (laboratory, technical, scientific)	13	ActAnders
	Beveiliging	Sécurité	Security	14	ActAnders
	Schoonmaak of tuinonderhoud (bijv. schoonmaakbedrijf, glazenwassers, hovenier)	Nettoyage ou entretien de jardins (par exemple, entreprise de nettoyage, laveurs de vitres, jardiniers)	Cleaning or grounds maintenance (e.g. cleaning company, window cleaner, gardener)	15	ActAnders
	Persoonlijke verzorging (bijv. kapper, schoonheidsspecialiste, masseur)	Soins personnels (par exemple, coiffeur, esthéticien, masseur)	Personal care (e.g. hairdresser, beautician, masseur)	16	ActAnders
	Andere	Autre	Other	17	SrtAnders_A
SrtAnders_A	Ik weet het niet	Je ne sais pas	I don't know	97	SrtAnders_A
	Geen antwoord	Pas de réponse	No answer	98	
	Om wat voor soort bedrijf gaat het dan?	De quel type d'entreprise s'agit-il ?	What type of company is it?		
ActAnders	Open veld (max 300 karakters)	Champ libre (max 300 caractères)	Open field (max 300 characters)		ActAnders
	Waarin is dit bedrijf gespecialiseerd? Probeer dit zo precies mogelijk te omschrijven. Bijv. juridisch advies geven op het gebied van werk, organiseren van reizen naar Azië, vormgeven van reclamefolders, verhuren van kleding, beveiligen van musea, particulier tuinonderhoud.	Quelle est la spécialité de cette entreprise ? Essayez de le décrire le plus précisément possible. Par exemple : conseils juridiques en matière de travail, organisation de voyages en Asie, conception de dépliants publicitaires, location de vêtements, sécurisation de musées, entretien de jardins privés.	What is this company's specialisation? Please try to describe this as precisely as possible. For example: providing labour-related legal advice, organising trips to Asia, designing advertising brochures, renting out clothes, providing security services to museums, private garden maintenance.		
	Open veld (max 300 karakters)	Champ libre (max 300 caractères)	Open field (max 300 characters)		Job2

Job2	IND5	<p>Wat is uw functie in dit bedrijf/deze organisatie? Wees zo precies mogelijk. Noteer bijvoorbeeld 'secretaresse' in plaats van 'medewerker', 'timmerman' in plaats van 'handarbeider', 'leraar op een middelbare school', enz.</p> <p>Deze vraag beantwoorden voor de job (of onderneming) waaraan u het grootste deel van uw werktijd besteedt als u meerdere jobs (of meerdere ondernemingen) hebt.</p>	<p>Quelle est votre fonction dans cette entreprise/organisation ? Soyez aussi précis que possible. Par exemple, notez "Secrétaire" au lieu de "Employé", "Charpentier" au lieu de "Travailleur manuel", "Enseignant à l'école secondaire", etc.</p> <p>Veillez répondre pour l'emploi (ou l'affaire) auquel vous consacrez le plus de temps si vous avez plusieurs emplois (ou plusieurs affaires).</p>	<p>What is your occupation in this business/ organisation? Please be as exact as possible. For example, note "Secretary" instead of "Employee", "Carpenter" instead of "Manual worker", "Teacher at secondary school", etc.</p> <p>Please answer this question for the job (or business) you spend most of your working time on if you have several jobs (or several businesses).</p>		
Job3	IND10_1	<p>Open veld (max 300 karakters)</p> <p>Hoeveel uren per week werkt u gewoonlijk in deze job?</p> <p>Als uw werkuren variëren van week tot week, gelieve dan een gemiddelde op te geven.</p> <p>Overuren meerekenen.</p> <p>Deze vraag beantwoorden voor de job (of onderneming) waaraan u het grootste deel van uw werktijd besteedt als u meerdere jobs (of meerdere ondernemingen) hebt.</p> <p>Aantal uren (0-168)</p>	<p>Champ libre (max 300 caractères)</p> <p>Pendant combien d'heures par semaine travaillez-vous habituellement dans cet emploi ?</p> <p>Si vos heures de travail varient d'une semaine à l'autre, merci de faire une moyenne.</p> <p>Tenez compte des heures supplémentaires.</p> <p>Veillez répondre pour l'emploi (ou l'affaire) auquel vous consacrez le plus de temps si vous avez plusieurs emplois (ou plusieurs affaires).</p> <p>Nombre d'heures (0-168)</p>	<p>Open field (max 300 characters)</p> <p>How many hours per week do you usually work in this job?</p> <p>If your working schedule varies considerably from week to week, please give an average value.</p> <p>Include overtime hours.</p> <p>Please answer this question for the job (or business) you spend most of your working time on if you have several jobs (or several businesses).</p> <p>Number of hours (0-168)</p>		Job3
Job5	IND44	<p>Geen antwoord</p> <p>Hebt u een vaste job of een contract van onbepaalde duur?</p> <p>Deze vraag beantwoorden voor de job (of onderneming) waaraan u het grootste deel van uw werktijd besteedt als u meerdere jobs (of meerdere ondernemingen) hebt.</p> <p>Ja, een vaste job of een contract van onbepaalde duur</p> <p>Neen, een tijdelijke job of een contract van bepaalde duur of voor een bepaald werk</p> <p>Geen antwoord</p>	<p>Pas de réponse</p> <p>Avez-vous un emploi permanent ou un contrat à durée indéterminée ?</p> <p>Veillez répondre pour l'emploi (ou l'affaire) auquel vous consacrez le plus de temps si vous avez plusieurs emplois (ou plusieurs affaires).</p> <p>Oui, un emploi permanent ou un contrat à durée indéterminée</p> <p>Non, un emploi temporaire ou un contrat à durée déterminée ou pour un travail déterminé</p> <p>Pas de réponse</p>	<p>No answer</p> <p>Do you have a permanent job or an open-ended work contract?</p> <p>Please answer this question for the job (or business) you spend most of your working time on if you have several jobs (or several businesses).</p> <p>Yes, a permanent job or open-ended work contract</p> <p>No, it is a temporary job or fixed-term contract or for specific work</p> <p>No answer</p>	98	
Job6	IND7	<p>Werkt u voltijds of deeltijds in deze job?</p> <p>Deze vraag beantwoorden voor de job (of onderneming) waaraan u het grootste deel van uw werktijd besteedt als u meerdere jobs (of meerdere ondernemingen) hebt.</p> <p>Voltijds</p> <p>Deeltijds</p> <p>Geen antwoord</p>	<p>Travaillez-vous à temps plein ou à temps partiel dans cet emploi ?</p> <p>Veillez répondre pour l'emploi (ou l'affaire) auquel vous consacrez le plus de temps si vous avez plusieurs emplois (ou plusieurs affaires).</p> <p>À temps plein</p> <p>À temps partiel</p> <p>Pas de réponse</p>	<p>In this job, do you work full-time or part-time?</p> <p>Please answer this question for the job (or business) you spend most of your working time on if you have several jobs (or several businesses).</p> <p>Full-time</p> <p>Part-time</p> <p>No answer</p>	98	
Job7	IND9	<p>Hoeveel betaalde vakantiedagen hebt u per jaar?</p> <p>De normale één of twee vrije dagen per week (in het weekend of door de week) worden niet beschouwd als betaalde vakantie.</p> <p>Deze vraag beantwoorden voor de job (of onderneming) waaraan u het grootste deel van uw werktijd besteedt als u meerdere jobs (of meerdere ondernemingen) hebt.</p> <p>Aantal dagen (0-100)</p> <p>Geen antwoord</p>	<p>Combien de jours de congés payés avez-vous par an ?</p> <p>Le congé normal d'un ou deux jours par semaine (le week-end ou en semaine) n'est pas considéré comme un congé payé.</p> <p>Veillez répondre pour l'emploi (ou l'affaire) auquel vous consacrez le plus de temps si vous avez plusieurs emplois (ou plusieurs affaires).</p> <p>Nombre de jours (0-100)</p> <p>Pas de réponse</p>	<p>How many days of paid holiday do you have a year?</p> <p>The normal one or two days off a week (either on weekend days or weekdays) are not regarded as contracted paid annual leave.</p> <p>Please answer this question for the job (or business) you spend most of your working time on if you have several jobs (or several businesses).</p> <p>No answer</p>	98	
Job8	IND14	<p>Hebt u, naast deze job, één of meerdere andere betaalde jobs?</p> <p>Ja</p> <p>Neen</p>	<p>En dehors de cet emploi, avez-vous un ou plusieurs autres emplois rémunérés ?</p> <p>Oui</p> <p>Non</p>	<p>In addition to this job, do you have any other paid job(s)?</p> <p>Yes</p> <p>No</p>	98	Job8
					1	Job9
					2	Activity6

IF Job4 = 1/2/3/4 THEN Job5
ELSE IF Job4 = 5/6 THEN Job
8 ELSE IF Job4 = 7 THEN
Activity4

Job9	IND38	Geen antwoord	Pas de réponse	No answer	98	
		Hoeveel uren per week besteedt u gewoonlijk aan deze andere job(s)?	Combien d'heures par semaine consacrez-vous habituellement à cet (ces) autre(s) emploi(s) ?	How many hours per week do you usually spend on this other job(s)?		
		Als uw werkuren variëren van week tot week, gelieve dan een gemiddelde op te geven.	Si vos heures de travail varient d'une semaine à l'autre, merci de faire une moyenne.	If your working schedule varies considerably from week to week, please give an average value.		
		Overuren meerekenen.	Tenez compte des heures supplémentaires.	Include overtime hours.		
		Aantal uren (0-168)	Nombre d'heures (0-168)	Number of hours (0-168)		Activity6
Activity4	IND15	Geen antwoord	Pas de réponse	No answer	98	
		Hebt u de afgelopen 4 weken iets ondernomen om een job te vinden, eventueel deeltijds of tijdelijk, of om een eigen zaak op te starten?	Au cours des quatre dernières semaines, avez-vous fait quelque chose pour trouver un emploi, même à temps partiel ou occasionnel, ou pour mettre sur pied votre propre entreprise ?	During the last 4 weeks, did you do anything to find a job, even part-time or occasional work, or to set up your own business?		
		Ja	Oui	Yes	1	Activity5
		Neen, ik heb al een job gevonden die start binnen nu en 3 maanden	Non, j'ai déjà trouvé un emploi qui commence dans les 3 mois	No, I already found a job that will start within 3 months	2	Activity6
		Neen	Non	No	3	Activity6
Activity5	IND16	Geen antwoord	Pas de réponse	Not applicable	98	
		Als u vandaag een job zou aangeboden worden, zou u dan binnen de twee weken kunnen beginnen werken?	Si on vous proposait un emploi aujourd'hui, pourriez-vous commencer à travailler dans un délai de deux semaines ?	If you were offered a job today, could you start working within two weeks?		
		Ja	Oui	Yes	1	Activity6
		Neen	Non	No	2	Activity6
Activity6	DDV6	Geen antwoord	Pas de réponse	No answer	98	
		Met welk van de hierna vermelde situaties komt uw (werk)situatie het best overeen?	Parmi les situations reprises ci-dessous, laquelle correspond le mieux à votre situation (professionnelle) ?	Which of the following situations corresponds best to your (work) situation ?		
		Ik heb een job	J'ai un emploi	I have a job	1	Comm_4
		Ik ben werkloos	Je suis au chômage	I am unemployed	2	Comm_4
		Ik ben (vervroegd) gepensioneerd	Je suis à la retraite (anticipée)	I am retired or in early retirement	3	Comm_4
		Ik ben arbeidsongeschikt als gevolg van langdurige gezondheidsproblemen	Je suis en incapacité de travail en raison de problèmes de santé de longue durée	I am unable to work due to long-term health problems	4	Comm_4
		Ik ben leerling, student of in beroepsopleiding	Je suis élève, étudiant(e) ou en formation professionnelle	I am a pupil, a student or in vocational training	5	Comm_4
		Ik verzorg het huishouden (huisvrouw/huisman)	Je m'occupe du ménage (femme/homme au foyer)	I take care of the household (housewife/househusband)	6	Comm_4
		Ik zit in een andere situatie zonder werk	Je suis dans une autre situation sans emploi	I am in another situation without work	7	Comm_4
Comm_4		Geen antwoord	Pas de réponse	No answer	98	
Comm_4		Nu volgen enkele vragen over uw schooltijd en uw studies.	Les questions suivantes portent sur votre scolarité et vos études.	Next are some questions about your school days and your studies.		Educ1
Educ1		Volgt u momenteel een opleiding, hetzij voltijds hetzij deeltijds?	Suivez-vous actuellement des études, à temps plein ou à temps partiel ?	Do you currently follow an education, full-time or part-time?		
		Ja	Oui	Yes	1	Educ2
		Neen	Non	No	2	Educ5
Educ2	IND20	Geen antwoord	Pas de réponse	No answer	98	
		Welk onderwijs volgt u momenteel?	Quel type d'enseignement suivez-vous actuellement ?	What education do you currently follow?		
		Lager onderwijs	Enseignement primaire	Primary education	1	Educ5
		Secundair onderwijs	Enseignement secondaire	Secondary education	2	Educ3
		Hoger onderwijs (hogeschool, universiteit of HBO5)	Enseignement supérieur (haute école, université ou BES)	Higher education (college for higher education, university or HBO5/BES)	3	Educ4
		Andere	Autre	Other	4	Educ5
Educ3	IND20	Geen antwoord	Pas de réponse	No answer	98	
		Gaat het om ...	Il s'agit de ...	Is it ...		
		Lager secundair onderwijs (1ste graad)	Enseignement secondaire inférieur du premier degré	Lower secondary education of the 1st degree	1	Educ5
		Algemeen secundair onderwijs 2de of 3de graad (ASO)	Enseignement secondaire général du 2e ou 3e degré	General secondary education of the 2nd or 3rd degree	2	Educ5
		Technisch onderwijs 2de of 3de graad (TSO)	Enseignement secondaire technique du 2e ou 3e degré	Technical secondary education of the 2nd or 3rd degree	3	Educ5

Educ4	IND20	Kunstsecundair onderwijs 2de of 3de graad (KSO)	Enseignement secondaire artistique du 2e ou 3e degré	Artistic secondary education of the 2nd or 3rd degree	4	Educ5
		Beroepssecundair onderwijs 2de of 3de graad (BSO)	Enseignement secondaire professionnel du 2e ou 3e degré	Vocational secondary education of the 2nd or 3rd degree	5	Educ5
		Postsecundair niet-hoger onderwijs	Enseignement post-secondaire non supérieur	Post-secondary non-higher education	6	Educ5
		Geen antwoord	Pas de réponse	No answer	98	
		Gaat het om ...	Il s'agit ...	Is it ...		
		Hoger Beroepsonderwijs (HBO5) of Brevet d'enseignement supérieur (BES) in het Franstalig onderwijs	Brevet d'enseignement supérieur (BES) ou Hoger Beroepsonderwijs (HBO5) dans l'enseignement néerlandophone	Higher professional education in adult education (HBO5/BES)	1	Educ5
		Hogescholenonderwijs van het korte type (1 cyclus) of graduaat (A1) of professionele bachelor	Enseignement en haute école de type court (1 cycle) ou graduat (A1) ou bachelier professionnalisant	Short-type higher education (1 cycle) or graduate (A1) or professional bachelor	2	Educ5
		Academische bachelor (hogeschool of universiteit)	Bachelier académique (dans une haute école ou à l'université)	Academic bachelor (at a college for higher education or university)	3	Educ5
		Voortgezette of aanvullende opleiding na graduaat of na bachelor (specialisatie, BanaBa, ...)	Formations complémentaires post-graduat ou post-bachelier (spécialisation, bachelier complémentaire, ...)	Advanced or additional training after graduate or bachelor's degree (specialisation, bachelor after bachelor, ...)	4	Educ5
		Schakelprogramma tussen professionele bachelor en master	Année préparatoire entre bachelier professionnel et master	Pre-master programme to bridge the gap between professional bachelor and master programmes	5	Educ5
Educ5	IND22_1	Master (hogeschool of universiteit)	Master (dans une haute école ou à l'université)	Master (at a college for higher education or university)	6	Educ5
		Voortgezette of aanvullende opleiding na master (specialisatie, ManaMa, ...)	Formations complémentaires post-master (spécialisation, master complémentaire, ...)	Advanced or additional training after master's degree (specialisation, master after master, ...)	7	Educ5
		Doctoraat met proefschrift	Doctorat avec thèse	Doctor's degree with doctor's thesis	8	Educ5
		Geen antwoord	Pas de réponse	No answer	98	
		Wat is het hoogste onderwijsniveau waarvoor u een diploma of certificaat hebt behaald?	Quel est le niveau d'études le plus élevé pour lequel vous avez obtenu un diplôme ou un certificat ?	What is the highest level of education for which you have obtained a diploma or certificate?		
		Iemand die een onderwijsniveau succesvol beëindigt, krijgt een diploma, getuigschrift of certificaat dat erkend wordt door de overheid. Noteer het hoogst behaalde niveau tot dusver indien u nog naar school gaat of momenteel een opleiding volgt.	Toute personne qui termine un niveau d'enseignement avec succès, reçoit un diplôme ou un certificat qui est reconnu par les autorités. Notez le niveau le plus élevé atteint jusqu'à présent si vous êtes actuellement encore scolarisé ou cours d'études.	A person who successfully completes a level of education receives a diploma or certificate acknowledged by the authorities. Note the highest level achieved so far if you are still attending school or currently in education.		
		Geen	Aucun	None	0	Comm_5
		Lager onderwijs	Enseignement primaire	Primary education	1	Comm_5
		Secundair onderwijs	Enseignement secondaire	Secondary education	2	Educ6
		Hoger onderwijs (hogeschool, universiteit of HBO5)	Enseignement supérieur (haute école, université ou BES)	Higher education (college for higher education, university or HBO5/BES)	3	Educ7
Educ6	IND22_1	Andere	Autre	Other	4	Educ8
		Ik weet het niet	Je ne sais pas	I don't know	97	Educ8
		Geen antwoord	Pas de réponse	No answer	98	
		Gaat het om ...	Il s'agit de ...	Is it ...		
		Lager secundair onderwijs (1ste graad)	Enseignement secondaire inférieur du premier degré	Lower secondary education of the 1st degree	1	Educ8
		Algemeen secundair onderwijs 2de of 3de graad (ASO)	Enseignement secondaire général du 2e ou 3e degré	General secondary education of the 2nd or 3rd degree	2	Educ8
		Technisch onderwijs 2de of 3de graad (TSO)	Enseignement secondaire technique du 2e ou 3e degré	Technical secondary education of the 2nd or 3rd degree	3	Educ8
		Kunstsecundair onderwijs 2de of 3de graad (KSO)	Enseignement secondaire artistique du 2e ou 3e degré	Artistic secondary education of the 2nd or 3rd degree	4	Educ8
		Beroepssecundair onderwijs 2de of 3de graad (BSO)	Enseignement secondaire professionnel du 2e ou 3e degré	Vocational secondary education of the 2nd or 3rd degree	5	Educ8
		Postsecundair niet-hoger onderwijs	Enseignement post-secondaire non supérieur	Post-secondary non-higher education	6	Educ8
Educ7	IND22_1	Ik weet het niet	Je ne sais pas	I don't know	97	Educ8
		Geen antwoord	Pas de réponse	No answer	98	
		Gaat het om ...	Il s'agit de ...	Is it ...		

Educ8	Hoger Beroepsonderwijs (HBO5) of Brevet d'enseignement supérieur (BES) in het Franstalig onderwijs	Brevet d'enseignement supérieur (BES) ou Hoger Beroepsonderwijs (HBO5) dans l'enseignement néerlandophone	Higher professional education in adult education (HBO5/BES)	1	Educ8
	Hogescholenonderwijs van het korte type (1 cyclus) of graduaat (A1) of professionele bachelor	Enseignement en haute école de type court (1 cycle) ou graduat (A1) ou bachelier professionnalisant	Short-type higher education (1 cycle) or graduate (A1) or professional bachelor	2	Educ8
	Academische bachelor (hogeschool of universiteit)	Bachelier académique (dans une haute école ou à l'université)	Academic bachelor (at a college for higher education or university)	3	Educ8
	Voortgezette of aanvullende opleiding na graduaat of na bachelor (specialisatie, BanaBa, ...)	Formations complémentaires post-graduat ou post-bachelier (spécialisation, bachelier complémentaire, ...)	Advanced or additional training after graduate or bachelor's degree (specialisation, bachelor after bachelor, ...)	4	Educ8
	Hogescholenonderwijs van het lange type (2 cycli) of master aan een hogeschool	Enseignement supérieur non universitaire de type long (2 cycles) ou master dans une haute école	Higher college education long programme (2 cycles) or a master at a college for higher education	5	Educ8
	Universitair onderwijs – licentiaat, master, ingenieur, dokter in de geneeskunde	Enseignement universitaire – licence, master, ingénieur, docteur en médecine	University degree - licentiate, master, engineer, doctor of medicine	6	Educ8
	Voortgezette of aanvullende opleiding na master (specialisatie, ManaMa, ...)	Formations complémentaires post-master (spécialisation, master complémentaire, ...)	Advanced or additional training after master's degree (specialisation, master after master, ...)	7	Educ8
	Doctoraat met proefschrift	Doctorat avec thèse	Doctor's degree with doctor's thesis	8	Educ8
	Ik weet het niet	Je ne sais pas	I don't know	97	Educ8
	Geen antwoord	Pas de réponse	No answer	98	Educ8
	Wat is het vakgebied of onderwerp van uw hoogste diploma?	Quel est le domaine ou la matière de diplôme le plus élevé ?	What is the field of study or discipline of your highest degree?		
	Algemeen programma	Programme de base	Generic programme	1	Educ9
	Onderwijs	Éducation	Education	2	Educ9
	Kunst, letteren en talen	Arts, lettres et langues	Arts, humanities and languages	3	Educ9
	Sociale wetenschappen, journalistiek en informatiewetenschappen	Sciences sociales, journalisme et information	Social sciences, journalism and information	4	Educ9
	Bedrijfskunde, bestuurskunde en recht	Commerce, administration et droit	Business, administration and law	5	Educ9
	Natuurwetenschappen, wiskunde en statistiek	Sciences naturelles, mathématiques et statistiques	Natural sciences, mathematics and statistics	6	Educ9
	Informatie- en communicatietechnologieën (ICT)	Technologies de l'information et des communications (TIC)	Information and Communication Technologies (ICTs)	7	Educ9
	Ingenieurswetenschappen, industrie en bouwnijverheid	Ingénierie, industries de transformation et construction	Engineering, manufacturing and construction	8	Educ9
Educ9	Landbouw, bosbouw, visserij en diergeneeskunde	Agriculture, sylviculture, halieutique et sciences vétérinaires	Agriculture, forestry, fisheries and veterinary	9	Educ9
	Gezondheid en welzijn	Santé et protection sociale	Health and welfare	10	Educ9
	Diensten	Services	Services	11	Educ9
	Ik weet het niet	Je ne sais pas	I don't know	97	Educ9
	Geen antwoord	Pas de réponse	No answer	98	Educ9
	Wat is de titel van uw hoogste diploma?	Quel est l'intitulé de votre diplôme le plus élevé ?	What is the title of your highest degree?		
	Vermeld de naam van de opleiding en uw eventuele major/minor.	Inclure le nom du programme et de votre majeur/mineur, le cas échéant.	Indicate the name of the programme and your major/minor, if any.		
	Open veld (max 300 karakters)	Champ libre (max 300 caractères)	Open field (max 300 characters)		Comm_5
	Ik weet het niet	Je ne sais pas	I don't know	97	
	Geen antwoord	Pas de réponse	No answer	98	
Comm_5	De volgende vragen gaan over uw tijdsbesteding in een gewone week en uw gedachten en gevoelens die daarmee gepaard gaan.	Les questions suivantes portent sur votre emploi du temps lors d'une semaine habituelle, ainsi que sur les pensées et les sentiments qui y sont liés.	The following questions relate to your time use during a typical week and your thoughts and feelings associated with it.		Val1
Val1	In hoeverre bent u het eens met de volgende uitspraken?	Dans quelle mesure êtes-vous d'accord avec les affirmations suivantes ?	To what extent do you agree with the following statements?		Val1_1
Val1_1	Er wordt te veel van mij verwacht	On attend trop de choses de moi	Too much is expected of me		
	Volledig oneens	Pas du tout d'accord	Totally disagree	1	Val1_2
	Oneens	Pas d'accord	Disagree	2	Val1_2
	Noch eens, noch oneens	Ni d'accord, ni pas d'accord	Neither agree nor disagree	3	Val1_2
	Eens	D'accord	Agree	4	Val1_2
	Volledig eens	Entièrement d'accord	Totally agree	5	
	Geen antwoord	Pas de réponse	No answer	98	
Val1_2	Ik raak nooit bijgewerkt	Je n'arriverai jamais à être à jour dans mon travail	I never catch up with my work		
	Volledig oneens	Pas du tout d'accord	Totally disagree	1	Val1_3

Val1_3	Oneens	Pas d'accord	Disagree	2	Val1_3
	Noch eens, noch oneens	Ni d'accord, ni pas d'accord	Neither agree nor disagree	3	Val1_3
	Eens	D'accord	Agree	4	Val1_3
	Volledig eens	Entièrement d'accord	Totally agree	5	Val1_3
	Geen antwoord	Pas de réponse	No answer	98	
	Ik heb nooit tijd voor mezelf	Je n'ai jamais de temps pour moi	I never have time for myself		
	Volledig oneens	Pas du tout d'accord	Totally disagree	1	Val1_4
	Oneens	Pas d'accord	Disagree	2	Val1_4
	Noch eens, noch oneens	Ni d'accord, ni pas d'accord	Neither agree nor disagree	3	Val1_4
	Eens	D'accord	Agree	4	Val1_4
Val1_4	Volledig eens	Entièrement d'accord	Totally agree	5	Val1_4
	Geen antwoord	Pas de réponse	No answer	98	
	Een dag heeft voor mij te weinig uren	Pour moi, un jour ne compte pas assez d'heures	There aren't enough hours in the day for me		
	Volledig oneens	Pas du tout d'accord	Totally disagree	1	Val1_5
	Oneens	Pas d'accord	Disagree	2	Val1_5
	Noch eens, noch oneens	Ni d'accord, ni pas d'accord	Neither agree nor disagree	3	Val1_5
	Eens	D'accord	Agree	4	Val1_5
	Volledig eens	Entièrement d'accord	Totally agree	5	Val1_5
	Geen antwoord	Pas de réponse	No answer	98	
	Ik moet meer doen dan ik wil doen	Je dois en faire plus que ce que je voudrais	I have to do more than I want to do		
Val1_5	Volledig oneens	Pas du tout d'accord	Totally disagree	1	Val1_6
	Oneens	Pas d'accord	Disagree	2	Val1_6
	Noch eens, noch oneens	Ni d'accord, ni pas d'accord	Neither agree nor disagree	3	Val1_6
	Eens	D'accord	Agree	4	Val1_6
	Volledig eens	Entièrement d'accord	Totally agree	5	Val1_6
	Geen antwoord	Pas de réponse	No answer	98	
	Ik heb geen tijd om de dingen te doen die ik moet doen	Je n'ai pas assez de temps pour faire tout ce que je dois faire	I have no time to do the things I have to do		
	Volledig oneens	Pas du tout d'accord	Totally disagree	1	Val1_7
	Oneens	Pas d'accord	Disagree	2	Val1_7
	Noch eens, noch oneens	Ni d'accord, ni pas d'accord	Neither agree nor disagree	3	Val1_7
Val1_6	Eens	D'accord	Agree	4	Val1_7
	Volledig eens	Entièrement d'accord	Totally agree	5	Val1_7
	Geen antwoord	Pas de réponse	No answer	98	
	Er wordt meer van mij verwacht dan ik aankan	On attend de moi plus que ce que je peux faire	More is expected from me than I can handle		
	Volledig oneens	Pas du tout d'accord	Totally disagree	1	Val1_8
	Oneens	Pas d'accord	Disagree	2	Val1_8
	Noch eens, noch oneens	Ni d'accord, ni pas d'accord	Neither agree nor disagree	3	Val1_8
	Eens	D'accord	Agree	4	Val1_8
	Volledig eens	Entièrement d'accord	Totally agree	5	Val1_8
	Geen antwoord	Pas de réponse	No answer	98	
Val1_7	Ik heb de indruk dat ik aan minder verplichtingen moet voldoen dan anderen	J'ai l'impression que j'ai moins à faire que d'autres	I have the impression that I have fewer obligations than others		
	Volledig oneens	Pas du tout d'accord	Totally disagree	1	Val2
	Oneens	Pas d'accord	Disagree	2	Val2
	Noch eens, noch oneens	Ni d'accord, ni pas d'accord	Neither agree nor disagree	3	Val2
	Eens	D'accord	Agree	4	Val2
	Volledig eens	Entièrement d'accord	Totally agree	5	Val2
	Geen antwoord	Pas de réponse	No answer	98	
	Hieronder staat een aantal beweringen over vrije tijd. Kan u aanduiden in welke mate elk van deze uitspraken overeenstemt met uw persoonlijke situatie?	Un certain nombre d'affirmations concernant le temps libre sont énumérées ci-dessous. Pourriez-vous indiquer dans quelle mesure ces affirmations correspondent à votre situation personnelle ?	You will find below a number of statements about free time. Indicate up to what extent each one applies to your personal situation.		Val2_1
	Vaak kom ik in mijn vrije tijd niet toe aan dingen die ik eigenlijk wil doen	Souvent, je ne parviens pas à faire ce que je veux vraiment faire pendant mon temps libre	During my free time, I often don't get around to doing the things I want to do		
	Volledig oneens	Pas du tout d'accord	Totally disagree	1	Val2_2
Val1_8	Oneens	Pas d'accord	Disagree	2	Val2_2
	Noch eens, noch oneens	Ni d'accord, ni pas d'accord	Neither agree nor disagree	3	Val2_2
	Eens	D'accord	Agree	4	Val2_2
	Volledig eens	Entièrement d'accord	Totally agree	5	Val2_2
	Geen antwoord	Pas de réponse	No answer	98	
	Hieronder staat een aantal beweringen over vrije tijd. Kan u aanduiden in welke mate elk van deze uitspraken overeenstemt met uw persoonlijke situatie?	Un certain nombre d'affirmations concernant le temps libre sont énumérées ci-dessous. Pourriez-vous indiquer dans quelle mesure ces affirmations correspondent à votre situation personnelle ?	You will find below a number of statements about free time. Indicate up to what extent each one applies to your personal situation.		
	Vaak kom ik in mijn vrije tijd niet toe aan dingen die ik eigenlijk wil doen	Souvent, je ne parviens pas à faire ce que je veux vraiment faire pendant mon temps libre	During my free time, I often don't get around to doing the things I want to do		
	Volledig oneens	Pas du tout d'accord	Totally disagree	1	Val2_2
	Oneens	Pas d'accord	Disagree	2	Val2_2
	Noch eens, noch oneens	Ni d'accord, ni pas d'accord	Neither agree nor disagree	3	Val2_2
Val2	Eens	D'accord	Agree	4	Val2_2
	Volledig eens	Entièrement d'accord	Totally agree	5	Val2_2
	Geen antwoord	Pas de réponse	No answer	98	
	Hieronder staat een aantal beweringen over vrije tijd. Kan u aanduiden in welke mate elk van deze uitspraken overeenstemt met uw persoonlijke situatie?	Un certain nombre d'affirmations concernant le temps libre sont énumérées ci-dessous. Pourriez-vous indiquer dans quelle mesure ces affirmations correspondent à votre situation personnelle ?	You will find below a number of statements about free time. Indicate up to what extent each one applies to your personal situation.		
	Vaak kom ik in mijn vrije tijd niet toe aan dingen die ik eigenlijk wil doen	Souvent, je ne parviens pas à faire ce que je veux vraiment faire pendant mon temps libre	During my free time, I often don't get around to doing the things I want to do		
	Volledig oneens	Pas du tout d'accord	Totally disagree	1	Val2_2
	Oneens	Pas d'accord	Disagree	2	Val2_2
	Noch eens, noch oneens	Ni d'accord, ni pas d'accord	Neither agree nor disagree	3	Val2_2
	Eens	D'accord	Agree	4	Val2_2
	Volledig eens	Entièrement d'accord	Totally agree	5	Val2_2
	Geen antwoord	Pas de réponse	No answer	98	

Val2_2	Geen antwoord		Pas de réponse	No answer	98	
	Ik moet in mijn vrije tijd te vaak rekening houden met anderen		Je dois trop souvent tenir compte des autres pendant mon temps libre	During my free time, I have to take other people into account too often		
	Volledig oneens		Pas du tout d'accord	Totally disagree	1	Val2_3
	Oneens		Pas d'accord	Disagree	2	Val2_3
	Noch eens, noch oneens		Ni d'accord, ni pas d'accord	Neither agree nor disagree	3	Val2_3
	Eens		D'accord	Agree	4	Val2_3
Val2_3	Volledig eens		Entièrement d'accord	Totally agree	5	Val2_3
	Geen antwoord		Pas de réponse	No answer	98	
	Ik kan me moeilijk ontspannen in mijn vrije tijd		Je peux difficilement me détendre pendant mon temps libre	I find it difficult to relax during my free time		
	Volledig oneens		Pas du tout d'accord	Totally disagree	1	Val2_4
	Oneens		Pas d'accord	Disagree	2	Val2_4
	Noch eens, noch oneens		Ni d'accord, ni pas d'accord	Neither agree nor disagree	3	Val2_4
Val2_4	Eens		D'accord	Agree	4	Val2_4
	Volledig eens		Entièrement d'accord	Totally agree	5	Val2_4
	Geen antwoord		Pas de réponse	No answer	98	
	Ik heb te veel vrije tijd		J'ai trop de temps libre	I have too much free time		
	Volledig oneens		Pas du tout d'accord	Totally disagree	1	Val2_5
	Oneens		Pas d'accord	Disagree	2	Val2_5
Val2_5	Noch eens, noch oneens		Ni d'accord, ni pas d'accord	Neither agree nor disagree	3	Val2_5
	Eens		D'accord	Agree	4	Val2_5
	Volledig eens		Entièrement d'accord	Totally agree	5	Val2_5
	Geen antwoord		Pas de réponse	No answer	98	
	Als ik vrij ben, zijn te veel vrijetijdsvoorzieningen (bv. zwembaden, clubs, musea ...) gesloten		Quand j'ai du temps libre, beaucoup de structures de loisirs sont fermées (par ex. bassin de natation, club sportif, musée, etc.)	When I am free, too many leisure facilities (e.g. swimming pools, clubs, museums, etc.) are closed.		
	Volledig oneens		Pas du tout d'accord	Totally disagree	1	Val2_6
Val2_6	Oneens		Pas d'accord	Disagree	2	Val2_6
	Noch eens, noch oneens		Ni d'accord, ni pas d'accord	Neither agree nor disagree	3	Val2_6
	Eens		D'accord	Agree	4	Val2_6
	Volledig eens		Entièrement d'accord	Totally agree	5	Val2_6
	Geen antwoord		Pas de réponse	No answer	98	
	Het kost me veel moeite om mijn vrijetijdsactiviteiten te plannen		C'est un grand effort pour moi de planifier des activités de loisir	It takes a lot of effort to plan my leisure activities		
Val2_7	Volledig oneens		Pas du tout d'accord	Totally disagree	1	Val2_7
	Oneens		Pas d'accord	Disagree	2	Val2_7
	Noch eens, noch oneens		Ni d'accord, ni pas d'accord	Neither agree nor disagree	3	Val2_7
	Eens		D'accord	Agree	4	Val2_7
	Volledig eens		Entièrement d'accord	Totally agree	5	Val2_7
	Geen antwoord		Pas de réponse	No answer	98	
Val2_8	Er zijn zoveel dingen die ik wil doen in mijn vrije tijd, dat ik vaak het gevoel heb tijd tekort te komen		Il y a tellement de choses que je veux faire dans mon temps libre que j'ai souvent l'impression de n'avoir pas assez de temps	There are so many things I want to do in my free time that I often feel like there is not enough time		
	Volledig oneens		Pas du tout d'accord	Totally disagree	1	Val2_8
	Oneens		Pas d'accord	Disagree	2	Val2_8
	Noch eens, noch oneens		Ni d'accord, ni pas d'accord	Neither agree nor disagree	3	Val2_8
	Eens		D'accord	Agree	4	Val2_8
	Volledig eens		Entièrement d'accord	Totally agree	5	Val2_8
Val2_8	Geen antwoord		Pas de réponse	No answer	98	
	Te veel van mijn vrijetijdsactiviteiten zijn versnipperd		Trop de mes activités de loisirs sont fragmentées	Too many of my leisure activities are fragmented		
	Volledig oneens		Pas du tout d'accord	Totally disagree	1	Comm_6
	Oneens		Pas d'accord	Disagree	2	Comm_6
	Noch eens, noch oneens		Ni d'accord, ni pas d'accord	Neither agree nor disagree	3	Comm_6
	Eens		D'accord	Agree	4	Comm_6
Comm_6	Volledig eens		Entièrement d'accord	Totally agree	5	Comm_6
	Geen antwoord		Pas de réponse	No answer	98	
	De laatste vragen gaan over uw gezondheid.		Les dernières questions portent sur votre santé.	The last set of questions relate to your health.		Health1
	Health1	IND26	Hoe vaak voelt u zich gehaast?	À quelle fréquence vous sentez-vous pressé(e) ?		
				How often do you feel rushed?		

Health2	IND23	Altijd gehaast	Toujours pressé(e)	Always rushed	1	Health2
		Soms gehaast	Parfois pressé(e)	Sometimes rushed	2	Health2
		Bijna nooit gehaast	Presque jamais pressé(e)	Almost never rushed	3	Health2
		Geen antwoord	Pas de réponse	No answer	98	
		Hoe is uw gezondheid in het algemeen?	Quel est votre état de santé en général ?	How is your health in general?		
Health3	IND24	Heel goed	Très bon	Very good	1	Health3
		Goed	Bon	Good	2	Health3
		Niet goed, niet slecht	Ni bon, ni mauvais	Fair (neither good nor bad)	3	Health3
		Slecht	Mauvais	Bad	4	Health3
		Heel slecht	Très mauvais	Very bad	5	Health3
Health4_1	IND25	Geen antwoord	Pas de réponse	No answer	98	
		Hebt u een ziekte of gezondheidsprobleem dat al minstens 6 maanden duurt of naar verwachting nog minstens 6 maanden zal duren?	Souffrez-vous d'une maladie ou d'un problème de santé qui dure depuis au moins 6 mois ou devrait durer au moins 6 mois ?	Do you have an illness or health problem that has been lasting, or is expected to last, for at least 6 months?		
		Ja	Oui	Yes	1	Health4_1
		Neen	Non	No	2	Health4_1
		Geen antwoord	Pas de réponse	No answer	98	
Health4_2	IND25	Bent u als gevolg van een gezondheidsprobleem beperkt in activiteiten die mensen gewoonlijk doen?	Êtes-vous limité(e), à cause d'un problème de santé, dans les activités que les gens font habituellement ?	Are you limited because of a health problem in activities people usually do?		
		Ja, ernstig beperkt	Oui, fortement limité(e)	Yes, severely limited	1	Health4_2
		Ja, beperkt, maar niet ernstig	Oui, limité(e), mais pas fortement	Yes, limited, but not severely	2	Health4_2
		Neen, helemaal niet beperkt	Non, pas limité(e) du tout	No, not limited at all	3	Health5
		Geen antwoord	Pas de réponse	No answer	98	
Health5	IND25	Bent u al minstens 6 maanden beperkt?	Êtes-vous limité depuis au moins six mois ?	Have you been limited for at least the past 6 months?		
		Ja	Oui	Yes	1	Health5
		Neen	Non	No	2	Health5
		Geen antwoord	Pas de réponse	No answer	98	
		Wat is uw lengte?	Quelle est votre taille ?	What is your height?		
Health6		Vul lengte in centimeters in.	Veillez noter votre taille en cm.	Please note your length in centimetres.		
		Lengte in cm (40-250)	Longueur en cm (40-250)	Length in cm (40-250)		Health6
		Geen antwoord	Pas de réponse	No answer	98	
		Wat is uw gewicht?	Quel est votre poids ?	What is your weight?		
		Vul gewicht in kilogram in.	Veillez noter votre poids en kg.	Please note your weight in kg.		
End		Gewicht in kg (20-300)	Poids en kg (20-300)	Weight in kg (20-300)		End
		Geen antwoord	Pas de réponse	No answer	98	
		Bedankt! Dit was voorlopig de laatste vraag. Klik op 'verzenden'.	Merci ! C'était la dernière question pour l'instant. Cliquez sur 'envoyer'.	Thank you! That's the last question for now. Click on 'submit'.		

CAWI - post-questionnaire

Variable name	HETUS-variable	N	F	E	Code	Logic	Remarks
Comm_0		Ter afronding van het onderzoek vragen we u om nog een laatste vragenlijst in te vullen. Deze vragenlijst is belangrijk omdat ze ons bijkomende informatie geeft over de dagen waarop u uw dagboekje hebt bijgehouden, uw tijdsbesteding en uw ervaring met de MOTUS-toepassing. Het invullen van deze vragenlijst duurt gemiddeld minder dan 10 minuten. Alvast bedankt voor uw medewerking.	Afin de clôturer l'enquête, nous vous demandons de répondre à un dernier questionnaire. Ce questionnaire est important car il nous fournit des informations supplémentaires sur les jours pendant lesquels vous avez tenu votre journal, sur votre emploi du temps et sur votre expérience avec l'application MOTUS. En moyenne, il faut moins de 10 minutes pour le compléter. Nous vous remercions d'avance de votre collaboration.	To conclude the study, please fill in one final questionnaire. This questionnaire is important because it gives us additional information about the days you kept your diary, your time use and your experience with the MOTUS application. On average, it takes less than 10 minutes to complete. Thank you very much for your cooperation.		Comm_1	
Comm_1		De volgende vragen gaan over de dagen waarop u uw dagboekje hebt bijgehouden.	Les questions suivantes portent sur les jours durant lesquels vous avez tenu votre journal.	The following questions refer to the days when you kept your diary.		Diary1	
Diary1		Is de week dat u het dagboekje moest invullen om één of andere reden een bijzondere week geweest?	La semaine durant laquelle vous deviez remplir votre journal était-elle une semaine particulière pour l'une ou l'autre raison ?	Was the week you had to fill in the diary a special week for some reason?			
		Ja	Oui	Yes	1	Diary2	
		Neen	Non	No	2	Diary3	
Diary2		Geen antwoord	Pas de réponse	No answer	98		
		Om welke reden was het een bijzondere week?	Pour quelle raison cette semaine était-elle particulière ?	Why was it a special week?			
		Open veld (max 600 karakters)	Champ libre (max 600 caractères)	Open field (max 600 characters)		Diary3	
		Geen antwoord	Pas de réponse	No answer	98		
Diary3		Hebt u gedurende de week van de dagboekregistraties één of meerdere dagen vrijaf gehad?	Durant la semaine où vous avez tenu votre journal, avez-vous eu un ou plusieurs jour(s) de congé ?	Did you have one or more days off during the week when you kept your diary?			
		Ja	Oui	Yes	1	Diary4	
		Neen	Non	No	2	Comm_2	
Diary4		Geen antwoord	Pas de réponse	No answer	98		
		Welke dag(en) had u in die week vrijaf?	Quel(s) jour(s) de cette semaine avez-vous eu congé ?	Which day(s) did you have off during that week?			
		Meerdere antwoorden mogelijk	Plusieurs réponses possibles	Multiple answers possible			
		Maandag	Lundi	Monday	1	Comm_2	
		Dinsdag	Mardi	Tuesday	2	Comm_2	
		Woensdag	Mercredi	Wednesday	3	Comm_2	
		Donderdag	Jeudi	Thursday	4	Comm_2	
		Vrijdag	Vendredi	Friday	5	Comm_2	
		Zaterdag	Samedi	Saturday	6	Comm_2	
		Zondag	Dimanche	Sunday	7	Comm_2	
		Geen antwoord	Pas de réponse	No answer	98		
Comm_2		De volgende vragen gaan over uw deelname aan vrijetijds- en culturele activiteiten tijdens het afgelopen jaar.	Les questions suivantes concernent votre participation à des activités de loisirs et culturelles au cours de l'année dernière.	The following questions concern your participation in leisure and cultural activities during the past year.		Cult1	
Cult1		Tijdens het afgelopen jaar, hoe vaak hebt u een bioscoop bezocht?	Au cours de l'année dernière, combien de fois êtes-vous allé(e) au cinéma ?	During the past year, how often did you visit a cinema?			
		Nooit	Jamais	Never	1	Cult2	
		Eén keer per jaar of minder	Une fois par an ou moins	Once a year or less	2	Cult2	
		Meerdere keren per jaar	Plusieurs fois par an	Several times a year	3	Cult2	
		Ongeveer één keer per maand	Environ une fois par mois	About once a month	4	Cult2	
		Meerdere keren per maand	Plusieurs fois par mois	Several times a month	5	Cult2	
		Ongeveer één keer per week	Environ une fois par semaine	About once a week	6	Cult2	
		Meerdere keren per week	Plusieurs fois par semaine	Several times a week	7	Cult2	
		Ik weet het niet	Je ne sais pas	I don't know	97		
		Geen antwoord	Pas de réponse	No answer	98		
Cult2		Tijdens het afgelopen jaar, hoe vaak hebt u een muziekconcert en/of -festival bijgewoond?	Au cours de l'année dernière, combien de fois avez-vous assisté à un concert et/ou à un festival de musique ?	During the past year, how often did you attend a music concert and/or festival?			
		Nooit	Jamais	Never	1	Cult3	
		Eén keer per jaar of minder	Une fois par an ou moins	Once a year or less	2	Cult3	
		Meerdere keren per jaar	Plusieurs fois par an	Several times a year	3	Cult3	
		Ongeveer één keer per maand	Environ une fois par mois	About once a month	4	Cult3	
		Meerdere keren per maand	Plusieurs fois par mois	Several times a month	5	Cult3	
		Ongeveer één keer per week	Environ une fois par semaine	About once a week	6	Cult3	
		Meerdere keren per week	Plusieurs fois par semaine	Several times a week	7	Cult3	
		Ik weet het niet	Je ne sais pas	I don't know	97		
		Geen antwoord	Pas de réponse	No answer	98		
Cult3		Tijdens het afgelopen jaar, hoe vaak hebt u podiumkunsten (ballet, dans, theater, toneel, stand-up comedy, musical, ...) bezocht?	Au cours de l'année dernière, combien de fois avez-vous assisté à un spectacle vivant (ballet, danse, théâtre, comédie, comédie musicale, ...) ?	During the past year, how often did you attend performing arts (ballet, dance, theatre, drama, stand-up comedy, musical, etc.)?			
		Nooit	Jamais	Never	1	Cult4	
		Eén keer per jaar of minder	Une fois par an ou moins	Once a year or less	2	Cult4	
		Meerdere keren per jaar	Plusieurs fois par an	Several times a year	3	Cult4	

Cult4	Ongeveer één keer per maand	Environ une fois par mois	About once a month	4	Cult4
	Meerdere keren per maand	Plusieurs fois par mois	Several times a month	5	Cult4
	Ongeveer één keer per week	Environ une fois par semaine	About once a week	6	Cult4
	Meerdere keren per week	Plusieurs fois par semaine	Several times a week	7	Cult4
	Ik weet het niet	Je ne sais pas	I don't know	97	
	Geen antwoord	Pas de réponse	No answer	98	
	Tijdens het afgelopen jaar, hoe vaak hebt u een bibliotheek bezocht?	Au cours de l'année dernière, combien de fois êtes-vous allé(e) à la bibliothèque ?	During the past year, how often did you visit the library?		
	Nooit	Jamais	Never	1	Cult5
	Eén keer per jaar of minder	Une fois par an ou moins	Once a year or less	2	Cult5
	Meerdere keren per jaar	Plusieurs fois par an	Several times a year	3	Cult5
Cult5	Ongeveer één keer per maand	Environ une fois par mois	About once a month	4	Cult5
	Meerdere keren per maand	Plusieurs fois par mois	Several times a month	5	Cult5
	Ongeveer één keer per week	Environ une fois par semaine	About once a week	6	Cult5
	Meerdere keren per week	Plusieurs fois par semaine	Several times a week	7	Cult5
	Ik weet het niet	Je ne sais pas	I don't know	97	
	Geen antwoord	Pas de réponse	No answer	98	
	Tijdens het afgelopen jaar, hoe vaak hebt u een sportwedstrijd als toeschouwer bijgewoond?	Au cours de l'année dernière, combien de fois avez-vous assisté en tant que spectateur à une compétition sportive?	During the past year, how often did you attend a sports event as a spectator?		
	Nooit	Jamais	Never	1	Cult6
	Eén keer per jaar of minder	Une fois par an ou moins	Once a year or less	2	Cult6
	Meerdere keren per jaar	Plusieurs fois par an	Several times a year	3	Cult6
Cult6	Ongeveer één keer per maand	Environ une fois par mois	About once a month	4	Cult6
	Meerdere keren per maand	Plusieurs fois par mois	Several times a month	5	Cult6
	Ongeveer één keer per week	Environ une fois par semaine	About once a week	6	Cult6
	Meerdere keren per week	Plusieurs fois par semaine	Several times a week	7	Cult6
	Ik weet het niet	Je ne sais pas	I don't know	97	
	Geen antwoord	Pas de réponse	No answer	98	
	Tijdens het afgelopen jaar, hoe vaak hebt u een tentoonstelling, museum bezocht?	Au cours de l'année dernière, combien de fois avez-vous visité une exposition, un musée ?	During the past year, how often did you visit art exhibitions and museums?		
	Nooit	Jamais	Never	1	Cult7
	Eén keer per jaar of minder	Une fois par an ou moins	Once a year or less	2	Cult7
	Meerdere keren per jaar	Plusieurs fois par an	Several times a year	3	Cult7
Cult7	Ongeveer één keer per maand	Environ une fois par mois	About once a month	4	Cult7
	Meerdere keren per maand	Plusieurs fois par mois	Several times a month	5	Cult7
	Ongeveer één keer per week	Environ une fois par semaine	About once a week	6	Cult7
	Meerdere keren per week	Plusieurs fois par semaine	Several times a week	7	Cult7
	Ik weet het niet	Je ne sais pas	I don't know	97	
	Geen antwoord	Pas de réponse	No answer	98	
	Tijdens het afgelopen jaar, hoe vaak bent u iets gaan eten en/of drinken in een restaurant, café, pub?	Au cours de l'année dernière, combien de fois êtes-vous allé(e) manger et/ou boire un verre dans un restaurant, café, bar ?	During the past year, how often did you go for a meal and/or a drink in a restaurant, café, pub?		
	Nooit	Jamais	Never	1	Cult8
	Eén keer per jaar of minder	Une fois par an ou moins	Once a year or less	2	Cult8
	Meerdere keren per jaar	Plusieurs fois par an	Several times a year	3	Cult8
Cult8	Ongeveer één keer per maand	Environ une fois par mois	About once a month	4	Cult8
	Meerdere keren per maand	Plusieurs fois par mois	Several times a month	5	Cult8
	Ongeveer één keer per week	Environ une fois par semaine	About once a week	6	Cult8
	Meerdere keren per week	Plusieurs fois par semaine	Several times a week	7	Cult8
	Ik weet het niet	Je ne sais pas	I don't know	97	
	Geen antwoord	Pas de réponse	No answer	98	
	Tijdens het afgelopen jaar, hoe vaak bent u naar een shoppingcenter geweest?	Au cours de l'année dernière, combien de fois vous êtes-vous rendu dans un centre commercial ?	During the past year, how often did you go to the shopping centre?		
	Nooit	Jamais	Never	1	Cult9
	Eén keer per jaar of minder	Une fois par an ou moins	Once a year or less	2	Cult9
	Meerdere keren per jaar	Plusieurs fois par an	Several times a year	3	Cult9
Cult9	Ongeveer één keer per maand	Environ une fois par mois	About once a month	4	Cult9
	Meerdere keren per maand	Plusieurs fois par mois	Several times a month	5	Cult9
	Ongeveer één keer per week	Environ une fois par semaine	About once a week	6	Cult9
	Meerdere keren per week	Plusieurs fois par semaine	Several times a week	7	Cult9
	Ik weet het niet	Je ne sais pas	I don't know	97	
	Geen antwoord	Pas de réponse	No answer	98	
	Tijdens het afgelopen jaar, hoe vaak hebt u een jaarmarkt, beurs bezocht?	Au cours de l'année dernière, combien de fois avez-vous visité une foire, un salon ?	During the past year, how often did you go to fairs, sales exhibitions?		
	Nooit	Jamais	Never	1	Cult10
	Eén keer per jaar of minder	Une fois par an ou moins	Once a year or less	2	Cult10
	Meerdere keren per jaar	Plusieurs fois par an	Several times a year	3	Cult10
	Ongeveer één keer per maand	Environ une fois par mois	About once a month	4	Cult10
	Meerdere keren per maand	Plusieurs fois par mois	Several times a month	5	Cult10
	Ongeveer één keer per week	Environ une fois par semaine	About once a week	6	Cult10
	Meerdere keren per week	Plusieurs fois par semaine	Several times a week	7	Cult10

Cult10	Ik weet het niet	Je ne sais pas	I don't know	97	
	Geen antwoord	Pas de réponse	No answer	98	
	Tijdens het afgelopen jaar, hoe vaak hebt u uitstapjes gemaakt, een stad, pretpark of dierentuin bezocht?	Au cours de l'année dernière, combien de fois avez-vous fait des excursions, visité une ville, un parc d'attractions ou un zoo ?	During the past year, how often did you take excursions or did you visit a city, amusement park or zoo?		
Cult11	Nooit	Jamais	Never	1	Cult11
	Eén keer per jaar of minder	Une fois par an ou moins	Once a year or less	2	Cult11
	Meerdere keren per jaar	Plusieurs fois par an	Several times a year	3	Cult11
	Ongeveer één keer per maand	Environ une fois par mois	About once a month	4	Cult11
	Meerdere keren per maand	Plusieurs fois par mois	Several times a month	5	Cult11
	Ongeveer één keer per week	Environ une fois par semaine	About once a week	6	Cult11
	Meerdere keren per week	Plusieurs fois par semaine	Several times a week	7	Cult11
	Ik weet het niet	Je ne sais pas	I don't know	97	
	Geen antwoord	Pas de réponse	No answer	98	
	Tijdens het afgelopen jaar, hoe vaak hebt u ander amusement bijgewoond?	Au cours de l'année dernière, combien de fois avez-vous assisté à d'autres divertissements ?	During the past year, how often have you attended other entertainment events?		
Comm_3	Nooit	Jamais	Never	1	Comm_3
	Eén keer per jaar of minder	Une fois par an ou moins	Once a year or less	2	Comm_3
	Meerdere keren per jaar	Plusieurs fois par an	Several times a year	3	Comm_3
	Ongeveer één keer per maand	Environ une fois par mois	About once a month	4	Comm_3
	Meerdere keren per maand	Plusieurs fois par mois	Several times a month	5	Comm_3
	Ongeveer één keer per week	Environ une fois par semaine	About once a week	6	Comm_3
	Meerdere keren per week	Plusieurs fois par semaine	Several times a week	7	Comm_3
	Ik weet het niet	Je ne sais pas	I don't know	97	
	Geen antwoord	Pas de réponse	No answer	98	
	De volgende vragen gaan over uw deelname aan sportactiviteiten tijdens het afgelopen jaar.	Les questions suivantes concernent votre participation à des activités sportives au cours de l'année dernière.	The following questions concern your participation in sports activities during the past year.		Sport1
Sport1	Tijdens het afgelopen jaar, hoe vaak hebt u joggen, hardlopen, atletiek of gymnastiek beoefend?	Au cours de l'année dernière, combien de fois avez-vous pratiqué le jogging, la course à pied, l'athlétisme ou la gymnastique ?	During the past year, how often have you practised jogging, running, athletics or gymnastics?		
	Nooit	Jamais	Never	1	Sport2
	Eén keer per jaar of minder	Une fois par an ou moins	Once a year or less	2	Sport2
	Meerdere keren per jaar	Plusieurs fois par an	Several times a year	3	Sport2
	Ongeveer één keer per maand	Environ une fois par mois	About once a month	4	Sport2
	Meerdere keren per maand	Plusieurs fois par mois	Several times a month	5	Sport2
	Ongeveer één keer per week	Environ une fois par semaine	About once a week	6	Sport2
	Meerdere keren per week	Plusieurs fois par semaine	Several times a week	7	Sport2
	Ik weet het niet	Je ne sais pas	I don't know	97	
	Geen antwoord	Pas de réponse	No answer	98	
Sport2	Tijdens het afgelopen jaar, hoe vaak hebt u wielersport beoefend of fietstochten gemaakt?	Au cours de l'année dernière, combien de fois avez-vous pratiqué le cyclisme ou fait des excursions à vélo ?	During the past year, how often have you practised cycling or taken bike trips?		
	Nooit	Jamais	Never	1	Sport3
	Eén keer per jaar of minder	Une fois par an ou moins	Once a year or less	2	Sport3
	Meerdere keren per jaar	Plusieurs fois par an	Several times a year	3	Sport3
	Ongeveer één keer per maand	Environ une fois par mois	About once a month	4	Sport3
	Meerdere keren per maand	Plusieurs fois par mois	Several times a month	5	Sport3
	Ongeveer één keer per week	Environ une fois par semaine	About once a week	6	Sport3
	Meerdere keren per week	Plusieurs fois par semaine	Several times a week	7	Sport3
	Ik weet het niet	Je ne sais pas	I don't know	97	
	Geen antwoord	Pas de réponse	No answer	98	
Sport3	Tijdens het afgelopen jaar, hoe vaak hebt u fitness, yoga of lichaamstraining beoefend?	Au cours de l'année dernière, combien de fois avez-vous pratiqué le fitness, le yoga ou l'entraînement corporel ?	During the past year, how often have you practised fitness, yoga or body training?		
	Nooit	Jamais	Never	1	Sport4
	Eén keer per jaar of minder	Une fois par an ou moins	Once a year or less	2	Sport4
	Meerdere keren per jaar	Plusieurs fois par an	Several times a year	3	Sport4
	Ongeveer één keer per maand	Environ une fois par mois	About once a month	4	Sport4
	Meerdere keren per maand	Plusieurs fois par mois	Several times a month	5	Sport4
	Ongeveer één keer per week	Environ une fois par semaine	About once a week	6	Sport4
	Meerdere keren per week	Plusieurs fois par semaine	Several times a week	7	Sport4
	Ik weet het niet	Je ne sais pas	I don't know	97	
	Geen antwoord	Pas de réponse	No answer	98	
Sport4	Tijdens het afgelopen jaar, hoe vaak hebt u zwemmen, aquajoggen of aquagym beoefend?	Au cours de l'année dernière, combien de fois avez-vous pratiqué la natation, l'aqua-jogging ou l'aquagym ?	During the past year, how often have you practised swimming, aqua jogging or aqua gym?		
	Nooit	Jamais	Never	1	Sport5
	Eén keer per jaar of minder	Une fois par an ou moins	Once a year or less	2	Sport5
	Meerdere keren per jaar	Plusieurs fois par an	Several times a year	3	Sport5
	Ongeveer één keer per maand	Environ une fois par mois	About once a month	4	Sport5
	Meerdere keren per maand	Plusieurs fois par mois	Several times a month	5	Sport5
	Ongeveer één keer per week	Environ une fois par semaine	About once a week	6	Sport5

Sport5	Meerdere keren per week	Plusieurs fois par semaine	Several times a week	7	Sport5
	Ik weet het niet	Je ne sais pas	I don't know	97	
	Geen antwoord	Pas de réponse	No answer	98	
	Tijdens het afgelopen jaar, hoe vaak hebt u andere watersporten (roeien, waterpolo, surfen, ...) beoefend?	Au cours de l'année dernière, combien de fois avez-vous pratiqué d'autres sports nautiques (aviron, water-polo, surf, ...) ?	During the past year, how often have you practised other water sports (rowing, water polo, surfing, etc.)?		
	Nooit	Jamais	Never	1	
	Eén keer per jaar of minder	Une fois par an ou moins	Once a year or less	2	
	Meerdere keren per jaar	Plusieurs fois par an	Several times a year	3	
	Ongeveer één keer per maand	Environ une fois par mois	About once a month	4	
	Meerdere keren per maand	Plusieurs fois par mois	Several times a month	5	
	Ongeveer één keer per week	Environ une fois par semaine	About once a week	6	
Sport6	Meerdere keren per week	Plusieurs fois par semaine	Several times a week	7	Sport6
	Ik weet het niet	Je ne sais pas	I don't know	97	
	Geen antwoord	Pas de réponse	No answer	98	
	Tijdens het afgelopen jaar, hoe vaak hebt u skiën, schaatsen of andere wintersporten beoefend?	Au cours de l'année dernière, combien de fois avez-vous pratiqué le ski, le patinage ou d'autres sports d'hiver ?	During the past year, how often have you practised skiing, skating or other winter sports?		
	Nooit	Jamais	Never	1	
	Eén keer per jaar of minder	Une fois par an ou moins	Once a year or less	2	
	Meerdere keren per jaar	Plusieurs fois par an	Several times a year	3	
	Ongeveer één keer per maand	Environ une fois par mois	About once a month	4	
	Meerdere keren per maand	Plusieurs fois par mois	Several times a month	5	
	Ongeveer één keer per week	Environ une fois par semaine	About once a week	6	
Sport7	Meerdere keren per week	Plusieurs fois par semaine	Several times a week	7	Sport7
	Ik weet het niet	Je ne sais pas	I don't know	97	
	Geen antwoord	Pas de réponse	No answer	98	
	Tijdens het afgelopen jaar, hoe vaak hebt u golf gespeeld?	Au cours de l'année dernière, combien de fois avez-vous joué au golf ?	During the past year, how often have you played golf?		
	Nooit	Jamais	Never	1	
	Eén keer per jaar of minder	Une fois par an ou moins	Once a year or less	2	
	Meerdere keren per jaar	Plusieurs fois par an	Several times a year	3	
	Ongeveer één keer per maand	Environ une fois par mois	About once a month	4	
	Meerdere keren per maand	Plusieurs fois par mois	Several times a month	5	
	Ongeveer één keer per week	Environ une fois par semaine	About once a week	6	
Sport8	Meerdere keren per week	Plusieurs fois par semaine	Several times a week	7	Sport8
	Ik weet het niet	Je ne sais pas	I don't know	97	
	Geen antwoord	Pas de réponse	No answer	98	
	Tijdens het afgelopen jaar, hoe vaak hebt u (tafel)tennis, badminton, squash of padel gespeeld?	Au cours de l'année dernière, combien de fois avez-vous joué au tennis (de table), badminton, squash ou padel ?	During the past year, how often have you played (table)tennis, badminton, squash or padel?		
	Nooit	Jamais	Never	1	
	Eén keer per jaar of minder	Une fois par an ou moins	Once a year or less	2	
	Meerdere keren per jaar	Plusieurs fois par an	Several times a year	3	
	Ongeveer één keer per maand	Environ une fois par mois	About once a month	4	
	Meerdere keren per maand	Plusieurs fois par mois	Several times a month	5	
	Ongeveer één keer per week	Environ une fois par semaine	About once a week	6	
Sport9	Meerdere keren per week	Plusieurs fois par semaine	Several times a week	7	Sport9
	Ik weet het niet	Je ne sais pas	I don't know	97	
	Geen antwoord	Pas de réponse	No answer	98	
	Tijdens het afgelopen jaar, hoe vaak hebt u balsport (voetbal, basketbal, volleybal, hockey, ...) beoefend?	Au cours de l'année dernière, combien de fois avez-vous pratiqué des sports de ballon (football, basket-ball, volley-ball, hockey...) ?	During the past year, how often have you played ball games (football, basketball, volleyball, hockey, etc.)?		
	Nooit	Jamais	Never	1	
	Eén keer per jaar of minder	Une fois par an ou moins	Once a year or less	2	
	Meerdere keren per jaar	Plusieurs fois par an	Several times a year	3	
	Ongeveer één keer per maand	Environ une fois par mois	About once a month	4	
	Meerdere keren per maand	Plusieurs fois par mois	Several times a month	5	
	Ongeveer één keer per week	Environ une fois par semaine	About once a week	6	
Sport10	Meerdere keren per week	Plusieurs fois par semaine	Several times a week	7	Sport10
	Ik weet het niet	Je ne sais pas	I don't know	97	
	Geen antwoord	Pas de réponse	No answer	98	
	Tijdens het afgelopen jaar, hoe vaak bent u gaan vissen, jagen of schieten?	Au cours de l'année dernière, combien de fois avez-vous pratiqué la pêche, la chasse ou le tir?	During the past year, how often have you gone fishing, hunting or shooting?		
	Nooit	Jamais	Never	1	
	Eén keer per jaar of minder	Une fois par an ou moins	Once a year or less	2	
	Meerdere keren per jaar	Plusieurs fois par an	Several times a year	3	
	Ongeveer één keer per maand	Environ une fois par mois	About once a month	4	
	Meerdere keren per maand	Plusieurs fois par mois	Several times a month	5	
	Ongeveer één keer per week	Environ une fois par semaine	About once a week	6	
Sport11	Meerdere keren per week	Plusieurs fois par semaine	Several times a week	7	Sport11
	Ik weet het niet	Je ne sais pas	I don't know	97	
	Geen antwoord	Pas de réponse	No answer	98	
	Eén keer per jaar of minder	Une fois par an ou moins	Once a year or less	2	
	Meerdere keren per jaar	Plusieurs fois par an	Several times a year	3	
	Ongeveer één keer per maand	Environ une fois par mois	About once a month	4	
	Meerdere keren per maand	Plusieurs fois par mois	Several times a month	5	
	Ongeveer één keer per week	Environ une fois par semaine	About once a week	6	
	Meerdere keren per week	Plusieurs fois par semaine	Several times a week	7	
	Ik weet het niet	Je ne sais pas	I don't know	97	

Sport11	Tijdens het afgelopen jaar, hoe vaak hebt u bowling, darts, pool, biljart of snooker gespeeld?	Au cours de l'année dernière, combien de fois avez-vous joué au bowling, aux fléchettes, au billard ou au snooker ?	During the past year, how often have you played bowling, darts, pool, billiards or snooker?		
	Nooit	Jamais	Never	1	Sport12
	Eén keer per jaar of minder	Une fois par an ou moins	Once a year or less	2	Sport12
	Meerdere keren per jaar	Plusieurs fois par an	Several times a year	3	Sport12
	Ongeveer één keer per maand	Environ une fois par mois	About once a month	4	Sport12
	Meerdere keren per maand	Plusieurs fois par mois	Several times a month	5	Sport12
	Ongeveer één keer per week	Environ une fois par semaine	About once a week	6	Sport12
	Meerdere keren per week	Plusieurs fois par semaine	Several times a week	7	Sport12
	Ik weet het niet	Je ne sais pas	I don't know	97	
	Geen antwoord	Pas de réponse	No answer	98	
Sport12	Tijdens het afgelopen jaar, hoe vaak bent u gaan wandelen (5 kilometer of meer)?	Au cours de l'année dernière, combien de fois avez-vous fait des randonnées (5 kilomètres ou plus) ?	During the past year, how often did you go hiking (5 kilometres or more)?		
	Nooit	Jamais	Never	1	Sport13
	Eén keer per jaar of minder	Une fois par an ou moins	Once a year or less	2	Sport13
	Meerdere keren per jaar	Plusieurs fois par an	Several times a year	3	Sport13
	Ongeveer één keer per maand	Environ une fois par mois	About once a month	4	Sport13
	Meerdere keren per maand	Plusieurs fois par mois	Several times a month	5	Sport13
	Ongeveer één keer per week	Environ une fois par semaine	About once a week	6	Sport13
	Meerdere keren per week	Plusieurs fois par semaine	Several times a week	7	Sport13
	Ik weet het niet	Je ne sais pas	I don't know	97	
	Geen antwoord	Pas de réponse	No answer	98	
Sport13	Tijdens het afgelopen jaar, hoe vaak hebt u andere sporten beoefend?	Au cours de l'année dernière, combien de fois avez-vous pratiqué d'autres sports ?	During the past year, how often have you practised other sports?		
	Nooit	Jamais	Never	1	Comm_4
	Eén keer per jaar of minder	Une fois par an ou moins	Once a year or less	2	Comm_4
	Meerdere keren per jaar	Plusieurs fois par an	Several times a year	3	Comm_4
	Ongeveer één keer per maand	Environ une fois par mois	About once a month	4	Comm_4
	Meerdere keren per maand	Plusieurs fois par mois	Several times a month	5	Comm_4
	Ongeveer één keer per week	Environ une fois par semaine	About once a week	6	Comm_4
	Meerdere keren per week	Plusieurs fois par semaine	Several times a week	7	Comm_4
	Ik weet het niet	Je ne sais pas	I don't know	97	
	Geen antwoord	Pas de réponse	No answer	98	
Comm_4	Bedankt om deel te nemen aan het Belgisch onderzoek naar tijdsbesteding! Ter evaluatie hadden we nog graag volgende vragen gesteld.	Merci d'avoir participé à l'enquête sur l'emploi du temps des belges ! En guise d'évaluation de cette enquête, nous aimerions vous poser les questions suivantes.	Thank you for taking part in the Belgian Time Use Survey! For the evaluation of this survey we would like to ask following questions.		Evaluation1
Evaluation1	Hebt u de MOTUS-app geïnstalleerd op uw smartphone of tablet?	Avez-vous téléchargé l'application mobile MOTUS sur votre smartphone ou tablette ?	Did you download the MOTUS mobile application from the app stores?		
	Ja, via de Play Store (Google)	Oui, depuis Play Store (Google)	Yes, via the Play Store (Google)	1	Evaluation6
	Ja, via de App store (Apple)	Oui, depuis l'App store (Apple)	Yes, via the App Store (Apple)	2	Evaluation6
	Neen, het installeren van de MOTUS-app is niet gelukt	Non, le téléchargement de l'application mobile MOTUS n'a pas abouti	No, the installation of the MOTUS app was unsuccessful	3	Evaluation2
Evaluation2	Neen, enkel en alleen deelgenomen via de MOTUS-webtoepassing	Non, j'ai utilisé seulement l'application web MOTUS	No, I only used the MOTUS web application	4	Evaluation4
	Geen antwoord	Pas de réponse	No answer	98	
	Waarom is het installeren van de MOTUS-app niet gelukt?	Pourquoi, selon vous, l'installation de l'application mobile MOTUS a-t-elle échoué ?	Why did you fail to install the MOTUS app?		
	Geef de belangrijkste reden aan.	Indiquez la raison principale.	Please indicate the main reason.		
Evaluation3	Problemen met de internetverbinding	Problèmes de connexion réseau	Problems with the internet connection	1	Evaluation13
	Te weinig beschikbare opslagruimte	Trop peu d'espace de stockage disponible	Too little available storage space	2	Evaluation13
	Verouderd besturingssysteem	Système d'exploitation obsolète	Outdated operating system	3	Evaluation13
	MOTUS-app niet gevonden	Je n'ai pas trouvé l'application mobile MOTUS	MOTUS app not found	4	Evaluation13
	Andere reden	Autre raison	Other reason	5	Evaluation3
	Ik weet het niet	Je ne sais pas	I don't know	97	Evaluation13
	Geen antwoord	Pas de réponse	No answer	98	
Evaluation4	Kan u kort iets meer vertellen waarom het installeren van de MOTUS-app niet gelukt is?	Pourquoi, selon vous, l'installation de l'application mobile MOTUS a-t-elle échoué ?	Can you briefly explain why the installation of the MOTUS app did not succeed?		
	Open veld (max 300 karakters)	Champ libre (max 300 caractères)	Open field (max 300 characters)		Evaluation13
Evaluation4	Geen antwoord	Pas de réponse	No answer	98	
	Waarom hebt u de MOTUS-app niet geïnstalleerd?	Pourquoi n'avez-vous pas téléchargé l'application mobile MOTUS ?	Why haven't you installed the MOTUS app?		
	Geef de belangrijkste reden aan.	Indiquez la raison principale.	Please indicate the main reason.		
	Geen smartphone of tablet met internet ter beschikking	Pas de smartphone ou de tablette avec internet	No smartphone or tablet with internet available	1	Evaluation13
	Niet digitaal vaardig (genoeg) om een app te downloaden	Pas ou pas suffisamment de compétences numériques pour télécharger une application	Not digitally proficient (enough) to download an app	2	Evaluation13
	Twijfel over de toegevoegde waarde van apps in het algemeen	Des doutes sur la valeur ajoutée des applications en général	Doubt about the added value of apps in general	3	Evaluation13
	Twijfel over de toegevoegde waarde van deze specifieke app	Des doutes sur la valeur ajoutée de cette application en particulier	Doubt about the added value of this app	4	Evaluation13

Evaluation5	Weinig of geen vertrouwen in hoe apps in het algemeen omgaan met persoonlijke gegevens	Peu ou pas de confiance dans la façon dont les applications traitent les données personnelles en général	Little or no confidence in how apps handle personal data in general	5	Evaluation13
	Weinig of geen vertrouwen in hoe deze app omgaat met persoonlijke gegevens	Peu ou pas de confiance dans la façon dont cette application traite les données personnelles	Little or no confidence in how this app handles personal data	6	Evaluation13
	Andere reden	Autre raison	Other Reason	7	Evaluation5
	Geen antwoord	Pas de réponse	No answer	98	
	Kan u kort iets meer vertellen waarom u de MOTUS-app niet hebt geïnstalleerd?	Pouvez-vous nous en dire plus pourquoi vous n'avez pas téléchargé l'application mobile MOTUS ?	Can you briefly explain why you have not installed the MOTUS app?		
Evaluation6	Open veld (max 300 karakters)	Champ libre (max 300 caractères)	Open field (max 300 characters)		Evaluation13
	Geen antwoord	Pas de réponse	No answer	98	
	Hoe gebruiksvriendelijk vindt u de MOTUS-app?	Comment jugez-vous l'ergonomie de l'application mobile MOTUS ? Est-elle facile à utiliser ?	How user-friendly do you find the MOTUS app?		
Evaluation7	Helemaal niet gebruiksvriendelijk	Pas du tout facile à utiliser	Not user-friendly at all	1	Evaluation7
	Niet erg gebruiksvriendelijk	Peu facile à utiliser	Not very user-friendly	2	Evaluation7
	Redelijk gebruiksvriendelijk	Assez facile à utiliser	Reasonably easy to use	3	Evaluation7
	Heel gebruiksvriendelijk	Très facile à utiliser	Very user-friendly	4	Evaluation7
	Uiterst gebruiksvriendelijk	Extrêmement facile à utiliser	Extremely user-friendly	5	Evaluation7
	Geen antwoord	Pas de réponse	No answer	98	
	Hoe beoordeelt u de MOTUS-app in het algemeen?	Comment jugez-vous l'application mobile MOTUS dans son ensemble ?	In general, how do you rate the MOTUS app?		
Evaluation8	Heel slecht	Très mauvais	Very Poor	1	Evaluation8
	Slecht	Mauvais	Bad	2	Evaluation8
	Redelijk	Ni bon ni mauvais	Fair	3	Evaluation8
	Goed	Bon	Good	4	Evaluation8
	Heel goed	Très bon	Very Good	5	Evaluation8
	Geen antwoord	Pas de réponse	No answer	98	
	Hebt u de MOTUS-webtoepassing gebruikt (www.motusresearch.io)?	Avez-vous utilisé l'application web MOTUS (www.motusresearch.io) ?	Have you used the MOTUS web application (www.motusresearch.io) ?		
Evaluation9	Ja	Oui	Yes	1	Evaluation9
	Nee	Non	No	2	Evaluation16
	Geen antwoord	Pas de réponse	No answer	98	
	De MOTUS-app en de MOTUS-webtoepassing zijn op elkaar afgestemd, wat het mogelijk maakt om beide afwisselend te gebruiken. Hebt u hiervan gebruik gemaakt?	L'application mobile MOTUS et l'application web MOTUS sont compatibles, ce qui permet de les utiliser en alternance. Est-ce que vous avez testé cet élément ?	The MOTUS app and the MOTUS web application are coordinated, which makes it possible to use them alternately. Have you used it?		
Evaluation10	Ja	Oui	Yes	1	Evaluation10
	Nee	Non	No	2	Evaluation10
	Geen antwoord	Pas de réponse	No answer	98	
	Hebt u een voorkeur voor één van beide toepassingen?	Avez-vous une préférence pour l'une des deux applications ?	Do you have a preference for one of the two applications?		
Evaluation11	Voorkeur voor de MOTUS-app	Préférence pour l'application mobile	Preference for the MOTUS app	1	Evaluation11
	Voorkeur voor de MOTUS-webtoepassing	Préférence pour l'application web	Preference for the MOTUS web application	2	Evaluation12
	Geen voorkeur	Aucune préférence	No preference	3	Evaluation13
	Geen antwoord	Pas de réponse	No answer	98	
	Kan u kort iets meer vertellen waarom u de MOTUS-app verkiest?	Pouvez-vous nous en dire plus pourquoi vous avez une préférence pour l'application mobile ?	Can you briefly tell us why you prefer the MOTUS app?		
Evaluation12	Open veld (max 300 karakters)	Champ libre (max 300 caractères)	Open field (max 300 characters)		Evaluation13
	Geen antwoord	Pas de réponse	No answer	98	
	Kan u kort iets meer vertellen waarom u de MOTUS-webtoepassing verkiest?	Pouvez-vous nous en dire plus pourquoi vous avez une préférence pour l'application web ?	Can you briefly tell us why you prefer the MOTUS web application?		
Evaluation13	Open veld (max 300 karakters)	Champ libre (max 300 caractères)	Open field (max 300 characters)		Evaluation13
	Geen antwoord	Pas de réponse	No answer	98	
	Welke webbrowser(s) hebt u gebruikt om de MOTUS-webtoepassing te bereiken?	Quel(s) navigateur(s) avez-vous utilisé(s) pour accéder à l'application web MOTUS?	What web browser(s) did you use to access the MOTUS web application?		
Evaluation14	Meerdere antwoorden mogelijk	Plusieurs réponses possibles	Multiple answers possible		
	Google Chrome	Google Chrome	Google Chrome	1	Evaluation14
	Microsoft Edge	Microsoft Edge	Microsoft Edge	2	Evaluation14
	Mozilla Firefox	Mozilla Firefox	Mozilla Firefox	3	Evaluation14
	Opera	Opera	Opera	4	Evaluation14
	Safari	Safari	Safari	5	Evaluation14
	Andere webbrowser	Autre	Other web browser	6	Evaluation14
	Geen antwoord	Pas de réponse	No answer	98	
	Hoe gebruiksvriendelijk vindt u de MOTUS-webtoepassing?	Comment jugez-vous l'ergonomie de l'application web MOTUS ? Est-elle facile à utiliser ?	How user-friendly do you find the MOTUS web application?		
	Helemaal niet gebruiksvriendelijk	Pas du tout facile à utiliser	Not user-friendly at all	1	Evaluation15
	Niet erg gebruiksvriendelijk	Peu facile à utiliser	Not very user-friendly	2	Evaluation15
	Redelijk gebruiksvriendelijk	Assez facile à utiliser	Reasonably easy to use	3	Evaluation15
	Heel gebruiksvriendelijk	Très facile à utiliser	Very user-friendly	4	Evaluation15
	Uiterst gebruiksvriendelijk	Extrêmement facile à utiliser	Extremely user-friendly	5	Evaluation15

Evaluation15	Geen antwoord Hoe beoordeelt u de MOTUS-webtoepassing in het algemeen? Heel slecht Slecht Redelijk Goed Heel goed	Pas de réponse Comment jugez-vous l'application web MOTUS dans son ensemble ? Très mauvais Mauvais Ni bon ni mauvais Bon Très bon	No answer In general, how do you rate the MOTUS web application? Very Poor Bad Fair Good Very Good	98 1 2 3 4 5		
						Evaluation16 Evaluation16 Evaluation16 Evaluation16 Evaluation16
Evaluation16	Geen antwoord Bij de start, na het lezen van de uitnodiging en de eerste instructies in de MOTUS-toepassing, was het duidelijk wat van u werd verwacht.	Pas de réponse Au début, après avoir lu l'invitation et les premières instructions dans l'application MOTUS, ce que l'on attendait de vous était clair.	No answer At the start, after reading the invitation and initial instructions in the MOTUS application, it was clear what was expected of you.	98		
	Helemaal niet mee eens Niet mee eens Mee eens Helemaal mee eens	Pas du tout d'accord Pas d'accord D'accord Tout à fait d'accord	Strongly disagree Disagree Agree Strongly agree	1 2 3 4		Evaluation17 Evaluation17 Evaluation18 Evaluation18
Evaluation17	Geen antwoord Welke onduidelijkheden waren er?	Pas de réponse Quelles étaient les ambiguïtés ?	No answer Which were the ambiguities?	98		
	Open veld (max 300 karakters) Geen antwoord	Champ libre (max 300 caractères) Pas de réponse	Open field (max 300 characters) No answer	98		Evaluation18
Evaluation18	De instructies voor het invullen van het dagboekje waren duidelijk.	Les instructions pour remplir le journal étaient claires.	The instructions for completing the diary were straightforward.			
	Helemaal niet mee eens Niet mee eens Mee eens Helemaal mee eens	Pas du tout d'accord Pas d'accord D'accord Tout à fait d'accord	Strongly disagree Disagree Agree Strongly agree	1 2 3 4		Evaluation19 Evaluation19 Evaluation20 Evaluation20
Evaluation19	Geen antwoord Welke onduidelijkheden waren er?	Pas de réponse Quelles étaient les ambiguïtés ?	No answer Which were the ambiguities?	98		
	Open veld (max 300 karakters) Geen antwoord	Champ libre (max 300 caractères) Pas de réponse	Open field (max 300 characters) No answer	98		Evaluation20
Evaluation20	De instructievideo waarin stap-voor-stap werd uitgelegd hoe het dagboekje moest worden ingevuld, was nuttig.	La vidéo d'instruction expliquant étape par étape comment remplir le journal a été utile.	The instructional video explaining step-by-step how to complete the diary was helpful.			
	Helemaal niet mee eens Niet mee eens Mee eens Helemaal mee eens	Pas du tout d'accord Pas d'accord D'accord Tout à fait d'accord	Strongly disagree Disagree Agree Strongly agree	1 2 3 4		Evaluation21 Evaluation21 Evaluation21 Evaluation21
Evaluation21	Ik heb de instructievideo niet bekeken Geen antwoord Het invullen van het dagboekje was eenvoudig.	Je n'ai pas regardé la vidéo d'instruction Pas de réponse Remplir le journal était facile.	I did not watch the instructional video No answer Completing the diary was easy.	5 98		Evaluation21
	Helemaal niet mee eens Niet mee eens Mee eens Helemaal mee eens	Pas du tout d'accord Pas d'accord D'accord Tout à fait d'accord	Strongly disagree Disagree Agree Strongly agree	1 2 3 4		Evaluation22 Evaluation22 Evaluation23 Evaluation23
Evaluation22	Geen antwoord Welke moeilijkheden waren er?	Pas de réponse Quels étaient les problèmes ?	No answer What difficulties were there?	98		
	Open veld (max 300 karakters) Geen antwoord	Champ libre (max 300 caractères) Pas de réponse	Open field (max 300 characters) No answer	98		Evaluation23
Evaluation23	De e-mails en berichten die automatisch werden verstuurd wanneer u een taak had afgerond en een nieuwe taak kon starten of wanneer u een taak uit het oog verloren had, zorgden voor een vlot verloop.	Les courriels et les messages qui étaient envoyés automatiquement lorsque vous aviez terminé une tâche et pouviez en commencer une nouvelle ou lorsque vous aviez perdu le fil d'une tâche, garantissaient une progression fluide.	The emails and messages that were automatically sent when you had completed a task and could start a new one or when you had lost track of a task ensured smooth progress.			
	Helemaal niet mee eens Niet mee eens Niet mee eens en niet mee oneens Mee eens Helemaal mee eens	Pas du tout d'accord Pas d'accord Ni d'accord, ni pas d'accord D'accord Tout à fait d'accord	Strongly disagree Disagree Neither agree nor disagree Agree Strongly agree	1 2 3 4 5		Evaluation24 Evaluation24 Evaluation24 Evaluation24 Evaluation24
Evaluation24	Geen antwoord De e-mails en berichten waren relevant.	Pas de réponse Les courriels et les messages étaient pertinents.	No answer The emails and messages were relevant.	98		
	Helemaal niet mee eens Niet mee eens Niet mee eens en niet mee oneens Mee eens Helemaal mee eens	Pas du tout d'accord Pas d'accord Ni d'accord, ni pas d'accord D'accord Tout à fait d'accord	Strongly disagree Disagree Neither agree nor disagree Agree Strongly agree	1 2 3 4 5		Evaluation25 Evaluation25 Evaluation25 Evaluation25 Evaluation25
Evaluation25	Geen antwoord De e-mails en berichten waren duidelijk.	Pas de réponse Les courriels et les messages étaient clairs.	No answer The emails and messages were straightforward.	98		
	Helemaal niet mee eens Niet mee eens Niet mee eens en niet mee oneens Mee eens Helemaal mee eens	Pas du tout d'accord Pas d'accord Ni d'accord, ni pas d'accord D'accord Tout à fait d'accord	Strongly disagree Disagree Neither agree nor disagree Agree Strongly agree	1 2 3 4 5		Evaluation26 Evaluation26 Evaluation26 Evaluation26 Evaluation26

Evaluation26	Geen antwoord	Pas de réponse	No answer	98	
	Het tijdsinterval tussen de e-mails en berichten was gepast.	L'intervalle de temps entre les courriels et les messages était approprié.	The time interval between emails and messages was appropriate.		
	Helemaal niet mee eens	Pas du tout d'accord	Strongly disagree	1	Evaluation27
	Niet mee eens	Pas d'accord	Disagree	2	Evaluation27
	Niet mee eens en niet mee oneens	Ni d'accord, ni pas d'accord	Neither agree nor disagree	3	Evaluation27
Evaluation27	Mee eens	D'accord	Agree	4	Evaluation27
	Helemaal mee eens	Tout à fait d'accord	Strongly agree	5	Evaluation27
	Geen antwoord	Pas de réponse	No answer	98	
	Meedoen aan het tijdsbestedingsonderzoek was interessante ervaring.	Participer à l'enquête sur l'emploi du temps a été une expérience intéressante.	Taking part in the time-use survey was interesting experience.		
	Helemaal niet mee eens	Pas du tout d'accord	Strongly disagree	1	Evaluation28
Evaluation28	Niet mee eens	Pas d'accord	Disagree	2	Evaluation28
	Niet mee eens en niet mee oneens	Ni d'accord, ni pas d'accord	Neither agree nor disagree	3	Evaluation28
	Mee eens	D'accord	Agree	4	Evaluation28
	Helemaal mee eens	Tout à fait d'accord	Strongly agree	5	Evaluation28
	Geen antwoord	Pas de réponse	No answer	98	
Evaluation29	Meedoen aan het tijdsbestedingsonderzoek was aangename ervaring.	Participer à l'enquête sur l'emploi du temps a été une expérience agréable.	Taking part in the time-use survey was pleasant experience.		
	Helemaal niet mee eens	Pas du tout d'accord	Strongly disagree	1	IF Evaluation1 = 1 or2 THEN Evaluation29 ELSE Evaluation33
	Niet mee eens	Pas d'accord	Disagree	2	IF Evaluation1 = 1 or2 THEN Evaluation29 ELSE Evaluation33
	Niet mee eens en niet mee oneens	Ni d'accord, ni pas d'accord	Neither agree nor disagree	3	IF Evaluation1 = 1 or2 THEN Evaluation29 ELSE Evaluation33
	Mee eens	D'accord	Agree	4	IF Evaluation1 = 1 or2 THEN Evaluation29 ELSE Evaluation33
Evaluation30	Helemaal mee eens	Tout à fait d'accord	Strongly agree	5	IF Evaluation1 = 1 or2 THEN Evaluation29 ELSE Evaluation33
	Geen antwoord	Pas de réponse	No answer	98	
	Zou u bereid zijn om deel te nemen aan een tijdsbestedingsonderzoek waarbij u wordt gevraagd uw locatie te delen via uw smartphone om op die manier uw verplaatsingsgedrag in kaart te brengen?	Seriez-vous prêt à participer à une enquête sur l'emploi du temps dans laquelle vous seriez invité à partager votre localisation via votre smartphone afin de cartographier votre comportement en matière de déplacements ?	Would you be willing to participate in a time-use survey where you are asked to share your location via your smartphone in order to map your travel behaviour?		
	Ja	Oui	Yes	1	Evaluation30
	Neen	Non	No	2	Evaluation31
Evaluation31	Geen antwoord	Pas de réponse	No answer	98	
	Kunnen we contact met u opnemen voor eventuele verdere vragen? Zo ja, geef dan a.u.b. uw e-mailadres.	Pouvons-nous vous contacter en cas de questions supplémentaires? Si oui, veuillez indiquer votre adresse e-mail.	Can we contact you for any further questions? If so, please provide your e-mail address.		
	Open veld (max 300 karakters)	Champ libre (max 300 caractères)	Open field (max 300 characters)		Evaluation33
	Geen antwoord	Pas de réponse	No answer	98	
	Waarom niet?	Pourquoi pas ?	Why not?		
Evaluation32	Geef de belangrijkste reden aan.	Indiquez la raison principale.	Please indicate the main reason.		
	Geen tijd	Pas le temps	No time	1	Evaluation33
	Geen interesse in het thema	Pas intéressé(e) par le thème	No interest in the topic	2	Evaluation33
	Niet digitaal vaardig (genoeg) om locatiegegevens te delen	Pas ou pas suffisamment de compétences numériques pour partager des données de localisation	Not digitally proficient (enough) to share location data	3	Evaluation33
	Niet bereid om locatiegegevens te delen wegens twijfel over de waarde van dit soort onderzoek	Refus de partager les données de localisation en raison de doutes sur la valeur de ce type de recherche	Unwilling to share location data due to doubts about the value of this type of research	4	Evaluation33
Evaluation33	Niet bereid om locatiegegevens te delen wegens gebrek aan vertrouwen in hoe apps omgaan met persoonlijke gegevens	Refus de partager les données de localisation en raison d'un manque de confiance dans la façon dont les applications traitent les données personnelles	Unwilling to share location data due to lack of trust in how apps handle personal data	5	Evaluation33
	Andere reden	Autre raison	Other Reason	6	Evaluation32
	Geen antwoord	Pas de réponse	No answer	98	
	Waarom niet?	Pourquoi pas ?	Why not?		
	Open veld (max 300 karakters)	Champ libre (max 300 caractères)	Open field (max 300 characters)		Evaluation33
Evaluation34	Geen antwoord	Pas de réponse	No answer	98	
	Nog opmerkingen, vragen of suggesties?	Des commentaires, des questions ou des suggestions ?	Any comments, questions or suggestions?		
	Open veld (max 600 karakters)	Champ libre (max 600 caractères)	Open field (max 600 characters)		End
	Geen antwoord	Pas de réponse	No answer	98	
	Dit was de laatste vraag. Als dank voor uw medewerking krijgt u een belastingvrije vergoeding van 15 euro. Er zal u een schuldvordering worden opgestuurd waarop u uw rekeningnummer kan noteren.	C'était la dernière question. En récompense du temps et de l'effort que vous avez consacrés à répondre aux questions, vous serez gratifié d'un dédommagement de 15 euros. Une déclaration de créance vous sera envoyée sur laquelle vous pourrez indiquer votre numéro de compte.	That was the last question. As a token of appreciation for your cooperation, you will receive a compensation of 15 euros. A debt-claim will be sent to you on which you can write down your account number.		
End	Klik op 'verzenden' om af te sluiten.	Cliquez sur 'envoyer'.	Click on 'submit'.		

Statbel organiseert een enquête bij de Belgische bevolking over hun tijdsbesteding. Uw deelname aan deze enquête is uitermate belangrijk voor de kwaliteit van het onderzoek.

Voornaam: |_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|

Naam: |_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|

Uw gebruikersnaam: |_|_|_|_|/|_|_|_|_|_|/|_|_|_| (*zie uitnodigingsbrief*)

Uw wachtwoord: |_|_|_|_|_|_|_|_| (*zie uitnodigingsbrief*)

De studie bestaat uit het beantwoorden van een vragenlijst en het bijhouden van uw dagelijkse activiteiten in een dagboek gedurende één week. Het dagboek is makkelijk in te vullen. Als u eerst de instructies leest en een blik werpt op de voorbeelden, wordt het zelfs kinderspel.

— — —

Instructies en voorbeeld dagboek invoegen

...

Dagboek invoegen

...

Gelieve deze vragenlijst in te vullen na de laatste dag waarop u uw dagboekje heb ingevuld!

Antwoorden doet u door de antwoordmogelijkheid die uw situatie het best weergeeft, aan te kruisen. In principe is er slechts één antwoord mogelijk per vraag. Als er meerdere antwoorden gegeven mogen worden, dan zal dit aangegeven worden in de vraag.

Mogen wij u vriendelijk vragen alle vragen te beantwoorden. Als in de verwijzing een vraagnummer vermeld staat (bijvoorbeeld “→ Ga naar vraag 25”), ga dan onmiddellijk over naar deze vraag. De tussenliggende vragen hoeft u niet te beantwoorden.

Informatie over uw huishouden

Vraag 1. Hoeveel volwassenen van 18 jaar of ouder maken deel uit van uw huishouden, uzelf inbegrepen? volwassenen

Vraag 2a. Hoeveel kinderen van 6 tot en met 17 jaar maken deel uit van uw huishouden? | | | kinderen van 6 tot en met 17 jaar

Vraag 2b. Hoeveel kinderen jonger of gelijk aan 5 jaar maken deel uit van uw huishouden? | | | kinderen jonger of gelijk aan 5 jaar

Vraag 3. In welk type huishouden woont u?

1. ☐ Eenpersoonshuishouden
2. ☐ Koppel zonder inwonende kinderen (+ eventueel inwonende andere personen)

Vraag 3a. Wat is uw positie in het huishouden? U bent ...

1. ☐ Partner (in koppel zonder inwonende kinderen)
2. ☐ Inwonende andere persoon

3. ☐ Koppel met inwonend(e) kind(eren) (+ eventueel inwonende andere personen)

Vraag 3b. Wat is uw positie in het huishouden? U bent ...

1. ☐ Partner (in koppel met inwonende kinderen)
2. ☐ Kind
3. ☐ Inwonende andere persoon

4. ☐ Eenoudergezin (+ eventueel inwonende andere personen)

Vraag 3c. Wat is uw positie in het huishouden? U bent ...

3. ☐ Inwonende andere persoon

5. ☐ Ander type huishouden (samenwonende broers/zussen, samenwonende vrienden, ...)

Informatie over uzelf

Vraag 4. Wat is uw geslacht?

1. ☐ Man
2. ☐ Vrouw
3. ☐ X
4. ☐ Zeg ik liever niet

Vraag 5. Wat is uw geboortedatum?

Uw arbeidsmarktsituatie

Vraag 6. Hebt u in de week die vorige zondag eindigde betaalde arbeid verricht, ook al was dit maar één uur?

(als werknemer of in uw eigen bedrijf of familiebedrijf of op de boerderij van uw gezin)

1. ☐ Ja → Ga naar vraag 9
2. ☐ Neen

Vraag 7. Was u tijdelijk afwezig wegens vakantie, ziekte of omwille van een andere reden?

1. ☐ Ja
2. ☐ Neen → Ga naar vraag 18

Vraag 8. Wat is de reden dat u die week helemaal niet werkte?

1. ☐ Eigen ziekte, letsel of tijdelijke arbeidsongeschiktheid
2. ☐ Vakantie
3. ☐ Moederschaps-, vaderschaps- of ouderschapsverlof
4. ☐ Verlof wegens studie
5. ☐ Arbeidsconflict
6. ☐ Technische of tijdelijke werkloosheid
7. ☐ Andere reden

Uw werksituatie

Als u meerdere jobs of meerdere ondernemingen hebt, is het de bedoeling dat u alleen over die job of onderneming vertelt waaraan u het grootste deel van uw werktijd besteedt. Indien voor alle jobs evenveel uren wordt gewerkt, dan wordt de job die het hoogste loon oplevert beschouwd als uw hoofdjob.

Vraag 9. Wat maakt of doet het bedrijf/de organisatie voornamelijk op de plaats waar u werkt?

Wees zo precies mogelijk. Bijvoorbeeld chemie, visserij, hotel/restaurant, gezondheidszorg en maatschappelijk werk, enz.

[illegible]

Vraag 10. Wat is uw functie in dit bedrijf/deze organisatie?

Wees zo precies mogelijk. Noteer bijvoorbeeld "secretaresse" in plaats van "medewerker", "timmerman" in plaats van "handarbeider", "leraar op een middelbare school", enz.

[illegible]

Vraag 11. Hoeveel uren per week werkt u gewoonlijk in deze job?

Overuren meerekenen. Als uw werkuren variëren van week tot week, gelieve dan een gemiddelde op te geven.

| | | | uren

Vraag 12. Tot welke beroepscategorie behoort u?

1. ☐ Private sector - arbeider
2. ☐ Private sector - bediende
3. ☐ Openbare sector - statutair ambtenaar
4. ☐ Openbare sector - contractueel
5. ☐ Vrij beroep / zelfstandige zonder personeel → Ga naar vraag 16
6. ☐ Vrij beroep / zelfstandige met personeel → Ga naar vraag 16
7. ☐ Helper / meewerkend familielid (zonder vergoeding) → Ga naar vraag 18

Vraag 13. Hebt u een vaste job of een contract van onbepaalde duur?

1. ☐ Ja, een vaste job of een contract van onbepaalde duur
2. ☐ Neen, een tijdelijke job of een contract van bepaalde duur of voor een bepaald werk

Vraag 14. Werkt u voltiids of deeltiids in deze job?

1. ☐ Voltijds
2. ☐ Deeltijds

Vraag 15. Hoeveel betaalde vakantiedagen hebt u per jaar?

De normale één of twee vrije dagen per week (in het weekend of door de week) worden niet beschouwd als betaalde vakantie.

| | | | dagen

Vraag 16. Hebt u, naast deze job, één of meerdere andere betaalde jobs?

1. ☐ Ja
2. ☐ Neen → Ga naar vraag 20

Vraag 17. Hoeveel uren per week besteedt u gewoonlijk aan deze andere job(s)?

Overuren meerekenen. Als uw werkuren variëren van week tot week, gelieve dan een gemiddelde op te geven.

|_|_| uren → Ga naar vraag 20

Werk zoeken

Vraag 18. Hebt u de afgelopen 4 weken iets ondernomen om een job te vinden, eventueel deeltijds of tijdelijk, of om een eigen zaak op te starten?

1. ☐ Ja
2. ☐ Neen, ik heb al een job gevonden die start binnen nu en 3 maanden → Ga naar vraag 20
3. ☐ Neen → Ga naar vraag 20

Vraag 19. Als u vandaag een job zou aangeboden worden, zou u dan binnen de twee weken kunnen beginnen werken?

1. ☐ Ja
2. ☐ Neen

Uw huidige dagelijkse activiteiten

Vraag 20. Met welk van de hierna vermelde situaties komt uw (werk)situatie het best overeen?

1. ☐ Ik heb een job
2. ☐ Ik ben werkloos
3. ☐ Ik ben (vervroegd) gepensioneerd
4. ☐ Ik ben arbeidsongeschikt als gevolg van langdurige gezondheidsproblemen
5. ☐ Ik ben leerling, student of in beroepsopleiding
6. ☐ Ik verzorg het huishouden (huisvrouw/huisman)
7. ☐ Ik zit in een andere situatie zonder werk

Uw schooltijd en studies

Vraag 21. Volgt u momenteel een opleiding, hetzij voltijds hetzij deeltijds?

1. ☐ Ja
2. ☐ Neen → Ga naar vraag 23

Vraag 22. Welk onderwijs volgt u momenteel?

1. ☐ Lager onderwijs
2. ☐ Secundair onderwijs

Vraag 22a. Gaat het om ...

1. ☐ Lager secundair onderwijs (1ste graad)
2. ☐ Algemeen secundair onderwijs 2de of 3de graad (ASO)
3. ☐ Technisch onderwijs 2de of 3de graad (TSO)
4. ☐ Kunstsecundair onderwijs 2de of 3de graad (KSO)
5. ☐ Beroepssecundair onderwijs 2de of 3de graad (BSO)
6. ☐ Postsecundair niet-hoger onderwijs

3. ☐ Hoger onderwijs

Vraag 22b. Gaat het om ...

1. ☐ Hoger Beroepsonderwijs (HBO5)
2. ☐ Hogescholenonderwijs van het korte type (1 cyclus) of graduaat (A1) of professionele bachelor
3. ☐ Academische bachelor (hogeschool of universiteit)
4. ☐ Voortgezette of aanvullende opleiding na graduaat of na bachelor (specialisatie, BanaBa, ...)
5. ☐ Schakelprogramma tussen professionele bachelor en master
6. ☐ Master (hogeschool of universiteit)
7. ☐ Voortgezette of aanvullende opleiding na master (specialisatie, ManaMa, ...)
8. ☐ Doctoraat met proefschrift

4. ☐ Andere

Vraag 23. Wat is het hoogste onderwijsniveau waarvoor u een diploma of certificaat hebt behaald?

Iemand die een onderwijsniveau succesvol beëindigt, krijgt een diploma, getuigschrift of certificaat dat erkend wordt door de overheid. Noteer het hoogst behaalde niveau tot dusver indien u nog naar school gaat of momenteel een opleiding volgt.

1. ☐ Geen → Ga naar vraag 26
2. ☐ Lager onderwijs → Ga naar vraag 26
3. ☐ Secundair onderwijs

Vraag 23a. Gaat het om ...

4. ☐ Hoger onderwijs
1. ☐ Lager secundair onderwijs (1ste graad)
 2. ☐ Algemeen secundair onderwijs 2de of 3de graad (ASO)
 3. ☐ Technisch onderwijs 2de of 3de graad (TSO)
 4. ☐ Kunstsecundair onderwijs 2de of 3de graad (KSO)
 5. ☐ Beroepssecundair onderwijs 2de of 3de graad (BSO)
 6. ☐ Postsecundair niet-hoger onderwijs

Vraag 23b. Gaat het om ...

2. ☐ Hoger Beroepsonderwijs (HBO5)
 2. ☐ Hogescholenonderwijs van het korte type (1 cyclus) of graduaat (A1) of professionele bachelor
 3. ☐ Academische bachelor (hogeschool of universiteit)
 4. ☐ Voortgezette of aanvullende opleiding na graduaat of na bachelor (specialisatie, BanaBa, ...)
 5. ☐ Hogescholenonderwijs van het lange type (2 cycli) of master aan een hogeschool
 6. ☐ Universitair onderwijs – licentiaat, master, ingenieur, dokter in de geneeskunde
 7. ☐ Voortgezette of aanvullende opleiding na master (specialisatie, ManaMa, ...)
 8. ☐ Doctoraat met proefschrift
5. ☐ Andere

Vraag 24. Wat is het vakgebied of onderwerp van uw hoogste diploma?

1. ☐ Algemeen programma
2. ☐ Onderwijs
3. ☐ Kunst, letteren en talen
4. ☐ Sociale wetenschappen, journalistiek en informatiewetenschappen
5. ☐ Bedrijfskunde, bestuurskunde en recht
6. ☐ Natuurwetenschappen, wiskunde en statistiek
7. ☐ Informatie- en communicatietechnologieën (ICT)
8. ☐ Ingenieurswetenschappen, industrie en bouwnijverheid
9. ☐ Landbouw, bosbouw, visserij en diergeneeskunde
10. ☐ Gezondheid en welzijn
11. ☐ Diensten

Vraag 25. Wat is de titel van uw hoogste diploma?

Vermeld de naam van de opleiding en uw eventuele major/minor.

[illegible]

Informatie over uw tijdsbesteding in een gewone week en uw gedachten en gevoelens die daarmee gepaard gaan

Vraag 26. In hoeverre bent u het eens met de volgende uitspraken?

	Volledig oneens	Oneens	Noch eens, noch oneens	Eens	Volledig eens
26a. Er wordt teveel van mij verwacht					
26b. Ik raak nooit bijgewerkt					
26c. Ik heb nooit tijd voor mezelf					
26d. Een dag heeft voor mij te weinig uren					
26e. Ik moet meer doen dan ik wil doen					
26f. Ik heb geen tijd om de dingen te doen die ik moet doen					
26g. Er wordt meer van mij verwacht dan ik aankan					
26h. Ik heb de indruk dat ik aan minder verplichtingen moet voldoen dan anderen					

Vraag 27. Hieronder staat een aantal beweringen over vrije tijd. Kan u aanduiden in welke mate elk van deze uitspraken overeenstemt met uw persoonlijke situatie?

	Volledig oneens	Oneens	Noch eens, noch oneens	Eens	Volledig eens
27a. Vaak kom ik in mijn vrije tijd niet toe aan dingen die ik eigenlijk wil doen					
27b. Ik moet in mijn vrije tijd te vaak rekening houden met anderen					
27c. Ik kan me moeilijk ontspannen in mijn vrije tijd					
27d. Ik heb teveel vrije tijd					
27e. Als ik vrij ben, zijn te veel vrijetijdsvoorzieningen (bv. zwembaden, clubs, musea, ...) gesloten					
27f. Het kost me veel moeite om mijn vrijetijdsactiviteiten te plannen					
27g. Er zijn zoveel dingen die ik wil doen in mijn vrije tijd, dat ik vaak het gevoel heb tijd tekort te komen					
27h. Teveel van mijn vrijetijdsactiviteiten zijn versnipperd					

Informatie over uw vrijetijdsbesteding tijdens het afgelopen jaar

Vraag 28. Hoe vaak hebt u deelgenomen aan onderstaande vrijetijds- en culturele activiteiten tijdens het afgelopen jaar?

	Nooit	Eén keer per jaar of minder	Meerdere keren per jaar	Ongeveer één keer per maand	Meerdere keren per maand	Ongeveer één keer per week	Meerdere keren per week
28a. Bioscoop bezoeken							
28b. Muziekconcert en/of -festival bijwonen							
28c. Podiumkunsten (ballet, dans, theater, toneel, stand-up comedy, musical, ...) bijwonen							
28d. Bibliotheek bezoeken							
28e. Sportwedstrijd als toeschouwer bijwonen							
28f. Tentoonstelling, museum bezoeken							
28g. Eten en/of drinken in een restaurant, café, pub							
28h. Naar een shoppingcenter gaan							
28i. Jaarmarkt, beurs bezoeken							
28j. Uitstapjes maken, een stad, pretpark of dierentuin bezoeken							
28k. Ander amusement bijwonen							

Vraag 29. Hoe vaak hebt u deelgenomen aan onderstaande sportactiviteiten tijdens het afgelopen jaar?

	Nooit	Eén keer per jaar of minder	Meerdere keren per jaar	Ongeveer één keer per maand	Meerdere keren per maand	Ongeveer één keer per week	Meerdere keren per week
29a. Joggen, hardlopen, atletiek of gymnastiek							
29b. Wielersport of fietstochten							
29c. Fitness, yoga of lichaamsstraining							
29d. Zwemmen, aquajoggen of aquagym							
29e. Andere watersporten (roeien, waterpolo, surfen, ...)							
29f. Skiën, schaatsen of andere wintersporten							

Vraag 30. Hoe vaak voelt u zich gehaast?

- Vraag 31. Hoe is uw gezondheid in het algemeen?**

- Vraag 32. Hebt u een ziekte of gezondheidsprobleem dat al minstens 6 maanden duurt of naar verwachting nog minstens 6 maanden zal duren?**

- Vraag 33. Bent u als gevolg van een gezondheidsprobleem beperkt in activiteiten die mensen gewoonlijk doen?**

- Vraag 34. Bent u al minstens 6 maanden beperkt?**

1. ☐ Ja
2. ☐ Neen

Vraag 35. Wat is uw lengte? Vul lengte in centimeters in. |_|_|_| cm

Vraag 36. Wat is uw gewicht? Vul gewicht in kilogram in. | | | kg

Informatie over de dagen waarop u uw dagboekje hebt bijgehouden

Vraag 37. Is de week dat u het dagboekje moest invullen om één of andere reden een bijzondere week geweest?

1. ☐ Ja
2. ☐ Neen → Ga naar vraag 39

Vraag 38. Om welke reden was het een bijzondere week?

[illegible]

Vraag 39. Hebt u gedurende de week van de dagboekregistraties één of meerdere dagen vrijaf gehad?

1. ☐ Ja
2. ☐ Neen → Ga naar vraag 41

Vraag 40. Welke dag(en) had u in die week vrijaf?

Meerdere antwoorden mogelijk

1. ☐ Maandag
2. ☐ Dinsdag
3. ☐ Woensdag
4. ☐ Donderdag
5. ☐ Vrijdag
6. ☐ Zaterdag
7. ☐ Zondag

Vraag 41. Wanneer hebt u het dagboekje ingevuld?

[illegible]

Bedankt om deel te nemen aan het Belgisch onderzoek naar tijdsbesteding! Als dank voor uw medewerking krijgt u een belastingvrije vergoeding van 15 euro. Er zal u een schuldvordering worden opgestuurd waarop u uw rekeningnummer kan noteren.

Final report: Large field test Germany

SSI Project

April 22, 2025

Maren Fritz (University of Mannheim), Florian Keusch (University of Mannheim), Johannes Volk (Destatis), & Lasse Häufglöckner (Destatis)

Executive Summary

As part of Workpackage 2.1 of the Smart Survey Implementation (SSI) project, funded by Eurostat, we conducted a large field test with a 14-day version of the Household Budget Survey (HBS) in Germany. The field test was conducted by the University of Mannheim in close collaboration with the Federal Statistical Office of Germany (Destatis). The field test took place between November 7 and December 19, 2024, with *AusgabenAtlas*, an app based on the MOTUS platform developed by hbits. The app had a smart feature as a microservice that allowed participants to scan their receipts for entry into the expenditure diary. Out of a random sample of 7,049 residents in Germany who were invited via postal mail to participate in the study, 126 completed the 14-day diary, making for an AAPOR RR1 of 1.8%. We find nonparticipation bias for age and nationality. The invitation letters included a manipulation of how the camera feature of the app and the effort required to participate in the study was presented. This experimental variation had no effect on the RR1. Highlighting the camera function in the invitation letter led to a significant increase in the use of receipt scanning in the diary.

Study Design

Population

The target population of this study were German residents born between September 1, 1953, and August 31, 2006.

Sampling

A two-stage, register-based sampling approach was used.¹

1. A stratified sample of 61² German municipalities was drawn as Primary Sample Units (PSUs) with the following requirements

¹Ipsos Public Affairs was commissioned to draw the sample and send out survey invitations.

²Originally 60 registrar offices of the selected municipalities were contacted via e-mail by Ipsos and asked to provide samples for the University of Mannheim based on §46 of the Bundesmeldegesetzes (BMG).

1. At least three PSUs in every state
 2. At least one municipality with less than 20,000 inhabitants, at least one with 20,000 to less than 100,000 and at least one with 100,000 and more inhabitants
 3. One PSU per municipality, except for Berlin (4 PSU), Hamburg (2 PSUs) and Bremen (PSUs)
2. 7,078 residents aged between 18 and 70 (as of August 31, 2024) were randomly selected from the municipality registers in the 61 PSUs (117 individuals per PSU). Eventually, 29 individuals were excluded because they were under 18 years of age, or the age was not known. This leaves a gross sample of 7,049 people who were invited to participate in the study.

Study invitation

All sampled individuals were invited to participate in the study via postal mail. Invited sample members received a two-page letter containing study information and an URL plus a QR code that linked to the Apple App Store for iPhones and the Google Play store for Android devices. The letter also included pictograms that showed a stylized process of entering expenditure data in the app. We also included personalized login credentials (username and initial password) and information about the conditional incentive.

1,000 people were invited on November 7, 2024, as part of the soft launch. The full launch was introduced on November 18, 2024. Those who did not activate their account in the app received a postal reminder 10 days after the first invitation letter. The reminder letters contained the information that the last day to start participating would be December 4, 2024.

Invitation and reminder letters can be found in the Appendix A.

Invitation Experiment

The sample was randomly assigned to one of three experimental groups, all receiving invitation letters with different wording according to their experimental group. In Group 1, the effort associated with participation was stressed in the text (“This [the documentation of purchases] includes the prices and product categories for all purchased items.”) and the scan-function was mentioned in the text (“You can use the app to conveniently take a picture of your receipts.”) and in the pictogram that described the data entry in the app (see letters in Appendix A). In Group 2 only the scan-function was mentioned in the text and the pictogram. In Group 3 neither the required effort for participants nor the scan-function was mentioned in the text and the pictogram also did not show the scanning function (see three versions of the letters below).

The reminder letters included the same experimental variation as the initial invitations.

Municipalities that did not respond were first reminded by e-mail and then repeatedly contacted by phone until contact was established with all of them. Two municipalities declined to provide data due to technical or staffing reasons and had to be replaced. A third sample point provided data after a phone call was interpreted as a refusal and a replacement municipality had already been selected. Since the originally intended replacement municipality later also delivered data, the sample ended up containing 61 instead of 60 PSUs.

App registration

Invited sample members who decided to participate in the study had to download the app to their smartphone and log in with their username and the initial password from the invitation letter, then give consent to the general data protection declaration (see Appendix G), enter their personal e-mail address, change their password according to the password requirements (at least 10 characters with at least one upper and one lower case letter, at least one number, and at least one special character), and verify their e-mail address via the link sent to the e-mail address provided in the registration.

In-app questionnaires

After completion of the registration, respondents were asked to fill out three questionnaires (see Appendix B for full question wording and Appendix C for app screenshots):

- First, a questionnaire regarding personal information comprising 27 questions
- Second, a questionnaire about personal income & expenses with 12 questions.
- Third, a questionnaire with household information with 7 questions.

Respondents could skip individual questions but were only moved to the next questionnaire once the prior questionnaire was completed.

Expenses diary

Having completed the three questionnaires, participants were directed to the diary, where they could enter their expenses for 14 days (see Appendix C for app screenshots). For each entry, they could choose between using the scanning function to take a picture of a receipt with their smartphone and entering the expenditures manually.

- When scanning a receipt, an OCR (optical character recognition) microservice developed as part of the SSI project was used for extracting the total sum of the receipt and the store. All entries from scans were automatically classified in the COICOP³ category “groceries”. The results of the scanned entries were shown back to the respondents, and they could make changes to the amount, store, and the classification. Once approved by the respondent, a scanned entry was stored in the diary.
- When entering expenditures manually, respondents had to provide information for all individual items, including price, quantity, and a COICOP category (see Appendix H for COICOP classification).

Respondents had the option of stating that they did not have any expenses on certain days.

The last day of data entry in the app was on December 19, 2024 (at 19:30 CET).

³Classification of Individual Consumption According to Purpose (COICOP) is a Reference Classification published by the United Nations Statistics Division.

Evaluation

After 14 days of diary, respondents were asked to complete a short evaluation survey with 7 questions (see Appendix A).

Notifications

Participants were sent in-app push notifications after two days if no activity in a questionnaire or diary was recorded, and e-mail notifications were sent after four days if no activity within a questionnaire or diary was registered. After the first week of the diary, reminders were sent to all participants.

All notification messages can be found in Appendices D and E.

Incentives

All respondents who entered expenses or stated that they did not have any expenses on at least 7 days received an Amazon voucher of 20€. Codes for Amazon vouchers were sent via e-mail (see Appendix D).

Respondent support

We provided participants with support via e-mail and telephone (Monday to Friday, 9:30 to 11:30 and 14:00 to 16:00 CET) during the field period, and we set up a website to inform sample members about the study and help answer frequently asked questions (see Appendix F).

Ethics approval

The study design was approved by the ethics committee of the University of Mannheim in June 2024 (EK Mannheim 27/2024).

Results

Participation behavior

To achieve an understanding of participation behavior and drop-out patterns, we employ rates as defined by the American Association for Public Opinion Research (AAPOR, 2023; see Stadtmüller et al., 2019 for an adaptation for Germany). We first assigned AAPOR Codes to all cases in the gross sample (see Table 1). While there are no standard definitions specifically for smart surveys, we can apply the logic of web-push surveys where specifically named persons are contacted in one mode (e.g., postal mail) and are then pushed to participate in another mode, here a smartphone app. When calculating response rates in such a hybrid approach, disposition codes related to participant ineligibility or unknown eligibility are determined by considerations related to the sample frame, and among those who are eligible, interview disposition codes are determined by the data collection mode. Since app-based

surveys are not specified in the AAPOR Standard Definitions, we use the logic of web surveys here.

Table 1. AAPOR Codes for all cases in the gross sample in the field test

AAPOR Code	Description	Number of persons (n)
1.1	Complete ¹	126
1.2	Partial ²	119
2.12	Break-off questionnaire too incomplete to process	163
2.112	Known-respondent-level refusal	17
2.9	Miscellaneous	4
2.332	Respondent language problem	1
2.32	Physically or mentally unable/incompetent	1
3.2	Unknown if eligible respondent in unit	20
3.9	Other	12
3.231	Deutsche Post category: Acceptance refused	2
3.311	Deutsche Post category: Recipient/company cannot be identified at the address given	268
3.32	Deutsche Post category: Recipient has moved. Consent to forward the new address has not been given.	36
3.19	Nothing ever returned	6,280
Total		7,049

¹Cases were counted as completed if expenses were entered on all 14 days of the diary.

²Cases were counted as partial if expenses were entered on 7 to 13 days of the diary.

The AAPOR Response Rate 1 (RR1) is the number of complete interviews (I) – defined as completing the diary on 14 days – divided by the total number of cases, including both complete (I) and partial interviews (P), non-interviews (such as refusals and break-offs (R), non-contacts (NC) and other non-interviews (O)), and cases of unknown eligibility (UH, UR, UO). This rate provides a comprehensive view of overall participation, accounting for all outcomes. AAPOR RR1 indicates the proportion of the target population that ultimately provided complete data.

$$RR1 = \frac{I}{(I + P) + (R + NC + O) + (UH + UR + UO)}$$

AAPOR RR2 includes partial interviews (P), which we define as respondents who provided entries in the diary on between seven and 13 days, in the numerator.

$$RR2 = \frac{(I + P)}{(I + P) + (R + NC + O) + (UH + UR + UO)}$$

Specifically for our studies, we additionally calculate an app activation rate (AAR). That AAR is the number of all cases in which the app was activated and data collection started divided by all cases. Using the AAPOR Standard Definitions, the number of started cases is the sum of completed interviews (I), partial interviews (P), and the number of break-offs (all cases with code 2.12 among the refusals (R)). The app activation rate informs us about how successful the recruitment was in pushing sample members to start using the app.

$$AAR = \frac{I + P + R_{2.12}}{(I + P) + (R + NC + O) + (UH + UR + UO)}$$

A second rate that is not specifically defined in the AAPOR definitions but informative for the large field tests, is the break-off rate. We define the break-off rate as one minus the number of completed interviews (I) divided by the sum of completed interviews, partial interviews (P), and break-offs (Rs with code 2.12). This rate informs us about the proportion of people who started in the app but did not complete the entire survey.

$$BOR = 1 - \frac{I}{I + P + R_{2.12}}$$

Table 2 provides the results for all calculated participation rates in the field test.

Table 2. Break-off, App activation, and Response Rates in the field test

App activation rate (AAR; %)	Break-off rate (BOR; %)	Response Rate 1 (RR1; %)	Response Rate 2 (RR2; %)
5.7%	68.3%	1.8%	3.5%

Figure 1 shows the participation in the different stages of the project. Out of the 7,049 invited sample members, 399 (5.7%) activated the app, 356 (5.1%) completed the first survey, 300 (4.3%) completed all three surveys, 288 (4.1%) provided at least one entry to the diary, 245 (3.5%) completed the diary on at least seven days, and 126 (1.8%) completed the diary on all 14 days.

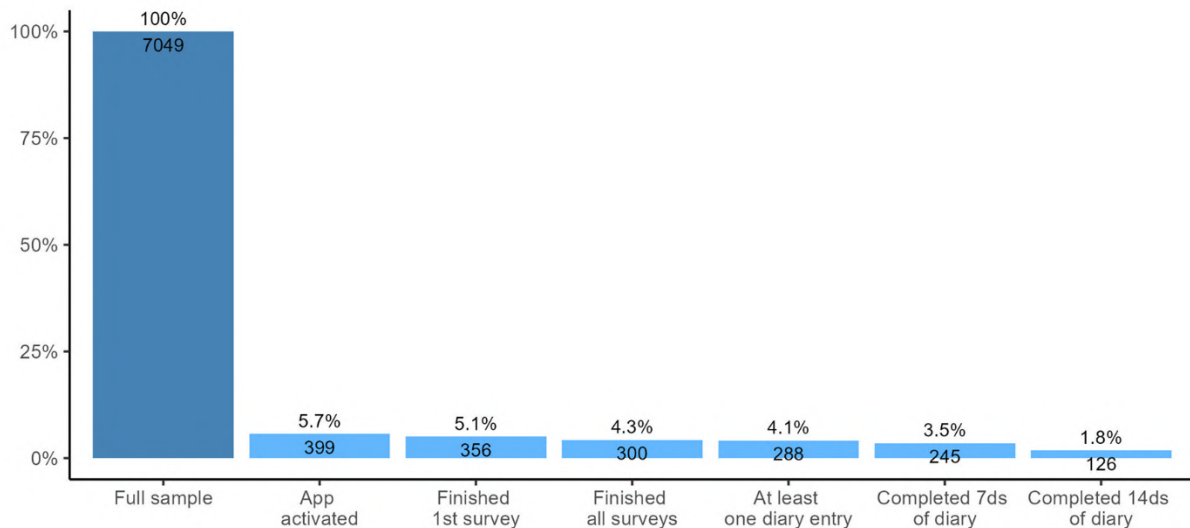


Figure 1. Participation rates in the app

Figure 2 shows the participation behavior broken down for the three experimental groups. Significantly fewer sample members activated the app when receiving the letter that mentioned the effort required to complete the study and the camera (4.4%) compared to the experimental groups that did not mention the effort but mentioned the camera (6.1%) and the group that did neither mention the effort nor the camera (6.5%; $\chi^2(2) = 10.35$, $p < 0.01$). However, there is no significant difference in the completion of the 14-day diary, that is, in RR1, between the three groups ($\chi^2(2) = 1.89$, $p > 0.05$).

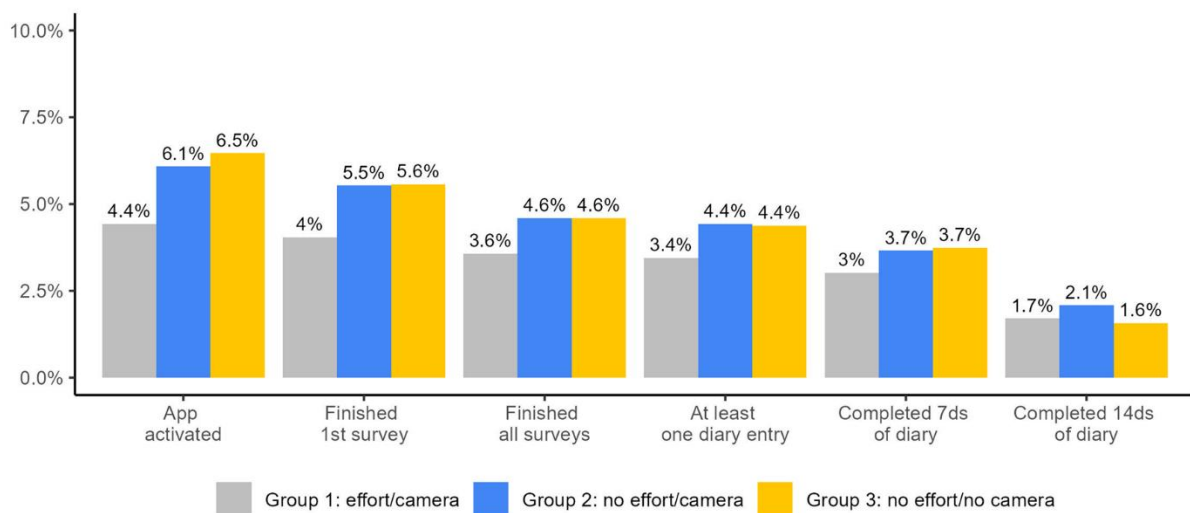


Figure 2. Participation rates in the app by experimental condition

Non-participation bias

Figure 3 shows the bias due to nonparticipation in the study, based on a comparison of the distribution of variables available on the sampling frame between the gross sample (i.e., sample members invited to participate in the study) and the net sample (i.e., those who completed seven days of the diary - left panel – and 14 days of the diary – right panel).

Following Couper et al. (2018), we calculate bias in a variable of interest (y) as the difference in proportions between the gross sample (g) and the net sample (n) as

$$bias(y) = y_n - y_g$$

The standard error (SE) of these differences is calculated as

$$SE(y_n - y_g) = \frac{n_g - n_n}{n_g} \sqrt{(var(y_n) + var(y_{np}))}$$

where np denotes non-participants.

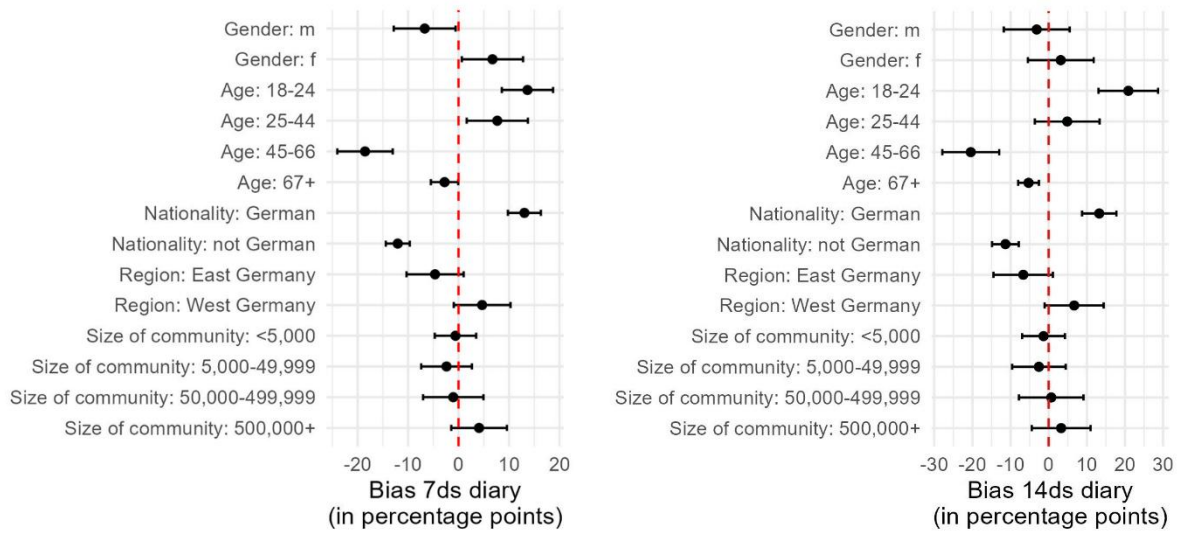


Figure 3. Non-participation bias in sociodemographic characteristics available on the sampling frame for seven- (left) and 14-day (right) diary completion (points: bias estimates; bars: 95% confidence intervals)

As for bias in seven-day diary completion (left panel Figure 3), we see women are 6.7 percentage points (p.p.) overrepresented in the net sample. People aged 18 to 24 years and 25 to 44 years are overrepresented by 13.6 p.p. and 7.7 p.p., respectively, and people aged 45 to 66 years and 67 years and older are underrepresented by -18.5 p.p. and -2.8 p.p., respectively. People with German nationality are 13.0 p.p. overrepresented in the net sample. No significant bias is found for region and size of the municipality. Looking at the non-participation bias for 14 days of diary completion (right panel in Figure 3), people aged 18 to 24 years are overrepresented by 20.9 p.p., while people aged 45 to 66 years and people aged 67 years and older are underrepresented by -20.4 p.p. and -5.3 p.p., respectively. People with

German nationality are 13.2 p.p. overrepresented in the net sample. No significant bias is found for the 14-day diary for gender, region, and size of the municipality.

Use of scanning function

59% of respondents who provided at least one diary entry used the camera at least once to enter a receipt. There are no significant differences between the different sociodemographic groups in the use of the camera function (gender: $\chi^2(1) = 1.13$, $p > 0.05$; age: $\chi^2(3) = 1.25$, $p > 0.05$; nationality: $\chi^2(1) = 0.04$, $p > 0.05$; region: $\chi^2(1) = 0.01$, $p > 0.05$; size of the municipality: $\chi^2(3) = 0.93$, $p > 0.05$).

As shown in Figure 4, mentioning the camera in the invitation letter significantly increased the likelihood of scanning at least one receipt (among all respondents who had provided at least one diary entry). 75.3% of respondents who had received the letter that mentioned the camera function and the effort used the scanning function at least once, 63.5% of respondents who had received the letter that mentioned the camera function but not the effort, and only 41.8% of respondents who had received the letter that did neither mention the camera function nor the effort ($\chi^2(2) = 22.44$, $p < 0.01$).

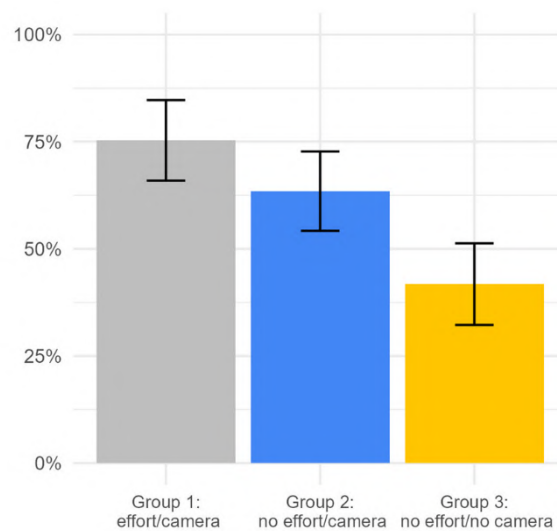


Figure 4. Percent of respondents who used the scanning function at least once by experimental condition (for all respondents who provided at least one diary entry)

Figure 5 shows that there were not only differences in whether or not the scanning function was used by experimental condition but also how often it was used. Respondents who had received the letter that mentioned the camera function and the effort scanned a total of 5.1 receipts on average, respondents who had received the letter that mentioned the camera function but not the effort scanned 3.8 receipts on average, and respondents who had received the letter that did neither mention the camera function nor the effort scanned only 1.9 receipts on average (ANOVA: $F(2, 171.73) = 10.47$, $p < 0.01$; see left panel in Figure 5).

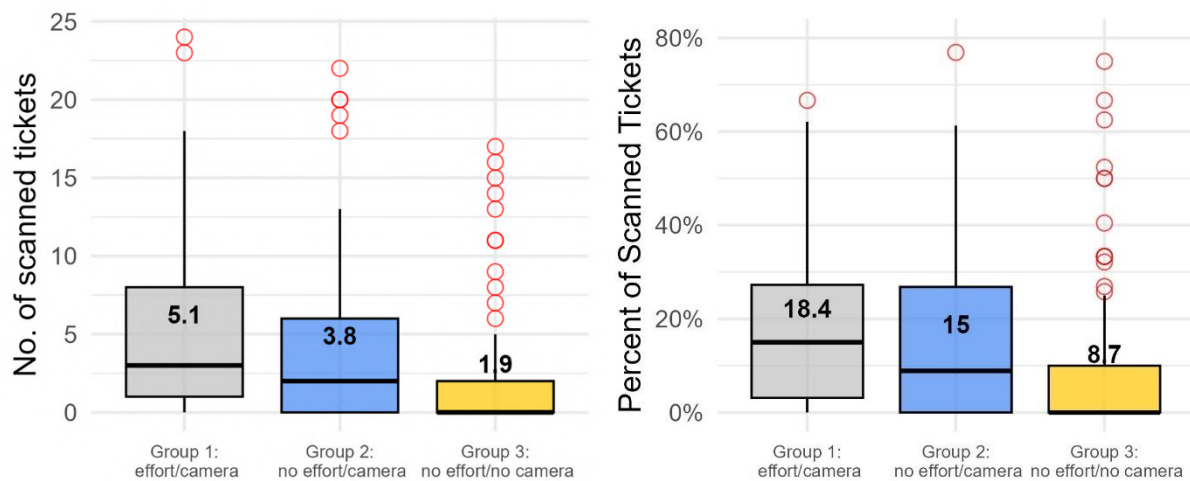


Figure 5. Number and percent of scanned tickets by experimental condition (for all respondents who provided at least one diary entry)

Mentioning the camera function in the invitation letter also influenced the share of scanned receipts among all diary entries (ANOVA: $F(2, 181.86) = 8.07, p < 0.01$; see right panel in Figure 5).

References:

- Couper M.P., Gremel, G., Axinn, W., Guyer, H., Wagner, J., & West, B.T. (2018). New options for national population surveys: The implications of internet and smartphone coverage. *Social Science Research* 73:221-235.
- Stadtmüller, S., Silber, H., Daikeler, J., Martin, S., Sand, M., Schmich, P., Schröder, J., Struminskaya, B., Weyandt, K. W., & Zabal, A. (2019). *Adaptation of the AAPOR Final Disposition Codes for the German Survey Context*. Mannheim, GESIS - Leibniz-Institute for the Social Sciences (GESIS - Survey Guidelines).
- The American Association for Public Opinion Research. 2023 *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*. 10th edition. AAPOR.

List of Appendices

Appendix A: Invitation and reminder letters

Appendix B: Questionnaires

Appendix C: App screenshots

Appendix D: E-mail notification messages

Appendix E: In-app push notification messages

Appendix F: Website

Appendix G: General data protection declaration

Appendix H: COICOP classification

Universität Mannheim, LS Keusch - A5, 6 · 68131 Mannheim

val_TITEL val_VORNAME val_NACHNAME_ZUSATZ
val_NACHNAME
val_STRASSE val_HAUSNUMMER val_HAUSNUMMER_ZUSATZ
val_PLZ val_ORT val_ORTSTEIL

Mannheim, 22. April 2025

Wie viel Geld geben die Menschen in Deutschland aus? Helfen Sie uns, es herauszufinden!

Guten Tag val_VORNAME val_NACHNAME_ZUSATZ val_NACHNAME,

die Lebenshaltungskosten in Deutschland steigen stetig. Doch wie viel Geld geben Bürgerinnen und Bürger wofür aus? Der **AusgabenAtlas** möchte diese Frage wissenschaftlich untersuchen.

Ihr Beitrag zu dieser Studie zählt! Als zufällig ausgewählte Person helfen Sie mit Ihrer Teilnahme ein detailliertes Bild der Ausgaben in Deutschland zu erstellen.

Dafür bitten wir Sie, für einen Zeitraum von zwei Wochen **genaue Angaben zu allen von Ihnen gekauften Produkten und Dienstleistungen in der AusgabenAtlas App zu erfassen**. Dies beinhaltet für jedes gekaufte Produkt den Preis und die Produktkategorie. Dazu können Sie Ihre **Kassenbons in der App bequem abfotografieren**.

Als Dankeschön erhalten Sie einen **20€ Amazon-Gutschein** nach Abschluss der Studie. Ihre Daten behandeln wir vertraulich gemäß den datenschutzrechtlichen Vorgaben.

So einfach machen Sie mit

1. App "AusgabenAtlas" herunterladen

- QR-Code scannen
- Oder app.ausgabenatlas.de in den Browser eintippen
- Oder im AppStore nach "AusgabenAtlas" suchen



2. Mit persönlichen Zugangsdaten anmelden



Kennung: val_Kennung
Passwort: val_Passwort

3. Ausgaben in der App erfassen



Kassenbon
mitnehmen



Foto machen



Ausgaben
überblicken

Wir danken für Ihre wertvolle Unterstützung!

Prof. Dr. Florian Keusch

Weitere Informationen auf der Rückseite – bitte wenden.

Weitere Informationen zur Studie

- Die Universität Mannheim führt die Studie „AusgabenAtlas“ im Rahmen eines Forschungsprojektes durch.
- Ziel ist es, das Konsumverhalten und die monatlichen Ausgaben der Menschen in Deutschland zu analysieren. Solche Daten dienen als wissenschaftliche Grundlage für die Entwicklung einer gerechten Sozial- und Familienpolitik und ermöglichen den Vergleich mit anderen EU-Ländern.

Warum gerade ich?

- Ihre Adresse wurde zufällig von einem Einwohnermeldeamt ausgewählt, um eine repräsentative Stichprobe der Bevölkerung zu gewährleisten.
- Ihre Teilnahme ist entscheidend, um aussagekräftige Ergebnisse über die Lebenssituation in Deutschland zu erhalten, welche die gesamte Bevölkerung widerspiegeln.
- Die Teilnahme ist freiwillig.

Danke für Ihre Teilnahme!

- Als Dankeschön für die Teilnahme erhalten Sie einen 20€ Amazon-Gutschein nach Abschluss der Studie, wenn Sie an mindestens 7 der insgesamt 14 Tage Ihre Ausgaben eingetragen haben.
- Sofern Sie an einem Tag keine Ausgaben hatten, geben Sie für diesen Tag bitte „Ich hatte heute keine Ausgaben“ an. Auch das zählt als Eintrag.

Datenschutzgarantie

- Wir garantieren die Einhaltung aller Datenschutzbestimmungen.
- Alle Angaben werden ausschließlich für wissenschaftliche Zwecke erhoben und streng vertraulich behandelt. Diese Studie verfolgt keine gewerblichen Interessen und verpflichtet Sie zu nichts.
- Ausführliche weitere Hinweise zur Studie, sowie Datenschutzhinweise finden Sie unter:
<https://www.uni-mannheim.de/ausgabenatlas>

Kontaktmöglichkeiten

Bei Fragen erreichen Sie uns

- per E-Mail unter ausgabenatlas@uni-mannheim.de
- telefonisch unter 0621 181-2203 (Montag bis Freitag zwischen 9:30 und 11:30 und zwischen 14 und 16 Uhr – außer Mittwoch Vormittag)

Weitere Informationen finden Sie auf unserer Webseite

<https://www.uni-mannheim.de/ausgabenatlas>

val_TITEL val_VORNAME val_NACHNAME_ZUSATZ
val_NACHNAME
val_STRASSE val_HAUSNUMMER val_HAUSNUMMER_ZUSATZ
val_PLZ val_ORT val_ORTSTEIL

Mannheim, 22. April 2025

Wie viel Geld geben die Menschen in Deutschland aus? Helfen Sie uns, es herauszufinden!

Guten Tag val_VORNAME val_NACHNAME_ZUSATZ val_NACHNAME,

die Lebenshaltungskosten in Deutschland steigen stetig. Doch wie viel Geld geben Bürgerinnen und Bürger wofür aus? Der **AusgabenAtlas** möchte diese Frage wissenschaftlich untersuchen.

Ihr Beitrag zu dieser Studie zählt! Als zufällig ausgewählte Person helfen Sie mit Ihrer Teilnahme ein detailliertes Bild der Ausgaben in Deutschland zu erstellen.

Dafür bitten wir Sie, für einen Zeitraum von zwei Wochen **genaue Angaben zu allen von Ihnen gekauften Produkten und Dienstleistungen in der AusgabenAtlas App zu erfassen**. Dazu können Sie Ihre **Kassenbons in der App bequem abfotografieren**.

Als Dankeschön erhalten Sie einen **20€ Amazon-Gutschein** nach Abschluss der Studie. Ihre Daten behandeln wir vertraulich gemäß den datenschutzrechtlichen Vorgaben.

So einfach machen Sie mit

1. App "AusgabenAtlas" herunterladen

- QR-Code scannen
- Oder app.ausgabenatlas.de in den Browser eintippen
- Oder im AppStore nach "AusgabenAtlas" suchen



2. Mit persönlichen Zugangsdaten anmelden



Kennung:	val_Kennung
Passwort:	val_Passwort

3. Ausgaben in der App erfassen



Kassenbon
mitnehmen



Foto machen



Ausgaben
überblicken

Wir danken für Ihre wertvolle Unterstützung!

Prof. Dr. Florian Keusch

Weitere Informationen auf der Rückseite – bitte wenden.

Weitere Informationen zur Studie

- Die Universität Mannheim führt die Studie „AusgabenAtlas“ im Rahmen eines Forschungsprojektes durch.
- Ziel ist es, das Konsumverhalten und die monatlichen Ausgaben der Menschen in Deutschland zu analysieren. Solche Daten dienen als wissenschaftliche Grundlage für die Entwicklung einer gerechten Sozial- und Familienpolitik und ermöglichen den Vergleich mit anderen EU-Ländern.

Warum gerade ich?

- Ihre Adresse wurde zufällig von einem Einwohnermeldeamt ausgewählt, um eine repräsentative Stichprobe der Bevölkerung zu gewährleisten.
- Ihre Teilnahme ist entscheidend, um aussagekräftige Ergebnisse über die Lebenssituation in Deutschland zu erhalten, welche die gesamte Bevölkerung widerspiegeln.
- Die Teilnahme ist freiwillig.

Danke für Ihre Teilnahme!

- Als Dankeschön für die Teilnahme erhalten Sie einen 20€ Amazon-Gutschein nach Abschluss der Studie, wenn Sie an mindestens 7 der insgesamt 14 Tage Ihre Ausgaben eingetragen haben.
- Sofern Sie an einem Tag keine Ausgaben hatten, geben Sie für diesen Tag bitte „Ich hatte heute keine Ausgaben“ an. Auch das zählt als Eintrag.

Datenschutzgarantie

- Wir garantieren die Einhaltung aller Datenschutzbestimmungen.
- Alle Angaben werden ausschließlich für wissenschaftliche Zwecke erhoben und streng vertraulich behandelt. Diese Studie verfolgt keine gewerblichen Interessen und verpflichtet Sie zu nichts.
- Ausführliche weitere Hinweise zur Studie, sowie Datenschutzhinweise finden Sie unter:
<https://www.uni-mannheim.de/ausgabenatlas>

Kontaktmöglichkeiten

Bei Fragen erreichen Sie uns

- per E-Mail unter ausgabenatlas@uni-mannheim.de
- telefonisch unter 0621 181-2203 (Montag bis Freitag zwischen 9:30 und 11:30 und zwischen 14 und 16 Uhr – außer Mittwoch Vormittag)

Weitere Informationen finden Sie auf unserer Webseite

<https://www.uni-mannheim.de/ausgabenatlas>

Universität Mannheim, LS Keusch - A5, 6 · 68131 Mannheim

val_TITEL val_VORNAME val_NACHNAME_ZUSATZ
val_NACHNAME
val_STRASSE val_HAUSNUMMER val_HAUSNUMMER_ZUSATZ
val_PLZ val_ORT val_ORTSTEIL

Mannheim, 22. April 2025

Wie viel Geld geben die Menschen in Deutschland aus? Helfen Sie uns, es herauszufinden!

Guten Tag val_VORNAME val_NACHNAME_ZUSATZ val_NACHNAME,

die Lebenshaltungskosten in Deutschland steigen stetig. Doch wie viel Geld geben Bürgerinnen und Bürger wofür aus? Der **AusgabenAtlas** möchte diese Frage wissenschaftlich untersuchen.

Ihr Beitrag zu dieser Studie zählt! Als zufällig ausgewählte Person helfen Sie mit Ihrer Teilnahme ein detailliertes Bild der Ausgaben in Deutschland zu erstellen.

Dafür bitten wir Sie, für einen Zeitraum von zwei Wochen **genaue Angaben zu allen von Ihnen gekauften Produkten und Dienstleistungen in der AusgabenAtlas App zu erfassen**.

Als Dankeschön erhalten Sie einen **20€ Amazon-Gutschein** nach Abschluss der Studie. Ihre Daten behandeln wir vertraulich gemäß den datenschutzrechtlichen Vorgaben.

So einfach machen Sie mit

1. App "AusgabenAtlas" herunterladen

- QR-Code scannen
- Oder app.ausgabenatlas.de in den Browser eintippen
- Oder im AppStore nach "AusgabenAtlas" suchen



2. Mit persönlichen Zugangsdaten anmelden



Kennung: val_Kennung
Passwort: val_Passwort

3. Ausgaben in der App erfassen



Kassenbon
mitnehmen



Ausgaben eintippen



Ausgaben
überblicken

Wir danken für Ihre wertvolle Unterstützung!

Prof. Dr. Florian Keusch

Weitere Informationen auf der Rückseite – bitte wenden.

Weitere Informationen zur Studie

- Die Universität Mannheim führt die Studie „AusgabenAtlas“ im Rahmen eines Forschungsprojektes durch.
- Ziel ist es, das Konsumverhalten und die monatlichen Ausgaben der Menschen in Deutschland zu analysieren. Solche Daten dienen als wissenschaftliche Grundlage für die Entwicklung einer gerechten Sozial- und Familienpolitik und ermöglichen den Vergleich mit anderen EU-Ländern.

Warum gerade ich?

- Ihre Adresse wurde zufällig von einem Einwohnermeldeamt ausgewählt, um eine repräsentative Stichprobe der Bevölkerung zu gewährleisten.
- Ihre Teilnahme ist entscheidend, um aussagekräftige Ergebnisse über die Lebenssituation in Deutschland zu erhalten, welche die gesamte Bevölkerung widerspiegeln.
- Die Teilnahme ist freiwillig.

Danke für Ihre Teilnahme!

- Als Dankeschön für die Teilnahme erhalten Sie einen 20€ Amazon-Gutschein nach Abschluss der Studie, wenn Sie an mindestens 7 der insgesamt 14 Tage Ihre Ausgaben eingetragen haben.
- Sofern Sie an einem Tag keine Ausgaben hatten, geben Sie für diesen Tag bitte „Ich hatte heute keine Ausgaben“ an. Auch das zählt als Eintrag.

Datenschutzgarantie

- Wir garantieren die Einhaltung aller Datenschutzbestimmungen.
- Alle Angaben werden ausschließlich für wissenschaftliche Zwecke erhoben und streng vertraulich behandelt. Diese Studie verfolgt keine gewerblichen Interessen und verpflichtet Sie zu nichts.
- Ausführliche weitere Hinweise zur Studie, sowie Datenschutzhinweise finden Sie unter:
<https://www.uni-mannheim.de/ausgabenatlas>

Kontaktmöglichkeiten

Bei Fragen erreichen Sie uns

- per E-Mail unter ausgabenatlas@uni-mannheim.de
- telefonisch unter 0621 181-2203 (Montag bis Freitag zwischen 9:30 und 11:30 und zwischen 14 und 16 Uhr – außer Mittwoch Vormittag)

Weitere Informationen finden Sie auf unserer Webseite

<https://www.uni-mannheim.de/ausgabenatlas>

val_TITEL val_VORNAME val_NACHNAME_ZUSATZ
val_NACHNAME
val_STRASSE val_HAUSNUMMER val_HAUSNUMMER_ZUSATZ
val_PLZ val_ORT val_ORTSTEIL

Mannheim, 22. April 2025

Wie viel Geld geben die Menschen in Deutschland aus? Helfen Sie uns, es herauszufinden!

Guten Tag val_VORNAME val_NACHNAME_ZUSATZ val_NACHNAME,

vor kurzem haben wir Sie zur Teilnahme an der wissenschaftlichen Studie „AusgabenAtlas“ eingeladen.

Falls Sie bereits teilgenommen haben – vielen Dank! Mit Ihrer Teilnahme unterstützen Sie die wissenschaftliche Untersuchung der Frage, wieviel Geld Bürgerinnen und Bürger wofür ausgeben.

Falls Sie noch keine Gelegenheit hatten teilzunehmen, möchten wir Sie hiermit erneut sehr herzlich zur Teilnahme an unserer Studie einladen. Sie können noch bis spätestens 04.12.2024 mit der Studie starten.

Dafür bitten wir Sie, für einen Zeitraum von zwei Wochen **genaue Angaben zu allen von Ihnen gekauften Produkten und Dienstleistungen in der AusgabenAtlas App zu erfassen**. Dies beinhaltet für jedes gekaufte Produkt den Preis und die Produktkategorie. Dazu können Sie Ihre **Kassenbons in der App bequem abfotografieren**.

Als Dankeschön erhalten Sie einen **20€ Amazon-Gutschein** nach Abschluss der Studie. Ihre Daten behandeln wir vertraulich gemäß den datenschutzrechtlichen Vorgaben.

So einfach machen Sie mit

1. App „AusgabenAtlas“ herunterladen

- QR-Code scannen
- Oder app.ausgabenatlas.de in den Browser eintippen
- Oder im AppStore nach „AusgabenAtlas“ suchen



2. Mit persönlichen Zugangsdaten anmelden



Kennung: val_Kennung
Passwort: val_Passwort

3. Ausgaben in der App erfassen



Kassenbon
mitnehmen



Foto machen



Ausgaben
überblicken

Wir danken für Ihre wertvolle Unterstützung!

Prof. Dr. Florian Keusch

Weitere Informationen auf der Rückseite – bitte wenden.

Weitere Informationen zur Studie

- Die Universität Mannheim führt die Studie „AusgabenAtlas“ im Rahmen eines Forschungsprojektes durch.
- Ziel ist es, das Konsumverhalten und die monatlichen Ausgaben der Menschen in Deutschland zu analysieren. Solche Daten dienen als wissenschaftliche Grundlage für die Entwicklung einer gerechten Sozial- und Familienpolitik und ermöglichen den Vergleich mit anderen EU-Ländern.

Warum gerade ich?

- Ihre Adresse wurde zufällig von einem Einwohnermeldeamt ausgewählt, um eine repräsentative Stichprobe der Bevölkerung zu gewährleisten.
- Ihre Teilnahme ist entscheidend, um aussagekräftige Ergebnisse über die Lebenssituation in Deutschland zu erhalten, welche die gesamte Bevölkerung widerspiegeln.
- Die Teilnahme ist freiwillig. Sollten Sie sich nach Erhalt dieses Schreibens gegen eine Teilnahme entscheiden, werden wir Sie nicht noch einmal kontaktieren.

Danke für Ihre Teilnahme!

- Als Dankeschön für die Teilnahme erhalten Sie einen 20€ Amazon-Gutschein nach Abschluss der Studie, wenn Sie an mindestens 7 der insgesamt 14 Tage Ihre Ausgaben eingetragen haben.
- Sofern Sie an einem Tag keine Ausgaben hatten, geben Sie für diesen Tag bitte „Ich hatte heute keine Ausgaben“ an. Auch das zählt als Eintrag.

Datenschutzgarantie

- Wir garantieren die Einhaltung aller Datenschutzbestimmungen.
- Alle Angaben werden ausschließlich für wissenschaftliche Zwecke erhoben und streng vertraulich behandelt. Diese Studie verfolgt keine gewerblichen Interessen und verpflichtet Sie zu nichts.
- Ausführliche weitere Hinweise zur Studie, sowie Datenschutzhinweise finden Sie unter:
<https://www.uni-mannheim.de/ausgabenatlas>

Kontaktmöglichkeiten

Bei Fragen erreichen Sie uns

- per E-Mail unter ausgabenatlas@uni-mannheim.de
- telefonisch unter 0621 181-2203 (Montag bis Freitag zwischen 9:30 und 11:30 und zwischen 14 und 16 Uhr – außer Mittwoch Vormittag)

Weitere Informationen finden Sie auf unserer Webseite

<https://www.uni-mannheim.de/ausgabenatlas>

val_TITEL val_VORNAME val_NACHNAME_ZUSATZ
val_NACHNAME
val_STRASSE val_HAUSNUMMER val_HAUSNUMMER_ZUSATZ
val_PLZ val_ORT val_ORTSTEIL

Mannheim, 22. April 2025

Wie viel Geld geben die Menschen in Deutschland aus? Helfen Sie uns, es herauszufinden!

Guten Tag val_VORNAME val_NACHNAME_ZUSATZ val_NACHNAME,

vor kurzem haben wir Sie zur Teilnahme an der wissenschaftlichen Studie „AusgabenAtlas“ eingeladen.

Falls Sie bereits teilgenommen haben – vielen Dank! Mit Ihrer Teilnahme unterstützen Sie die wissenschaftliche Untersuchung der Frage, wieviel Geld Bürgerinnen und Bürger wofür ausgeben.

Falls Sie noch keine Gelegenheit hatten teilzunehmen, möchten wir Sie hiermit erneut sehr herzlich zur Teilnahme an unserer Studie einladen. Sie können noch bis spätestens 04.12.2024 mit der Studie starten.

Dafür bitten wir Sie, für einen Zeitraum von zwei Wochen **genaue Angaben zu allen von Ihnen gekauften Produkten und Dienstleistungen in der AusgabenAtlas App zu erfassen**. Dazu können Sie Ihre **Kassenbons in der App bequem abfotografieren**.

Als Dankeschön erhalten Sie einen **20€ Amazon-Gutschein** nach Abschluss der Studie. Ihre Daten behandeln wir vertraulich gemäß den datenschutzrechtlichen Vorgaben.

So einfach machen Sie mit

1. App „AusgabenAtlas“ herunterladen

- QR-Code scannen
- Oder app.ausgabenatlas.de in den Browser eintippen
- Oder im AppStore nach „AusgabenAtlas“ suchen



2. Mit persönlichen Zugangsdaten anmelden



Kennung: val_Kennung
Passwort: val_Passwort

3. Ausgaben in der App erfassen



Kassenbon
mitnehmen



Foto machen



Ausgaben
überblicken

Wir danken für Ihre wertvolle Unterstützung!

Prof. Dr. Florian Keusch

Weitere Informationen auf der Rückseite – bitte wenden.

Weitere Informationen zur Studie

- Die Universität Mannheim führt die Studie „AusgabenAtlas“ im Rahmen eines Forschungsprojektes durch.
- Ziel ist es, das Konsumverhalten und die monatlichen Ausgaben der Menschen in Deutschland zu analysieren. Solche Daten dienen als wissenschaftliche Grundlage für die Entwicklung einer gerechten Sozial- und Familienpolitik und ermöglichen den Vergleich mit anderen EU-Ländern.

Warum gerade ich?

- Ihre Adresse wurde zufällig von einem Einwohnermeldeamt ausgewählt, um eine repräsentative Stichprobe der Bevölkerung zu gewährleisten.
- Ihre Teilnahme ist entscheidend, um aussagekräftige Ergebnisse über die Lebenssituation in Deutschland zu erhalten, welche die gesamte Bevölkerung widerspiegeln.
- Die Teilnahme ist freiwillig. Sollten Sie sich nach Erhalt dieses Schreibens gegen eine Teilnahme entscheiden, werden wir Sie nicht noch einmal kontaktieren.

Danke für Ihre Teilnahme!

- Als Dankeschön für die Teilnahme erhalten Sie einen 20€ Amazon-Gutschein nach Abschluss der Studie, wenn Sie an mindestens 7 der insgesamt 14 Tage Ihre Ausgaben eingetragen haben.
- Sofern Sie an einem Tag keine Ausgaben hatten, geben Sie für diesen Tag bitte „Ich hatte heute keine Ausgaben“ an. Auch das zählt als Eintrag.

Datenschutzgarantie

- Wir garantieren die Einhaltung aller Datenschutzbestimmungen.
- Alle Angaben werden ausschließlich für wissenschaftliche Zwecke erhoben und streng vertraulich behandelt. Diese Studie verfolgt keine gewerblichen Interessen und verpflichtet Sie zu nichts.
- Ausführliche weitere Hinweise zur Studie, sowie Datenschutzhinweise finden Sie unter:
<https://www.uni-mannheim.de/ausgabenatlas>

Kontaktmöglichkeiten

Bei Fragen erreichen Sie uns

- per E-Mail unter ausgabenatlas@uni-mannheim.de
- telefonisch unter 0621 181-2203 (Montag bis Freitag zwischen 9:30 und 11:30 und zwischen 14 und 16 Uhr – außer Mittwoch Vormittag)

Weitere Informationen finden Sie auf unserer Webseite

<https://www.uni-mannheim.de/ausgabenatlas>

val_TITEL val_VORNAME val_NACHNAME_ZUSATZ
val_NACHNAME
val_STRASSE val_HAUSNUMMER val_HAUSNUMMER_ZUSATZ
val_PLZ val_ORT val_ORTSTEIL

Mannheim, 22. April 2025

Wie viel Geld geben die Menschen in Deutschland aus? Helfen Sie uns, es herauszufinden!

Guten Tag val_VORNAME val_NACHNAME_ZUSATZ val_NACHNAME,

vor kurzem haben wir Sie zur Teilnahme an der wissenschaftlichen Studie „AusgabenAtlas“ eingeladen.

Falls Sie bereits teilgenommen haben – vielen Dank! Mit Ihrer Teilnahme unterstützen Sie die wissenschaftliche Untersuchung der Frage, wieviel Geld Bürgerinnen und Bürger wofür ausgeben.

Falls Sie noch keine Gelegenheit hatten teilzunehmen, möchten wir Sie hiermit erneut sehr herzlich zur Teilnahme an unserer Studie einladen. Sie können noch bis spätestens 04.12.2024 mit der Studie starten.

Dafür bitten wir Sie, für einen Zeitraum von zwei Wochen **genaue Angaben zu allen von Ihnen gekauften Produkten und Dienstleistungen in der AusgabenAtlas App zu erfassen.**

Als Dankeschön erhalten Sie einen **20€ Amazon-Gutschein** nach Abschluss der Studie. Ihre Daten behandeln wir vertraulich gemäß den datenschutzrechtlichen Vorgaben.

So einfach machen Sie mit

1. App „AusgabenAtlas“ herunterladen

- QR-Code scannen
- Oder app.ausgabenatlas.de in den Browser eintippen
- Oder im AppStore nach „AusgabenAtlas“ suchen



2. Mit persönlichen Zugangsdaten anmelden



Kennung: val_Kennung
Passwort: val_Passwort

3. Ausgaben in der App erfassen



Kassenbon
mitnehmen



Ausgaben eintippen



Ausgaben
überblicken

Wir danken für Ihre wertvolle Unterstützung!

Prof. Dr. Florian Keusch

Weitere Informationen auf der Rückseite – bitte wenden.

Weitere Informationen zur Studie

- Die Universität Mannheim führt die Studie „AusgabenAtlas“ im Rahmen eines Forschungsprojektes durch.
- Ziel ist es, das Konsumverhalten und die monatlichen Ausgaben der Menschen in Deutschland zu analysieren. Solche Daten dienen als wissenschaftliche Grundlage für die Entwicklung einer gerechten Sozial- und Familienpolitik und ermöglichen den Vergleich mit anderen EU-Ländern.

Warum gerade ich?

- Ihre Adresse wurde zufällig von einem Einwohnermeldeamt ausgewählt, um eine repräsentative Stichprobe der Bevölkerung zu gewährleisten.
- Ihre Teilnahme ist entscheidend, um aussagekräftige Ergebnisse über die Lebenssituation in Deutschland zu erhalten, welche die gesamte Bevölkerung widerspiegeln.
- Die Teilnahme ist freiwillig. Sollten Sie sich nach Erhalt dieses Schreibens gegen eine Teilnahme entscheiden, werden wir Sie nicht noch einmal kontaktieren.

Danke für Ihre Teilnahme!

- Als Dankeschön für die Teilnahme erhalten Sie einen 20€ Amazon-Gutschein nach Abschluss der Studie, wenn Sie an mindestens 7 der insgesamt 14 Tage Ihre Ausgaben eingetragen haben.
- Sofern Sie an einem Tag keine Ausgaben hatten, geben Sie für diesen Tag bitte „Ich hatte heute keine Ausgaben“ an. Auch das zählt als Eintrag.

Datenschutzgarantie

- Wir garantieren die Einhaltung aller Datenschutzbestimmungen.
- Alle Angaben werden ausschließlich für wissenschaftliche Zwecke erhoben und streng vertraulich behandelt. Diese Studie verfolgt keine gewerblichen Interessen und verpflichtet Sie zu nichts.
- Ausführliche weitere Hinweise zur Studie, sowie Datenschutzhinweise finden Sie unter:
<https://www.uni-mannheim.de/ausgabenatlas>

Kontaktmöglichkeiten

Bei Fragen erreichen Sie uns

- per E-Mail unter ausgabenatlas@uni-mannheim.de
- telefonisch unter 0621 181-2203 (Montag bis Freitag zwischen 9:30 und 11:30 und zwischen 14 und 16 Uhr – außer Mittwoch Vormittag)

Weitere Informationen finden Sie auf unserer Webseite

<https://www.uni-mannheim.de/ausgabenatlas>

Appendix B: Questionnaires

Questionnaire - Personal Information

Sequence	Question	Question type	Logic	Response options
1	Herzlich Willkommen zum AusgabenAtlas! Vielen Dank, dass Sie an unserer Studie teilnehmen und für einen Zeitraum von zwei Wochen alle Ihre persönlichen Ausgaben dokumentieren.	Static text		
2	Page break	Page break		
3	Zu Beginn möchten wir Ihnen einige Fragen zu Ihrer Person stellen.	Static text		
4	Die Ergebnisse dieser Befragung werden auch für unterschiedliche Altersgruppen ausgewertet. Bitte nennen Sie dazu Tag, Monat und Jahr Ihrer Geburt.	Year-month-day		
5	Welches Geschlecht haben Sie?	Select		Männlich Weiblich Divers
6	Wie lautet die Postleitzahl Ihres Wohnortes?	Numeric		
7	Welchen Familienstand haben Sie?	Select		Verheiratet und lebe mit meinem/meiner Ehepartner/-in zusammen In eingetragener Lebenspartnerschaft zusammenlebend (gleichgeschlechtlich) Verheiratet und lebe von meinem/meiner Ehepartner/-in getrennt Ledig Geschieden Verwitwet Eingetragene Lebenspartnerschaft, getrennt lebend (gleichgeschlechtlich) Eingetragene Lebenspartnerschaft aufgehoben (gleichgeschlechtlich) Eingetragene/r Lebenspartner/-in verstorben (gleichgeschlechtlich)

8	Page break	Page break		
9	Welchen höchsten allgemeinbildenden Schulabschluss haben Sie?	Select	filter1	Schüler/in, besuche eine allgemeinbildende Vollzeitschule Von der Schule abgegangen ohne Schulabschluss Hauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss Polytechnische Oberschule der DDR mit Abschluss der 8. oder 9. Klasse Realschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss Polytechnische Oberschule der DDR mit Abschluss der 10. Klasse Fachhochschulreife Abitur/Allgemeine oder fachgebundene Hochschulreife (Gymnasium bzw. EOS, auch EOS mit Lehre) Einen anderen Schulabschluss, und zwar:
10	Bitte geben Sie an, welchen anderen Schulabschluss Sie haben.	Short text		
11	Welchen allgemeinbildenden Schulabschluss streben Sie an?	Select	filter-2	Ich strebe aktuell keinen weiteren Schulabschluss an Hauptschulabschluss oder gleichwertigen Abschluss Realschulabschluss (Mittlere Reife) oder gleichwertigen Abschluss Fachhochschulreife Abitur/Allgemeine oder fachgebundene Hochschulreife Einen anderen Schulabschluss, und zwar:
12	Bitte geben Sie an, welchen anderen Schulabschluss Sie anstreben.	Short text		

13	Welche beruflichen Ausbildungsabschlüsse haben Sie? (Mehrfachnennungen möglich.)	Multiple choice	filter-3	<p>Noch in beruflicher Ausbildung (Berufsvorbereitungsjahr,Auszubildende/-r, Praktikant/-in, Student/-in)</p> <p>Schüler/-in und besuche eine berufsorientierte Aufbau-, Fachschule oder Ähnliches</p> <p>Keinen beruflichen Abschluss und bin nicht in beruflicher Ausbildung</p> <p>Beruflich-betriebliche Berufsausbildung (Lehre) abgeschlossen</p> <p>Berufsqualifizierender Abschluss einer beruflich-schulischen Ausbildung (Berufsfachschule, Kollegschule)</p> <p>Vorbereitungsdienst für den mittleren Dienst in der öffentlichen Verwaltung</p> <p>Abschluss einer einjährigen Ausbildung an einer Schule des Gesundheitswesens</p> <p>Abschluss einer zwei- bis dreijährigen Ausbildung an einer Schule des Gesundheitswesens</p> <p>Abschluss einer Ausbildung zum Erzieher/zur Erzieherin</p> <p>Abschluss einer Fachschule der DDR</p> <p>Abschluss einer Fach-, Meister-, Technikerschule, Verwaltungs- und Wirtschaftsakademie oder Fachakademie</p> <p>Bachelor</p> <p>Diplom</p> <p>Master, Magister, Staatsexamen</p> <p>Promotion</p> <p>Einen anderen beruflichen Abschluss, und zwar:</p>
14	Bitte geben Sie an, welchen weiteren beruflichen Ausbildungsabschluss Sie haben.	Short text		
15	Page break	Page break		

16	Welche Erwerbssituation passt für Sie? Bitte beachten Sie, dass unter Erwerbstätigkeit jede bezahlte bzw. mit einem Einkommen verbundene Tätigkeit verstanden wird.	Select	filter	Vollzeiterwerbstätig Teilzeiterwerbstätig Altersteilzeit (unabhängig davon, ob in der Arbeits- oder Freistellungsphase befindlich) Geringfügig erwerbstätig, 450-Euro-Job, Minijob Ein-Euro-Job (bei Bezug von Arbeitslosengeld II) Gelegentlich oder unregelmäßig beschäftigt In einer beruflichen Ausbildung/Lehre In Umschulung Freiwilliger Wehrdienst Bundesfreiwilligendienst oder Freiwilliges Soziales Jahr Mutterschafts-, Erziehungsurlaub, Elternzeit oder sonstige Beurlaubung Nicht erwerbstätig (einschließlich: Schüler/-innen oder Studierende, die nicht gegen Geld arbeiten, Arbeitslose, Vorruheständler/-innen, Rentner/-innen ohne Nebenverdienst)
17	Page break	Page break		
18	Wenn Sie nicht vollzeit- oder teilzeiterwerbstätig sind: Zu welcher Gruppe auf dieser Liste gehören Sie?	Select	filter	Schüler/-innen an einer allgemeinbildenden Schule Studenten/-innen Rentner/-innen, Pensionäre/-innen, im Vorruhestand Arbeitslose Dauerhaft Erwerbsunfähige Hausfrauen/Hausmänner Sonstiges, und zwar:
19	Bitte geben Sie Ihre sonstige Gruppe an.	Short text		
20	Welche berufliche Tätigkeit üben Sie derzeit hauptsächlich aus? Wenn Sie nicht mehr erwerbstätig sind, welche Tätigkeit haben Sie bei Ihrer früheren hauptsächlichsten Erwerbstätigkeit zuletzt ausgeübt? Bitte beschreiben Sie diese berufliche Tätigkeit genau.	Short text		
21	Hat dieser Beruf noch eine besondere Bezeichnung?	Select	filter-4	Nein Ja und zwar:

22	Bitte geben Sie die Bezeichnung Ihres Berufes an.	Short text		
23	Page break	Page break		
24	Welche berufliche Stellung haben oder hatten Sie in Ihrer hauptsächlich ausgeübten Erwerbstätigkeit? (Bitte nur eine Antwortoption wählen)	Select	filter1 filter2 filter3 filter4 filter5 filter6 filter7	Selbstständige/r Landwirt/-in bzw. Genossenschaftsbauer/-bäuerin Akademiker/-in in freiem Beruf (Arzt/Ärztin, Rechtsanwalt/-anwältin, Steuerberater/-in o.Ä..) und habe/hatte ... Selbstständige im Handel, im Gastgewerbe, im Handwerk, in der Industrie, der Dienstleistung, auch Ich-AG oder PGH-Mitglied und habe/hatte... Beamter/Beamtin, Richter/-in, Berufssoldat/-in, und zwar... Angestellte/-r, und zwar... Arbeiter/-in, und zwar... Andere, und zwar:
25	Page break	Page break		
26	Sie sind/waren Selbstständige/r Landwirt/-in bzw. Genossenschaftsbauer/-bäuerin. Was trifft auf Sie zu?	Select		Ich verfüge über eine landwirtschaftlich genutzte Fläche bis unter 10 ha Ich verfüge über eine landwirtschaftlich genutzte Fläche von 10 und mehr ha Ich bin/war Genossenschaftsbauer/-bäuerin (ehemals LPG)
27	Sie sind Akademiker/-in in freiem Beruf (Arzt/Ärztin, Rechtsanwalt/-anwältin, Steuerberater/-in o.Ä..) und haben/hatten ...	Select		keine weiteren Mitarbeiter/-innen 1 bis 4 Mitarbeiter/-innen 5 und mehr Mitarbeiter/-innen
28	Sie sind Selbstständige/r und haben/hatten . . .	Select		keine weiteren Mitarbeiter/-innen 1 bis 4 Mitarbeiter/-innen 5 und mehr Mitarbeiter/-innen Mitglied in der Produktionsgenossenschaft des Handwerks (PGH)

29	Sie sind Beamter/Beamtin, Richter/-in, Berufssoldat/-in, und zwar . . .	Select		im einfachen Dienst oder in einer vergleichbaren Laufbahngruppe im mittleren Dienst oder in einer vergleichbaren Laufbahngruppe im gehobenen Dienst oder in einer vergleichbaren Laufbahngruppe im höheren Dienst oder in einer vergleichbaren Laufbahngruppe
30	Sie sind Angestellte/-r, und zwar . . .	Select		mit ausführender Tätigkeit nach allgemeiner Anweisung (z. B. Verkäufer/-in, Datentypist/-in, Sekretariatsassistent/-in, Pflegehelfer/-in) mit einer qualifizierten Tätigkeit, die nach Anweisung erledigt wird (z. B. Sachbearbeiter/-in, Buchhalter/-in, technische/r Zeichner/-in) mit eigenständiger Leistung in verantwortlicher Tätigkeit oder Verantwortung für Personal (z. B. wissenschaftliche/r Mitarbeiter/-in, Abteilungsleiter/-in, bzw. Meister/-in im Angestelltenverhältnis) mit umfassenden Führungsaufgaben und Entscheidungsbefugnissen (z. B. Direktor/-in, Geschäftsführer/-in, Mitglied des Vorstandes)
31	Sie sind Arbeiter/-in, und zwar . . .	Select		ungelernt angelernt Facharbeiter/-in Vorarbeiter/-in, Kolonnenführer/-in Meister/-in, Polier/-in, Brigadier/-in
32	Welche andere berufliche Stellung haben Sie?	Select		In einer beruflichen Ausbildung/Lehre Mithelfende/r Familienangehörige/-r
33	Page break	Page break		

34	Bitte ordnen Sie den Betrieb, in dem Sie in Ihrer Haupterwerbstätigkeit tätig sind, einer Branche/einem Wirtschaftszweig zu. Richten Sie sich bitte nach dem wirtschaftlichen Schwerpunkt des Betriebes (nicht des gesamten Unternehmens).Für Selbstständige: Geben Sie bitte die Branche/den Wirtschaftszweig an, in dem Sie als Selbstständige/-r schwerpunktmäßig tätig sind.	Select		Land- und Forstwirtschaft, Fischerei Bergbau und Gewinnung von Erdöl, Erdgas, Steinen und Erden Verarbeitendes Gewerbe/Herstellung von Waren Energieversorgung Wasserversorgung; Abwasser- und Abfallentsorgung und Beseitigung von Umweltverschmutzung Baugewerbe, Hoch- und Tiefbau Groß- und Einzelhandel; Instandhaltung und Reparatur von Kraftfahrzeugen Personen- und Güterverkehr; Lagerei (auch Post- und Kurierdienst) Gastgewerbe/Beherbergung und Gastronomie Information und Kommunikation Banken/Finanz- und Versicherungsdienstleister Grundstücks- und Wohnungswesen Freiberufliche, wissenschaftliche und technische Dienstleistungen Sonstige wirtschaftliche Dienstleistungen für Unternehmen und Privatpersonen Öffentliche Verwaltung, Gerichte, Öffentliche Sicherheit und Ordnung, Verteidigung, Sozialversicherung Erziehung und Unterricht, Gesundheits- und Sozialwesen Sonstige überwiegend personenbezogene Dienstleistungen; allgemeine Reparaturen von Waren und Geräten Kunst, Unterhaltung, Sport und Erholung Gewerkschaft, Verband, Partei und sonstige Interessenvertretung, kirchliche und religiöse Vereinigung Konsulat, Botschaft, internationale und supranationale
35	Page break	Page break		
36	Wie viele Stunden arbeiten Sie normalerweise pro Woche?	Numeric		
37	Arbeiten Sie in Ihrer Haupterwerbstätigkeit im öffentlichen Dienst oder in der Privatwirtschaft?	Select		Öffentlicher Dienst Privatwirtschaft

Questionnaire - Income & Expenses

Sequence	Question	Question type	Logic	Response options
1	Angaben zu persönlichen Einkünften & Ausgaben. Nachdem Sie Angaben zu Ihrer Person gemacht haben, stellen wir Ihnen im Folgenden einige Fragen zu Ihren monatlichen Einkünften & Ausgaben.	Static text		
2	Page break	Page break		
3	Einkünfte aus selbstständiger bzw. landwirtschaftlicher Tätigkeit Haben Sie im vergangenen Monat Einkünfte aus selbstständiger oder landwirtschaftlicher Tätigkeit erzielt?	Select		Nein Ja
4	Page break	Page break		
5	Haben Sie diese Einkünfte als Selbstständige/-r oder als Landwirt/-in erzielt?	Select	filter	Selbstständige/-r, Landwirt/-in
6	Haben Sie diese Einkünfte haupt- oder nebenberuflich erzielt?	Select		Hauptberuflich, Nebenberuflich
7	Wie hoch waren die Entnahmen aus dem Betriebs-/Geschäftsvermögen für den Eigenverbrauch? Geben Sie bitte den Monatsbetrag in Brutto (Einkommen vor eventuellem Abzug von Steuern und Sozialversicherungsbeiträgen) an. Entnahmen (bar oder per Überweisung) aus dem Betriebs-/Geschäftsvermögen für den privaten Gebrauch. Bitte denken Sie bei Entnahmen auch an Ausgaben wie Steuern, Sozialversicherungsbeiträge, Versicherungsprämien.	Numeric		volle Euro
8	Page break	Page break		

9	Haben Sie sonstige Einkommen aus selbstständiger Erwerbstätigkeit (z. B. Honorare aus freiberuflicher Tätigkeit) erzielt?	Select	filter	Ja Nein
10	Beschreiben Sie diese Einkommen bitte genau:	Short text		
11	Wie hoch waren Ihre sonstigen Einkommen aus selbstständiger Erwerbstätigkeit? Geben Sie bitte den Monatsbetrag in Brutto (Einkommen vor eventuellem Abzug von Steuern und Sozialversicherungsbeiträgen) an.	Numeric		volle Euro
12	Page break	Page break		
13	Einkommen aus nicht selbstständiger Tätigkeit	Static text		
14	Haben Sie im vergangenen Monat Einkommen als Arbeitnehmer/-in erhalten? Hierzu gehören auch Minijobs und Besoldungen für Beamte und Beamtinnen.	Select	filter	Nein Ja
15	Page break	Page break		
16	Haben Sie im vergangenen Monat folgende Einkommen (Lohn/Gehalt) als Arbeitnehmer/-in erhalten? Geben Sie bitte das monatliche Einkommen in Brutto (Einkommen vor eventuellem Abzug von Steuern und Sozialversicherungsbeiträgen) an. Einschließlich Orts-, Familien-, Sonntags-, Feiertags-, Nachtarbeits-, Überstundenzuschlag, Wechselschicht-, Erschwerniszulagen, Nachzahlungen, Trinkgelder. Ohne einmalige Zahlungen, Arbeitgeberanteil zur Sozialversicherung, vermögenswirksame Leistungen des Arbeitgebers, Kindergeld.	Static text		
17	Grundlohn/-gehalt aus Haupterwerbstätigkeit	Select	filter	Nein Ja

18	Höhe Grundlohn/-gehalt aus Haupterwerbstätigkeit	Numeric		volle Euro
19	Grundlohn/-gehalt aus Nebenerwerbstätigkeit	Select	filter	Nein Ja
20	Höhe Grundlohn/-gehalt aus Nebenerwerbstätigkeit	Numeric		volle Euro
21	Altersteilzeitentgelt (Grundgehalt und Aufstockungsbetrag)	Select	filter	Nein Ja
22	Höhe Altersteilzeitentgelt	Numeric		volle Euro
23	Page break	Page break		
24	Haben Sie im vergangenen Monat eine oder mehrere der folgenden Sondervergütungen erhalten?Geben Sie bitte den Monatsbetrag in Brutto (Einkommen vor eventuellem Abzug von Steuern und Sozialversicherungsbeiträgen) an.	Static text		
25	Einmalige Zahlungen (z. B. Weihnachts-, Urlaubsgeld)	Select	filter	Nein Ja
26	Höhe Einmalige Zahlungen (z. B. Weihnachts-, Urlaubsgeld)	Numeric		volle Euro
27	Abfindungen, Entlassungs- und Übergangsgelder	Select	filter	Nein Ja
28	Höhe Abfindungen, Entlassungs- und Übergangsgelder	Numeric		volle Euro
29	Gewinnbeteiligungen (z. B. Bonuszahlung, Erfolgsprämien)	Select	filter	Nein Ja
30	Höhe Gewinnbeteiligungen (z. B. Bonuszahlung, Erfolgsprämien)	Numeric		volle Euro
31	Page break	Page break		
32	Einkünfte aus Renten/Pensionen	Static text		
33	Haben Sie im vergangenen Monat Renten/Pensionen aus eigenen Ansprüchen erhalten?	Select	filter	Nein Ja
34	Page break	Page break		

35	Welche Einkommen aus Renten/Pensionen aus eigenen Ansprüchen haben Sie im vergangenen Monat erhalten? Geben Sie bitte den Monatsbetrag in Brutto (Einkommen vor eventuellem Abzug von Steuern und Sozialversicherungsbeiträgen) an.	Static text		
36	Renten der gesetzlichen Rentenversicherung	Select	filter	Nein Ja
37	Höhe Renten der gesetzlichen Rentenversicherung	Numeric		volle Euro
38	Pensionen (einschließlich einmaliger Zahlungen, z. B. Weihnachtsgeld)	Select	filter	Nein Ja
39	Höhe Pensionen (einschließlich einmaliger Zahlungen, z. B. Weihnachtsgeld)	Numeric		volle Euro
40	Renten der Zusatzversorgungskassen des öffentlichen Dienstes	Select	filter	Nein Ja
41	Höhe Renten der Zusatzversorgungskassen des öffentlichen Dienstes	Numeric		volle Euro
42	Werks- bzw. Betriebsrenten, betriebliche Vorruhestandsgelder	Select	filter	Nein Ja
43	Höhe Werks- bzw. Betriebsrenten, betriebliche Vorruhestandsgelder	Numeric		volle Euro
44	Renten berufsständischer Versorgungswerke, landwirtschaftlicher Alterskassen, Landabgabereuten	Select	filter	Nein Ja
45	Höhe Renten berufsständischer Versorgungswerke, landwirtschaftlicher Alterskassen, Landabgabereuten	Numeric		volle Euro
46	Renten der gesetzlichen Unfallversicherung	Select	filter	Nein Ja
47	Höhe Renten der gesetzlichen Unfallversicherung	Numeric		volle Euro
48	Renten aus privater Lebens-/Rentenversicherung	Select	filter	Nein Ja
49	Höhe Renten aus privater Lebens-/Rentenversicherung	Numeric		volle Euro

50	Renten aus privaten Unfallversicherungen oder Haftpflichtansprüchen	Select	filter	Nein Ja
51	Höhe Renten aus privaten Unfallversicherungen oder Haftpflichtansprüchen	Numeric		volle Euro
52	Sonstige Renten/Pensionen (z. B. Auslandsrenten)	Select	filter	Nein Ja
53	Beschreiben Sie diese Rente bitte genau:	Short text		
54	Höhe Sonstige Renten/Pensionen (z. B. Auslandsrenten)	Numeric		volle Euro
55	Seitentrennung	Page break		
56	Abzüge und Beiträge	Static text		
57	Haben Sie im vergangenen Monat folgende Abzüge und Sozialversicherungsbeiträge gezahlt? Geben Sie bitte den gezahlten Monatsbetrag an.	Static text		
58	Einkommensteuer/Lohnsteuer (einschließlich Steuervorauszahlungen/ -nachzahlungen)	Select	filter	Nein Ja
59	Höhe Einkommensteuer/Lohnsteuer (einschließlich Steuervorauszahlungen/ -nachzahlungen)	Numeric		volle Euro
60	Kirchensteuer (auch Nachzahlungen)	Select	filter	Nein Ja
61	Höhe Kirchensteuer (auch Nachzahlungen)	Numeric		volle Euro
62	Solidaritätszuschlag (auch Nachzahlungen)	Select	filter	Nein Ja
63	Höhe Solidaritätszuschlag (auch Nachzahlungen)	Numeric		volle Euro
64	Gesetzliche Rentenversicherung: Pflichtbeiträge	Select	filter	Nein Ja
65	Höhe Gesetzliche Rentenversicherung: Pflichtbeiträge	Numeric		volle Euro
66	Gesetzliche Rentenversicherung: Freiwillige Beiträge	Select	filter	Nein Ja
67	Höhe Gesetzliche Rentenversicherung: Freiwillige Beiträge	Numeric		volle Euro

68	Beiträge zur Zusatzversorgung im öffentlichen Dienst (ZÖD) z. B. VBL-Arbeitnehmeranteil oder zur betrieblichen Altersversorgung als Entgeltumwandlung (Alters-/Pensionskassen, Pensionsfonds, Direktversicherungen)	Select	filter	Nein Ja
69	Höhe Beiträge zur Zusatzversorgung im öffentlichen Dienst (ZÖD) z. B. VBL-Arbeitnehmeranteil oder zur betrieblichen Altersversorgung als Entgeltumwandlung (Alters-/Pensionskassen, Pensionsfonds, Direktversicherungen)	Numeric		volle Euro
70	Gesetzliche Krankenversicherung (GKV): Pflichtbeiträge einschließlich Zusatzbeiträge (auch Rentner/-in)	Select	filter	Nein Ja
71	Höhe Gesetzliche Krankenversicherung (GKV): Pflichtbeiträge einschließlich Zusatzbeiträge (auch Rentner/-in)	Numeric		volle Euro
72	Gesetzliche Krankenversicherung (GKV): Freiwillige Beiträge (auch Rentner/-in)	Select	filter	Nein Ja
73	Höhe Gesetzliche Krankenversicherung (GKV): Freiwillige Beiträge (auch Rentner/-in)	Numeric		volle Euro
74	Beiträge zur privaten Krankenversicherung einschließlich Zuschuss des Arbeitgebers (auch Rentner/-in)	Select	filter	Nein Ja
75	Höhe Beiträge zur privaten Krankenversicherung einschließlich Zuschuss des Arbeitgebers (auch Rentner/-in)	Numeric		volle Euro
76	Pflichtbeiträge zur sozialen oder privaten Pflegeversicherung (auch Rentner/-in)	Select	filter	Nein Ja
77	Höhe Pflichtbeiträge zur sozialen oder privaten Pflegeversicherung (auch Rentner/-in)	Numeric		volle Euro
78	Arbeitslosenversicherung (auch freiwillige Beiträge)	Select	filter	Nein Ja

79	Höhe Arbeitslosenversicherung (auch freiwillige Beiträge)	Numeric		volle Euro
80	Beiträge vermögenswirksamer Leistungen (Arbeitgeberanteil)	Select	filter	Nein Ja
81	Höhe Beiträge vermögenswirksamer Leistungen (Arbeitgeberanteil)	Numeric		volle Euro
82	Beiträge vermögenswirksamer Leistungen (Arbeitnehmeranteil)	Select	filter	Nein Ja
83	Höhe Beiträge vermögenswirksamer Leistungen (Arbeitnehmeranteil)	Numeric		volle Euro
84	Sonstige Abzüge (z. B. Lohn-/Gehaltspfändungen)	Select	filter	Nein Ja
85	Höhe Sonstige Abzüge (z. B. Lohn-/Gehaltspfändungen)	Numeric		volle Euro
86	Seitentrennung	Page break		
87	Sie haben alle Fragen zu Ihren monatlichen Einkünften & Ausgaben beantwortet. Im Anschluss folgt ein letzter Fragebogen zu Ihrer persönlichen Haushalts- und Wohnsituation.	Static text		

Questionnaire - Household Information

Sequence	Question	Question type	Logic	Response options
1	Angaben zum Haushalt Nachdem Sie Angaben zu Ihrer Person sowie zu Ihren Einkünften & Ausgaben gemacht haben, stellen wir Ihnen im letzten Fragebogen einige Fragen zu Ihrer Haushalts- und Wohnsituation.	Static text		
2	Page break	Page break		
3	Die folgenden Fragen beziehen sich auf Ihre aktuelle, persönliche Wohnsituation und auf die von Ihrem Haushalt überwiegend genutzte Wohnung (Hauptwohnung).	Static text		
4	In welcher Art von Gebäude wohnen Sie? (Wohngebäude dienen überwiegend zu Wohnzwecken. Immobilien gelten auch dann als Einfamilienhäuser, wenn sich darin eine Einliegerwohnung befindet. Andere Arten von Gebäuden sind überwiegend für Nichtwohnzwecke, nämlich für gewerbliche, soziale, kulturelle oder Verwaltungszwecke bestimmte Gebäude mit mindestens einer Wohneinheit.)	Select		Freistehendes Einfamilienhaus, Einfamilienhaus als Doppelhaushälfte oder Reihenhaushaus, Einfamilienhaus mit zusätzlicher Einliegerwohnung oder Zweifamilienhaus, Wohngebäude mit 3 bis 9 Wohnungen, Wohngebäude mit 10 oder mehr Wohnungen, Wohnheim, Andere Art von Gebäude

5	In welcher Wohnform nutzen Sie Ihre Hauptwohnung? (Zu „Hypothekenschulden“ zählen auch die Rückzahlungen von Bauspardarlehen. „Vergünstigte Miete“ bedeutet, dass der Haushalt z. B. mit einem Wohnberechtigungsschein in einer Sozialwohnung oder in einem Studierendenwohnheim wohnt. Mietfrei trifft nicht zu, wenn die Miete für die Hauptwohnung von Dritten (z. B. Arbeitsagentur, Sozialamt, Eltern für ihre Kinder) gezahlt wird.)	Select		"Als Eigentümer/-in (ohne Hypothekenschulden)", "Als Eigentümer/-in (mit Hypothekenschulden)", "Als Mieter/-in, Untermieter/-in (Miete zu Marktpreisen)", "Als Mieter/-in, Untermieter/-in (vergünstigte Miete)", "Mietfrei in einer Werkswohnung", "Mietfrei in einer sonstigen Wohnung bzw. einem Haus"
6	Wie viele Quadratmeter Wohnfläche hat Ihre Hauptwohnung? (Wohnfläche Zur Wohnfläche zählen die Flächen folgender Räume: Wohn- und Schlafräum, Küchen, Nebenräume (Bad, Toilette, Flur usw.), Balkone, Terrassen, Loggien Nicht zu berücksichtigen sind: Ausschließlich gewerblich genutzte Flächen sowie Keller-, Boden- und Wirtschaftsräume, die nicht zu Wohnzwecken genutzt werden.)	Numeric		
7	Wie viele Personen leben ständig in Ihrem Haushalt, Sie selbst eingeschlossen? Zu diesem Haushalt zählen alle Personen, die hier gemeinsam wohnen und wirtschaften. Denken Sie dabei bitte auch an alle im Haushalt lebenden Kinder.	Select	filter,filter 2	"Eine Person", "Mehrere Personen, und zwar:"
8	Anzahl Personen im Haushalt	Numeric		
9	Wie viele der in Ihrem Haushalt lebenden Personen sind unter 18 Jahre alt?	Numeric		

10	Was ist Ihre Rolle in Ihrem Haushalt?	Select		Mutter, Vater, Partner, Ehepartner, Kind, Andere
11	Welche der nachfolgend gelisteten Versicherungen sind in Ihrem Haushalt vorhanden? Bei „Kombiversicherungen“ kreuzen Sie bitte die Versicherungen an, die in der „Kombiversicherung“ enthalten sind, z. B. bei einer „Familierversicherung“: „Private Haftpflichtversicherung“, „Hausratversicherung“ und „Private Unfallversicherung“.Bitte alles Zutreffende ankreuzen.	Multiple choice		"Berufsunfähigkeitsversicherung (auch als Zusatzversicherung)", "Risikolebensversicherung", "Kapitalbildende Lebensversicherung (auch Sterbegeld- oder Ausbildungsversicherung)", "Private Rentenversicherung (ohne Riester-, Basis- bzw. Rürup-Rentenversicherung)", "Riester-, Basis- bzw. Rürup- Rentenversicherung", "Kfz-Haftpflicht- und/oder Kaskoversicherung", "Private Haftpflichtversicherung", "Hausratversicherung", "Rechtsschutzversicherung", "Zusätzliche private Krankenversicherung", "Zusätzliche private Pflegeversicherung", "Private Unfallversicherung"
12	Seitentrennung	Page break		
13	Vielen Dank, dass Sie Angaben zu Ihrer Wohnsituation gemacht haben.Als nächstes folgt das Ausgabentagebuch, in dem Sie für die nächsten zwei Wochen Ihre persönlichen Ausgaben dokumentieren können.	Static text		

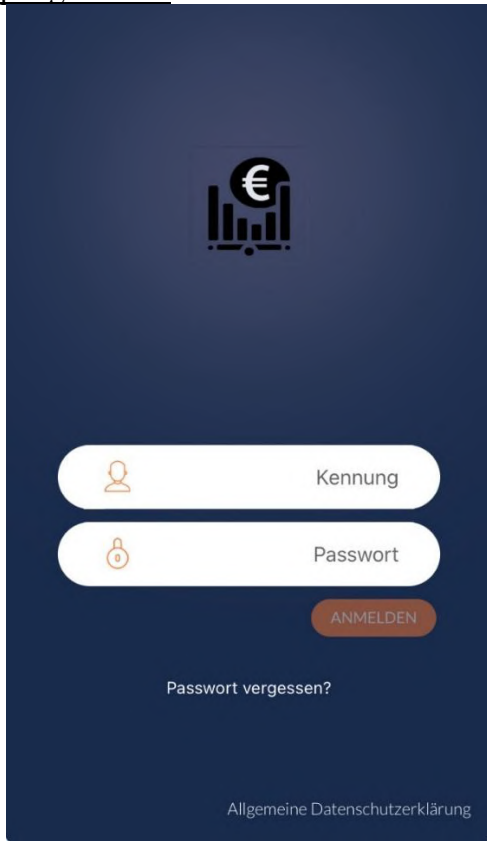
Questionnaire - Evaluation of the app

Sequence	Question	Question type	Logic	Response options
1	Vielen Dank für Ihre Teilnahme an der Studie AusgabenAtlas! Zum Abschluss haben Sie im folgenden die Möglichkeit, uns Feedback zur App zu geben.	Static text		
2	Page break	Page break		
3	Wie bewerten Sie die App insgesamt?	Scale		1 Sehr schlecht, 2 Schlecht, 3 Neutral, 4 Gut, 5 Sehr gut
4	Was hat Ihnen an der App gefallen? Und warum?	Long text		
5	Was würden Sie gerne an der App ändern?	Long text		
6	Page break	Page break		
7	Wie einfach oder schwierig war für Sie die Bedienung der App?	Scale		1 Sehr einfach, 2 Einfach, 3 Teils/teils, 4 Schwierig, 5 Sehr schwierig
8	Wie aufwändig war es für Sie, die Fragen am Anfang zu beantworten?	Scale		1 Überhaupt nicht aufwändig, 2 Kaum aufwändig, 3 Teils/teils, 4 Aufwändig, 5 Sehr aufwändig
9	Wie aufwändig war es für Sie, Ihre Ausgaben in der App zu erfassen?	Scale		1 Überhaupt nicht aufwändig, 2 Kaum aufwändig, 3 Teils/teils, 4 Aufwändig, 5 Sehr aufwändig

10	Wie bewerten Sie die Foto-Funktion für Kassenbons?	Scale	1 Sehr schlecht, 2 Schlecht, 3 Neutral, 4 Gut, 5 Sehr gut
----	--	-------	---

Appendix C: App screenshots

App registration

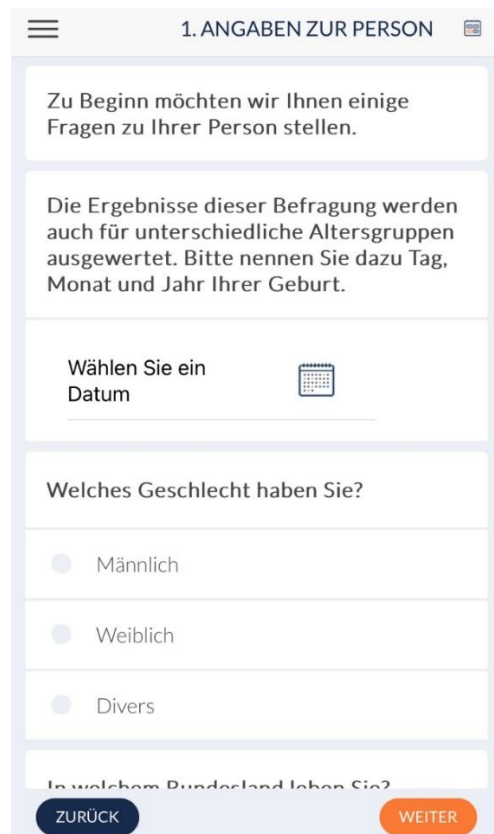


The registration screen has a dark blue background. At the top center is a logo consisting of a bar chart with a Euro symbol (€) above it. Below the logo are two white input fields with rounded corners. The first field has a person icon and is labeled 'Kennung'. The second field has a lock icon and is labeled 'Passwort'. Below these fields is an orange button labeled 'ANMELDEN'. Underneath the button is a link that says 'Passwort vergessen?'. At the bottom right, there is a small link for 'Allgemeine Datenschutzerklärung'.

Survey 1: Personal information





This is the first screen of the survey, titled '1. ANGABEN ZUR PERSON'. It has a light blue background. A white box at the top contains the text: 'Herzlich Willkommen zum AusgabenAtlas!' followed by 'Vielen Dank, dass Sie an unserer Studie teilnehmen und für einen Zeitraum von zwei Wochen alle Ihre persönlichen Ausgaben dokumentieren.' At the bottom right is an orange button labeled 'WEITER'.



This is the second screen of the survey, also titled '1. ANGABEN ZUR PERSON'. It has a light blue background. The text says: 'Zu Beginn möchten wir Ihnen einige Fragen zu Ihrer Person stellen.' followed by 'Die Ergebnisse dieser Befragung werden auch für unterschiedliche Altersgruppen ausgewertet. Bitte nennen Sie dazu Tag, Monat und Jahr Ihrer Geburt.' Below this is a section titled 'Wählen Sie ein Datum' with a calendar icon and a date picker. The next section is 'Welches Geschlecht haben Sie?' with three radio button options: 'Männlich', 'Weiblich', and 'Divers'. The final section is 'In welchem Bundesland leben Sie?'. At the bottom are two buttons: a dark blue 'ZURÜCK' button on the left and an orange 'WEITER' button on the right.

Survey 2: Personal income & expenses

 2. PERSÖNLICHE EINKÜNFTE ... 

Angaben zu persönlichen Einkünften & Ausgaben


Einkünfte aus selbstständiger bzw. landwirtschaftlicher Tätigkeit

Haben Sie im vergangenen Monat Einkünfte aus selbstständiger oder landwirtschaftlicher Tätigkeit erzielt?



Bitte beantworten Sie die Frage.

☐ Nein

☐ Ja




Survey 3: Household information

 3. ANGABEN ZUM HAUSHALT 

Angaben zum Haushalt

Die folgenden Fragen beziehen sich auf Ihre aktuelle, persönliche Wohnsituation und auf die von Ihrem Haushalt überwiegend genutzte Wohnung (Hauptwohnung).


In welcher Art von Gebäude wohnen Sie? 

☐ Freistehendes Einfamilienhaus

☐ Einfamilienhaus als Doppelhaushälfte oder Reihenhäuser

☐ Einfamilienhaus mit zusätzlicher Einliegerwohnung oder Zweifamilienhaus

☐ Wohngebäude mit 3 bis 9 Wohnungen



Diary: manual entry

HAUSHALTSBUCH

Qualität des Haushaltsbuch
Tippen Sie auf , um zu erfahren, was Sie tun können, um Ihr Haushaltsbuch zu verbessern

< Juni 2024 >

M	D	M	D	F	S	S
27	28	29	30	31	01	02
03	04	05	06	07	08	09
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30

30 Juni 2024 0,00 €

Insgesamt (ohne feste Kosten) 0,00 €

Feste Kosten 0,00 €

Hinzufügen oder bearbeiten

Ausgaben tag 30 Juni

Kassenbon scannen

Neue Ausgabe

Ich hatte heute keine Ausgaben

+

Rewe

Wählen Sie eine Unterkategorie

Supermärkte / Discounter

Lokaler Markt

Spezialitätengeschäfte

Discountläden

Jahrmärkte & Messen

Sonstige

ABBRECHEN

BESTÄTIGEN

NEUE AUSGABE

Datum der Ausgabe?

Sonntag
30 Juni
2024

Wo haben Sie die Ausgabe getätigt?

Gesamtbetrag: 0,00 €

SPEICHERN

NEUE AUSGABE

Datum der Ausgabe?

Sonntag
30 Juni
2024

Wo haben Sie die Ausgabe getätigt?

Rewe
Supermärkte / Discounter

Was war Ihre Ausgabe?

Ausgabe oder Rückgabe hinzufügen

Gesamtbetrag: 0,00 €

SPEICHERN

Äpfel

Stein- und Kernobst – frisch oder gekühlt

Frucht- und Gemüsesäfte

Obst und Nüsse, gemahlen oder anders zubereitet

Erfrischungsgetränke ohne Mineralwasser

Wein aus anderen Früchten

Restaurants, Cafes und Ähnliches mit Bedienung

Reparatur, Installation und Miete von Haushaltsgeräten

Gemüse, Knollengewächse, Kochbananen und Hülsenfrüchte, gemahlen oder andere Zubereitungen

Nach Kategorien filtern

Filter zurücksetzen

ABBRECHEN

BESTÄTIGEN

14:42

NEUE AUSGABE

Ausgabe

Schokolade

Kategorie

Lebensmittel & Getränke (z.B. Supermarkt & Discounter)

Preis

5,00

€

Ich habe den Artikel zurückgegeben (Reklamation o. ä.)

Gesamtpreis: 5,00 €

HINZUFÜGEN

NEUE AUSGABE

Datum der Ausgabe?

Sonntag
30 Juni
2024

Wo haben Sie die Ausgabe getätigt?

Rewe
Supermärkte / Discounter

Was war Ihre Ausgabe?

Äpfel

1,50 €

Ausgabe oder Rückgabe hinzufügen

Rabattaktionen, Coupons, Pfand?

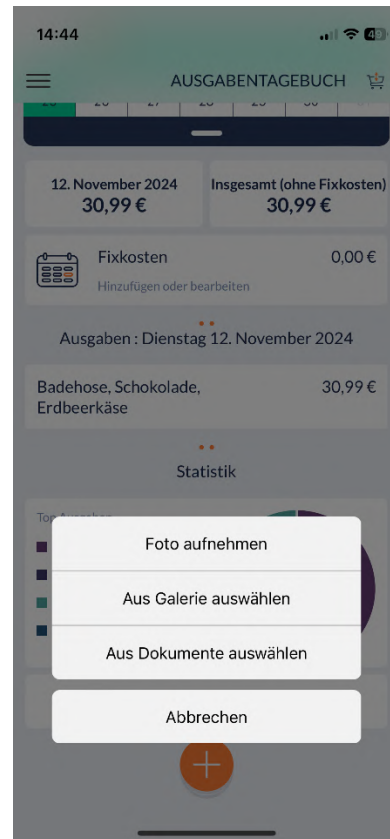
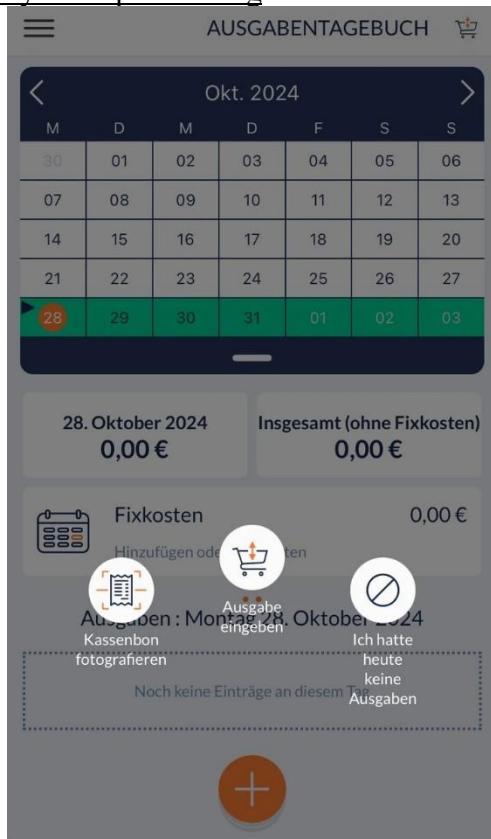
Geschäftliche Ausgabe?

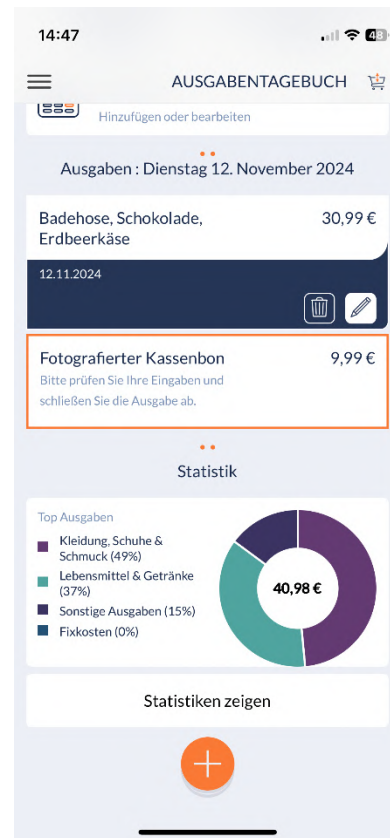
Mit Kreditkarte gezahlt?

Gesamtbetrag: 1,50 €

SPEICHERN

Diary: receipt scanning





Video of the scanning process: [Ticket scanning demo - YouTube](#)

Editing a scanned expense

MEDIONmobile 15:15 4G 45%

NEUE AUSGABE

Ausgabe
Gesamtsumme: REWE

Kategorie
Lebensmittel & Getränke (z.B. Supermarkt & Discounter)

Preis 9,99 €

Ich habe den Artikel zurückgegeben (Reklamation o. ä.)

Gesamtpreis: 9,99 € AKTUALISIEREN

Evaluation survey



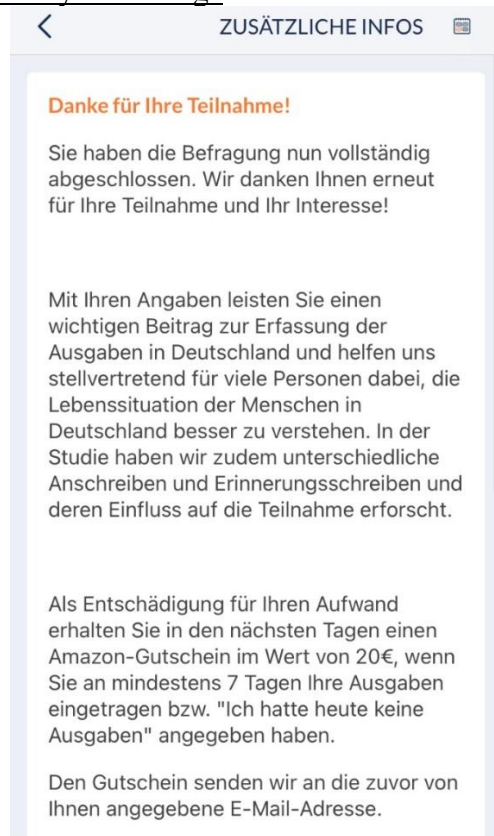
5. FEEDBACK ZUR APP

Vielen Dank für Ihre Teilnahme an der Studie AusgabenAtlas!

Zum Abschluss haben Sie die Möglichkeit, uns Feedback zur App zu geben.

WEITER

Thank you message



ZUSÄTZLICHE INFOS

Danke für Ihre Teilnahme!

Sie haben die Befragung nun vollständig abgeschlossen. Wir danken Ihnen erneut für Ihre Teilnahme und Ihr Interesse!

Mit Ihren Angaben leisten Sie einen wichtigen Beitrag zur Erfassung der Ausgaben in Deutschland und helfen uns stellvertretend für viele Personen dabei, die Lebenssituation der Menschen in Deutschland besser zu verstehen. In der Studie haben wir zudem unterschiedliche Anschreiben und Erinnerungsschreiben und deren Einfluss auf die Teilnahme erforscht.

Als Entschädigung für Ihren Aufwand erhalten Sie in den nächsten Tagen einen Amazon-Gutschein im Wert von 20€, wenn Sie an mindestens 7 Tagen Ihre Ausgaben eingetragen bzw. "Ich hatte heute keine Ausgaben" angegeben haben.

Den Gutschein senden wir an die zuvor von Ihnen angegebene E-Mail-Adresse.

Appendix D: E-mail notification messages

Activation e-mail:

Guten Tag,

vielen Dank, dass Sie an der Studie AusgabenAtlas teilnehmen!

Sie erhalten diese E-Mail, um die folgende Kennung für Sie freizuschalten:

Kennung: username

Mit dem folgenden Link wird Ihr Zugang aktiviert und Sie können sich ab sofort mit dem von Ihnen gewählten Passwort sowie Ihrer Kennung in der App einloggen:

activation url

Sollten Sie die Aktivierung nicht veranlasst haben, können Sie diese E-Mail ignorieren.

Viele Grüße

Das Team des AusgabenAtlas

(Universität Mannheim)

Welcome-e-mail

Guten Tag,

vielen Dank, dass Sie sich dafür entschieden haben, an der Studie AusgabenAtlas teilzunehmen!

Ab sofort können Sie mit der Teilnahme beginnen. Sobald Sie sich das erste Mal angemeldet haben, bitten wir Sie, einige Fragen zu Ihrer Person, Ihren monatlichen Einkünften und Ausgaben und zu Ihrer Wohnsituation zu beantworten. Danach bitten wir Sie, im Ausgabentagebuch für zwei Wochen Ihre Ausgaben (z.B. für Lebensmitteleinkäufe) zu dokumentieren.

Ihre Kennung: username

Ihr Passwort haben Sie bei der ersten Anmeldung in der App selbst vergeben.

Sollten Sie noch Fragen zur Studie oder der App AusgabenAtlas haben, finden Sie hier (LINK TO: <https://www.uni-mannheim.de/ausgabenatlas>) weiterführende Informationen. Dort finden Sie auch Kontaktmöglichkeiten, falls Ihre Frage nicht beantwortet wurde.

Wir bedanken uns ganz herzlich für Ihre Unterstützung!

Viele Grüße

Das Team des AusgabenAtlas

(Universität Mannheim)

Password reset

Guten Tag,

vielen Dank, dass Sie an der Studie AusgabenAtlas teilnehmen!

Sie erhalten diese E-Mail, weil Sie ein neues Passwort für Ihre Kennung angefragt haben.

Mit einem Klick auf den folgenden Link können Sie das Passwort für Ihre Kennung username ändern:

reset password url

Bitte beachten Sie unsere Passwortanforderungen und wählen Sie ein Passwort, dass:

- min. 10 Zeichen lang ist
- min. 1 Groß- und Kleinbuchstaben
- min. 1 Zahl
- min. 1 Sonderzeichen

... enthält.

Sollten Sie keine Änderung Ihres Passworts vornehmen wollen, können Sie diese E-Mail ignorieren.

Viele Grüße

Das Team des AusgabenAtlas

(Universität Mannheim)

Questionnaire reminder (after 4 days of inactivity)

Guten Tag,

vielen Dank, dass Sie an der AusgabenAtlas Studie der Universität Mannheim teilnehmen!

Ab sofort können Sie mit der Teilnahme beginnen und die kurze Vorbefragung abschließen.

Danach folgt das Ausgabentagebuch, in dem Sie Ihre Ausgaben dokumentieren können.

Ihre Kennung: username

Ihr Passwort haben Sie bei der ersten Anmeldung in der App selbst vergeben.

Sollten Sie noch Fragen zur Studie oder der App AusgabenAtlas haben, finden Sie hier (LINK TO: <https://www.uni-mannheim.de/ausgabenatlas>) weiterführende Informationen. Dort finden Sie auch Kontaktmöglichkeiten, falls Ihre Frage nicht beantwortet wurde.

Wir bedanken uns ganz herzlich für Ihre Unterstützung!

Viele Grüße

Das Team des AusgabenAtlas

(Universität Mannheim)

Diary reminder (after 4 days of inactivity)

Guten Tag,

vielen Dank, dass Sie an der AusgabenAtlas Studie der Universität Mannheim teilnehmen!

Sie sind im Ausgabentagebuch angekommen und haben in den letzten vier Tagen keine Ausgaben dokumentiert. Wir bitten Sie, weiterhin Ihre täglichen Ausgaben zu erfassen. Sofern Sie an einem Tag keine Ausgaben hatten, geben Sie für diesen Tag bitte „Ich hatte heute keine Ausgaben“ an. Auch das zählt als Eintrag.

Ihre Kennung: username

Ihr Passwort haben Sie bei der ersten Anmeldung in der App selbst vergeben.

Vielen Dank für Ihre Teilnahme!

Viele Grüße

Das Team des AusgabenAtlas

(Universität Mannheim)

Diary reminder (after 1 week after diary start)

Guten Tag,

wir freuen uns, dass Sie sich entschieden haben, an unserer Studie teilzunehmen.

Die Hälfte der Befragungszeit ist bereits vorbei, bitte tragen Sie auch in den kommenden 7 Tagen Ihre Ausgaben in der App ein. Wenn Sie an einem Tag keine Ausgaben hatten, tragen Sie bitte "Ich hatte heute keine Ausgaben" ein.

Ihre Kennung: username

Ihr Passwort haben Sie bei der ersten Anmeldung in der App selbst vergeben.

Vielen Dank für Ihre Teilnahme!

Viele Grüße

Das Team des AusgabenAtlas

(Universität Mannheim)

Information that diary closes in 2 days

Guten Tag,

vielen Dank, dass Sie im Rahmen der Studie AusgabenAtlas zwei Wochen lang Ihre täglichen Ausgaben erfasst haben.

Das Ausgabentagebuch schließt bald. Sie haben noch zwei Tage Zeit um Ausgaben in der App einzutragen bzw. nachzutragen.

Sie können das Ausgabentagebuch ab sofort jederzeit schließen.

Ihre Kennung: username

Ihr Passwort haben Sie bei der ersten Anmeldung in der App selbst vergeben.

Wir bedanken uns ganz herzlich für Ihre Unterstützung!

Viele Grüße

Das Team des AusgabenAtlas

(Universität Mannheim)

Diary closed automatically

Guten Tag,

vielen Dank, dass Sie an der AusgabenAtlas Studie der Universität Mannheim teilnehmen!

Ihre Täglichen Ausgaben wurden automatisch geschlossen.

Zum Abschluss der Studie haben wir im letzten Fragebogen ein paar Fragen zu Ihren Erfahrungen mit der App. Anschließend haben Sie es geschafft!

Ihre Kennung: username

Ihr Passwort haben Sie bei der ersten Anmeldung in der App selbst vergeben.

Sollten Sie noch Fragen zur Studie oder der App haben, finden Sie auf LINK TO:
<https://www.uni-mannheim.de/ausgabenatlas> weiterführende Informationen.

Wir bedanken uns ganz herzlich für Ihre Unterstützung!

Viele Grüße

Das Team des AusgabenAtlas

(Universität Mannheim)

Pre-Announcement end of survey

Guten Tag,

vielen Dank für Ihre Teilnahme an der Studie AusgabenAtlas.

Sie können noch bis Donnerstag, den 19.12.2024 Ihre täglichen Ausgaben eintragen bzw. nachtragen. Danach schließt die Studie.

Sollten Sie noch Fragen zur Studie oder App haben, finden Sie auf LINK TO:
<https://www.uni-mannheim.de/ausgabenatlas> weiterführende Informationen.

Wir bedanken uns ganz herzlich für Ihre Unterstützung!

Viele Grüße

Das Team des AusgabenAtlas

(Universität Mannheim)

Information cut-off survey

Guten Tag,

vielen Dank für Ihre Teilnahme an der Studie AusgabenAtlas. Die Studie wird am 19.12.2024 final geschlossen.

Bis dahin können Sie noch Ausgaben in der App eintragen bzw. nachtragen.

Ihre Kennung: username

Ihr Passwort haben Sie bei der ersten Anmeldung in der App selbst vergeben.

Sollten Sie noch Fragen zur Studie oder der App haben, finden Sie auf LINK TO: <https://www.uni-mannheim.de/ausgabenatlas> weiterführende Informationen.

Wir bedanken uns ganz herzlich für Ihre Unterstützung!

Viele Grüße

Das Team des AusgabenAtlas

(Universität Mannheim)

Data collection has ended

Guten Tag,

Sie hatten sich für die Studie AusgabenAtlas der Universität Mannheim registriert. Vielen Dank dafür!

Die Studie wurde in der Zwischenzeit beendet. Die Teilnahme ist nicht weiter möglich.

Wenn Sie an mindestens 7 Tagen Ihre Ausgaben eingetragen, bzw. "Ich hatte heute keine Ausgaben" angegeben haben, erhalten Sie als Entschädigung für Ihren Aufwand in den nächsten Tagen einen Amazon-Gutschein im Wert von 20€ per E-Mail.

Mit Ihren Angaben leisten Sie einen wichtigen Beitrag zur Erfassung der Ausgaben in Deutschland und helfen uns stellvertretend für viele Personen dabei, die Lebenssituation der Menschen in Deutschland besser zu verstehen. In der Studie haben wir zudem unterschiedliche Anschreiben und Erinnerungsschreiben und deren Einfluss auf die Teilnahme erforscht.

Viele Grüße

Das Team des AusgabenAtlas

(Universität Mannheim)

Thank you message after evaluation questionnaire

Guten Tag,

Sie haben die Studie AusgabenAtlas erfolgreich beendet. Dafür möchten wir uns ganz herzlich bei Ihnen bedanken! Mit Ihren Angaben leisten Sie einen wichtigen Beitrag zur Erfassung der Ausgaben in Deutschland und helfen uns stellvertretend für viele Personen dabei, die Lebenssituation der Menschen in Deutschland besser zu verstehen. In der Studie haben wir zudem unterschiedliche Anschreiben und Erinnerungsschreiben und deren Einfluss auf die Teilnahme erforscht.

Als Entschädigung für Ihren Aufwand erhalten Sie per E-Mail in den nächsten Tagen einen Amazon-Gutschein im Wert von 20€, wenn Sie an mindestens 7 Tagen Ihre Ausgaben eingetragen bzw. "Ich hatte heute keine Ausgaben" angegeben haben.

Sollten Sie noch Fragen zur Studie oder der App haben, finden Sie auf LINK TO: <https://www.uni-mannheim.de/ausgabenatlas> weiterführende Informationen.

Viele Grüße

Das Team des AusgabenAtlas

(Universität Mannheim)

Thank you message

Guten Tag,

Sie haben die Studie AusgabenAtlas erfolgreich beendet. Dafür möchten wir uns ganz herzlich bei Ihnen bedanken! Mit Ihren Angaben leisten Sie einen wichtigen Beitrag zur Erfassung der Ausgaben in Deutschland und helfen uns stellvertretend für viele Personen dabei, die Lebenssituation der Menschen in Deutschland besser zu verstehen. In der Studie haben wir zudem unterschiedliche Anschreiben und Erinnerungsschreiben und deren Einfluss auf die Teilnahme erforscht.

Als Entschädigung für Ihren Aufwand erhalten Sie per E-Mail in den nächsten Tagen einen Amazon-Gutschein im Wert von 20€, wenn Sie an mindestens 7 Tagen Ihre Ausgaben eingetragen bzw. "Ich hatte heute keine Ausgaben" angegeben haben.

Sollten Sie noch Fragen zur Studie oder der App haben, finden Sie auf LINK TO: <https://www.uni-mannheim.de/ausgabenatlas> weiterführende Informationen.

Viele Grüße

Das Team des AusgabenAtlas

(Universität Mannheim)

Incentive



Vielen Dank für die Teilnahme an unserer Studie!

Ihr AusgabenAtlas Team der Universität Mannheim

20,00 €

Auf Amazon-Konto gutschreiben

Sobald der Geschenkgutschein zu Ihrem Amazon-Konto hinzugefügt wurde, wird dessen Betrag zu Ihrem Geschenkgutschein-Guthaben hinzugefügt. Ihr

Geschenkgutschein-Guthaben kann nicht auf andere Konten übertragen, zum Kauf anderer Geschenkgutscheine verwendet, oder sofern dies nicht gesetzlich vorgeschrieben ist, für Bargeld eingelöst werden.

Ihr Geschenkgutschein-Guthaben wird automatisch mit berechtigten Bestellungen während des Bezahlvorgangs sowie bei der Verwendung von 1-Click verrechnet. Wenn Sie Ihren Geschenkgutschein-Guthaben nicht verwenden möchten, können Sie ihn an der Kasse als Zahlungsmittel deaktivieren.

Wenden Sie sich bitte an die Gutscheinmarke, falls Sie Probleme oder Fragen zur Einlösung, Verwendung oder dem aktuellen Guthabenstatus haben.

Falls Sie Probleme mit oder allgemeine Fragen zu Ihrer Bestellung haben, wenden Sie sich an den Kundendienst und stellen Sie folgende Informationen bereit:

Bestellnummer: 303-9477890-1145908

Amazon.de ist ein Handelsname für Amazon EU Sarl, für Amazon Europe Core Sarl, für Amazon

Media EU Sarl und für Amazon EU Sarl, die alle ihren eingetragenen Sitz unter 38 avenue John F. Kennedy, L-1855 Luxemburg haben und für die Amazon Digital Germany GmbH, mit eingetragenem Sitz in der Domagkstr. 28, 80807 München, Deutschland. 02024 Amazon.com, Inc. oder verbundene Unternehmen. Amazon sowie alle zugehörigen Marken sind Marken der

Amazon.com, Inc. oder deren verbundenen Unternehmen. Es können Versandkosten anfallen. Für Ihren Kauf gilt der Preis und die Verfügbarkeit, die zum Zeitpunkt des Kaufs auf amazon.de angezeigt werden.

Appendix E: In-app push notification messages

Reminder at questionnaire

Sie haben offene Fragen in der AusgabenAtlas Studie

Jetzt teilnehmen und beantworten!

Ticket ready message: Receipt processing done

Kassenbon erfolgreich hochgeladen!

Ihr Kassenbon wurde erfolgreich erfasst und steht im Ausgabentagebuch zur Verfügung.

Diary reminder

Eintragungen AusgabenAtlas

In den vergangenen zwei Tagen haben Sie keine Eintragungen im Ausgabentagebuch vorgenommen. Bitte denken Sie daran, alle Ihre Ausgaben zu erfassen. Wenn Sie keine Ausgaben hatten, wählen Sie bitte *Ich hatte heute keine Ausgaben*. Vielen Dank!

Information that diary closes in 2 days

AusgabenAtlas schließt bald

Sie haben noch zwei Tage Zeit um Ausgaben einzutragen. Sie können das Ausgabentagebuch ab sofort jederzeit schließen.



Universität Mannheim ■ AusgabenAtlas

– Die Studie AusgabenAtlas ist beendet. Vielen Dank an alle Teilnehmenden! -



Herzlich Willkommen beim AusgabenAtlas! Wir freuen uns über Ihr Interesse an unserer Befragung und möchten Ihnen auf den folgenden Seiten umfassende Informationen zur Verfügung stellen.

Was ist der AusgabenAtlas?



Der AusgabenAtlas ist eine Studie, mit dem Ziel zu untersuchen, wofür Menschen in Deutschland ihr Geld ausgeben. Mit diesen Daten wollen wir das Konsumverhalten der Menschen in Deutschland wissenschaftlich untersuchen und herausfinden, wie viel Geld Bürgerinnen und Bürgern zum Leben zur Verfügung steht und wie hoch monatliche Ausgaben

sind. Hierfür stellen wir Ihnen Fragen zu Ihren Einnahmen und Ausgaben sowie zu Ihren Wohnverhältnissen.

Warum sollte ich am AusgabenAtlas teilnehmen?



Uns ist bewusst, dass Sie für die Teilnahme an unserer Umfrage einen Teil Ihrer Zeit opfern. Zeit ist knapp und deshalb ist Ihre Teilnahme alles andere als selbstverständlich.

Wir können Ihnen aber versichern, dass es sich lohnt.

Für politische und gesellschaftliche Entscheidungen braucht es zuverlässige Daten. Diese Daten zeigen, wofür die Menschen in Deutschland ihr Geld ausgeben. Sie leisten damit einen wichtigen Beitrag zur Erforschung des Konsumverhaltens der Bürgerinnen und Bürger in Deutschland. Vielen Dank!

[Alle Infos zur Teilnahme](#)

Zweck der Studie

Die Studie liegt im öffentlichen Interesse und dient der wissenschaftlichen Sozialforschung.

Auftraggeber und Durchführung der Studie

Verantwortlich für die Studie und damit auch für die Verarbeitung Ihrer Daten ist die Universität Mannheim.

Um die Studie durchzuführen, arbeiten wir mit Auftragsverarbeitern wie beispielsweise Dienstleistern für die Adressprüfung, Druck- und Versandarbeiten sowie Datenerfassung und Datenergänzung zusammen. Die erforderlichen Datenschutzvereinbarungen nach Art. 28 DSGVO wurden abgeschlossen. Die Dienstleister arbeiten hiernach streng weisungsgebunden. Wir werden Ihren Namen und Ihre Kontaktdaten nicht an Dritte außer an die eingebundenen Dienstleister weitergeben.

Freiwilligkeit der Teilnahme

Die Teilnahme an unserer Befragung und die Beantwortung der Fragen ist freiwillig. Erhaltene Aufwandsentschädigungen sind keine Bezahlung, sondern ein Dankeschön für Ihre Unterstützung der wissenschaftlichen Forschung. Sie gehen damit keine Verpflichtung ein. Die Beendigung oder das Abbrechen der Teilnahme ist jederzeit möglich und mit keinen Nachteilen für Sie verbunden. Falls Sie teilnehmen, können Sie auch danach noch Ihre Einwilligung ohne Angabe von Gründen widerrufen.

Kontaktieren Sie uns gerne

Per Post:

Universität Mannheim

Lehrstuhl für Social Data Science und Methodology

AusgabenAtlas

A5, 6

68159 Mannheim

Schreiben Sie uns eine E-Mail:

ausgabenatlas@uni-mannheim.de

Oder rufen Sie uns an:

[+49 621 – 181-2203](tel:+496211812203)

Wir sind telefonisch von Montag bis Freitag zwischen 9:30 und 11:30 und zwischen 14 und 16 Uhr zu erreichen (außer Mittwoch Vormittag).



[Universität Mannheim](#) ■ [AusgabenAtlas](#) ■ [Teilnahme](#)

Alle Informationen zur Teilnahme

Warum ist meine Teilnahme wichtig?

Die Ergebnisse unserer Befragung sind nur dann für die gesamte Bevölkerung aussagekräftig („repräsentativ“), wenn sich möglichst viele der zufällig ausgewählten Bürgerinnen und Bürger an unserer Umfrage beteiligen. Sie sind eine dieser ausgesuchten Personen. Daher ist Ihre Teilnahme so wichtig!

Wie lange dauert die Teilnahme an der Befragung?

Unsere Befragung hat eine Laufzeit von zwei Wochen und ist als eine Tagebuchstudie angelegt. Zunächst füllen Sie einen kurzen Fragebogen aus (Dauer ca. 20 Minuten). Nach dem Ausfüllen des Fragebogens können Sie Ihre Ausgaben jederzeit in der App erfassen.

Bis wann ist eine Teilnahme möglich?

Die Studie endet am 19.12.2024 um 20 Uhr. Bis zu diesem Zeitpunkt können Sie sich in die App einloggen und Ihre Ausgaben eingeben sowie die Fragebögen beantworten.

Woher hat die Universität Mannheim meine Kontaktdaten?

Um ein möglichst aussagekräftiges Bild über das Konsumverhalten der Menschen in Deutschland zu erhalten, haben wir zufällig ca. 7.000 Bürgerinnen und Bürger im Alter von 18 bis 70 Jahren dazu eingeladen, sich an unserer Umfrage zu beteiligen. Dazu haben wir im Einklang mit den bestehenden gesetzlichen Rahmenbedingungen Stichproben aus den Melderegistern der Gemeinden und Städte

in Deutschland gezogen. Konkret kommen die Kontaktinformationen also von den jeweiligen Einwohnermeldeämtern. Diese Informationen werden nur für die Einladung zu dieser Studie genutzt und danach von uns vernichtet.

Sollten Sie an der Studie teilnehmen, hinterlegen Sie selbst zu Beginn eine gültige E-Mail-Adresse in der App. Diese E-Mail-Adresse nutzen wir nur um mit Ihnen zu kommunizieren und für die Zusendung der Aufwandsentschädigung.

Warum ist die Teilnahme ausschließlich über eine App möglich?

Das Smartphone gewinnt für das Leben der Menschen immer größere Bedeutung. Auch in Deutschland besitzen rund 80% der Bürgerinnen und Bürger ein Smartphone. Die Teilnahme am AusgabenAtlas über unsere App ist einfach und hat einige Vorteile für Sie. Ebenso unterstützen Sie die Wissenschaftlerinnen und Wissenschaftler, die an den Ergebnissen der Studie interessiert sind.

Aus den folgenden Gründen ist die Teilnahme per Smartphone App vorteilhaft:

1. Sie können die Beantwortung beginnen, wann und wo es Ihnen persönlich am besten passt.
2. Sie sparen sich Zeit bei der Eingabe der Daten: Unsere App zeigt Ihnen nur für Sie relevante Fragen an.
3. Sie können die Kamera Ihres Smartphones für eine Foto-Funktion nutzen, um Kassenbons zu fotografieren und müssen diese nicht händisch eingeben.
4. Ihre Einnahmen und Ausgaben sind in der App dokumentiert und werden Ihnen visuell dargestellt. Somit erhalten Sie einen guten Überblick über Ihre finanzielle Situation.
5. Ihre Daten werden sicher und schnell an uns übermittelt.
6. Durch die computergestützte Erfassung Ihrer Antworten stehen die Ergebnisse der Studie interessierten Wissenschaftlerinnen und Wissenschaftlern schnell zur Verfügung. Die Studienergebnisse können daher kurzfristig die wissenschaftliche und politische Diskussion beeinflussen.

Bin ich verpflichtet, an der Befragung teilzunehmen?

Die Teilnahme an unserer Umfrage und die Beantwortung der Fragen ist freiwillig. Erhaltene Aufwandsentschädigungen sind keine Bezahlung, sondern ein Dankeschön für Ihre Unterstützung der wissenschaftlichen Forschung. Sie gehen damit keine Verpflichtung ein. Die Beendigung oder das

Abbrechen der Teilnahme ist jederzeit möglich und mit keinen Nachteilen für Sie verbunden. Falls Sie teilnehmen, können Sie auch danach noch Ihre Einwilligung ohne Angabe von Gründen widerrufen.

Kann ich an der Befragung in einer anderen Sprache als Deutsch teilnehmen?

Leider ist eine Teilnahme nur auf Deutsch möglich.

Ich möchte nicht mehr teilnehmen. Wie kann ich meine Teilnahme widerrufen?

Sollten Sie nicht weiter von uns kontaktiert werden wollen, dann senden Sie uns bitte eine persönliche Mitteilung per E-Mail an ausgabenatlas@uni-mannheim.de, rufen uns unter der Telefonnummer +49 621 – 181-2203 an oder senden uns eine schriftliche Mitteilung per Brief zu. Die Beendigung der Teilnahme hat keinerlei negative Konsequenzen für Sie. Dennoch möchten wir Sie natürlich bitten, möglichst vollständig an unserer Befragung teilzunehmen, um unsere wissenschaftliche Forschung zu unterstützen.

In welcher Form und wann erhalte ich meine Aufwandsentschädigung?

Als Entschädigung für Ihren Aufwand erhalten Sie einen Amazon-Gutschein im Wert von 20€, wenn Sie an mindestens 7 der insgesamt 14 Tage Ihre Ausgaben eingetragen haben. Sofern Sie an einem Tag keine Ausgaben hatten, geben Sie für diesen Tag bitte „Ich hatte heute keine Ausgaben“ an. Auch das zählt als Eintrag.

Für das Einlösen des Gutscheins ist ein Amazon-Konto notwendig.

Wenige Tage nach erfolgreicher Teilnahme erhalten Sie den Amazon-Gutschein in Form eines Gutscheincodes. Der Gutscheincode wird an eine von Ihnen gewählte E-Mail-Adresse versendet und steht Ihnen danach frei zur Verfügung.



[Universität Mannheim](#) ■ [AusgabenAtlas](#) ■ [App](#)

Wie funktioniert die App?

Wie installiere ich die App?

Die App ist im Google Play Store und im App Store von Apple unter dem Namen „AusgabenAtlas“ verfügbar. In Ihrem persönlichen Einladungsschreiben finden Sie Informationen zum App Download per QR-Code sowie Ihre Zugangsdaten. Alternativ finden Sie [hier](#) den Link zur App im Apple App Store und [hier](#) den Link zur App im Google Play Store.

Welche Systemvoraussetzungen gibt es?

Die Smartphone-App ist für Android-Geräte mit Version 10 und höher erfolgreich getestet. Für Apple-Geräte (iOS-Betriebssystem) ist die Version iOS 13 oder höher erforderlich.

Auf welche Funktionen/Sensoren meines Smartphones hat die App Zugriff?

Wenn Sie die App auf Ihrem Smartphone installieren, wird die Berechtigung erfragt, Push-Benachrichtigungen zu versenden (beachten Sie hierzu auch die Frage „[Warum sollte ich Benachrichtigungen \(Push, E-Mail\) aktivieren?](#)“). Zudem erfragt die App die Berechtigung, auf Ihre Kamera zuzugreifen. Wenn Sie der App erlauben, auf die Kamera zuzugreifen, können Sie diese zum Fotografieren Ihrer Kassenbons nutzen. Möchten Sie Ihre Kassenbons nicht fotografieren, können Sie jederzeit die Berechtigung zum Zugriff auf die Kamera deaktivieren. Weitere Berechtigungen, z.B. der Zugriff auf Ihr Mikrofon oder Ihren Standort, erfragt die App zu keinem Zeitpunkt. Android und iOS führen gegebenenfalls standardmäßig in ihren Listen zu möglichen App-Berechtigungen die Zugriffe auf Mikrofon oder Standort auf. Die AusgabenAtlas App greift jedoch ausdrücklich nicht auf diese Funktionen zu und nutzt diese auch nicht.

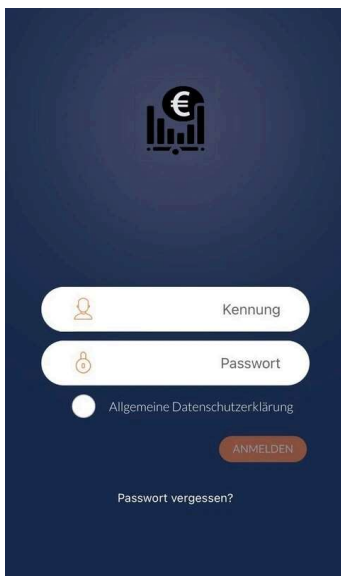
Kann ich die App auch offline nutzen?

Nein. Die App funktioniert offline nicht. Sie nutzt die Verbindung ins Internet, um Ihnen z.B. die Anmeldung mit Ihrer Kennung und Ihrem Passwort zu ermöglichen. Außerdem nutzen die Suchfunktion sowie die Foto-Funktion in den „Täglichen Ausgaben“ Ihre aktive Internetverbindung, um Ergebnisse darstellen zu können.

Wie erhalte ich meine Zugangsdaten?

Ihre Zugangsdaten (Kennung und Passwort) für die App „AusgabenAtlas“ werden Ihnen per Post zugeschickt.

Wie aktiviere ich den Zugang?



Öffnen Sie die App und geben Sie Ihre Kennung und das Passwort ein. Die Zugangsdaten für die erste Anmeldung erhalten Sie per Brief von der Universität Mannheim. Nach der erstmaligen Anmeldung können Sie Ihr Passwort jederzeit selbständig ändern. Die Änderung Ihres Passworts ist unter dem Reiter „Profil“ möglich.

Was mache ich, wenn ich mein Passwort vergessen habe?

Wenn Sie Ihr Passwort einmal vergessen haben, können Sie unter Angabe Ihrer Kennung die Funktion „Passwort vergessen?“ nutzen. Sie erhalten im Anschluss eine E-Mail und können ein neues Passwort vergeben.

Sie finden die „Passwort vergessen?“-Funktion sowohl im Anmeldebildschirm der Smartphone-App als auch auf der Anmeldeseite im Browser. Sollten Sie weitere Unterstützung benötigen, wenden Sie sich an uns.

Was mache ich, wenn sich meine E-Mail-Adresse geändert hat?

Sie können Ihre E-Mail-Adresse jederzeit im Menü der App unter „Mein Profil“ ändern.

Wenn sich seit dem Absenden der Teilnahmeerklärung zur Befragung und Ihrer Teilnahme an der Befragung Ihre E-Mail-Adresse geändert hat, können Sie bei der Erstanmeldung in der App auch einfach die neue E-Mail-Adresse eintragen.

Warum sollte ich Benachrichtigungen (Push, E-Mail) aktivieren?

Um Sie beispielsweise an den Start der 2-wöchigen Dokumentation der Ausgaben zu erinnern, versenden wir gelegentlich Push-Benachrichtigungen bzw. E-Mails. Sie erhalten von uns keine über die Teilnahme hinausgehenden Benachrichtigungen oder Werbung. Zum Betrieb der App und zur Teilnahme an der Befragung sind Benachrichtigungen nicht zwingend notwendig. Wenn Sie keine Erinnerungen erhalten möchten, können Sie diese daher auch im Menü der App ausschalten.

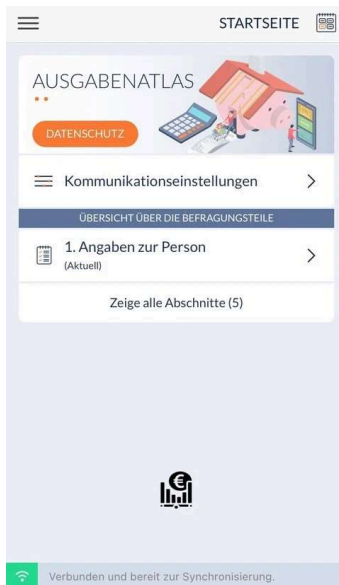
Ich habe weitere Fragen zur App. An wen kann ich mich wenden?

Gerne können Sie sich bei weiteren Fragen telefonisch oder per E-Mail an die Mitarbeiterinnen und Mitarbeiter der Universität Mannheim wenden. Die Kontaktdaten finden Sie [hier](#).

Wie ist die App aufgebaut?

Die Startseite

Auf dem Startbildschirm der App finden Sie den Personenfragebogen, den Sie bitte zuerst ausfüllen. Sie können den jeweils nächsten Teil der Studie immer erst dann ausfüllen, wenn der vorhergegangene abgeschlossen wurde. Durch die Tafel „Zeige alle Abschnitte“ können Sie sich eine Übersicht aller kommenden und bereits ausgefüllten Abschnitte anzeigen lassen.



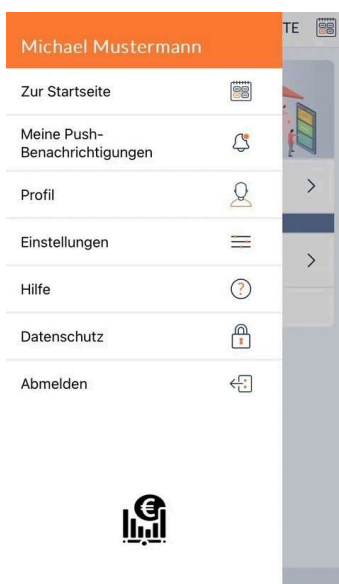
Nachdem Sie die ersten Fragen zu Ihnen, Ihren Einnahmen und Ausgaben sowie Ihrem Haushalt beantwortet haben, erhalten Sie Zugriff auf das Ausgabentagebuch, in welches Sie Ihre täglichen Ausgaben eintragen können.

Die Befragung

Die Befragung gliedert sich in der App in zwei Teile:

1. Sie beantworten anfangs einige Fragen zu Ihrer Person und dem Haushalt, in dem Sie leben. Dieser erste Fragebogen beinhaltet auch Fragen zu Ihren üblichen Einnahmen und Ausgaben.
2. Sie füllen in dem vorgegebenen 2-wöchigen Zeitraum die „Täglichen Ausgaben“ aus.

Das Menü



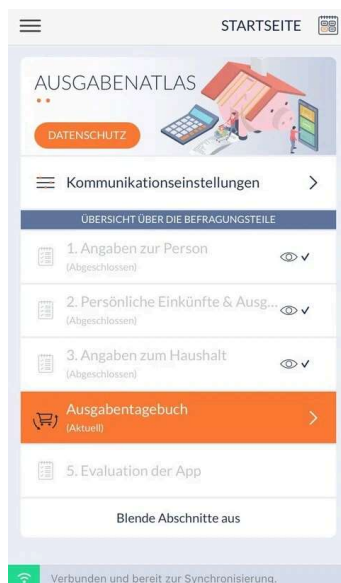
Klicken Sie auf die drei horizontalen Balken in der oberen linken Ecke, um das Menü zu öffnen. Im Menü haben Sie Zugriff auf folgende Menüpunkte:

- **Startseite:** Hierüber gelangen Sie zur Übersichtseite mit Ihren Befragungsteilen.
- **Push-Benachrichtigungen:** Hier finden Sie eine Übersicht Ihrer erhaltenen Push-Benachrichtigungen.

- **Profil:** Hier haben Sie die Möglichkeit, Ihre E-Mail-Adresse und Ihr Passwort zu ändern. Klicken Sie dazu einfach auf den Stift in der rechten oberen Ecke.
- **Einstellungen:** Hier können Sie die Push-Benachrichtigungen sowie Benachrichtigungen per E-Mail aktivieren bzw. deaktivieren.
- **Hilfe:** Hier finden Sie unsere Kontaktdaten sowie Antworten auf häufig gestellte Fragen (FAQs).
- **Datenschutz:** Hier finden Sie die Datenschutzerklärung.
- **Abmelden:** Hier können Sie sich aus der App abmelden und gelangen so wieder auf die Anmeldeseite. Um sich wieder anzumelden, müssen Sie nach betätigen der „Abmelden“-Funktion Ihre Zugangsdaten (Kennung, Passwort) erneut eingeben.

Wie gebe ich die Ausgaben ein und was muss ich dabei beachten?

Funktion Tägliche Ausgaben



Wir bitten Sie, zwei Wochen lang Ihre Ausgaben zu dokumentieren. Klicken Sie dazu im Startbildschirm auf „Ausgabentagebuch“. Bitte beachten Sie, dass Sie diesen Befragungsteil erst ausfüllen können, wenn Sie zuvor den Personenfragebogen ausgefüllt und abgesendet haben.

Für diese Studie müssen Sie hier im Bereich Fixkosten nichts eintragen. Falls in den zwei Wochen Ihrer Studienzeit Fixkosten (wie z.B. Miete, Versicherungen, Gebühren, etc) anfallen, dann geben Sie diese bitte wie alle anderen Ausgaben auch über das orangene Plus-Symbol ein.



Neue Ausgabe eintragen



Um eine Ausgabe einzutragen, klicken Sie auf das orangefarbige Plus-Symbol (+) und wählen Sie „Ausgabe eingeben“ aus. Anschließend können Sie bei Bedarf das Datum anpassen, z.B. wenn Ihre Ausgabe einige Tage zurückliegt. Alternativ können Sie die Funktion „Kassenbon fotografieren“ wählen. Eine Erläuterung zu dieser Funktion finden Sie [hier](#).

Wählen Sie „Ausgabe hinzufügen“. Geben Sie Ihre Ausgabe in das Suchfeld oben ein. Basierend auf Ihrem Eintrag werden Ihnen dann mögliche Ausgabekategorien angezeigt. Wählen Sie aus dieser Liste die passende Kategorie aus. Sollten es eventuell zu viele Suchergebnisse sein, können Sie die Liste filtern über „Nach Kategorien filtern“. Beispiele, wie Sie Ihre Ausgaben in der App in den „Täglichen Ausgaben“ eintragen, finden Sie [hier](#).

 NEUE AUSGABE 

Datum der Ausgabe?

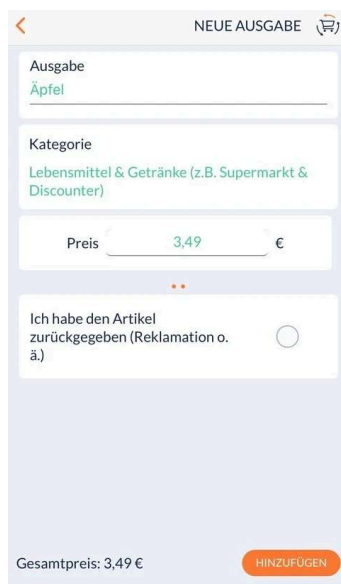
Montag 28 Oktober 2024

Was war Ihre Ausgabe?

Ausgabe hinzufügen

Gesamtbetrag: 0,00 €

SPEICHERN



Sollte Ihnen für Ihren Sucheintrag eventuell einmal keine passende Ausgabekategorie vorgeschlagen werden, wird Ihnen die Kategorie „Sonstiges [Suchbegriff ohne genaue Auswahl speichern]“ angezeigt. Bitte ergänzen Sie in diesem Fall Ihren Suchbegriff um möglichst viele Details. Damit helfen Sie uns zu verstehen, welche Suchbegriffe fehlen und wie wir die Suchfunktion für alle Teilnehmenden verbessern können.

Nachdem Sie eine passende Produktkategorie ausgewählt haben, können Sie den Preis des gekauften Artikels eintragen. Fügen Sie dann Ihre Ausgabe mit dem orangenen „Hinzufügen“ Button unten rechts hinzu. Falls dieser nicht sichtbar ist, während die Bildschirmtastatur angezeigt wird, können Sie die Tastatur durch das Betätigen der „zurück“ Taste auf Ihrem Smartphone bzw. durch Wischen vom

linken Bildschirmrand ausblenden.

Wenn Sie das Produkt erfasst haben, können Sie über die Schaltfläche „Ausgabe hinzufügen“ weitere Produkte hinzufügen.

Speichern Sie anschließend die gesamte Ausgabe mit dem orangenen „Speichern“ Button unten rechts.

Nutzen der Foto-Funktion

Um eine Ausgabe per Foto-Funktion einzutragen, klicken Sie auf das orangefarbige Plus-Symbol (+) und wählen Sie „Kassenbon fotografieren“ aus. Sie werden gefragt, ob Sie ein bestehendes Bild auswählen möchten oder die Kamera für eine neue Aufnahme benutzen wollen.

Bitte achten Sie darauf, dass die Fotos stets den Namen des Geschäfts sowie die Gesamtsumme beinhalten. Zwei oder mehrere Kassenbons fotografieren Sie bitte

einzelnd (während die App einen fotografierten Kassenbon bearbeitet, können Sie bereits ein neues Foto aufnehmen).

Die Foto-Funktion ist für Kassenbons von Supermärkten gut erprobt. Sollten Sie auf Probleme beim Fotografieren von Kassenbons anderer Geschäfte stoßen, können Sie die Ausgaben jederzeit manuell eintippen.

Ausgaben bearbeiten



Wenn Sie auf eine bereits eingetragene Ausgabe klicken, sehen Sie Details zu dieser Ausgabe. Sie können eine Ausgabe im Nachhinein auch ändern oder löschen. Klicken Sie dafür zunächst die betreffende Ausgabe an und wählen Sie dann auf der rechten Seite das Stift- bzw. Papierkorb-Symbol aus.

„Tägliche Ausgaben“ abschließen

Sobald der für Sie vorgesehene Zeitraum von zwei Wochen vorbei ist, können Sie den Befragungsteil „Tägliche Ausgaben“ schließen und damit die Daten senden. Nachträgliche Änderungen sind dann nicht mehr möglich. Sie können sich Ihre Angaben noch ansehen, indem Sie auf das Auge-Symbol neben dem Befragungsteil klicken, diese aber nicht mehr verändern. Sie erkennen dies auch daran, dass der abgeschlossene Befragungsteil in grauer Schrift dargestellt wird.

Wie trage ich meine Ausgaben in der App in den „Täglichen Ausgaben“ ein? – Beispiel mit Anleitung

Hier finden Sie einige Beispiele, wie Sie Ihre Ausgaben in der App im Befragungsteil „Tägliche Ausgaben“ eintragen.

Ausgaben beim Einkaufen im Supermarkt eintragen

Angenommen, Sie haben bei einem Einkauf die folgenden Produkte gekauft:

- 1 Packung Weizenmischbrot für 2,19 Euro
- 3 Packungen Vollmilch für jeweils 1,15 Euro
- 1 Packung geschnittener Käse für 2,09 Euro

- 6 Äpfel für jeweils 0,50 Euro
- 1 Kiste Mineralwasser für 5,99 Euro zzgl. 3,30 Euro Pfand
- 2 Packungen Orangensaft für jeweils 1,79 Euro
- 1 Flasche Weißwein für 4,99 Euro
- 1 Packung Toilettenpapier für 3,89 Euro
- 1 Flasche Duschgel für 1,29 Euro
- 1 T-Shirt für ein 5-jähriges Mädchen für 3,99 Euro

Insgesamt haben Sie 37,76 Euro für den Einkauf bezahlt. Nun fragen Sie sich vielleicht, wie Sie diesen Einkauf richtig eintragen. Ganz einfach: Produkte, die zur gleichen Kategorie gehören, können Sie zusammenfassen und als eine Ausgabe eintragen.

Im Beispiel addieren Sie alle Ausgaben für...

- Lebensmittel & Getränke (z.B. Supermarkt & Discounter):

umfasst: Weizenmischbrot, Vollmilch, Käse, Äpfel, Mineralwasser, Orangensaft, Weißwein – ohne Pfand

- Pfand (inkl. Flaschen-, Dosen-, und Kistenpfand):

umfasst: Pfand

- Körperpflege (Verbrauchsgüter, Geräte, Dienstleistungen):

umfasst: Toilettenpapier, Duschgel

- Bekleidung:

umfasst: T-Shirt

und tragen Sie die Summe der Ausgaben je Kategorie in den „Täglichen Ausgaben“ ein.

Bitte tragen Sie nicht die Gesamtsumme des Einkaufs als eine Position ein, denn der Einkauf umfasst Artikel aus unterschiedlichen Kategorien. Dadurch würden die Ausgaben für einzelne Bereiche überschätzt und für andere Bereiche unterschätzt werden. Dies gilt auch für andere Einkäufe, z.B. im Baumarkt oder der Drogerie.

(Flaschen-) Pfand eintragen

Ausgaben für Pfand tragen Sie bitte unter der Kategorie „Pfand (inkl. Flaschen-, Dosen-, und Kistenpfand)“ ein. Wenn Sie z.B. eine Kiste Mineralwasser für 5,99 Euro zzgl. 3,30 Euro Pfand kaufen, tragen Sie bitte 3,30 Euro unter dieser Kategorie ein.

Wenn Sie Pfand zurückgeben und dafür Geld wiederbekommen, können Sie dies genauso eintragen wie Ausgaben für Pfand und zusätzlich das Feld auswählen „Ich habe den Artikel zurückgegeben“. Der Pfandbetrag fließt dann als negativer Betrag in den „Täglichen Ausgaben“ ein.

Ausgaben für Gastronomie (z.B. Restaurant, Lieferdienst) eintragen

Angenommen, Sie haben im Restaurant oder bei einem Lieferdienst folgende Ausgaben:

- Pizza für 11,90 Euro
- Nudelgericht für 9,50 Euro
- Salat für 8 Euro
- 3 Getränke für jeweils 2,50 Euro
- Trinkgeld 3,10 Euro

Insgesamt haben Sie 40 Euro inklusive Trinkgeld bezahlt. Nun fragen Sie sich vielleicht, wie Sie diese Ausgabe richtig eintragen. Ganz einfach: Addieren Sie alle Ausgaben für Speisen und Getränke sowie Trinkgeld und tragen Sie die Summe (40 Euro) bei der Kategorie „Gastronomiedienstleistungen“ ein.

Diese Einteilung für Ausgaben in der Gastronomie gilt bei allen Ausgaben für Verzehr außer Haus bzw. bei Bestellungen von Speisen und Getränken, z.B. auch für den Verzehr von Getränken in einer Bar oder für den Kaffee zum Mitnehmen vom Bäcker.

Habe ich auch mit meinem Laptop/Computer die Möglichkeit, den Befragungsteil „Tägliche Ausgaben“ zu führen?

Nein. Eine Teilnahme am AusgabenAtlas ist nur über die App möglich.



Universität Mannheim ■ AusgabenAtlas ■ Datenschutz

Datenschutz

Datenschutzerklärung AusgabenAtlas

Vielen Dank für Ihr Interesse an der wissenschaftlichen Studie AusgabenAtlas. Die Studie wird von der Universität Mannheim durchgeführt. Ziel der Studie ist es, das Konsumverhalten der Menschen in Deutschland wissenschaftlich zu untersuchen und herauszufinden, wie viel Geld Bürgerinnen und Bürgern zum Leben zur Verfügung steht und wie hoch die monatlichen Ausgaben sind. Hierfür stellen wir Ihnen Fragen zu Ihren Einnahmen und Ausgaben sowie zu Ihren Wohnverhältnissen.

Die Universität Mannheim ist für die Durchführung der Erhebung und die Verarbeitung der Daten datenschutzrechtlich verantwortlich und zuständig für die Erhebung und Aufbereitung Ihrer Daten. Die Plattform zur Verarbeitung der Daten wird von der Firma hbits CV bereitgestellt. hbits CV fungiert als Datenverarbeiter. Ein entsprechender Vertrag zwischen der Universität Mannheim und hbits CV wurde auf Grundlage der EU-Verordnung 2016/679 geschlossen.

Die App AusgabenAtlas verarbeitet zu Forschungszwecken Ihre App-Nutzungsdaten. Im Folgenden erfahren Sie, welche personenbezogenen Daten wir zu welchen Zwecken verarbeiten, welche Rechtsgrundlage uns das erlaubt, wie lange wir die Daten verarbeiten, an wen die Daten ggf. weitergegeben werden und welche Rechte Sie geltend machen können.

1. Verantwortliche und Datenschutzbeauftragte

Universität Mannheim
Schloss
68161 Mannheim
Deutschland
rektor@uni-mannheim.de

Ausführende Stelle:

Prof. Dr. Florian Keusch

Lehrstuhl für Social Data Science & sozialwissenschaftliche Methodenlehre

A5, 6

68131 Mannheim

Datenschutzbeauftragter Universität Mannheim

Jan Morgenstern

Rechtsanwalt und Fachanwalt für IT-Recht, Datenschutzbeauftragter

datenschutzbeauftragter@uni-mannheim.de

2. Art und Zweck der Datenverarbeitung

Die Studie AusgabenAtlas besteht aus zwei Erhebungsteilen. Im ersten Teil bekommen Sie zu Beginn der Studie Fragen zu Ihren soziodemografischen Merkmalen, Ihren fixen monatlichen Einnahmen & Ausgaben, sowie Ihren Wohnverhältnissen gestellt. Der zweite Erhebungsteil bezieht sich auf Ihre täglichen, variablen Ausgaben. Hier bitten wir Sie, über einen Zeitraum von zwei Wochen Ihre täglichen Ausgaben zu dokumentieren.

Insgesamt wurden 7.000 Personen der deutschen Wohnbevölkerung im Alter zwischen 18 und 70 Jahren zufällig ausgewählt, um an der Befragung teilzunehmen.

Zweck der Datenerhebung ist die wissenschaftliche Untersuchung von Lebensbedingungen von in Deutschland lebenden Personen. Im Rahmen eines Forschungsprojekts der Universität Mannheim soll in diesem Zusammenhang auch untersucht werden, inwieweit die Nutzung einer Befragungssapp die Dateneingabe erleichtern kann.

3. Rechtsgrundlage und Freiwilligkeit

Die Datenverarbeitung erfolgt aufgrund einer Einwilligung gemäß Art. 6 Abs. 1 lit. a DS-GVO.

Damit Sie am Forschungsprojekt teilnehmen können und wir Ihre Daten zu Forschungszwecken verarbeiten dürfen, ist Ihre Einwilligung nötig. Erst mit Ihrer Einwilligung dürfen Ihre Daten zu dem jeweils angegebenen Zweck verarbeitet werden. Die Einwilligung erscheint beim ersten Öffnen der App und ist Teil des Einrichtungsprozesses.

Die Erteilung der Auskunft ist freiwillig. Die Angabe und Verarbeitung der Daten ist erforderlich, um an der Studie AusgabenAtlas teilzunehmen. Ohne diese Angaben ist eine Teilnahme nicht möglich. Nachteile im Falle einer Nichtteilnahme entstehen nicht.

Die Einwilligung in die Verarbeitung der personenbezogenen Angaben kann jederzeit widerrufen werden. Der Widerruf wirkt erst für die Zukunft. Verarbeitungen, die vor dem Widerruf erfolgt sind, sind davon nicht betroffen.

Davon ausgenommen sind die für die Aufwandsentschädigung erforderlichen Daten (Vor- und Nachname, E-Mail-Adresse). Die Verarbeitung ist zur Erfüllung einer rechtlichen Verpflichtung erforderlich, der der Verantwortliche unterliegt (Art. 6 Abs. 1 lit. c i. V. m. Abs. 3 DS-GVO i. V. m. §§ 70, 75 Landeshaushaltsordnung).

4. Arten der durch die App zur technischen Durchführung der Erhebung verarbeiteten Daten und Verarbeitungszwecke

Ihre Daten geben Sie über die mobile App ein, die mit den Betriebssystemen Android und iOS bedient werden kann. Damit Sie endgeräteübergreifend auf Ihre Daten zugreifen können, werden Ihre Angaben immer dann, wenn Sie online (also mit dem Internet verbunden) sind, an eine zentrale Datenbank der App übermittelt. Die Datenbank wird in einem Netzwerk gespeichert, das entsprechend (ISO27000 zertifiziert) geschützt ist und von unserem Datenverarbeiter hbits CV bereitgestellt wird.

Es werden nur die Daten gesammelt, die nötig sind. Persönliche Daten werden für die Anmeldung und Einladungen, die Teilnahme an der Studie und die Unterstützung der Teilnehmenden bei der Beantwortung von Fragen verwendet. Die mobile App verarbeitet die folgenden Daten:

- **Persönliche Daten:** Für Identifizierung, Kontaktaufnahme und Zugang. Dazu gehören E-Mail-Adressen und Gerätetokens sowie Benutzername und Kennwort (verschlüsselt).
- **Studiendaten:** Alle weiteren Daten, die Sie in der App eingeben.
- **Sensordaten:** Nutzung von Sensoren in Geräten. Diese Daten werden den Teilnehmenden in der App als vorläufige Einträge für Ausgabentagebücher gezeigt. Die App kann um Erlaubnis bitten:
 - die Kamera zu benutzen, um Fotos zu machen.
 - auf die Foto-App und die Dateien-App zuzugreifen, um Fotos und Dokumente oder Belege hochzuladen.
- **Nutzerdaten:** Werden während der Nutzung der App gesammelt, um die Servicequalität und Kommunikation zu verbessern. Dazu gehören Art des genutzten mobilen Endgeräts sowie Art und Version des verwendeten Betriebssystems, Datum und Uhrzeit der Abrufe, Meldung, ob der Abruf erfolgreich war sowie innerhalb der App genutzte Suchbegriffe.

Die Zustimmung kann jederzeit über die Einstellungen der App auf dem Gerät des Teilnehmenden an- oder ausgeschaltet werden.

5. Bereitstellung über die App Stores und Nutzung von Push-Benachrichtigungen

Die mobile App wird Ihnen über den Google Play Store oder den Apple App Store zur Verfügung gestellt. Die Nutzung der Stores und die damit verbundene Verarbeitung personenbezogener Daten unterliegt der Vereinbarung zwischen den Betreibern der Stores und Ihnen. Die dort stattfindende Datenverarbeitung unterliegt also nicht der Verantwortung der Universität Mannheim.

Um Ihnen Push-Benachrichtigungen (beispielsweise zur Erinnerung an Ihre tägliche Ausgaben) zusenden zu können, verwenden wir das Notification-Framework Google Firebase Cloud Messaging. Dazu wird bei diesem Service bei der Installation der App ein gerätebezogener pseudonymisierter Schlüssel (Token) als Referenz erstellt. Der Token wird an die Server unseres Datenverarbeiters hbits CV, Brüssel übermittelt und in dem Datensatz gespeichert, der Ihre Zugangsinformationen beinhaltet. Die versendeten Push-Benachrichtigungen werden mit Hilfe dieses Tokens über die Firebase-Server an Ihr Smartphone weitergeleitet. Die Firebase-Server dienen ausschließlich als Übermittler. Die in diesem Zusammenhang gespeicherten Daten werden von uns nicht weiterverarbeitet. Der Empfang von Push-Benachrichtigungen kann von Ihnen jederzeit in den App-Einstellungen Ihres Endgeräts deaktiviert oder aktiviert werden.

6. Cookies

Wir verwenden Erstanbieter-Cookies für die Funktionalität der App und das Benutzererlebnis. Zu den Arten gehören notwendige Cookies (Sitzungsscookies und permanente Cookies). Drittanbieter-Cookies werden nicht verwendet.

7. Empfänger und Kategorien von Empfängern der Daten

Der technische Betrieb unserer Datenverarbeitungssysteme erfolgt durch die Firma hbits CV, (Witte Patersstraat 4, 1040 Etterbeek, Belgien). hbits CV ist ein Spin-off der Forschungsgruppe BRISPO der Vrije Universiteit Brussel (VUB), Pleinlaan 2, Brüssel, Belgien, unter der MOTUS-Lizenzvereinbarung. Ein entsprechender Auftragsverarbeitungsvertrag wurde geschlossen.

Die Universität Mannheim vermietet, verkauft oder teilt keine persönlichen Daten zu finanziellen Zwecken.

Die für die Aufwandsentschädigung erforderlichen Daten werden an das für den Haushalt zuständige Dezernat der Universität Mannheim übermittelt.

8. Trennung und Löschung der Daten

E-Mail-Adressen und Benutzerkennungen sind Hilfsmerkmale, die lediglich der technischen Durchführung der Erhebung dienen. Sie werden von den Angaben zu den Erhebungsmerkmalen getrennt und gesondert aufbewahrt oder gespeichert und nach Abschluss der Überprüfung der

Erhebungs- und Hilfsmerkmale auf ihre Schlüssigkeit und Vollständigkeit vernichtet beziehungsweise gelöscht. Die in elektronischer Form gespeicherten personenbezogenen Daten werden spätestens nach Abschluss der Datenaufbereitung gelöscht beziehungsweise vernichtet. Dies ist spätestens zwei Monatenach der Beendigung der Studie (30.04.25) der Fall.

Vor-/Nachname und E-Mail-Adresse werden für die Zusendung der Aufwandsentschädigung und der Abwicklung der Incentivierung verwendet, die für die freiwillige Teilnahme am AusgabenAtlas gezahlt wird. Die Daten werden zehn Jahre lang aufbewahrt.

Die Rohdaten aus den täglichen Ausgaben werden zunächst erfasst und gespeichert und anschließend im Rahmen der Datenaufbereitung nach einer Systematik codiert. Nur in codierter Form finden sie Eingang in die dauerhaft gespeicherten formal anonymisierten Datensätze, die die Grundlage für die Auswertungen und Ergebnisveröffentlichungen sind.

Im Falle eines Widerrufs der Einwilligung werden die personenbezogenen Daten unverzüglich gelöscht; ausgenommen sind die personenbezogenen Daten, die im Falle der Zahlung einer Aufwandsentschädigung verarbeitet werden.

Anonymisierte Datensätze, die keinen Rückschluss auf Ihre Person enthalten, werden für wissenschaftliche Zwecke verarbeitet und sind in diesem Rahmen für wissenschaftliche Forschungszwecke frei verfügbar (open data).

Ggf. werden die Unterlagen vom zuständigen Universitätsarchiv übernommen und dort in der Regel unbegrenzt aufbewahrt.

9. Persönliche Daten von Kindern

Die Universität Mannheim sammelt zu keinem Zeitpunkt Daten von Kindern und Jugendlichen unter 18 Jahren. Wenn solche Daten entdeckt werden, werden sie entfernt.

10. Ihre Rechte als Betroffener der Datenverarbeitung

Sie haben das Recht:

- gemäß Art. 15 DS-GVO Auskunft über Ihre von uns verarbeiteten personenbezogenen Daten zu verlangen
- gemäß Art. 16 DS-GVO unverzüglich die Berichtigung unrichtiger oder Vervollständigung Ihrer bei uns gespeicherten personenbezogenen Daten zu verlangen
- gemäß Art. 17 DS-GVO die Löschung Ihrer bei uns gespeicherten personenbezogenen Daten zu verlangen

- gemäß Art. 18 DS-GVO die Einschränkung der Verarbeitung Ihrer personenbezogenen Daten zu verlangen
- gemäß Art. 20 DS-GVO Ihre personenbezogenen Daten, die Sie uns bereitgestellt haben, in einem strukturierten, gängigen und maschinenlesbaren Format zu erhalten oder die Übermittlung an einen anderen Verantwortlichen zu verlangen
- Außerdem haben Sie in dem Fall, in dem Sie die Einwilligung gegeben haben, das Recht, Ihre Einwilligung jederzeit ohne Angabe von Gründen zu widerrufen, wobei die Rechtmäßigkeit der aufgrund der Einwilligung bis zum Widerruf erfolgten Verarbeitung nicht berührt wird.

Bitte wenden Sie sich zur Ausübung Ihrer Betroffenenrechte an

ausgabenatlas@uni-mannheim.de.

- gemäß Art. 77 DS-GVO sich bei einer Aufsichtsbehörde zu beschweren. Die Aufsichtsbehörde in Baden-Württemberg ist der Landesbeauftragte für den Datenschutz und die Informationsfreiheit Baden-Württemberg.

Widerspruchsrecht

Sofern Ihre personenbezogenen Daten auf Grundlage von Art. 6 Abs. 1 lit. e DS-GVO (Datenverarbeitung im öffentlichen Interesse) verarbeitet werden, haben Sie das Recht, gemäß Art. 21 DS-GVO Widerspruch gegen die Verarbeitung Ihrer personenbezogenen Daten einzulegen, soweit dafür Gründe vorliegen, die sich aus Ihrer besonderen Situation ergeben. Möchten Sie von Ihrem Widerspruchsrecht Gebrauch machen, senden Sie bitte eine E-Mail an ausgabenatlas@uni-mannheim.de.

Weitere Fragen

Für weitere Fragen zur Studie, Ihren Rechten oder die Verarbeitung Ihrer Daten kontaktieren Sie bitte die Studienverantwortlichen der Universität Mannheim unter ausgabenatlas@uni-mannheim.de.



[Universität Mannheim](#) ■ [AusgabenAtlas](#) ■ [Kontakt](#)

Haben Sie weitere Fragen? Dann schreiben Sie uns eine E-Mail oder rufen Sie uns an.

Telefonisch sind wir von Montag bis Freitag zwischen 9:30 und 11:30 und zwischen 14 und 16 Uhr zu erreichen (außer Mittwoch Vormittag).

Studie AusgabenAtlas

Universität Mannheim
Lehrstuhl für Social Data Science und
Methodology
A 5, 6
68159 Mannheim

Tel.: [+49 621 181-2203](tel:+496211812203)

E-Mail: ausgabenatlas@uni-mannheim.de

Web: www.sowi.uni-mannheim.de/keusch

Appendix G: General data protection declaration

Datenschutzerklärung AusgabenAtlas

Vielen Dank für Ihr Interesse an der wissenschaftlichen Studie AusgabenAtlas. Die Studie wird von der Universität Mannheim durchgeführt. Ziel der Studie ist es, das Konsumverhalten der Menschen in Deutschland wissenschaftlich zu untersuchen und herauszufinden, wie viel Geld Bürgerinnen und Bürgern zum Leben zur Verfügung steht und wie hoch die monatlichen Ausgaben sind. Hierfür stellen wir Ihnen Fragen zu Ihren Einnahmen und Ausgaben sowie zu Ihren Wohnverhältnissen.

Die Universität Mannheim ist für die Durchführung der Erhebung und die Verarbeitung der Daten datenschutzrechtlich verantwortlich und zuständig für die Erhebung und Aufbereitung Ihrer Daten. Die Plattform zur Verarbeitung der Daten wird von der Firma hbits CV bereitgestellt. hbits CV fungiert als Datenverarbeiter. Ein entsprechender Vertrag zwischen der Universität Mannheim und hbits CV wurde auf Grundlage der EU-Verordnung 2016/679 geschlossen.

Die App AusgabenAtlas verarbeitet zu Forschungszwecken Ihre App-Nutzungsdaten. Im Folgenden erfahren Sie, welche personenbezogenen Daten wir zu welchen Zwecken verarbeiten, welche Rechtsgrundlage uns das erlaubt, wie lange wir die Daten verarbeiten, an wen die Daten ggf. weitergegeben werden und welche Rechte Sie geltend machen können.

1. Verantwortliche und Datenschutzbeauftragte

Universität Mannheim
Schloss
68161 Mannheim
Deutschland
rektor@uni-mannheim.de

Ausführende Stelle:

Prof. Dr. Florian Keusch
Lehrstuhl für Social Data Science & sozialwissenschaftliche Methodenlehre
A5, 6
68131 Mannheim

Datenschutzbeauftragter Universität Mannheim

Jan Morgenstern
Rechtsanwalt und Fachanwalt für IT-Recht, Datenschutzbeauftragter
datenschutzbeauftragter@uni-mannheim.de

2. Art und Zweck der Datenverarbeitung

Die Studie AusgabenAtlas besteht aus zwei Erhebungsteilen. Im ersten Teil bekommen Sie zu Beginn der Studie Fragen zu Ihren soziodemografischen Merkmalen, Ihren fixen monatlichen Einnahmen & Ausgaben, sowie Ihren Wohnverhältnissen gestellt. Der zweite Erhebungsteil

bezieht sich auf Ihre täglichen, variablen Ausgaben. Hier bitten wir Sie, über einen Zeitraum von zwei Wochen Ihre täglichen Ausgaben zu dokumentieren.

Insgesamt wurden 7.000 Personen der deutschen Wohnbevölkerung im Alter zwischen 18 und 70 Jahren zufällig ausgewählt, um an der Befragung teilzunehmen.

Zweck der Datenerhebung ist die wissenschaftliche Untersuchung von Lebensbedingungen von in Deutschland lebenden Personen. Im Rahmen eines Forschungsprojekts der Universität Mannheim soll in diesem Zusammenhang auch untersucht werden, inwieweit die Nutzung einer Befragungssapp die Dateneingabe erleichtern kann.

3. Rechtsgrundlage und Freiwilligkeit

Die Datenverarbeitung erfolgt aufgrund einer Einwilligung gemäß Art. 6 Abs. 1 lit. a DS-GVO.

Damit Sie am Forschungsprojekt teilnehmen können und wir Ihre Daten zu Forschungszwecken verarbeiten dürfen, ist Ihre Einwilligung nötig. Erst mit Ihrer Einwilligung dürfen Ihre Daten zu dem jeweils angegebenen Zweck verarbeitet werden. Die Einwilligung erscheint beim ersten Öffnen der App und ist Teil des Einrichtungsprozesses.

Die Erteilung der Auskunft ist freiwillig. Die Angabe und Verarbeitung der Daten ist erforderlich, um an der Studie AusgabenAtlas teilzunehmen. Ohne diese Angaben ist eine Teilnahme nicht möglich. Nachteile im Falle einer Nichtteilnahme entstehen nicht.

Die Einwilligung in die Verarbeitung der personenbezogenen Angaben kann jederzeit widerrufen werden. Der Widerruf wirkt erst für die Zukunft. Verarbeitungen, die vor dem Widerruf erfolgt sind, sind davon nicht betroffen.

Davon ausgenommen sind die für die Aufwandsentschädigung erforderlichen Daten (Vor- und Nachname, E-Mail-Adresse). Die Verarbeitung ist zur Erfüllung einer rechtlichen Verpflichtung erforderlich, der der Verantwortliche unterliegt (Art. 6 Abs. 1 lit. c i. V. m. Abs. 3 DS-GVO i. V. m. §§ 70, 75 Landeshaushaltsordnung).

4. Arten der durch die App zur technischen Durchführung der Erhebung verarbeiteten Daten und Verarbeitungszwecke

Ihre Daten geben Sie über die mobile App ein, die mit den Betriebssystemen Android und iOS bedient werden kann. Damit Sie endgeräteübergreifend auf Ihre Daten zugreifen können, werden Ihre Angaben immer dann, wenn Sie online (also mit dem Internet verbunden) sind, an eine zentrale Datenbank der App übermittelt. Die Datenbank wird in einem Netzwerk gespeichert, das entsprechend (ISO27000 zertifiziert) geschützt ist und von unserem Datenverarbeiter hbits CV bereitgestellt wird.

Es werden nur die Daten gesammelt, die nötig sind. Persönliche Daten werden für die Anmeldung und Einladungen, die Teilnahme an der Studie und die Unterstützung der Teilnehmenden bei der Beantwortung von Fragen verwendet. Die mobile App verarbeitet die folgenden Daten:

- **Persönliche Daten:** Für Identifizierung, Kontaktaufnahme und Zugang. Dazu gehören E-Mail-Adressen und Gerätetokens sowie Benutzername und Kennwort (verschlüsselt).
- **Studiendaten:** Alle weiteren Daten, die Sie in der App eingeben.
- **Sensordaten:** Nutzung von Sensoren in Geräten. Diese Daten werden den Teilnehmenden in der App als vorläufige Einträge für Ausgabentagebücher gezeigt. Die App kann um Erlaubnis bitten:
 - die Kamera zu benutzen, um Fotos zu machen.
 - auf die Foto-App und die Dateien-App zuzugreifen, um Fotos und Dokumente oder Belege hochzuladen.
- **Nutzerdaten:** Werden während der Nutzung der App gesammelt, um die Servicequalität und Kommunikation zu verbessern. Dazu gehören Art des genutzten mobilen Endgeräts sowie Art und Version des verwendeten Betriebssystems, Datum und Uhrzeit der Abrufe, Meldung, ob der Abruf erfolgreich war sowie innerhalb der App genutzte Suchbegriffe.

Die Zustimmung kann jederzeit über die Einstellungen der App auf dem Gerät des Teilnehmenden an- oder ausgeschaltet werden.

5. Bereitstellung über die App Stores und Nutzung von Push-Benachrichtigungen

Die mobile App wird Ihnen über den Google Play Store oder den Apple App Store zur Verfügung gestellt. Die Nutzung der Stores und die damit verbundene Verarbeitung personenbezogener Daten unterliegt der Vereinbarung zwischen den Betreibern der Stores und Ihnen. Die dort stattfindende Datenverarbeitung unterliegt also nicht der Verantwortung der Universität Mannheim.

Um Ihnen Push-Benachrichtigungen (beispielsweise zur Erinnerung an Ihre tägliche Ausgaben) zusenden zu können, verwenden wir das Notification-Framework Google Firebase Cloud Messaging. Dazu wird bei diesem Service bei der Installation der App ein gerätebezogener pseudonymisierter Schlüssel (Token) als Referenz erstellt. Der Token wird an die Server unseres Datenverarbeiters hbits CV, Brüssel übermittelt und in dem Datensatz gespeichert, der Ihre Zugangsinformationen beinhaltet. Die versendeten Push-Benachrichtigungen werden mit Hilfe dieses Tokens über die Firebase-Server an Ihr Smartphone weitergeleitet. Die Firebase-Server dienen ausschließlich als Übermittler. Die in diesem Zusammenhang gespeicherten Daten werden von uns nicht weiterverarbeitet. Der Empfang von Push-Benachrichtigungen kann von Ihnen jederzeit in den App-Einstellungen Ihres Endgeräts deaktiviert oder aktiviert werden.

6. Cookies

Wir verwenden Erstanbieter-Cookies für die Funktionalität der App und das Benutzererlebnis. Zu den Arten gehören notwendige Cookies (Sitzungsscookies und permanente Cookies). Drittanbieter-Cookies werden nicht verwendet.

7. Empfänger und Kategorien von Empfängern der Daten

Der technische Betrieb unserer Datenverarbeitungssysteme erfolgt durch die Firma hbits CV, (Witte Patersstraat 4, 1040 Etterbeek, Belgien). hbits CV ist ein Spin-off der Forschungsgruppe

BRISPO der Vrije Universiteit Brussel (VUB), Pleinlaan 2, Brüssel, Belgien, unter der MOTUS-Lizenzvereinbarung. Ein entsprechender Auftragsverarbeitungsvertrag wurde geschlossen.

Die Universität Mannheim vermietet, verkauft oder teilt keine persönlichen Daten zu finanziellen Zwecken.

Die für die Aufwandsentschädigung erforderlichen Daten werden an das für den Haushalt zuständige Dezernat der Universität Mannheim übermittelt.

8. Trennung und Löschung der Daten

E-Mail-Adressen und Benutzerkennungen sind Hilfsmerkmale, die lediglich der technischen Durchführung der Erhebung dienen. Sie werden von den Angaben zu den Erhebungsmerkmalen getrennt und gesondert aufbewahrt oder gespeichert und nach Abschluss der Überprüfung der Erhebungs- und Hilfsmerkmale auf ihre Schlüssigkeit und Vollständigkeit vernichtet beziehungsweise gelöscht. Die in elektronischer Form gespeicherten personenbezogenen Daten werden spätestens nach Abschluss der Datenaufbereitung gelöscht beziehungsweise vernichtet. Dies ist spätestens zwei Monate nach der Beendigung der Studie (30.04.25) der Fall.

Vor-/Nachname und E-Mail-Adresse werden für die Zusendung der Aufwandsentschädigung und der Abwicklung der Incentivierung verwendet, die für die freiwillige Teilnahme am AusgabenAtlas gezahlt wird. Die Daten werden zehn Jahre lang aufbewahrt.

Die Rohdaten aus den täglichen Ausgaben werden zunächst erfasst und gespeichert und anschließend im Rahmen der Datenaufbereitung nach einer Systematik codiert. Nur in codierter Form finden sie Eingang in die dauerhaft gespeicherten formal anonymisierten Datensätze, die die Grundlage für die Auswertungen und Ergebnisveröffentlichungen sind.

Im Falle eines Widerrufs der Einwilligung werden die personenbezogenen Daten unverzüglich gelöscht; ausgenommen sind die personenbezogenen Daten, die im Falle der Zahlung einer Aufwandsentschädigung verarbeitet werden.

Anonymisierte Datensätze, die keinen Rückschluss auf Ihre Person enthalten, werden für wissenschaftliche Zwecke verarbeitet und sind in diesem Rahmen für wissenschaftliche Forschungszwecke frei verfügbar (open data).

Ggf. werden die Unterlagen vom zuständigen Universitätsarchiv übernommen und dort in der Regel unbegrenzt aufbewahrt.

9. Persönliche Daten von Kindern

Die Universität Mannheim sammelt zu keinem Zeitpunkt Daten von Kindern und Jugendlichen unter 18 Jahren. Wenn solche Daten entdeckt werden, werden sie entfernt.

10. Ihre Rechte als Betroffener der Datenverarbeitung

Sie haben das Recht:

- gemäß Art. 15 DS-GVO Auskunft über Ihre von uns verarbeiteten personenbezogenen Daten zu verlangen
- gemäß Art. 16 DS-GVO unverzüglich die Berichtigung unrichtiger oder Vervollständigung Ihrer bei uns gespeicherten personenbezogenen Daten zu verlangen
- gemäß Art. 17 DS-GVO die Löschung Ihrer bei uns gespeicherten personenbezogenen Daten zu verlangen
- gemäß Art. 18 DS-GVO die Einschränkung der Verarbeitung Ihrer personenbezogenen Daten zu verlangen
- gemäß Art. 20 DS-GVO Ihre personenbezogenen Daten, die Sie uns bereitgestellt haben, in einem strukturierten, gängigen und maschinenlesbaren Format zu erhalten oder die Übermittlung an einen anderen Verantwortlichen zu verlangen
- Außerdem haben Sie in dem Fall, in dem Sie die Einwilligung gegeben haben, das Recht, Ihre Einwilligung jederzeit ohne Angabe von Gründen zu widerrufen, wobei die Rechtmäßigkeit der aufgrund der Einwilligung bis zum Widerruf erfolgten Verarbeitung nicht berührt wird.

Bitte wenden Sie sich zur Ausübung Ihrer Betroffenenrechte an ausgabenatlas@uni-mannheim.de.

- gemäß Art. 77 DS-GVO sich bei einer Aufsichtsbehörde zu beschweren. Die Aufsichtsbehörde in Baden-Württemberg ist der Landesbeauftragte für den Datenschutz und die Informationsfreiheit Baden-Württemberg.

Widerspruchsrecht

Sofern Ihre personenbezogenen Daten auf Grundlage von Art. 6 Abs. 1 lit. e DS-GVO (Datenverarbeitung im öffentlichen Interesse) verarbeitet werden, haben Sie das Recht, gemäß Art. 21 DS-GVO Widerspruch gegen die Verarbeitung Ihrer personenbezogenen Daten einzulegen, soweit dafür Gründe vorliegen, die sich aus Ihrer besonderen Situation ergeben. Möchten Sie von Ihrem Widerspruchsrecht Gebrauch machen, senden Sie bitte eine E-Mail an ausgabenatlas@uni-mannheim.de.

Weitere Fragen

Für weitere Fragen zur Studie, Ihren Rechten oder die Verarbeitung Ihrer Daten kontaktieren Sie bitte die Studienverantwortlichen der Universität Mannheim unter ausgabenatlas@uni-mannheim.de

Appendix H: COICOP classification

Code	Description (default language)	Filters
000	[Ihr Suchbegriff ist leider zu ungenau]	various_other_expenses
010	Lebensmittel & Getränke (z.B. Supermarkt & Discounter)	foodstuffs
023	Tabakwaren auch für Shishas, E-Zigaretten (ohne Raucherartikel)	various_other_expenses
031	Bekleidung (einschl. Reparatur, Reinigung, Miete) für Herren, Damen und Kinder, Bekleidungsstoffe, -zubehör, -artikel	clothes_accessories
032	Schuhe und Zubehör (einschl. Reparatur) für Herren, Damen und Kinder	clothes_accessories
041	Mietzahlungen (ohne Nebenkosten)	variable_property_costs
045	Strom, Gas, Wasser (u.a. flüssige und feste Brennstoffe, Energieträger zum Heizen und Kühlen)	variable_property_costs
051	Wohn- und Gartenmöbel, Beleuchtung, Einrichtungsgegenstände (einschl. Zubehör, Lieferung, Install., Miete, Reparatur)	furniture_household_goods
052	Heimtextilien, Möbelsstoffe, (einschl. Miete, Reparatur), auch Gardinen, Bettzeug, -wäsche	furniture_household_goods
053	Elektrische Haushalts-, Reinigungs- und Küchengeräte (einschl. Reparatur, Installation, Miete und Leasing)	furniture_household_goods
054	Glaswaren, Geschirr, Besteck und nicht elektrische Gebrauchsgüter für den Haushalt (einschl. Reparatur und Miete)	furniture_household_goods
055	Elektrische und nicht elektrische Werkzeuge, und Geräte und andere Gebrauchsgüter für Haus und Garten (einschl. Miete, Leasing, Reparatur)	furniture_household_goods
056	Waren- und Dienstleistungen und Verbrauchsgüter, z.B. Reinigungs- und Pflegemittel für den Haushalt	furniture_household_goods
061	Arzneimittel, Medikamente, Hilfsmittel und Gesundheitsprodukte (einschl. Eigenanteil/Zuzahlung, Miete, Reparatur, Instandhaltung, medi. Alarmsysteme)	health_personal_care
062	Ambulante kurz und langfristige Gesundheitsdienstleistungen (einschl. Eigenanteil/Zuzahlung, Vorsorge, Impfungen, Zahnarzt, Heilbehandlungen)	health_personal_care
063	Stationäre kurz und langfristige Gesundheitsdienstleistungen (einschl. Eigenanteil/Zuzahlung, Rehabilitation)	health_personal_care
064	Andere Dienstleistungen für die Gesundheit (einschl. Eigenanteil/Zuzahlung, Labordienste, Röntgen, Ultraschall, Notfalltransporte)	health_personal_care
069	Vorauszahlungen für pauschalisierte Zuzahlungen zur gesetzlichen Krankenversicherung	health_personal_care
071	Kauf von Fahrzeugen Pkws, Kraft- und Fahrrädern, Kutschen für Zugtiere (ohne für Freizeit)	transport
072	Waren, Dienstleistungen und Zubehör für den Betrieb von Privatfahrzeugen (einschl. Reparatur, Wartung, Pflege, Leasing, Miete), Parkplatz-, Maut-, Fahrunterrichts- und Zulassungsgebühren (ohne Kfz-Steuer)	transport
073	Personenbeförderungsdienstleistungen im Schienen-, Straßen-, Flug-, (ohne Pauschalreise) See- und Binnenschiffsverkehr (einschl. kombinierte Tickets, Fahrrad- und Seilbahnen, Sessellifte)	transport
074	Transportdienstleistungen für Waren z.B. Post- und Kurierdienst-, Umzugs- und Lagerungsdienstleistungen, Zustellung von Waren (ohne Warenwert)	transport
081	Geräte der Informations- und Kommunikationstechnik z.B. Telefone, Computer, Peripheriegeräte, Ton-, Bildempfangs-, -aufnahme- und -wiedergabegeräte, Aufzeichnungsträger einschl. Zubehör, Einzel-, und Ersatzteile	information_communication
082	Software (ohne Spiele)	information_communication

083	Informations- und Kommunikationsdienstleistungen (einschl. Bereitstellung, Reparatur, Miete, Install., Fernseh-, Kabel-, Pay-TV-, Rundfunk-, Streaming- und Abogebühren)	information_communication
091	Langlebige Freizeitgüter z.B. Film-, Fotoausrüstung, Optische Geräte, Wohnmobile, Boote, Flugzeuge, Pferde, Ponys (ohne Zugtiere), Schwimmbecken (ohne aufblasbare)	culture_sport_free_time
092	Andere Freizeitgüter z.B. Spiele, Spielzeug, Hobbywaren, Spiele Apps, -konsolen, -computer, Artikel f. Feierlichkeiten (ohne Adventskränze), Geräte- u. Ausrüstung für Sport- u. Camping	culture_sport_free_time
093	Gartenartikel, Pflanzen, Blumen, Dienstleistung von Friedhofsgärtnern, und Haus-, Klein- und Wildtiere (einschl. Gebrauchs- und Verbrauchsgüter)	culture_sport_free_time
094	Freizeitdienstleistungen z.B. Miete, Leasing, Reparatur, Wartung, Überwintern von Freizeitausrüstung, -geräten f.d. Freizeit z.B. Foto-, Film-, Camping-, Sportgeräte, Wohnmobil, Spiele, -zeug, Veterinär-, Freizeit-, Sport-, Erholungs- u.a. Dienstleistungen	culture_sport_free_time
095	Gebrauchsgüter für Kultur z.B. Musikinstrumente, Audiovisuelle Medien	culture_sport_free_time
096	Kulturdienstleistungen z.B. Kino, Theater, Konzert, Museen, Bibliotheken, Musik-, Tanzunterricht, Dienstleistungen v. Fotografen, Raummieten	culture_sport_free_time
097	Zeitungen, Bücher, Magazine, und Schreibwaren und Zeichenmaterial, Büroartikel, Druckerzeugnisse	culture_sport_free_time
098	Pauschalreisen ins In- oder Ausland	culture_sport_free_time
101	Frühkindliche Bildung und Grundschulziehung (Kita 0-3 Jahren und Grundschulklasse 1-4, ohne Verpflegung)	various_other_expenses
102	Bildungseinrichtungen des Sekundarbereichs (Klassenstufe 5-13, Berufsschulen, ohne Verpflegung)	various_other_expenses
103	Bildungseinrichtungen des postsekundären, nicht tertiären Bildungsbereichs (Kunst-, Hoch-, Fachhoch-, Berufs- und technische Oberschulen, Fach-, Berufsakademien, Abendgymnasien, Kollegs, ohne Verpflegung)	various_other_expenses
104	Bildungseinrichtungen des Tertiärbereichs (Meister-, Technikerausbildung, Bachelor-, Masterstudiengänge an Unis, wissenschaftl. Hochschulen, ohne Verpflegung)	various_other_expenses
105	Nicht einstuftbare Bildungseinrichtungen z.B. Nachhilfeunterricht, Hauslehrer, Sprach-, EDV-Kurse, Buchführungskurse, Erste-Hilfe-Kurse (nicht für den Erwerb einer Fahrerlaubnis), Volkshochschulkurse	various_other_expenses
111	Gastronomiedienstleistungen z.B. in Restaurants, Cafes, Straßenverkauf mit und ohne Bedienung	foodstuffs
112	Beherbergungsdienstleistungen z.B. in Hotels, Gasthöfen, Resorts, Jugendherbergen, Pensionen, Ferienwohnungen, Campingplätzen	services_education
121	Versicherungsdienstleistungen z.B. Abschluss-, Makler- und Hebegebühren für Versicherungsverträge, a. n. g. Finanzdienstleistungen, z.B. Konto-, Kredit-, Überweisungsgebühren, Gebühren und Courtagen für Darlehen,	various_other_expenses
122	Finanzanlagenberatung	various_other_expenses
131	Körperpflege (Verbrauchsgüter, Geräte, Dienstleistungen)	various_other_expenses
132	Andere persönliche Gebrauchsgegenstände z.B. Schmuck, Uhren, Devotionalien, Reise-, Begräbnis, Raucher-, Babyartikel (ohne Verbrauchsgüter, einschl. Reparatur, Umarbeitung, Miete)	various_other_expenses

133	Dienstleistungen sozialer Einrichtungen z.B. Kinderheime, -betreuung, vor oder nach der Schule, Ammen, Alten-, Behindertenpflege, Essen auf Rädern, Schiedsstellen	various_other_expenses
139	Andere Dienstleistungen z.B. v. Kirchen, Verwaltungs-, Rechtsberatungs-, Immobilien-, Begräbnisgebühren, Prostitution Steuern (ohne Grundsteuer, Grunderwerbsteuer) Einkommen-, Erbschafts-, Schenkungs, Kfz-, Hunde-, Kulturabgabe-, Betten-, Zoll-, Fischerei-, Jagd- und Börsenumsatzsteuer	services_education
161	Beiträge für die betriebliche Altersversorgung z.B. Pensionsfonds, Direktversicherung und für private Versicherung z.B. Kranken-, Pflege-, Kfz-, u.a. Schaden- u. Unfallversicherungen	various_other_expenses
163	Mitgliedsbeiträge, Geldspenden (u.a. Beiträge, Unterhaltszahlungen, Bußgelder, Gerichtskosten, Spieleinsätze)	various_other_expenses
164	Ausgaben für die Bildung von Sachvermögen z.B. Kauf, Um-, Ausbau, werterhöhende baul. Maßnahmen an Gebäuden, Grundstücken, Eigentumswohnungen, Wärmeisolierung, Kernsanierung, Nichtgewerblicher Gold- und Silberkauf, -barren, Private Einlagen von Selbstständigen	various_other_expenses
166	Ausgaben für die Unterhaltung von Grundstücken, Gebäuden und Eigentumswohnungen z.B. Grundsteuer, Instandhaltung, werterhaltende baul. Maßnahmen, Wohngebäudeversicherungen	various_other_expenses
167	Ausgaben für die Bildung von Geldvermögen	various_other_expenses
168	Pfand (inkl. Flaschen-, Dosen-, und Kistenpfand)	various_other_expenses
169	Sonstiges [Suchbegriff ohne genaue Auswahl speichern]	various_other_expenses
999		
99999999	Bitte geben Sie ein passendes Suchwort ein.	

Conducting a test of an app-based smart survey project: learnings & recommendations from the German field test

Documentation as part of the Smart Survey Implementation (SSI) Project
(work package 2.1)

Prepared by:

Destatis

Lasse Häufglöckner
Johannes Volk

University of Mannheim

Florian Keusch
Maren Fritz

hbits

Joeri Minnen
Pieter Beyens

This document summarises steps and tasks to consider for mastering an app-based survey project. Destatis conducted a field test together with the University of Mannheim as part of the SSI project. To set up the smart survey, the software Motus by hbits was used, which is why the recommendations refer to conducting an app test using Motus. However, we believe the insights gained are transferrable to the use of other survey software as well. The study was meant to test the use of a receipt scanner. To facilitate this, a microservice was developed, which had to be trained for use in Germany. The document is organized in three main parts: the preparation of the study before the field phase (1), conducting the study (2), and the follow-up after the field phase (3). For every phase, the most important steps and tasks are summarised.

Content

1.	Preparation: before the field phase	4
1.1	Organising the project	4
1.1.1	Clarify goals	4
1.1.2	Set up roadmap	4
1.1.3	Define roles and responsibilities	5
1.1.4	Project communication	5
1.1.5	Budget	5
1.2	Conceptual survey preparation	6
1.2.1	Basic survey design	6
1.2.2	Plan study setup/flow	6
1.2.3	Plan communication measures	7
1.2.4	Plan incentives	8
1.3	Plan participant support during field phase	9
1.3.1	Ways of contact	9
1.3.2	Staff training	9
1.4	Technical preparation	9
1.4.1	Understand the backoffice	9
1.4.2	Create individual study elements	10
1.4.3	Adjust language & country specific elements	10
1.4.4	Set up communication measures	11
1.4.5	Set up accounts for respondents	11
1.4.6	App Store upload	11
1.4.7	Testing & bug fixing	12
1.4.8	Servers & namespace	12
1.4.9	Define roles in backoffice	13
1.4.10	Train algorithm	13
1.5	Legal preparations	14
2	Conducting the study: during the field phase	15
2.1	Organisational considerations	15
2.2	Support of respondents	15
2.3	Technical monitoring	15
3	Follow-up: after the field phase	16
3.1	Data export	16
3.2	Closing the study in the backoffice	16
3.3	Closing the namespace	16

3.4	Feedback round.....	16
4	Appendix.....	17
4.1	Example of a bug report form	17

1. Preparation: before the field phase

1.1 Organising the project

1.1.1 Clarify goals

- Clearly define the goals of the app test
 - o Explain (and document) what is meant by which goal as concrete as possible
- Define SMART goals
 - o Specific
 - o Measurable
 - o Achievable
 - o Reasonable
 - o Time-bound
- Define concrete technical goals for the app
 - o Define the state of development of the app and of parts of the app in detail
 - o Define what functionalities should be developed/ integrated
 - o Use concrete examples explaining the functionalities in the app
 - o Explain (and document) technical terms so that everybody involved understands them

1.1.2 Set up roadmap

- Create a roadmap with timeline
 - o What steps need to be done to reach the goals?
 - o By whom and by when?
 - o Set up a clear time schedule for all tasks and subtasks throughout the project with clear deadlines
 - o Plan with at least 10% time buffer (e.g. three weeks before fieldwork start; everything needs to be ready)
- Include all working areas in the roadmap
 - o Technical development
 - o Organisational topics
 - o Recruitment of participants
 - o Legal processes
- Discuss the roadmap with all project partners to check if it is correct and to identify internal dependencies (e.g., between project partners, work packages, tasks...)
- Identify possible dependencies concerning other (external) persons or institutions, e.g., sample vendors, app stores
 - o Make sure that dependencies do not cause problems in conducting the study in the envisioned time frame
- Also consider holidays of team members
 - o Who is available when?
 - o Who can be the substitute for whom?
 - o Especially when it comes to times of field work, it is important that all necessary persons are available
- Check regularly (e.g., every 4 weeks) where you are in the project and if the timeline is still realistic

1.1.3 Define roles and responsibilities

- Discuss and document in detail who is responsible for which tasks and subtasks
 - o What exactly is included in which role and responsibility?
 - o Who is responsible to decide about which questions?
 - o Who has to inform whom?
 - o Who needs to be in which meetings?
- Consider prior experiences and expertise of team members
 - o Include everyone's expertise: assign roles according to experiences and skills
 - o Identify who has specific knowledge on topics (e.g., an expert on the use of the applied app/survey software like Motus)
 - o Identify gaps in expertise and where to find the necessary skills or information
 - Find out how to solve critical tasks for which no one has any previous experience
 - Identify other colleagues who may have previous knowledge
 - Try to get experienced colleagues in the team

1.1.4 Project communication

- Decide with all partners how often you want to meet in which constellation
 - o E.g., a weekly meeting every Friday
 - o Block a regular meeting in all necessary team members' calendars
 - o Define one person in charge of the meeting, who will create an agenda and moderate the meeting
 - o Write minutes of each meeting: Define a person who writes minutes and sends it around as a common ground for all team members
- Establish a joint platform (e.g., MS Teams/Sharepoint, Dropbox, Google Drive) that every project member has access to, in order to share and save documents
- Standardise ways of communication
 - o Discuss how to communicate in individual situations, e.g. how to report bugs (make use of a bug report form (see example attached))

1.1.5 Budget

- Set up a budget plan
- Which parts of the survey cost what amount of money?
 - o Incentives
 - o Personnel costs
 - o Paper and stamps (for sending letter and reminder)
 - What response rate do you expect?
 - o Server for app
 - o Server for information web site
 - o E-Mail server (to communicate with respondents, e.g. for reminders, password change)
 - o Sampling process
 - o
- Plan with buffer

1.2 Conceptual survey preparation

Before actually setting up the study in the survey software (e.g., Motus), it is advisable to create a detailed plan of the entire study, including all study parts. Such a study protocol could also be used for ethical requests or other requests that may be necessary.

1.2.1 Basic survey design

- Who do you want to recruit for the survey?
- How do you get in touch?
 - o Postal Mail
 - o E-mail
 - o Interviewer at the door
 - o Telephone call
 - o Access panel
 - o Advertisements
- How do you get necessary contact information?
- How many persons do you want to complete your survey?
 - o What response rate do you expect?
 - o How many persons do you need to/want to contact?
- Survey period:
 - o Which time frame do you reserve for the study?
 - o How long will the diary period take?
- How long should it be possible to answer the survey?
- How many days does a person have to report information?
 - o E.g., 2 weeks diary of all products they buy
- When do you want to start?
 - o Is it a good time to do a survey? Any holidays? Christmas?
- When should be the last day for answers?
- Do you want to do softlaunch/test?
 - o Before contacting potentially thousands of persons, you definitely should do some kind of test with only a few hundred persons to check if everything is working as it should?
 - o Conduct an internal pretest within the project team or with external participants (if resources are available)

1.2.2 Plan study setup/flow

Create a plan in advance on how the study should look like, this includes:

Flow: Sequences of surveys, pages & diary

- Which elements/screens are needed in the app?
 - o Registration screen including forgot password screen
 - o Consent screens
 - o Tutorial screens
 - o Legal information screens, e.g., data protection, GDPR
 - o Information screens; e.g., after uploading a receipt & profile information screen
 - o Questionnaires
 - o Diaries
 - o
- in which order should the elements occur?

Individual elements:

- Set up texts in Word
 - o Write down the texts and check it with colleagues
- Set up questionnaires in Word/Excel
 - o How many questions?
 - o Which questions?
 - o Which filters?
 - o Which kind of question types?
 - o Response types: How are respondents supposed to answer (e.g., only numerical responses possible for age question)
 - o
- Plan diary
 - o Which information do you want to collect?
 - o Which level of detail?
 - o What data collection methods should be applied (manual entry and/or smart feature, e.g., a receipt scanner)
 - o Which classification?
 - o How do you want to collect the information (with a smart feature/manual entries)?

1.2.3 Plan communication measures

Conceptual:

How and what do you want to communicate with respondents?

- Which communications channels do you want to use?
 - o Postal mail,
 - Who is responsible for postal invitations that are sent back (for example, because the recruited person has moved)?
 - Who will classify why the invitations returned (moved, not findable, active refusal, etc.)?
 - o email,
 - o personal communication via phone or in person by interviewers,
 - o push messages within the app,
 - o notifications,
 - o SMS,
 - o ...
- Which messages do you want to send?
 - o Welcome information (e.g., email)
 - o Information for registration (e.g., email)
 - o Reminder after inactivity (e.g., push notification and/or email)
 - o Information after successful upload (e.g., push notification)
 - o Information that survey will be closed in X days (e.g., email)
 - o "Thank you" information after completion (e.g., email)
 - o Information about incentive (e.g., email)
 - o Error messages (e.g., push notification)
 - o Message with link to change password (e.g., email)
 - o Confirmation on successful password change (e.g., email)
 - o ...
- At what point and under which circumstances should which communication be sent out?
 - o e.g., reminders per email after not adding entries to a diary after 3 days
- Create a plan regarding all necessary communication measures towards the participants during their participation in the field phase (e.g., reminders)
- Write down texts in word and check them with colleagues

- Implement texts as communication measures in the backoffice (e.g. create the necessary email & push notification templates)
 - Test whether communication measures are properly sent out as planned at the correct point in time and under the correct circumstances during the survey period
- Should there be an option for respondents to get in touch with you? → See point 1.3.

Technical

Which options are offered by the backoffice of your survey software to communicate with respondents?

- Look into the possibilities to send out communication measures from the backoffice
 - Be sure about which messages can be sent automatically via the backoffice (e.g., reminders, password reset emails) and which have to be sent out manually via an external email software (e.g., emails with incentives, answering support requests) and prepare accordingly
- Check the study flow and check the (technical) conditions under which emails & push messages can be sent out in each study stage
 - What is technically possible?
 - What needs to be done to set it up as you want?
- Can respondents deactivate communication (entirely or only some communication channels) via the app?
 - What other communication measures do you implement for these cases, if any?

1.2.4 Plan incentives

- Decide whether an incentive should be given to respondents for participating
- How much should the incentive be?
 - Predict how many incentives in total will approximately paid out to estimate the total sum needed
- How should the incentive be paid?
 - Prepaid? Postpaid?
 - Bank transfer, voucher or cash?
 - From which bank account will the money be sent?
 - Which voucher? From which company?
 - The supply may take some time
 - It may be necessary to check with your legal department whether it is okay to use vouchers of certain companies
 - How would the cash be sent out to the respondents?
 - How do you get necessary information from respondents (e.g., bank account, email address) to send the incentives?
 - The easiest way to do so would be using the activation process on initial login (respondents have to provide and confirm a personal email address, under which they can be contacted), but it is an additional burden to participate if you have to register first
- Define concrete and verifiable conditions under which the incentive will be paid
 - e.g., participate for a certain amount of time while providing a certain amount of data inputs (like expense entries in an expense diary)
 - Make sure the conditions can be feasibly checked for in order to send out incentives correctly and in time: make sure that the data to check this is included in the data export of the study

1.3 Plan participant support during field phase

As respondents may encounter (technical) problems, a way of contact and providing support in case of questions or problems is necessary

1.3.1 Ways of contact

- How will (technical) support of respondents be managed?
- Decide for ways of contact:
 - o Email? Which email address? Who is responsible to answer?
 - o Telephone hotline? Which number? At which times? Also during evenings and weekends? Who is responsible to answer?
 - o Website with FAQ and contact form (and additional e-mail address & telephone number)
 - Which information should be provided on the website?
 - Who is setting up the website? Which server?
- Document all incoming questions
 - o Prepare an overview list over all requests, including the problem/topic, its status (solved/open), the respondent ID/username and who of the support team had contact (see appendix).
 - o It can be used as an overview for all staff members to keep track of support requests, also in case of follow-up questions.

1.3.2 Staff training

- Staff is necessary to answer the support phone hotline and emails
- Make sure that the staff knows all necessary details about the use of the survey software (in this case: Motus), as well as about respondent management & background of the study itself
 - o A training of the support staff is absolutely necessary
 - Who is doing the training? When? How long?
 - o Including small exercises on typical respondent requests and how to solve them in the backoffice
- How to deal with technical problems?
 - o Make sure that the staff has access to the backoffice
 - o Can it be solved by the staff managing the backoffice/study (e.g., via adjustments in the backoffice) or is reporting to another person necessary?
 - o Prepare procedure for issues/bugs that could not be resolved,
 - o Prepare a form for bug reporting that the staff can use if an issue could not be resolved (see example form attached)
 - o Have a clear structure/workflow where to or to whom to escalate bugs

1.4 Technical preparation

To set up the whole survey technically in the survey software (in this case in Motus) you need to plan enough time to get familiar with the backoffice and to do all the tasks

1.4.1 Understand the backoffice

Define staff that will set up the study in the backoffice. The staff that will set up the backoffice needs to understand the backoffice and should become familiar with it as soon as possible

- Make sure to have early access to the backoffice of the applied survey platform/software
- Get an overview of the backoffice and its functionalities
 - o Take your time to find your way around the backoffice
 - o Experiment with the options in the survey flow and get an overview of the options for creating questionnaires and diaries

- If possible, ask a colleague with previous experience in the survey software to give you an introduction to all functionalities

In the case of Motus:

- Understand its modular design
 - Questionnaires, Diaries, ... can be created individually
 - And can then be implemented into the “Flow” in order to actually define the sequence of the study
 - Understand the logic of how Flow, diary and classifications are interrelated
 - In the flow, a diary can be defined to appear at a certain point in the study procedure. This diary has to be set up and adjusted before and a classification (e.g., COICOP, HETUS) can be implemented
- Anticipate technical limitations, problems that may arise from discrepancies between what is planned and what is possible to implement in Motus

1.4.2 Create individual study elements

- Questionnaires

There are two ways of creating questionnaires:

 - Import existing questionnaires from a previous study as .mpk file
 - If necessary, adjust the questionnaire (individual questions, question types, sequence of questions, etc.)
 - Create questionnaire from scratch
 - Create it in Excel or Word first, including question text, answer options and filters.
 - Create the questionnaire in Motus afterwards (in order to not have to change things often)
- Diaries (including classifications, COICOP)
 - Import existing diary if possible
 - Make sure that the used classification fits the purpose and that categories apply
- Make use of the export/import function to save time
 - All individual elements (questionnaires, diaries, classifications, communication) can be exported and imported into the study in the backoffice as a .mpk file
 - Also, the entire study can be exported
 - In case that existing questionnaires or diaries from former studies can be used, it is useful to import those
- Define start and end time of survey, condition when a survey can be ended by respondents

1.4.3 Adjust language & country specific elements

- Adjust app for the applicable languages/countries
 - Get overview over what labels and texts have to be changed or translated
 - Check how and where to actually adjust texts
 - Either directly in the backoffice
 - Or via XLIFF files
- XLIFF files
 - Make sure how to use and adjust labels via XLIFFs
 - Ask colleagues with previous experience
 - Use handy programme, e.g., poedit instead of txt
 - Have the necessary XLIFF files ready in time

1.4.4 Set up communication measures

- For communicating via email:
 - o It is necessary to ask for the contact email address of respondents (e.g., during registration)
 - o This means the activation process has to be applied
 - respondents enter their email address when first logging into the app & have to click a confirmation link
 - this of course does not apply if the personal email addresses are already available for the whole sample (e.g. in case of samples from previous studies, panels, ...)
- Motus also needs an email address for automatic responses (e.g., noreply@studyname.com) and from which any emails are sent out to respondents
 - o If you want to use a specific email address (i.e., not the Motus standard one) you have to set up this email address and need an email server which has enough power to send out hundreds of emails at the same time, and potentially also receive lots of emails at the same time.
 - o In Motus the project email address can be specified, as well as the reply address, in case respondents want to reply directly on the email. The same goes with the contact details of the institution (name, address, telephone number)
 - o When no email address of respondents is known an anonymous email address has to be used at read in of the respondents. This email address is specified in the settings of the back-office.
- Think about how persons should be able to start the survey:
 - o Activation process, yes/ no?
 - o Mandatory change of password? Yes/ no?
 - If yes: which rules for password creation do you want to use?
 - o Check box for GDPR conditions? Yes/ No?

1.4.5 Set up accounts for respondents

- Accounts have to be set up for respondents to be able to log in & participate
- Accounts can be created by importing respondent accounts via a .csv import (under "Respondents" in the Motus backoffice)
 - o A .csv file has to be created, containing all accounts in one file, one account per line
 - o Basic information has to be added for each account in the columns
 - o Necessary information: initial email address (does not need to actually exist), username, password

1.4.6 App Store upload

- In case an own app is used, consider that getting the app into the app store can take some time (i.e., several weeks) and that there are some hurdles
- It is important to comply with the current rules of the app stores in order to not be rejected
 - o Check current rules of app stores (iOS, Android) on what information is needed that the app is accepted in the stores
 - o Check if changes to these rules may occur during the field phase (it might be announced beforehand by the app store operators)
- Plan enough time (4 weeks at least), the app may be rejected the first time
- Screenshots (of the app running on certain devices) are needed in the correct aspect ratio
 - o There is the option to not only upload plain screenshots but to use specifically designed ones that contain explanation and text on it, in addition to showing the app screenshot itself (see example attached). Those are shown in the app store entry and can be seen by the users

- Decide whether a specifically designed screenshot should be used or if plain screenshots are sufficient (plain screenshots are sufficient for acceptance in the stores but may convey less trust, reputation and interest for respondents, and therefore may lead to less response)
- Screenshots for at least 2 different screen sizes are needed for iOS and Android (= at least 4 sets of screenshots)
- Create an icon for the app
- If you don't want to start with an app without any rating or comments, download the app before you start the field phase with different devices and write comments and ratings
 - Ask friends/colleagues to also do so
 - A positive rating might be important for some persons to participate

1.4.7 Testing & bug fixing

- Testing in advance is crucial to assure that the study and app setup is correct and to avoid a high amount of support requests during the field phase
- Test in advance with enough time to fix issues and bugs
 - Plan enough time so that you could also do another upload in the app store if necessary
 - Have frontoffice & backoffice ready for the test
 - Have different test devices ready (at least 1 Android, 1 iOS)
- Have test accounts ready for all testing team members
 - Make use of the .csv import to create users (see chapter 1.4.4), for fast & easy test account setup
- For reporting bugs:
 - Use bug report form (see example attached) for standardised and easy collection of bugs, including the same information on every issue
 - Create screenshots and write down how to reproduce the issue as concrete and detailed as possible (e.g., what was done before the bug happened, which steps exactly led to the issue occurring, how many participants reported the bug, which study parts are affected)
 - Create two bug lists:
 - 1) bugs that have to be resolved BEFORE the field phase/app usage (= must haves)
 - 2) bugs that are optional for the field test, which can be resolved down the road

1.4.8 Servers & namespace

- Decide for a domain name (e.g. AusgabenAtlas) for Motus frontoffice, backoffice and core, e.g.
 - app.ausgabenatlas.de
 - backoffice.ausgabenatlas.de
 - AusgabenAtlas app in iOS and Android store
- Typically, it is required to change/add DNS entries for email
- Configure email: sending bulk email must be possible, receiving inbox must have enough storage size
- Decide on how to deploy Motus: at your own premises, or at hbits
 - At your own premises: full control, data is at your side, on your servers
 - At hbits: dedicated namespace required in cluster for most privacy, app with own styling etc. With a namespace the back-office, front-office and database are

separated through containerisation. The institution has administrator rights on the back-office. The database is on hbits servers. Hbits has only a technical entry to the database.

- Email for automatic response is needed (and email server which has enough power)

1.4.9 Define roles in backoffice

- Define (different) roles in the backoffice
 - o Who has the right to do what?
 - o Motus offers quite a lot of different options (e.g., right to edit the flow of the study, to access respondent data, to only view the study setup, etc.). Use the various rights that can be assigned to each individual user. Not every person working in the help desks needs to have the right to change the survey flow.

1.4.10 Train algorithm

If you want to use a general algorithm (e.g., OCR algorithm to scan receipts) you might have to train this algorithm for your national country. In our case we trained a receipt scanning microservice and collected the necessary training data (receipts collected from supermarkets) to do so.

- Talk early in the project with developers about the possible need of training
- Plan enough time to do trainings. You might need a lot of time (several months) for the training
- You might need a lot of material/data (e.g., receipts) to adjust/train the algorithm for your country
- You might need a concept on how to train the algorithm:
 - o Which data is needed?
 - o How much training data is needed? How large does the training data set need to be?
 - o How long will collecting the data need?
 - o How do you collect the data?
 - o On which platform do you label / train the data?
 - o ...
- You might need persons who are...
 - o Collecting the data (e.g., collect receipts)
 - Bear in mind that receipts have to be collected from real supermarket purchases, e.g., it makes sense to instruct as many team members as possible (if possible also from outside of the project team) to collect their receipts
 - Plan with several weeks to months for that
 - o Providing the data (e.g., scan all the receipts)
 - Instruct several team members specifically for that task
 - Plan with several weeks here too, depending on how many receipts have to be scanned by how many team members
 - o Checking and correcting the results of the algorithm
- In our case, the situation was as follows:
 - o Several team members annotated the receipts over a time span of several weeks in order to train the algorithm
 - o Collected receipts were being photographed and analysed by Optical Character Recognition (OCR) to identify the plain text on the receipts. The photos of the receipts and the respective OCR results of each receipt were uploaded into an annotation tool
 - o In the annotation tool, team members defined which text parts contain which information (e.g., total sum, product name, product price)

1.5 Legal preparations

- Get in contact with your data protection office as soon as possible
 - o Take your time and explain in detail
 - What you want to do
 - What data should be collected and how
 - How many respondents are expected and how they will be recruited
 - Who has access to what type of data
 - How the data is protected within the Motus system
 - What happens to the data once the field period ends
 - Who will work with the data
 - Analyse data anonymized
 -
- Plan a constant exchange and involve also technical persons early to make sure that all technical aspects are reported and understood correctly
- Which information needs to be given (both to the DPO and respondents)? GDPR?
- How exactly should the respondent give their active consent in the app? (e.g., by selecting a checkbox)
- Who of the support staff team is allowed to do what/ to see which information/ has access to which data via the backoffice?
- Is an additional contract between different project partners needed to make sure to fulfil the legal criteria of data protection offices?
 - o Bear in mind that your used namespace must also fulfil the legal requirements

2 Conducting the study: during the field phase

2.1 Organisational considerations

- Keep weekly meetings between project partners also during field phase
 - o Fixed meeting dates help to coordinate

2.2 Support of respondents

- There needs to be support staff during the field phase who answers requests and questions by respondents (See chapter 1.3)
 - o This should be handled by the staff trained in advance (see 1.3)
- Typical requests by respondents are:
 - o Help with technical problems (e.g., login, reset password etc.)
 - o Answer questions regarding the study itself (e.g., questions on privacy, why am I being contacted, where does my contact data come from, who can see my answers, what is the purpose of the study, ...)
 - o Information on non-participation (e.g.: “I do not want to participate because I do not have enough time“)
- For bugs/problems that cannot be resolved immediately: Create a bug ticket or fill out a bug form and send it to the app developers
 - o App developers (e.g., hbits) have then to decide on the basis of the ticket if the bug is critical and has to be resolved
 - o Support staff should stay in close contact with the developers about possible bugs

2.3 Technical monitoring

- Regularly check the back office to make sure that everything is going well. That includes:
 - o Does the activation process work?
 - o Do logged in respondent accounts correctly follow the study flow?
 - o Can individual study steps in the flow (questionnaires, diaries) be completed? Are accounts not forwarded to the next step?
 - o ...
- This can also be done by the field phase support staff
 - o In case an issue occurs, the support staff needs to immediately report it to the rest of the team (and app developers)
- Do a daily export of the data

3 Follow-up: after the field phase

3.1 Data export

- Export respondent data and answers as a JSON and check data for completeness
- Conduct first exploratory analysis to make sure that all required data are available

3.2 Closing the study in the backoffice

- Make sure respondents cannot enter the study any more after a certain closing date
 - o This can be assured by setting the study time frame to certain start and end dates
 - o Make sure that respondents are informed about the closing date
- Secure .mpk files (of questionnaires, diaries, communication measures, ...) for future use
 - o If necessary, make screenshots of all screens on the app (and/or create video) for documentation
- Export the entire study for future use
- Record demonstration videos about the study design, showing the front- and backoffice (in case necessary)

3.3 Closing the namespace

- Close the namespace and delete all data
- Delete / remove access of accounts that can manage or edit the namespace
- ...

3.4 Feedback round

- Do a review and feedback meeting on the entire project, to discuss things that went good, things that could be improved for future projects
 - o Create a document on lessons learned
- Discuss things to do after the field phase:
 - o Download an anonymize response data
 - o Data evaluation & result preparation
 - o Checking data quality
 - o Additional data, which is not directly analysed for the research question may be available (e.g., survey response data, photo metadata, paradata)
 - Discuss how to save this data (anonymously)
 - Discuss use cases and what the data could be used for

4 Appendix

4.1 Example of a bug report form

Serial no. No. (from list of requests):

Support staff member who is/was handling this bug:

1. account name (login):

2. time (date, time) / period of the error:

3. what is going wrong:

4. what should actually happen:

5. end device used:

6. version of the app:

7. connection (WLAN / mobile data):

8. attempts made so far to rectify the error:

9. how many users are affected by the error:

10. screenshots

11. further details:

The SSI Recruitment Field Test in Norway

Documentation Report 2024

Aina Holmøy and Nina Berg

TALL

SOM FORTELLER

NOTATER / DOCUMENTS

2025/3

In the series Documents, documentation, method descriptions, model descriptions and standards are published.

© Statistics Norway

Published: 21 January 2025

ISBN 978-82-587-1085-8 (electronic)

ISSN 2535-7271 (electronic)

Symbols in tables	Symbol
Category not applicable Figures do not exist at this time, because the category was not in use when the figures were collected.	.
Not available Figures have not been entered into our databases or are too unreliable to be published.	..
Confidential Figures are not published to avoid identifying persons or companies.	:
Decimal punctuation mark	.

Preface

This publication by Statistics Norway (SSB) provides documentation of the SSI Recruitment Experiment Survey 2024 in Norway, which is part of WP2.1 of the Smart Survey Implementation (SSI) grant ESS-SSI 2023-25.

The field test and associated qualitative interviews with respondents and telephone interviewers were planned and conducted by Aina Holmøy from the Division for Social Surveys and Nina Berg from the Division for Methods. The delivery of the project was made possible through the contributions of the data collection team and interviewers at Statistics Norway, as well as the contribution of Professor Florian Keusch, University of Mannheim, Professor Peter Lugtig, Utrecht University, and their respective teams.

Statistics Norway, 13.01.25

Arvid Olav Lysø

Abstract

This report documents the activities carried out by Statistics Norway as part of the WP2.1 SSI Recruitment Experiment Survey in Norway, under the 2023–2025 Smart Survey Implementation ESSnet grant. The primary aim of the field test and related activities was to evaluate different methods for recruiting participants to a smart survey, focusing on the use of telephone interviewers (CATI) and login solutions with different levels of assumed trust. The survey engaged 1 973 households, testing with and without interviewer-led recruitment in combinations with self-reporting via SMS links directly or the national Altinn platform. Additionally, the survey was supplemented by a Post-Q survey, in-depth interviews with participants and drop-outs, and a workshop with the telephone interviewers.

The results from the field test indicate that CATI recruitment significantly increased response rates, nearly doubling participation compared to self-administered modes without interviewers. However, response rates were notably low among seniors and individuals with lower education levels. The field test also highlighted the trade-offs between ease of use and security: SMS with hyper link to survey being the most user-friendly solution, while the Altinn platform the more secure, associated with greater trust. When interviewers were involved, response rates were similar across the two login methods. However, in the absence of interviewers, SMS links delivered the highest response rates compared to Altinn. Response bias tend to increase without interviewers, particularly for Altinn.

From the user experience of the participants, we learned that the use of app technology and receipt scanning were well-received, except by the oldest age group, who are less familiar with technology. The contact strategy effectively legitimized the survey across all age groups, though usability issues were noted with saving the progressive web application (PWA) on mobile home screens, which caused uncertainty and may have contributed to survey dropout. When asked, users did not express a clear preference for login methods, though security measures such as national ID-Porten/BankID were expected. We also note that telephone interviewers took on a larger role than traditionally, in verifying authenticity, motivating respondents, and providing support, and were especially appreciated by seniors.

The interviewers noted that Statistics Norway's strong reputation was a key motivator for participation. However, the survey's complexity and high response burden, along with concerns about security and opening hyperlinks, especially for seniors, negatively impacted completion rates. Despite the app's generally high usability, interviewers suggested making the app available in app store to enhance trust. The multi-channel contact strategy¹ proved effective, although respondents often overlooked information and instructions. Interviewers emphasized that their support was instrumental in reducing dropouts and that their cohesive, small team structure allowed for efficient technical support and follow-up. While the back-office system ensured smooth coordination, improvements to the Service Desk's response time and accuracy were recommended.

Drawing from the field test and the qualitative insights from the users; future surveys should prioritize maintaining interviewer support and a secure login solution. Additionally, efforts should continue to simplify survey and enhance application further to ensure participation rates, improve data quality, and to maintain Statistics Norway's trust within society.

¹ By multi-channel contact strategy, we refer to several contact forms in the recruitment process, such as the information letter, recruitment call, sms, etc.

Contents

Preface	3
Abstract.....	4
1. Background.....	6
2. Sample.....	7
3. Survey instruments	8
4. Field test design	9
4.1. Contact and registration period	10
5. Data Collection	11
5.1. Sub-sample 1) CATI	11
5.2. Sub-sample 2) No CATI	12
5.3. Incentives	13
5.4. Response rate.....	13
5.5. Type of devices	17
6. Data quality	19
6.1. Non-response	19
6.2. Bias.....	22
6.3. Summary	26
7. Data delivery	28
7.1. Data agreement	28
8. User experience	29
8.1. Involvement and engagement	29
8.2. Usability and user experience	30
8.3. Trust and privacy safeguards	31
8.4. Qualitative insights on recruitment.....	32
9. Interviewer feedback	34
9.1. Engagement and trust.....	34
9.2. Response burden and usability.....	35
9.3. Survey communication.....	35
10. Recommendations for future surveys	36
10.1. Key learnings	36
10.2. Actions for future surveys.....	37
References	38
Appendix A: Altinn letters to sub-sample 1) CATI.....	39
Appendix B: SMS texts to sub-sample 1) CATI.....	41
Appendix C: Altinn letters to sub-sample 2) NO CATI	43
Appendix D: SMS texts to sub-sample 2) NO CATI	45
Appendix E: Screenshots from the web application	47
Appendix F: Recruitment interview CATI (Blaise)	50
Appendix G: Web questionnaire	54
Appendix H: Data files	63
Appendix I: Overview qualitative data sources	68
Appendix J: Non-response.....	69
Appendix K: Bias	73

1. Background

The ESS-SSI 2023-25-grant, titled the Smart Survey Implementation (SSI) project, is a collaboration between several European national statistical institutes and universities. The project started on 1 May 2023 and will run to May 2025. The SSI project aims to develop a data collection process that provides reliable and comparable statistical information through the use of smart surveys.

Smart surveys intelligently combine self-report questions with data collected from sensor-enabled devices such as smartphones, wearables, and other technologies, aiming to enhance data quality, reduce burden on participants, and provide more timely and granular data.

The SSI project focuses on three key areas: 1) citizen engagement and participation, 2) usability and user experience, and 3) trust and privacy safeguards in smart surveys. The long-term goal is to enable the seamless integration of new smart features into different applications by using microservices and best-practice guidelines developed during the project.

Statistics Norway's contribution to the project involves conducting a field test, the Recruitment Experiment Survey with 2 000 respondents. The aim of the field test is to gain more insight into effective recruitment methods for smart surveys. For the field test, we used the survey application (diary) from the Norwegian Household Budget Survey in 2022. This diary incorporates a smart feature, that allows respondents to photograph receipts with their smartphones, which are then automatically processed using optical character recognition (OCR) technology.

Chapters 2-5 will detail the sample and survey instruments used in the field test, the test design, the data collection process, and data delivery. Chapter 6 presents an analysis of data quality, addressing on non-response and bias. Chapter 7 describes the data delivery format from Statistics Norway to the SSI project. Additionally, Chapter 8 provides a qualitative assessment of user experience, while Chapter 9 focuses on feedback from the interviewers. Finally, in Chapter 10, we summarize our findings and present recommendations to guide future surveys.

2. Sample

The target population in the Household budget survey includes private households in Norway. A household is defined as all individuals permanently residing at the same address who share a common food budget. Institutional residents are not included in the survey. Statistics Norway's household register covers this population, which serves as the survey population.

For the purposes of data collection, the survey population was restricted to households that contained at least one person in the age group 18–79 years.

The household register divides households into 23 different categories. This classification can be used to define six types of households: 1) living alone, 2) couples with no resident children, 3) couples with small children (0–5 years), 4) couples with older children (6–17 years), 5) lone parent with children (0–17 years), 6) other types of households. This in turn allows us to oversample for household types that would otherwise be too few in the sample.

The gross sample consists of 1 973 households distributed across the six household types mentioned above.

In each household, a reference/contact person was selected. This is the individual we follow if the household splits, and the person we contact for recruitment to the survey.

In families with children living at home (single-family household), the contact person was randomly selected among the adults in the household, defined as members of the parents' generation.

In families without children living at home (single-family households) or in multi-family households, the contact person was randomly selected, with the one condition that the person had to be between 18 to 69 years old.

If no one in the household was younger than 70, the youngest person in the household was selected as the contact person.

3. Survey instruments

The survey instrument, based on the Household Budget Survey 2022, is a Progressive Web Application (PWA), that can be used either as a mobile application or a website on any device, regardless of screen size. One of the key advantages of the PWA is its responsiveness, which was a major factor in choosing this solution. It functions both online and offline, allowing participants to access it on a tablet or desktop computer if they did not want to install the app on their mobile device.

The households used the web application to record their private expenses for a 7-day period, similar to a diary. The app offered the option to take photos of receipts using optical character recognition (OCR), also called scanning, or to manually register their purchases and products items. Additionally, the web application included an integrated web questionnaire.

For this field test, two key adjustments were done to the web application from the Norwegian Household Budget Survey (HBS) in 2022: The households not contacted by telephone interviewers, could directly change their diary week in the application, and the questionnaire was reduced from approximately 40 minutes in HBS 2022 to 15-20 minutes in the field test.

Additionally, the telephone recruitment interview (CATI) was reduced in length compared to HBS 2022, and household composition data was uploaded from the household register instead of being confirmed during the interview, as was done in HBS 2022.

For a detailed description and illustration of the diary, see Appendix E, which includes screenshots and a video walkthrough of the workflow of the web application. Appendix F shows the CATI interview and Appendix G the web questionnaire.

4. Field test design

The aim of the field test was to examine the impact of using CATI interviewers and two different login methods for the survey.

The sample was divided into two sub-samples. In sub-sample 1 (CATI), households were recruited through a brief telephone interview (CATI), and interviewers followed up with respondents during the diary week. In sub-sample 2 (No CATI), households were neither recruited nor followed-up by interviewers.

Each sub-sample was further divided evenly based on different login methods. Group A) Low trust and group B) High trust. Group A received an SMS on the diary start date with a direct link to the web application. Group B received an SMS on the diary start day directing them to go to Altinn's homepage, where they could access a letter with more information about the survey and a link to the web application.

Altinn is a government platform widely used for official communication with citizens in Norway. It is used for submitting the tax form and other public services. It is highly trusted by the population and employs a two-step-authentication login system via ID-porten/BankID which is standard for accessing government, banking, health services, etc. in Norway.

Participants were directed to Altinn's homepage to login with ID-porten/BankID, which automatically logged them into the web application via single-sign-on. Those who received the web application link directly via SMS also logged in with two-step authentication using ID-porten/BankID, but only once, at the first login. While both methods used secure authentication, hyperlinks via SMS might be perceived as less secure and less trustworthy compared to the Altinn login process. Based on these assumed perceptions, the SMS login method was classified as the "low trust" solution, while the Altinn login method was classified as the "high trust" solution.

This design resulted in four experimental groups, as shown in figure 4.1.

Figure 4.1 Sample size in each experimental group

Survey mode:	Login solution:	
	A) LOW TRUST	B) HIGH TRUST
	SMS with link to survey	SMS: "Go to Altinn platform"
1) CATI	1A) CATI and LOW trust (500 households)	1B) CATI and HIGH trust (496 households)
2) NO CATI	2B) No CATI and LOW trust (500 households)	2A) CATI and HIGH trust (477 households)

Source: The SSI Recruitment Experiment Survey 2024. Statistics Norway

4.1. Contact and registration period

The registration period refers to the time during which a household records its expenses using the web application. In HBS in Norway this period spans 7 days, referred to as the registration week.

For sub-sample 1) CATI, households were randomly assigned a contact period, i.e., a week during which the households were to be contacted by an interviewer for a recruitment/telephone interview (CATI). During the recruitment interview, households were assigned a registration week, initially set at the first full week after the recruitment interview. However, households had the option to postpone the registration period by up to three weeks.

For sub-sample 2) No CATI, the households were randomly assigned a registration week without prior contact by an interviewer. They also had the option to postpone their registration week by up to three weeks, which they could do directly in the diary.

The reference or contact person could be changed in agreement with respondents during the telephone interview for sub-sample 1) CATI. For sub-sample 2) No CATI, the contact person could not be changed.

5. Data Collection

The data collection period for the field test ran from April 2 to May 27, 2024, with the registration period spanning from April 8 to May 19.

As outlined in Chapter 4, the sample was divided into two sub-samples. 1. CATI – Households were contacted by an interviewer for a short telephone recruitment interview. 2. No CATI – Households received electronic information about the survey and got a link to the web application on the start day of their registration.

5.1. Sub-sample 1) CATI

The field test for sub-sample 1) CATI consisted of three parts:

- A telephone interview of approximately 5-10 minutes.
- Registration of the households' private expenses using a web application for 7 days.
- A household web questionnaire integrated into the web application (approx. 20 minutes).

Three days before the interviewers began calling, the respondents received a letter in the Altinn portal with information about the survey and privacy details. This letter (Appendix A) explained that a telephone interviewer would call to conduct a short interview. It also described the purpose of the survey, the importance of participating, the gift certificate of NOK 500 for those completing the survey (see section 5.3) and gave assurance regarding data security. The households also received a text message referring to the letter in Altinn.

Following the telephone interview, on the registration start day, the household received a new letter in Altinn with instructions and a link to the survey (Appendix A).

Group 1A) CATI and low trust (from figure 4.1) also received an SMS at the registration start day with a direct link to the survey, while group 1B) CATI and high trust, received an SMS directing them to Altinn.no to access the survey link there.

One or two days into the registration period, an interviewer called the contact person in the household to offer assistance, e.g., with installing the web application on the mobile. The contact person also received a couple of text messages during the registration week, regardless of whether they had started the registration or not. The content of the SMS messages differed for households that had started or not in the web application. Households not started, got a reminder to start registration, and those started got tips about the registration.

After the registration period, respondents received an SMS reminding them about the gift certificate or completing all parts of the survey. Towards the end of the first week after the registration period, interviewers contacted those who had not yet completed the survey to remind them the need to complete all parts to qualify for the gift certificate.

In all SMS reminders to those not started, group 1A) CATI and low trust, received a direct link to the survey in the SMS, while group 1B) CATI and high trust, got an SMS with instructions to go to altinn.no to access the link.

The communication plan for sub-sample 1) CATI, is illustrated in table 5.1.

Both Altinn letters and SMS templates for sub-sample 1) CATI can be found in Appendix A and C.

Table 5.1 Survey communication plan for sub-sample 1) CATI

Week	Weekday	Contact channel	Category	Respondent's status
W-2	Thursday	Altinn	Information letter about survey, and CATI contact from SSB	Not started
W-1	Monday	SMS	Information about start of CATI recruitment (Registration week could be changed in CATI*)	Not started
	Thursday	SMS	Reminder	Not started
W-0 Registration week	Monday	Altinn	Information about registration and login information	Recruited
	Monday	SMS	Login information**	Recruited
	Tuesday	CATI	Call to those not started***	Login information sent
	Wednesday	SMS	Reminder to start registration**	Login information sent
	Wednesday	SMS	Tips for collecting receipts	Started
	Friday	SMS	Reminder to start registration**	Login information sent
	Friday	SMS	Tips regarding weekend expenses	Started
W+1	Monday	SMS	Thank you for participating	Finished and started
	Thursday	CATI	Call to those not finished	Started
W+2	Monday	SMS	Reminder to finish survey	Started

* The registration week was possible to change in the CATI recruitment interview, within a 3-week period.

** Different path to the hyperlink for group 1A) CATI and Low trust and group 1B) CATI and High trust (figure 4.1)

*** If the interviewer could not reach the household this day, they tried to call several times during the registration week.

Source: The SSI Recruitment Experiment Survey 2024. Statistics Norway

5.2. Sub-sample 2) No CATI

The field test for sub-sample 2) No CATI consisted of two parts:

- Registration of the households' private expenses using a web application for 7 days.
- A household questionnaire integrated in the web application (approx. 20 minutes).

Three days before the registration week started, households received a letter in the Altinn portal with information about the survey, privacy information, and instructions for completing the survey. This letter (Appendix B) also outlined the specific period for which the household should record their expenses.

On the start day of the registration, households received another letter in Altinn (Appendix B) with additional details about expense registration in the web application and a link to the survey.

Group 2A) No CATI and low trust (from table 2.1) also received an SMS on the start day with a direct link to the survey, while group 2B) no CATI and high trust, received an SMS instructing them to visit Altinn.no and access the link.

Like for sub-sample 1) CATI, the contact person in each household received two text messages during the registration week. After the registration week, respondents were sent an SMS reminding them that they would receive a gift certificate of NOK 500 upon completing all parts of the survey.

In all SMS reminders to those not started, group 2A) No CATI and low trust, received a direct link to the survey, while group 2B) No CATI and high trust, got instructions to visit Altinn.no to access the link.

The communication plan for sub-sample 2) No CATI, is illustrated Table 5.2.

Templates for both Altinn letters and SMS messages for subsample 2) No CATI can be found in Appendix B and D.

Table 5.2 Survey communication plan for sub-sample 2) NO CATI

Week	Weekday	Contact channel	Category	Respondent's status
W-1	Thursday	Altinn	Information letter about survey and registration week (registration week selected by SSB*)	Not started
W-0 Registration week	Monday	Altinn	Letter with login information	Not started
	Monday	SMS	Login information**	Not started
	Wednesday	SMS	Tips for collecting receipts	Started
	Wednesday	SMS	Reminder to start registration**	Login information sent
	Friday	SMS	Tips regarding weekend expenses	Started
	Friday	SMS	Reminder to start registration**	Login information sent
W+1	Monday	SMS	Thank you for participating	Finished and started
W+2	Monday	SMS	Reminder to finish survey	Started

*It was possible for the respondent to change the Registration in the app.

** Different path to the hyperlink for group 2A) No CATI and Low trust and group 2B) No CATI and high trust (figure 4.1)

Source: The SSI Recruitment Experiment Survey 2024. Statistics Norway

5.3. Incentives

All participants received a gift certificate to the value of NOK 500 upon completion of the survey. The gift certificate can be used in almost 6 000 stores in Norway. This incentive was mentioned in all letters and reminders and was communicated by interviewers for those participating in sub-sample 1 (CATI).

To receive the gift certificate, the respondent had to fulfil the web questionnaire and submit at least one expense over the whole 7-day registration period.

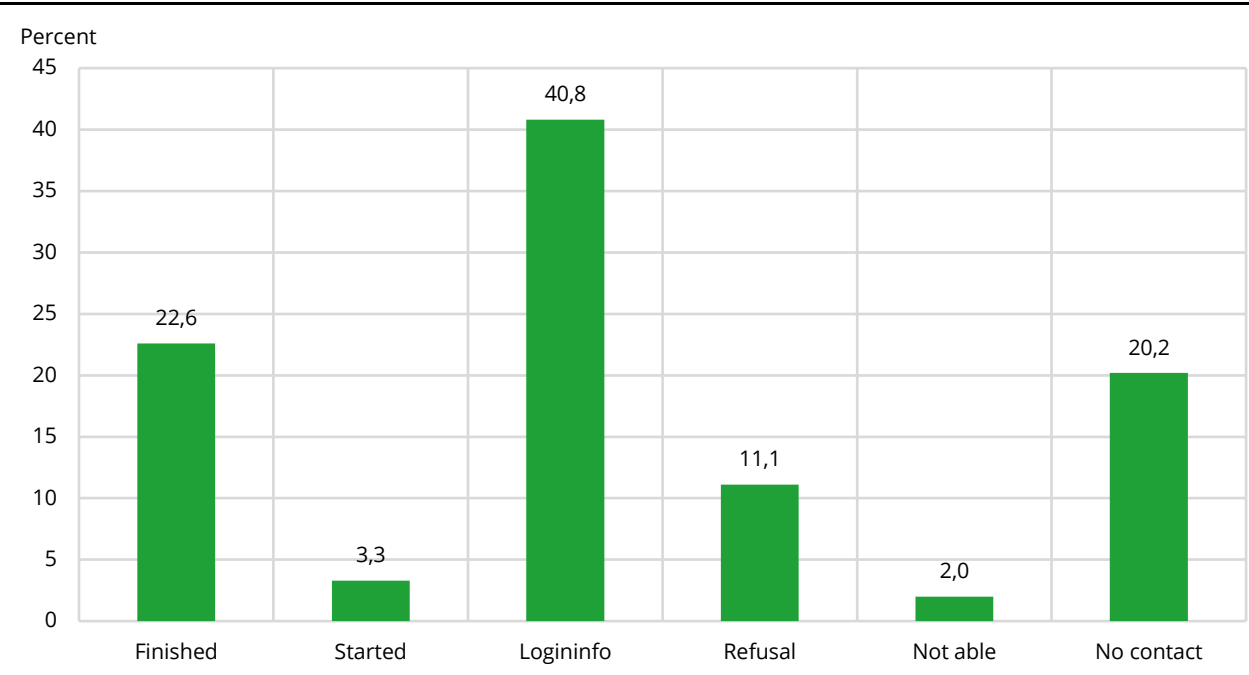
5.4. Response rate

As illustrated in figure 5.1, a total of 22,6 percent of the sample finished the survey. To finish the survey means that the respondent has completed all parts of the survey. For sub-sample 1A) CATI, this includes a short telephone recruitment interview, recording of expenses for one week, and completing the web questionnaire. For sub-sample 1B) No CATI, it entails only the registration of the expenses and completion of the web questionnaire. The term started refers to households that started registering expenses or filling out the web questionnaire but did not complete all parts.

The term "logininfo" in figure 5.1, means that the respondent has received a link to the web application, but did not start recording or filling out the web questionnaire.

The "No contact" group consists of households in sub-sample 1) CATI that the interviewers did not reach by phone. Since these households did not conduct the recruitment interview, they did not receive a link to the web application either.

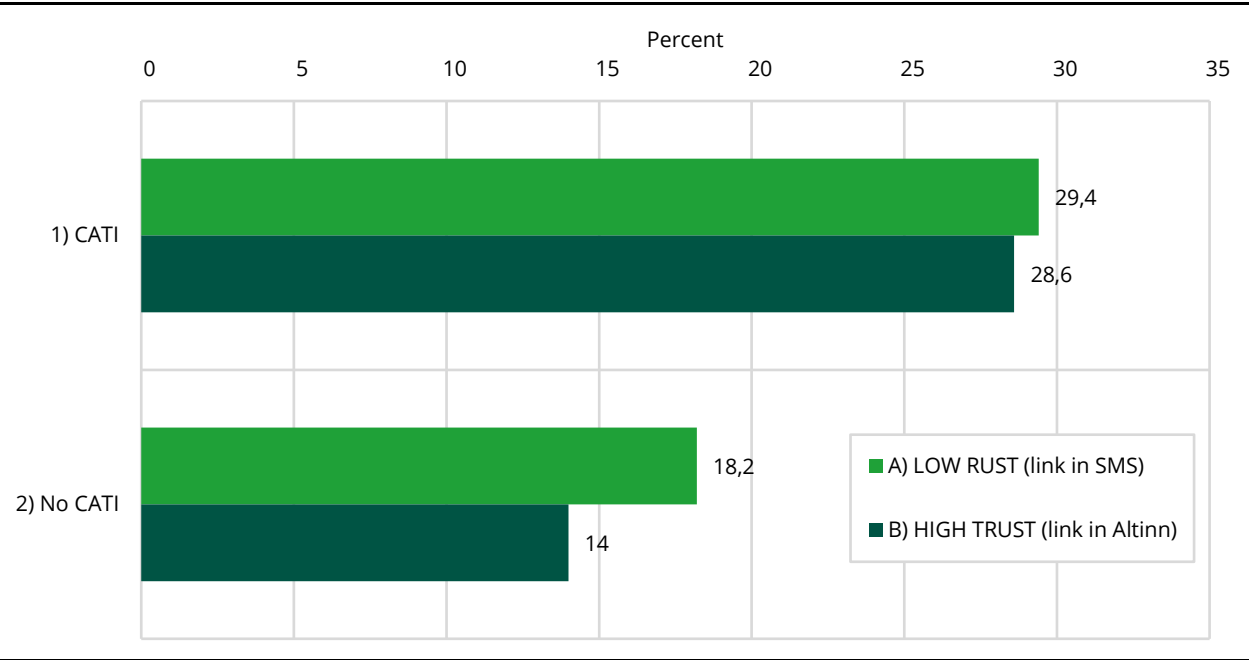
Figure 5.1 Response rate and non-response



Source: The SSI Recruitment Experiment Survey 2024. Statistics Norway

Figure 5.2 shows the response rate across the four experimental groups. The group with the highest response rate is 1A which used CATI and included a direct link to the survey in SMS (Low trust). This approach is nearly identical to the strategy used in the Norwegian HBS 2022, and the response rate is equivalent (29.5 % in HBS 2022). In contrast, the group with the lowest response rate is 2B, which used No CATI recruitment and provided the survey link only via Altinn (High trust). The response rate for this group is less than half of that observed in group 1A.

Figure 5.2 Response rate in the four experiment groups



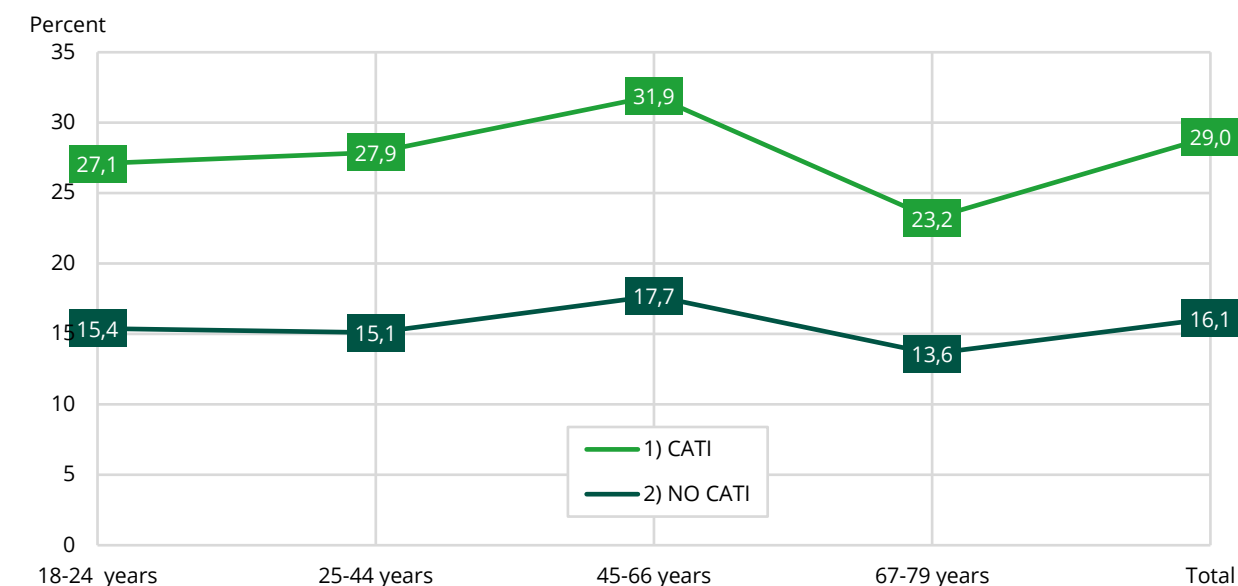
Source: The SSI Recruitment Experiment Survey 2024. Statistics Norway

A total of 41 percent of the respondents in sub-sample 1) CATI were recruited to the survey by an interviewer. Figure 5.3 shows that 29 percent of the respondents in this sub-sample finished all

parts of the survey. This means that almost 30 percent dropped out of the survey after they were recruited. In sub-sample 2) No CATI, 16,1 percent of respondents finished the survey.

When examining the response rate in sub-sample 1) CATI and 2) No CATI by age-groups, figure 5.3 exhibits that the age group 45-66 years has the highest response rate, while the 67 to 79 age group has the lowest, for both sub-samples.

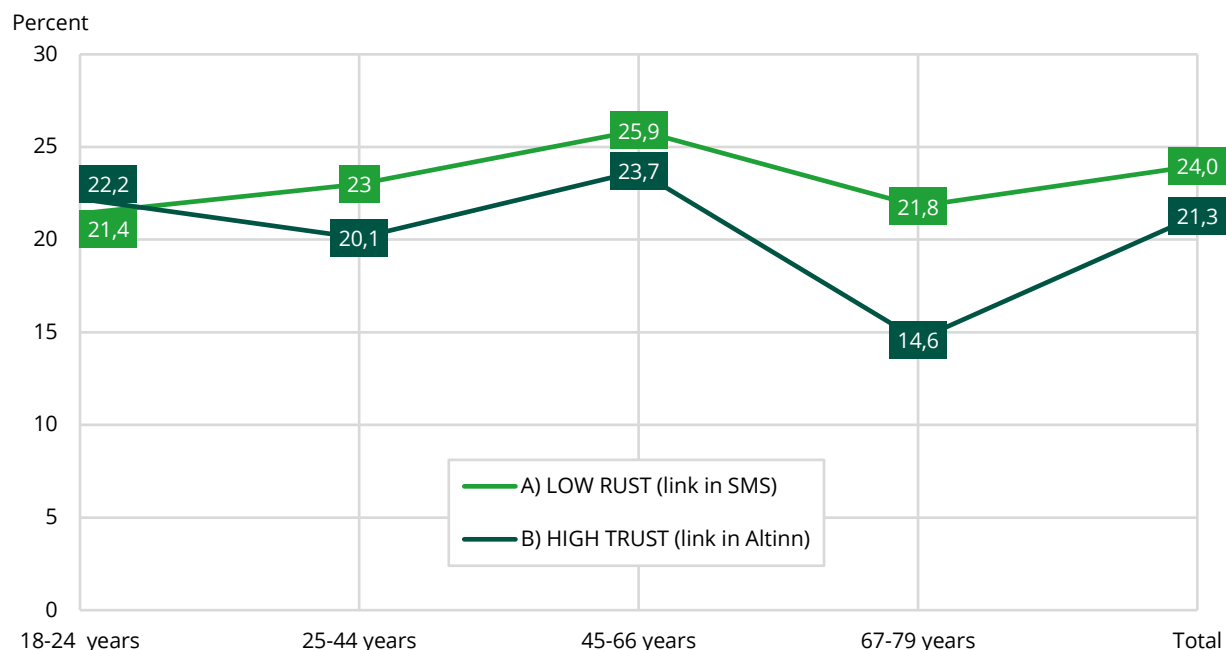
Figure 5.3 Response-rate by age, 1) CATI versus 2) No CATI



Source: The SSI Recruitment Experiment Survey 2024. Statistics Norway

Figure 5.4 compares the response rate between group A) Low trust (households receiving link to the survey in SMS and group B) High trust (households receiving link to the survey only in Altinn, without taking into account whether the households were recruited by CATI or not. In group A 24 percent finished the survey and 21,3 percent finished in group B.

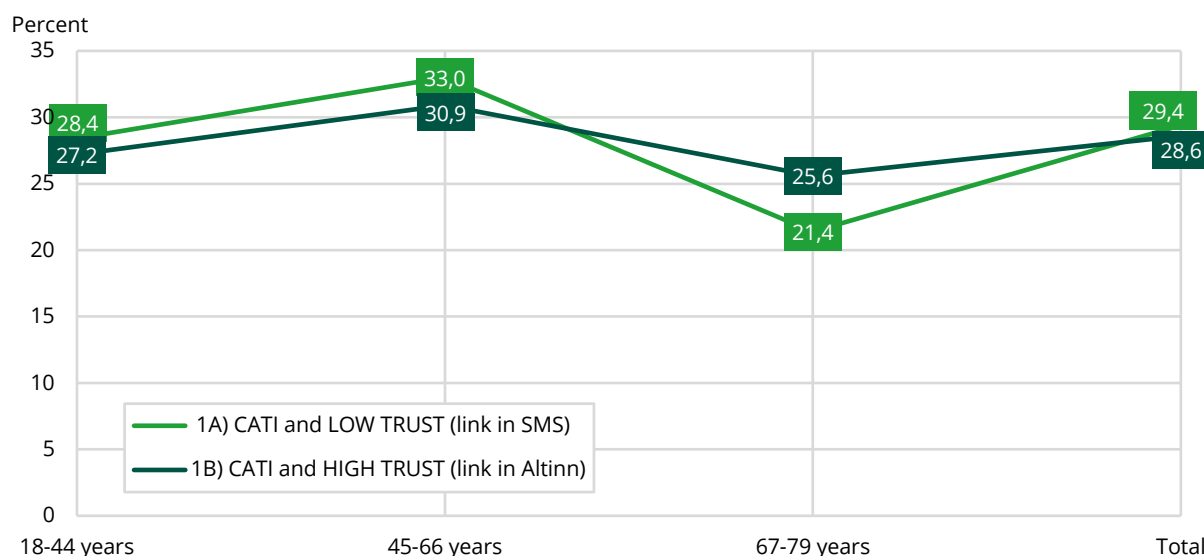
The age group 45-66 years has the highest response rate in both groups, while the age group 67 to 79 years has the lowest response rate for those who only receive a link to the survey in Altinn, and the age group 18-25 years has the lowest response rate for those who also receive a link in SMS.

Figure 5.4 Response-rate by age, A) Low trust (link in SMS) versus B) High trust (link only in Altinn)

Source: The SSI Recruitment Experiment Survey 2024. Statistics Norway

Figure 5.5 shows that the response rates in the experimental group 1A) CATI and Low trust and group 1B) CATI and High trust are 29,4 and 28,6 percent, respectively. Only the oldest age group have higher response rate when the survey link is provided solely via Altinn, rather than both Altinn and an SMS. It is important to note that the oldest age group had few observations, which may introduce some uncertainty in the numbers.

This result may suggest that ease of responding may outweigh the concerns about the perceived safety of clicking on links. It is also worth noting that all respondents in both groups had prior contact with an interviewer, who informed them that they would receive an SMS with a link to the web application. This likely contributed to increased confidence in clicking on the link.

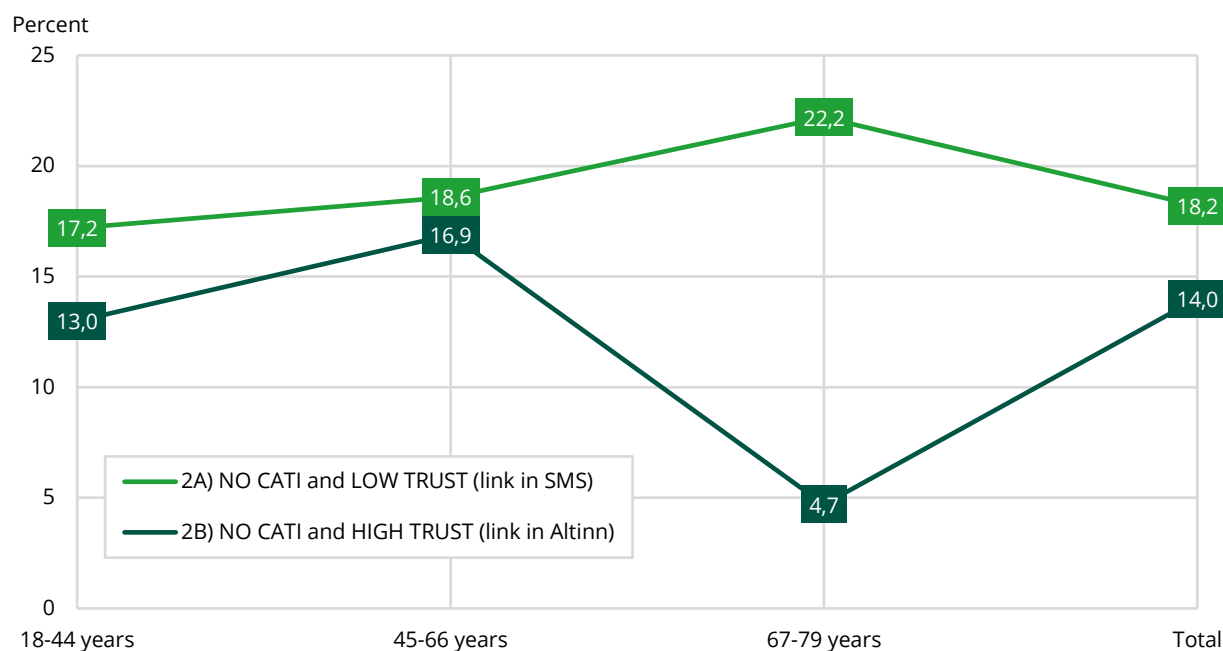
Figure 5.5 Response-rate by age*, 1A) CATI and Low trust versus 1B) CATI and High trust

*Because the group 18-24 years has very few observations, we have combined it with the group 25-44 years.

Source: The SSI Recruitment Experiment Survey 2024. Statistics Norway

When comparing experimental group 2A) No CATI and Low trust (link to the survey in SMS) and group 2B) No CATI and High trust (link to the survey in Altinn) in Figure 5.6, the response rates are 18,2 and 14 percentage, respectively. In all age groups, the link provided in SMS yielded the highest response rate. For the oldest age group, the difference between group 2A and 2B is particularly large. However, the small number of observations has likely impacted the results. This finding reinforces the notion that the ease of responding outweighs concerns about safety of clicking on links.

Figure 5.6 Response-rate by age. 2A) No CATI and Low trust versus 2B) no CATI and High trust

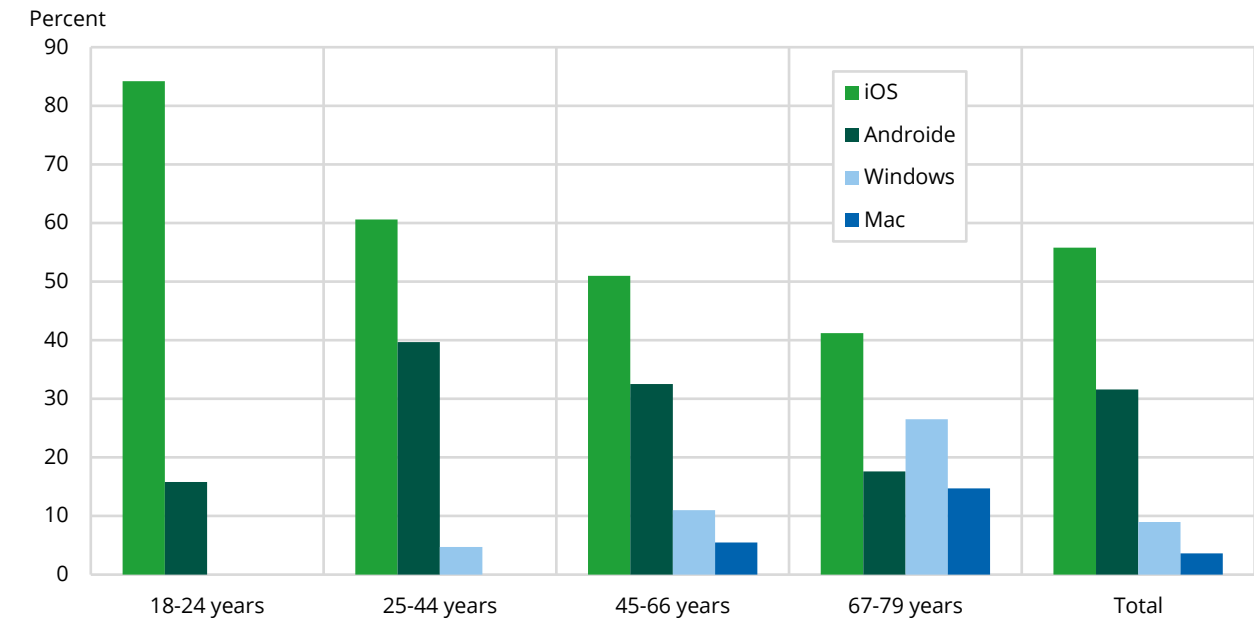


Source: The SSI Recruitment Experiment Survey 2024. Statistics Norway

5.5. Type of devices

The web application was accessible on PC, tablet, and mobile phone. As illustrated in figure 5.7, most respondents used the application on a mobile phone. This option allowed them to install an app icon, making it convenient to take pictures of receipts for OCR. However, a number of the oldest age group preferred to use computer or tablet.

This suggests that relying exclusively on an app available through App store or Google play may risk excluding some respondents in this age group, an age group which already has a low response rate. An alternative for this group could be offering paper diaries.

Figure 5.7 Type of device used for registration

Source: The SSI Recruitment Experiment Survey 2024. Statistics Norway

6. Data quality

The quality of the data collected in a sample survey is determined by various factors. This section examines key elements that may impact data quality.

6.1. Non-response

In a voluntary survey, it is expected that not all selected participants will respond. Some individuals may be unreachable, unwilling to participate, or unable to participate for various reasons. These instances contribute to the non-response rate.

In this section, we compare response rate and non-response in the two sub-samples 1) CATI and 2) No CATI, as well as the four experimental groups mentioned in table 4.1 (chapter 4). Please note that some of the figures may carry uncertainty due to limited number of observations.

Additional details regarding response rate and non-response across the sub-samples and experimental groups can be found in Appendix J.

Sub-samples 1) CATI and 2) No CATI

Table 6.1 provides an overview of the response rate and non-response for the households in sub-sample 1) with CATI interviewing and 2) with no CATI. These figures are based on the gross sample and analyse how response and non-response vary according to the contact person's gender, age, and education level. The table distinguishes between finished survey and started survey. Finished survey refers to households who have completed all parts of the survey; telephone interview (only sub-sample 1) CATI), web questionnaire, and registration of expenses for a week. In sub-sample 1) CATI, started survey refers to households who have completed the telephone interview but have not started or completed registration or web questionnaire. In sub-sample 2) No CATI, started survey refers to household who have started registration or web questionnaire, but did not finish.

Table 6.1 Response rate and non-response by the contact person's gender, age and level of education. Sub-sample 1) CATI and 2) No CATI

	Finished survey		Started survey		Non-response	
	1) CATI	2) NO CATI	1) CATI	2) NO CATI	1) CATI	2) NO CATI
All	29,0	16,1	8,8	4,9	62,2	79,0
Gender						
Men	26,6	15,2	11,8	5,3	61,6	79,5
Women	31,0	16,8	6,4	4,6	62,5	78,6
Age						
18-24 years	27,1	15,4	8,3	7,7	64,6	76,9
25-44 years	27,9	15,1	10,6	5,6	61,5	79,2
45-66 years	31,9	17,7	7,7	4,2	60,3	78,1
67-79 years	23,2	13,6	5,3	3,4	71,6	83,0
Level of education						
Below upper secondary school	17,4	8,6	12,2	4,3	70,3	87,0
Upper secondary school	28,6	12,6	7,6	4,5	63,8	82,9
Higher education	36,9	23,0	7,8	6,2	55,4	70,8
Unspecified	12,7	16,7	14,5	0,0	72,7	83,3

Source: The SSI Recruitment Experiment Survey 2024. Statistics Norway

Table 6.1 indicates that the overall response rate is about twice as high among households recruited through CATI compared to those who were not. This trend is consistent across all background

variables in the table, suggesting that the involvement of an interviewer plays a role in achieving higher response rates in this type of surveys.

The proportion of households that started but did not finish the survey is lowest in sub-sample 2) without CATI. If we look at those who actually started registration of expenses or the web questionnaire in sub-sample 1) with CATI, only 1,8 percent started but did not finish (table J.1 in Appendix J).²

In both sub-samples, households where the contact person is in the oldest age group or have a low level of education show the lowest response rate. These findings align with the results from the Norwegian HBS in 2022 and other web surveys conducted by Statistics Norway (ref Berg et al. 2024).

From analysis of status codes in the CATI groups and discussing with interviewers, the low response rate among households with an old contact person, is due to several factors, including lack of willingness to participate, health issues, and challenges associated with the survey's exclusively digital format. Additionally, some individuals in this group do not use the Altinn portal.

In households where the contact person has a low or unspecified education level, the primary reason for non-response was difficulty in reaching them. In sub-sample 1) CATI, these households also had a notable portion that started but did not finish the survey, suggesting that the digital format may have posed challenges. Among those with unspecified education level, there is a clear overrepresentation by immigrants, many of whom lack proficiency in Norwegian.³

Table 6.2 shows the response rate and non-response for the households who received a link to the survey in both SMS and Altinn (Group A Low trust) and households who received a link to the survey only in Altinn (group B High trust), without taking into account whether the households were recruited by CATI or not. These figures are based on the gross sample and analyse how response and non-response vary according to the contact person's gender, age, and education level.

The primary goal of the field experiment was to split the sample evenly in two sub-samples, where one half was recruited to the survey by an interviewer (CATI), while the other half was not (No CATI). Each sub-samples were then further divided in two groups based on login method for accessing the survey: Group A (low trust) got a link to the survey in both SMS and Altinn, while group B (high trust) got a link to the survey only in Altinn. This set up resulted in four experimental groups, as illustrated in Figure 4.1.

We have compared sub-sample 1) CATI with sub-sample 2) No CATI, but comparing Group A (low trust) with Group B (high trust introduces) introduces certain complications because households within the same group were subject to different recruitment strategies. Notably, half of the households in each group were recruited by an interviewer (CATI), while the other half was not (No CATI). This overlap of recruitment modes across login methods makes it difficult to isolate the effect of the login solution. Consequently, detailed analyses of response rate and non-response by login method, groups A and B, will not be presented in this report.

² In table 6,1 8,8 percent started which is 7.0 percent received login information plus 1,8 started registration,

³ "Unspecified" refers to a group primarily consisting of contact persons with *no recorded education* in Norway. Most of them are likely immigrants. A group often underrepresented in social surveys

Table 6.2 Response rate and non-response by the contact person's gender, age and level of education. Group A) Low trust and B) High trust

	Finished survey		Started survey		Non-response	
	A) LOW TRUST	B) HIGH TRUST	A) LOW TRUST	B) HIGH TRUST	A) LOW TRUST	B) HIGH TRUST
All	24,0	21,3	3,5	3,2	72,5	75,5
Gender						
Men	21,8	20,2	4,2	4,0	74,0	75,8
Women	25,7	22,2	2,9	2,6	71,4	75,2
Age						
18-24 years	21,4	22,2	4,8	2,2	73,8	75,2
25-44 years	23,0	20,1	4,4	3,2	56,9	76,7
45-66 years	25,9	23,7	2,6	3,5	71,5	72,8
67-79 years	21,8	14,6	2,0	2,4	76,2	83,0
Level of education						
Below upper secondary school	15,2	11,4	3,8	2,8	81,0	85,8
Upper secondary school	21,2	19,7	3,3	3,1	75,1	77,2
Higher education	30,7	29,4	3,7	4,0	65,6	66,6
Unspecified	21,3	8,4	2,0	0,0	76,7	91,6

Source: The SSI Recruitment Experiment Survey 2024. Statistics Norway

The four experimental groups

Table 6.2 provides an overview of the response rate and non-response for the four experimental groups described in figure 4.1 (chapter 4).

- 1A) CATI and Low trust (link to the survey in SMS)
- 1B) CATI and High trust (link to the survey only in Altinn)
- 2A) No CATI and Low trust (link to the survey in SMS)
- 2B) No CATI and High trust (link to the survey only in Altinn)

The figures are based on the gross sample, with an analysis of how non-response varies by the contact persons gender, age, and level of education. More detailed information about non-response in the experimental groups is available in Appendix J.

Table 6.3 demonstrates that the overall response rate is highest in the groups with CATI. Please note that the results may carry some uncertainty due to limited numbers of observations.

Table 6.3 Response rate and non-response by the contact person's gender, age and level of education. Experimental group 1A), 1B), 2A) and 2B)

	Finished survey				Started survey				Non-response			
	1A)	1B)	2A)	2B)	1A)	1B)	2A)	2B)	1A)	1B)	2A)	2B)
All	29,4	28,6	18,2	14,0	7,6	10,1	5,9	4,0	63,0	61,3	75,9	82,0
Gender												
Men	26,9	26,2	16,4	14,0	10,0	13,5	7,0	3,6	63,1	60,3	76,5	82,4
Women	31,3	30,7	19,7	14,0	5,7	7,1	4,9	4,3	63,0	62,2	75,4	81,7
Age												
18-44 years	28,4	27,2	17,2	13,0	8,0	12,8	7,4	4,2	63,6	60,0	75,4	82,8
45-66 years	33,0	30,9	18,6	16,9	7,7	7,7	4,3	4,1	59,3	61,4	77,1	79,0
67-79 years	21,4	25,6	22,2	4,7	5,4	5,1	4,4	2,3	73,2	69,3	73,3	93,0
Level of education												
Below upper secondary school	18,2	14,8	15,1	5,8	9,1	15,6	5,7	1,0	72,7	69,6	79,2	93,3
Upper secondary school	29,5	27,7	13,1	12,1	6,7	8,4	5,1	4,0	63,8	63,9	81,8	83,9
Higher education	34,6	39,5	26,0	20,3	7,7	7,9	6,9	5,6	57,7	52,2	67,1	74,1

Source: The SSI Recruitment Experiment Survey 2024. Statistics Norway

Overall, the response rate is highest in group 1A) with CATI recruitment and link to the survey in SMS and lowest in group 2B) with no CATI and link to the survey solely through Altinn. Among individuals with low education the response rate is slightly higher in group 2A) with no CATI and SMS compared to group 1B) with CATI and Altinn. Additionally, for the age group 67-79, group 2A) with no CATI and SMS seem to outperform group 1A) with CATI and SMS. However, for all other demographic characteristics, the response rates are highest in the two CATI groups.

Table 6.3. shows that using a more secure login solution (Altinn) generally seems to result in a lower response rate than providing a link to the survey directly via SMS in all the experimental groups. Nonetheless, for the oldest age group and those with higher education, experimental group 1B) CATI with link in Altinn achieved the highest response rate.

The oldest age group and those with low education have the highest non-response rate across most of the experimental groups. Non-response is particularly pronounced in group 2B) where no CATI was conducted, and the survey link was provided only through Altinn. In 2B only 5.8 percent of households where the contact person had low education and 4.7 percent of households where the contact person was in the oldest age group finished the survey. It is important to note that these results may be uncertain due to small number of observations.

In group 1B) with CATI and the link provided via Altinn, 10 percent of households started but did not finish the survey, suggesting that the digital format may have posed challenges.

6.2. Bias

The proportion of individuals remaining after excluding those who are not eligible is referred to as the gross sample. The gross sample represents the population targeted for this survey. Conversely, the proportion of individuals who finished or completed the survey is termed the net sample. The difference between the gross and net sample is classified as non-response. Non-response can introduce bias when the distribution of a specific characteristics differs between respondents and non-respondents. Bias indicates that the net sample may not accurately represent the overall population.

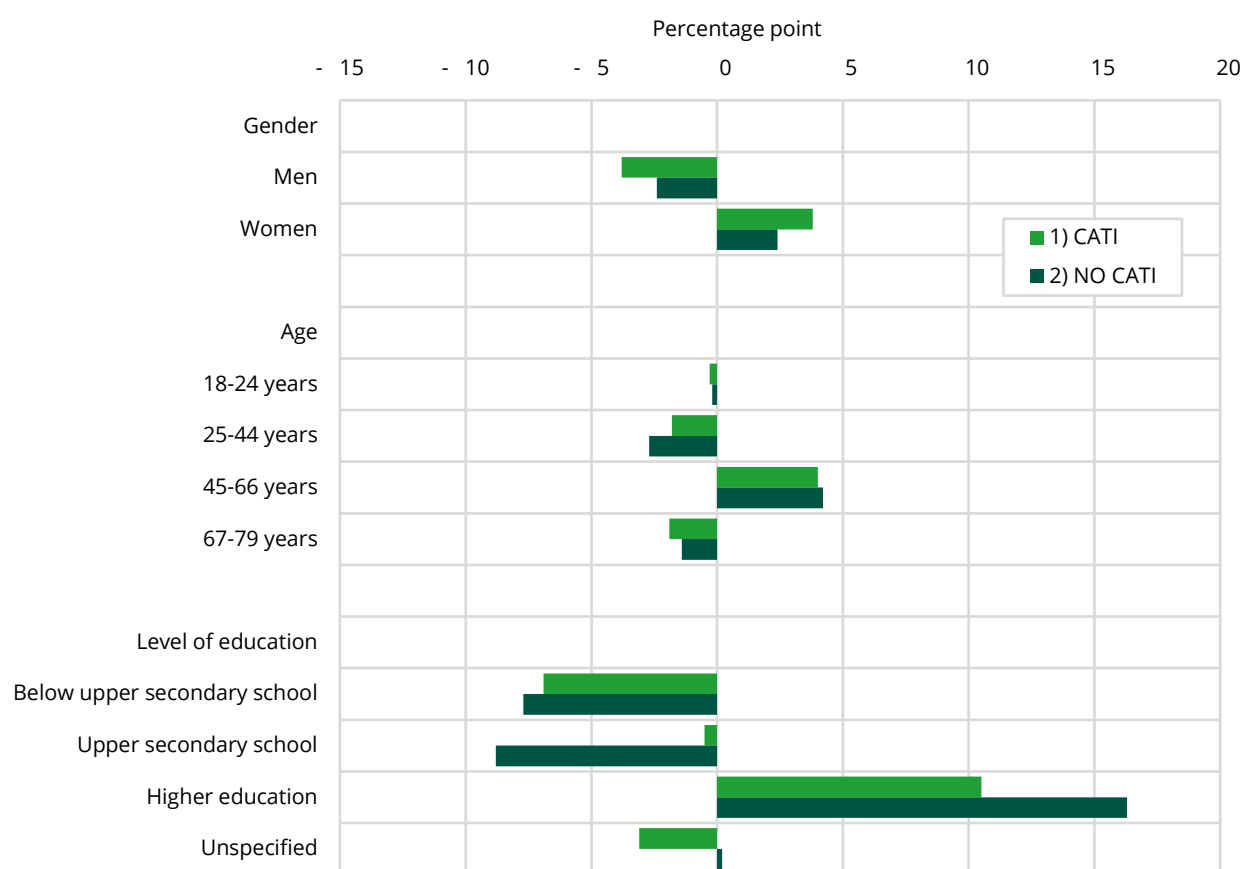
However, bias in one characteristic does not necessarily imply bias in other characteristics. Conversely, a close correspondence between the distributions of the net and the gross sample for

certain characteristics does not guarantee absence of bias for other. In this survey, the non-response rate is nearly 80 percent, which increases the likelihood of bias compared to scenarios with lower non-response rate.

Figure 6.1 illustrates the difference between the net and gross sample in the two sub-samples 1) with CATI and 2) with No CATI. These differences are analysed based on the contact persons gender, age and level of education. Additional details about the distribution in both the gross and net sample for both sub-samples can be found in Appendix K.

The figure reveals some deviations between the distribution in the net and gross samples across all the characteristics, with the most noticeable discrepancies observed in age and education.

Figure 6.1 Deviation between net and gross sample by contact person's gender, age and level of education. Sub-sample 1) CATI and sub-sample 2) No CATI



Source: The SSI Recruitment Experiment Survey 2024. Statistics Norway

When examining the contact person's age, the age groups 25-44 and 67-79 years appear to be the most underrepresented in both sub-samples, while the 45-66 age group is the most overrepresented in both sub-samples (1) CATI and 2) No CATI). Notably, the 67-79 age group is particularly underrepresented in the CATI sub-sample.

The underrepresentation of the oldest age group in the gross sample is especially concerning, given the already small proportion of the oldest age group. Table 6.4 shows the difference between the gross and net sample of each age group, in both percentage point and percent. Households where the contact person are aged 67-79 years are underrepresented by about 20 percent in 1) CATI and 15,5 percent in 2) No CATI. In sub-sample 1) this bias was 12,8 percent following the telephone interview, indicating a considerable dropout rate in this age group after recruitment.

Table 6.4 Differences between net and gross sample, across different age groups. Sub-sample 1) CATI and 2) No CATI

	1) CATI Difference in net and gross sample, percentage point	2) No CATI. Difference in net and gross sample, percentage point	1) CATI Difference in net and gross sample, percent	2) No CATI. Difference in net and gross sample, percent
18-24 years	-0.3	-0.2	- 6,3 %	-5,0 %
25-44 years	-1.8	-2.7	- 4,0 %	-6,0 %
45-66 years	4.0	4.2	9,9 %	10,1 %
67-79 years	-1.9	-1.4	- 20 %	-15,5 %

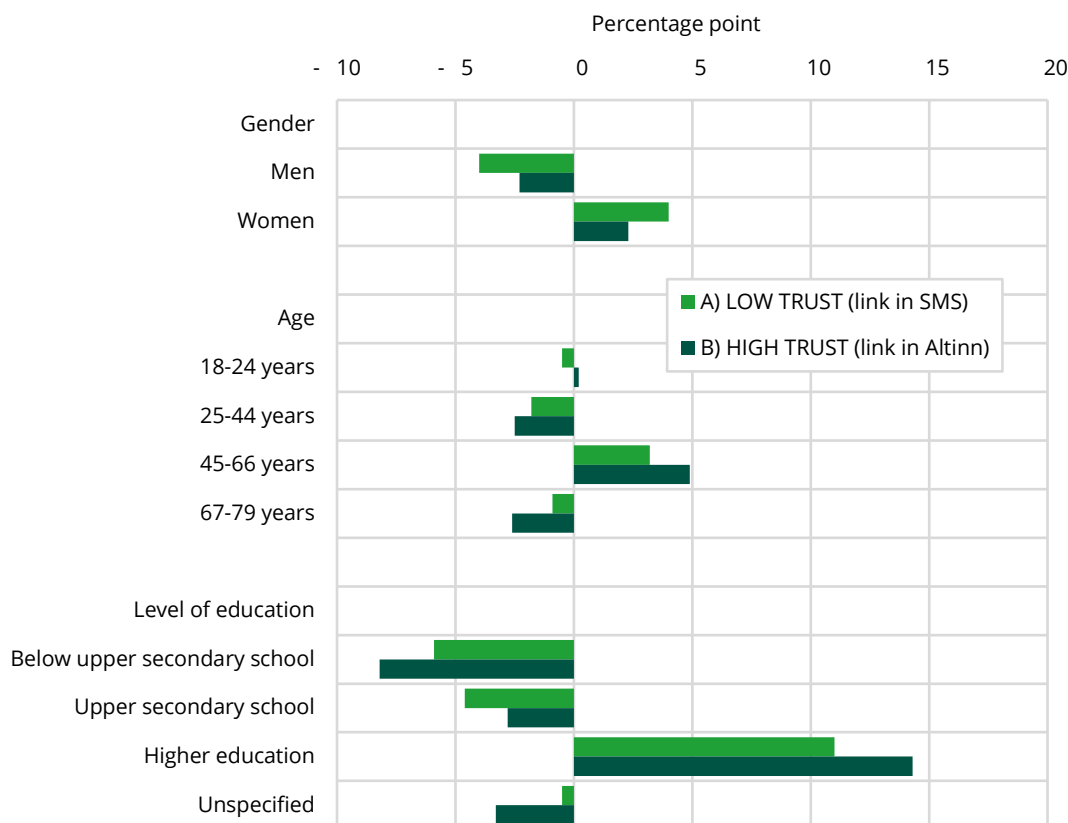
Source: The SSI Recruitment Experiment Survey 2024. Statistics Norway

Figure 6.1 illustrates that households where the contact person only has basis schooling or lower level of education are underrepresented in the net sample in both sub-samples, while those with higher education are clearly overrepresented. Contact persons with upper secondary education are not underrepresented in sub-sample 1) CATI, but they are underrepresented in sub-sample 2) No CATI.

This educational bias increases after the telephone interview, as a notable portion of participants who did not finish the survey had lower education levels. This pattern aligns findings from the Quality of life survey (Pettersen and Engvik, 2022) and other web surveys. However, the overrepresentation of respondents with higher education is even more pronounced this field test compared to the HBS 2022.

Figure 6.2 illustrates the difference between the net and gross sample in the two groups A) Low trust (link in Altinn and SMS) and B) High trust (link only in Altinn). These differences are analysed based on the contact persons gender, age and level of education. Additional details about the distribution in both the gross and net sample will not be given because comparing these two groups poses certain challenges, as each group was split evenly: half were recruited to the survey by an interviewer (CATI), while the other half was not (No CATI).

Figure 6.2 Deviation between net and gross sample by contact person's gender, age and level of education. Groups A) Low trust and B) High trust



Source: The SSI Recruitment Experiment Survey 2024. Statistics Norway

Figure 6.2 reveals some deviations between the distribution in the net and gross samples across all the characteristics, with the most noticeable discrepancies observed in age and education. When examining the contact person's age, the age groups 25-44 and 67-79 years appear to be the most underrepresented in both groups, but the underrepresentation is most prominent in group B) High trust. The 45-66 age group is the most overrepresented in both groups A and B. Contact persons with only basis schooling or lower level of education are underrepresented in the net sample in both groups, but the deviations are most notable in group B where the contact persons got a link to the survey only in Altinn.

Figure 6.3 illustrates the difference between the net and gross sample across the four experimental groups. We examine deviations concerning the contact persons gender, age and level of education. More details about the distribution in both the gross and net sample for the four groups, can be found in Appendix K.

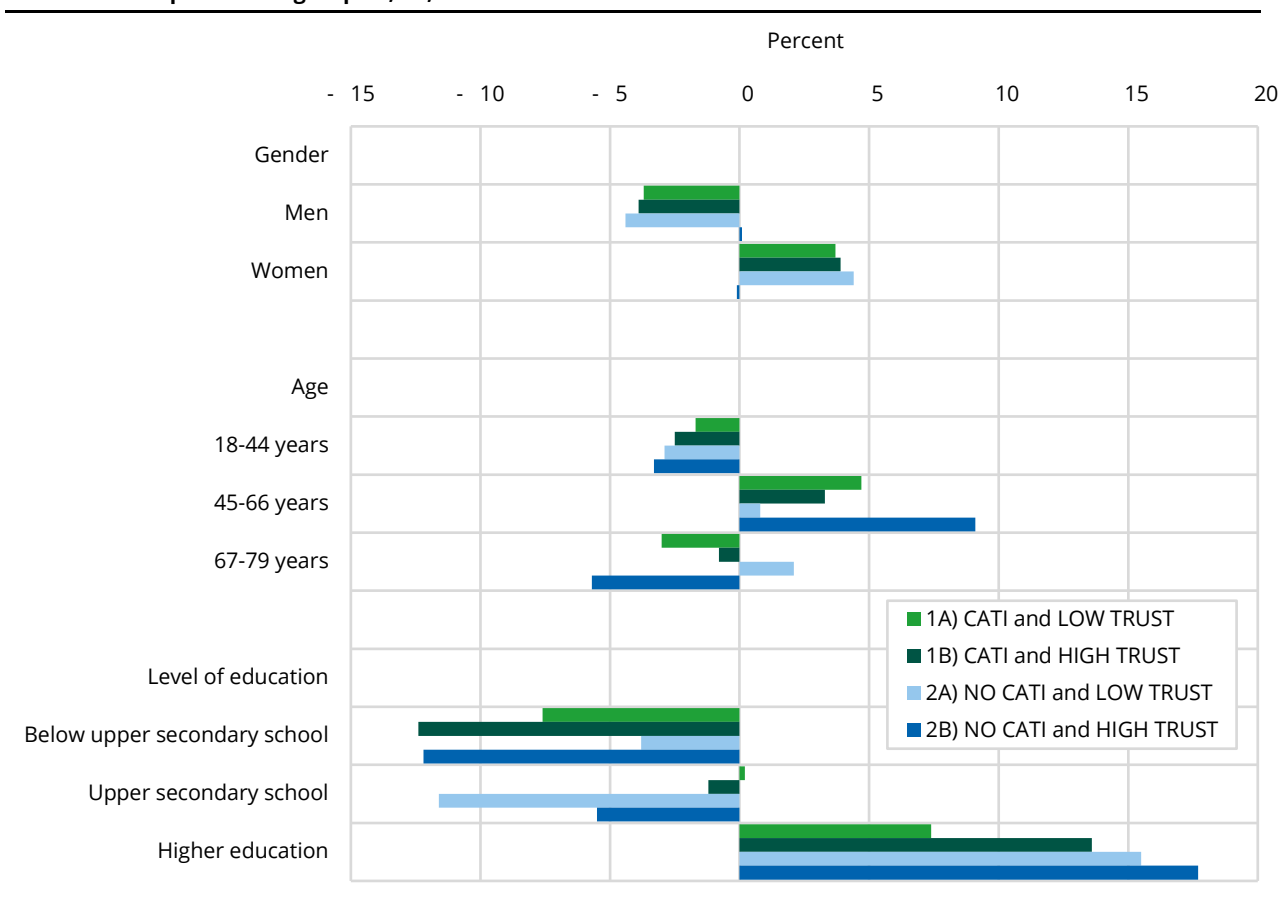
Except for gender, the bias appears highest in group 2B) with no CATI and link to the survey only in Atinn. In this group, very few among individuals aged 67-79 or with only basic schools finished the survey, resulting in a particularly low representation of these groups in the net sample.

Men are underrepresented in all experimental groups except group 2B. Across all groups, contact persons aged 18-44 are underrepresented, those aged 45-66 are consistently overrepresented. The

oldest age group (67-79 years) is underrepresented in all groups except in group 2B) with no CATI and link to the survey in SMS.

Regarding education, contact persons with only basic school is underrepresented in all experimental groups, while those with higher education are overrepresented. In group 2B, contact persons with higher education are overrepresented by as much as 17,7 percent.

Figure 6.3 Deviation between net and gross sample by contact person's gender, age and level of education.
Experimental group 1A, 1B, 2A and 2B



Source: The SSI Recruitment Experiment Survey 2024. Statistics Norway

6.3. Summary

The findings presented in Chapter 6 indicate the following insights in respect to:

Use of interviewers

- The use of CATI for recruitment nearly doubled the response rate in this survey. The response rate was consistently higher with CATI across all examined characteristics.
- The oldest age group and individuals with low education level showed the highest non-response across most of experimental groups. Non-response was especially pronounced in group 2B which had no CATI recruitment and provided the survey link only via Altinn.
- Use of CATI appears to reduce bias, especially concerning education level. However, its effect on other factors such as age, is less consistent.

Login solution

- Using a more secure login solution (Altinn) generally resulted in a lower response rate compared to a direct survey link sent via SMS.
- Despite this trend, experimental group 1B) CATI and link to the survey in Altinn achieved the highest response rate among the oldest age group and those with higher education.

7. Data delivery

Statistics Norway have delivered two datafiles to researchers at University of Utrecht and the University of Mannheim, who are partners in the SSI project. All delivered data are anonymised.

The data primarily consists of paradata. From the web application the only included information pertains to the type of device used and number of receipts scanned by OCR or manual registration.

We have not been able to deliver date and time for when letters and SMS messages were sent out connected to each household, however, an overview of all mailings and their respective send dates can be obtained on request.

An overview of the variables on the two files is included in Appendix H.

7.1. Data agreement

Statistics Norway and the consortium have an overarching agreement regarding the project and SSB contributions. A separate data handling agreement is not required for the delivery, as all data provided is anonymised.

Note that Statistics Norway expects data to be analysed by the University of Utrecht and the University of Mannheim and used exclusively for the SSI project. Any use, publication, or dissemination of statistics derived from the data must include appropriate credit to Statistics Norway as the data provider.

8. User experience

In relation to the WP2.1 SSI Recruitment Experiment Survey, our goal was to gain qualitative insights into the user experience, focusing on the use of telephone interviewers and use of different login solutions in the recruitment process. We explored the respondents' perspectives through two separate initiatives:

- **Post-Q evaluation survey** (n=148). While primarily quantitative, this survey complements our qualitative analysis due to the nature of its questions and its purpose. Therefore, it is included in our qualitative examination.
- **Follow-up interviews** (n=8). These interviews provided deeper insights into the user experience.

Additionally, we conducted cognitive interviews for the WP2.3 Small Scales User Tests on OCR (n=11) as part of the broader SSI project. These interviews contributed further insights on trust and data security⁴. For more details on data, please refer to Table I in the appendix.

We present the qualitative findings on recruitment collectively, as the results from the various studies align closely. Our focus is on the overall user experience, with comments on specific group deviations when the material allows. To provide relevant context and a clear understanding of the user perspective on recruitment, we have organized the material into three sections: 1) Involvement and engagement, 2) Usability and user experience, and 3) Trust and privacy safeguards. Finally, we summarise key learnings from the recruitment experiment related to the use of telephone interviewers and the type of login solutions. However, limited material constrains our discussion of the four groups tested in the quantitative recruitment experiment (ref. Figure 4.1).

8.1. Involvement and engagement

Motivation to participate in the survey was driven by similar factors across demographic groups. Most important motivator for all respondents were a sense of social responsibility and personal interest in the survey's topic, and more so with age. Among young adults, the gift certificate served as the key incentive, whereas for adults and seniors, incentives were an appreciated bonus.

Recognition of Statistics Norway as the sender played a crucial role in encouraging participation. The bureau's high level of trust motivated respondents to click on the survey link, whether it was included directly in an SMS or accessed via the national Altinn portal. Respondents generally did not distinguish much between the two methods of opening the survey link. However, consistent with the experiment survey results, seniors were more ready to engage with a direct survey link (SMS) than through Altinn, presumably due to its perceived simplicity or user friendliness.

Seniors aged 70 and above, however, expressed concerns about scams and phishing likely due to national campaigns warning against fraudulent SMS links and misuse of national ID numbers. Young adults typically verified the survey legitimacy independently. And once reassured that the process was straightforward and not time-consuming, they proceeded without further consultation. In contrast, seniors often lacked the skills to distinguish legitimate links from fraudulent ones, making them hesitant to click. For them telephone interviewers made a difference and provided reassurance. Notably, seniors appreciated the opportunity to consult with spouse or family members before receiving the recruitment call and deciding to participate. The oldest respondents,

⁴ See report: [Report WP2,3 Norway 22,08,24](#) (Readers outside the project group can get access by contacting the authors.).

even though they often felt obliged to Statistics Norway, more often declined participation due to lack of device, difficulties with technology, or the overall response burden of the survey.

For most participants, the use of what they perceived as a regular “app” and scanning technology was motivating rather than discouraging. Though for some the variety of technological options added complexity to an already demanding survey. Had we been able to include more dropouts and rejectors in our interviews, we might have observed greater reluctance to participate due to issues like lack of device, inexperience, or hesitation to engage with technology. It is reasonable to assume that these barriers are more pronounced among the elderly, individuals outside the workforce, and immigrants.

Once respondents started using the web app, dropout rates were low. Dropout typically occurred between the recruitment call and the registration week and were rarely related to usability, except for the oldest participants. The primary reasons for dropout were response burden and lack of time, particularly due to the diary and the lengthy questionnaire on fixed expenses. Many respondents struggled to answer these questions because they were unfamiliar with the required information and often needed to consult family or other sources.

Respondents who received support from telephone interviewers found it helpful. Some indicated that they might not have completed the survey without their reassurance and assistance, as they had lost engagement or encountered obstacles they couldn’t overcome alone. While participants without interviewer support generally didn’t express a need for interviewers. Recruitment data, however, shows that interviewer support increased survey participation, likely due to their ability to validate legitimacy, resolve issues, and offer friendly nudging. Both respondent groups with and without telephone interviewers experienced the SMS-reminders during registration week appropriate, helpful, and neither excessive nor invasive.

8.2. Usability and user experience

As expected, young adults were in general more comfortable with digital tools and demonstrated greater ease in using smartphones and apps. Adults also managed well, and most seniors navigated the technology adequately, though they were less proficient in fully utilizing its potential. Seniors often required initial assistance from a spouse or interviewer. This support was typically required only at the start, highlighting the critical role of interviewers during the onboarding process.

Participants frequently overlooked much of the onboarding information and instructions provided via email and in-app. They paid minimal attention to the login process, not reading, but making sure they passed the steps, and went straight for receipt registration – which they perceived as their primary task. The majority accessed the survey using their mobile device. However, many expected the ‘app’ to be available through an app store rather than as a web-based application. This misunderstanding led to initial confusion and may have caused some participants to abandon the survey. A key challenge after opening the survey link was a lengthy instruction on options how to save the link to the mobile home screen. The unfamiliarity with this process for many respondents added to their uncertainty. Despite these challenges, login was not reported as a significant obstacle in retrospective interviews. Although, as noted, some seniors required initial support to get started.

Regarding usability, distributing the survey link via SMS outperformed Altinn. Although both methods involved authentication through ID-Porten or BankID, Altinn was perceived as more cumbersome. Neither login methods seem to be important in the respondents set of minds when they decide to participate or drop out.

Again, it is important to note that our interviewees were respondents who had either started or completed the survey, excluding rejectors and non-respondents. It's likely that usability and technology-related challenges might be more pronounced among these groups.

Across all age groups, participants preferred scanning receipts over manual entry, finding scanning more efficient while manual registration felt burdensome. Still, manual registration occurred more frequently than desirable from a data qualitative perspective. Reasons for not scanning included lack of receipts, missing receipts from other household members, or the perception that manual entry was easier for certain purchases. The lack of receipts was more pronounced in young adults and adults, as they are not accustomed to collecting receipts. The youngest tended to resort to manual data entry, satisficing, and short cuts more frequently. Seniors, by contrast, demonstrated more patience with the process, and once familiar with the web application, managed scanning without encountering more problems than others. Seniors were also most diligent in verifying fixed costs using secondary sources, such as bank statements and paper records. But they were less likely to explore the interactivity of the web application and found it more difficult to switch between sites on their device.

Since the survey link or information about it was initially received on a smartphone and Statistics Norway promoted mobile device as the most efficient option, most participants accessed and completed the survey on their smartphones. Nonetheless, seniors encountered more often usability issues due to the small screen size. Particularly with text size, navigation, and touch functionality. They generally prefer larger screens, especially for reviewing data, and more often transitioned from using smartphones to larger devices like iPads or PCs.

8.3. Trust and privacy safeguards

Trust was essential for all participants in their decision to take part in the survey. High confidence in Statistics Norway reassured participants and positively influenced their willingness to contribute to the survey. Communication through multiple channels (SMS, email, Altinn and Statistics Norway's homepage), combined with a professional-looking application and secure authentication via ID-Porten/BankID, helped build trust. Use of telephone interviewers further reassured uncertain respondents, playing a key role in strengthening trust.

A critical indicator of trust was participants' willingness to click on the survey link (or more precisely the link to the web application of the survey). Most participants did so without hesitation, whether the link was provided directly in an SMS or via an SMS with information to go through the Altinn portal to find it. Only a few chose a more cautious approach, such as typing the web address manually and verifying its legitimacy⁵. In retrospective follow-up interviews, most respondents acknowledged initial concerns about scams. Young adults valued having multiple verification options to ensure the survey's legitimacy, relying on their ability to assess the trustworthiness on their own. While seniors more often were in doubt and likely to consult relatives or be reassured by telephone interviewers.

While most participants considered ID-Porten/BankID and Altinn cumbersome, they generally accepted these as necessary security measures. All respondents in the qualitative interviews were familiar with these digital solutions and capable of using them. Nevertheless, it's important to remember that not all citizens, particularly the oldest, are equally familiar. Some participants voiced concerns about using their national ID number, required regardless of the login solution tested, due

⁵ 93 % answered that they opened link in Altinn or sms, 7 % searched for the survey online, ref. Post-Q Evaluation Survey, SSB.

to fears of fraud. A few respondents expressed a preference for reserving the use of their national ID number for serious matters, rather than surveys.

Privacy and data security

Respondents generally did not consider sharing expenses or scanning of their own receipts as sensitive and were comfortable providing this information. Receipts rarely contain personal details, and respondents in the small-scale OCR user tests did not review receipts for such information before scanning. Retrospective interviews indicated that participants did not feel the need to edit or alter by cropping or masking information for privacy reasons. However, some participants mentioned that they might withhold receipts for purchases they perceived as subject to judgment, such as alcohol, candy, energy drinks, or more private items (e.g., from sex shops). In these cases, they would prefer to omit the receipts rather than edit them.

Our test did not suggest that the use of new technology, with what many perceived as a standard app with scanning capabilities, negatively affected respondents' trust in Statistics Norway. On the contrary, respondents expected Statistics Norway to utilize modern technology to improve efficiency.

Seniors demonstrated the highest levels of trust in Statistics Norway, confident that their data would remain secure and not misused. Young adults also trusted Statistics Norway but were more likely to question the appropriateness of the amount and type of information being collected. Across all age groups, respondents had strong trust in Statistics Norway's compliance with privacy laws and ability to handle and store information. However, this trust is somewhat fragile, as privacy is a complex issue that many found difficult to fully comprehend or discuss during the interviews. Respondents made it clear that a breach of this trust would be difficult to repair.

It is important to note that trust in Statistics Norway is likely lower among rejectors than among the participants included in this study.

8.4. Qualitative insights on recruitment

We can summarize our key qualitative insights into the recruitment experience as follows:

Use of new technology

The use of smart technology, sensors, and receipt scanning in a web application for data collection was well-received by most respondents serving as an incentive for participation. The web application itself was not considered a barrier to participation, except among the oldest respondents. The contact strategy, which included multiple contacts through different channels, effectively informed respondents that Statistics Norway was reaching out and ensured the survey's legitimacy. For younger respondents, it reassured them that the survey link was safe to open, while for seniors, it allowed time to consult with family or friends to verify the authenticity of the survey.

Still, we need to evaluate whether the oldest age group requires alternative participation modes to ensure they are not excluded.

Login

The process of adding the web application to a mobile home screen proved challenging, as many users are unfamiliar with web app technology, causing confusion. It remains unclear whether using an app store would improve usability, as some respondents are also hesitant to download apps. This usability issue occurs at the start of the survey process, making it a critical point that could affect response rates, and therefore needs further considerations.

There is no clear user preference between a login solution that provides the survey link directly in the SMS or accessing it through Altinn. While a direct link in an SMS offers the greatest usability from a user's perspective, it raises security concerns in the current environment, even when secured by ID-porten/BankID. Respondents expect Statistics Norway to prioritize security and adhere to best practices. And they anticipate needing to use ID-porten/BankID and are generally willing to use Altinn, even if they find it cumbersome.

Interviewer support

Despite the shift towards a self-administered data collection mode, interviewers remain indispensable from a user perspective. They play a critical role in verifying the survey's authenticity, motivating respondents, assisting with start help, reminding non-starters and potential dropouts, and ensuring the completion of registrations. Interviewers are also central to onboarding communication, clarifying the survey's purpose, explaining respondents' tasks, interpreting questions, and emphasizing the importance of accurate and detailed reporting. Additionally, they offer user guidance, provide technical support to those less familiar with web applications, and handle administrative inquiries.

We conclude that interviewer involvement is essential, particularly for ensuring the representation of groups such as seniors, individuals outside the workforce, and immigrants. Without their involvement, these groups would likely be even more underrepresented than they have been so far.

9. Interviewer feedback

After concluding the fieldwork of the recruitment experiment, we held a workshop with three key telephone interviewers involved in data collection for the recruitment experiment survey, along with the lead project planner. The main objective was to identify what worked well, what didn't, and what could be improved regarding the contact strategy, survey communication, recruitment activities, web application, and back-office system for data collection for future surveys. Consequently, the following summary reflects the general feedback from the interviewers on the overall process as they see it and has a broader focus than the SSI research scope.

9.1. Engagement and trust

Motivation

According to the interviewers, Statistics Norway's reputation is the most effective recruitment argument. Many respondents feel special being selected and are happy "to serve" when Statistics Norway calls. Additionally, gaining insight into their own economy and contributing to understanding consumption patterns in society at large encourages participation. Therefore, the interviewers recommend displaying a graphical overview of each household's expenses continuously within the app during the registration week. For younger respondents, a gift certificate of 500 NOK is also considered substantial and important to offer.

Trust-building and response burden

Statistics Norway enjoys a high level of trust in society. Respondents are generally willing to engage with the survey and click on SMS links, especially when they recognise Statistics Norway as the sender. However, seniors aged 70 and above exhibit more scepticism, particularly regarding use of their national identification number to log-in. This reluctance can hinder recruitment efforts and is not all alleviated using the national Altinn-portal. To mitigate use of national identification number, interviewers recommend referring only to "ID-Porten," and not "bankID", and emphasizing that it is the same secure login used for taxes and government services such as Altinn.

Seniors also express concern about phishing likely due to national campaigns highlighting fraudulent SMS links. They may lack the skills to detect legitimate links, making them hesitant to click on them compared to younger respondents. While some seniors benefit from reassurance and assistance from interviewers in this matter, many find the overall response burden of the survey task too high to participate. Once they perceive the task as too demanding, they are often difficult to sway.

Drop-out reasons

Close follow-up by interviewers, combined with SMS reminders during data collection, is essential to ensure respondents complete the survey. This is crucial for both data quality and response rates. The primary reasons for dropouts include response burden and lack of time, particularly due to the lengthy questionnaire on fixed expenses. Many respondents find it challenging to answer questions because they are unfamiliar with the required information and don't know the exact answers. However, features like the use of app technology with scanning for the registration week of household spending have a new appeal so far and are not drop-out reasons for groups below retirement age.

9.2. Response burden and usability

Many respondents are unfamiliar with the distinctions between regular apps, that is native apps, and web applications (PWAs) or web apps, which can lead to confusion about usage. Interviewers believe that making the app available in the App Store or Google Play would enhance trust and encourage more respondents to engage, as these platforms are associated with legitimacy.

Once respondents have opened and started the web application, it is generally easy to navigate, though seniors face more difficulties than younger respondents. A significant challenge for many is forgetting to collect physical receipts. While most older respondents can navigate the web application with some initial assistance, the oldest individuals often fear struggling with technology and are hesitant to explore it.

9.3. Survey communication

Survey communication

The contact strategy for the survey was efficient, utilizing multiple contact points and formats such as email, the national portal Altinn, Statistics Norway homepage, and SMS notifications. As we know from the user tests and earlier work, interviewer feedback also indicates that respondents do not read all in the provided materials and instructions. However, the interviewers themselves found the instruction texts to be clear and precise, though they felt there is a tendency to overwhelm respondents with too much text.

The interviewers received positive feedback regarding the assistance they provided to the respondents. They felt that their explanation of the main tasks of the survey and important features of the web application was appreciated and crucial for getting respondents to report according to Statistics Norway requirements. The interviewer support helped ease respondents through the registration week and encouraged them to complete the survey.

Organisation of data collection

Engagement is high in this interview team. They enjoyed working together, dividing tasks, and collaborating effectively. They have gained a substantial knowledge across what works best for the HBS. It has been important that they know the web application and have saved it on their own mobiles. Within the team, they were able to handle and answer most inquiries, including technical support. They do not want to expand the team or tap into the field corps at large, as the team works well as a small unit. They believe the contact and recruitment strategy is good and want to maintain this plan with strong attention to details and a prompt schedule. There should never be uncertainty about what groups to follow up on and when.

The back-office system worked well and provided good support for efficient follow up without redundant work. However, they were not as happy with the support from the Service Desk. It was too often too late, and the information provided about the survey could be incorrect. For the next run of the Household Budget Survey, the interviewers have detailed suggestions for improvements of the back-office system (see recommendations Appendix I).

10. Recommendations for future surveys

This section outlines recommendations based on insights from the Recruitment Field Test and related activities of the experiment, supported by insights from the Household Budget Survey 2022. Several key lessons have emerged from this work, providing guidance for future enhancements to data collection in the Household Budget Survey.

10.1. Key learnings

Despite implementing new technologies, like app-based data collection with scanning, intended to improve efficiency for both users and data collectors, the survey remains highly complex and continues to impose a significant response burden. While the web application introduced in 2022 seemingly has made the process more user-friendly for most respondents, the questionnaire on fixed and large expenses remains lengthy and difficult, presenting a persistent challenge for future surveys.

Our main concerns about response rates, potential bias, and data quality are driven primarily by this response burden rather than by the new technology itself. (Though it should not be forgotten that technology is a significant barrier for participation for the very oldest individuals (70/80 years+), which need to be addressed for future surveys.) Further, we are concerned about the quality of self-reported data in a time when attention spans and willingness to provide detailed responses are diminishing. This trend, together with reduced respondent engagement in social surveys observed over the past decade or so, raises concerns about underreporting, data accuracy, and a higher in-house editing workload.

Respondents often do not fully engage with onboarding and instructional materials, leading to inconsistent recording practices, insufficient details, and missing data. Scanning is particularly challenged by receipt unavailability, while manual entry lacks product item details. Also, omitted expenses from other household members are a concern we believe might be hard to control. In the Household Budget Survey 2022, we observed a lower-than-expected share of scanning, with only about 50 percent of all receipts but 85 percent of all product items scanned. Surprisingly, scanning compliance appears to increase with respondent age, as mature adults tend to follow instructions more closely than younger ones.

While editing, imputation, and machine learning can address many data quality issues, these solutions currently require significant resources. However, as machine learning models continue to advance, we expect this resource demand to decrease. Enhancing data quality will involve increasing the proportion of scanned receipts since manual data entry poses greater challenges due to missing information and fewer data points needed for accurate and automated product categorization.

A critical question is whether the web application's current digital assistance effectively communicates task requirements and convey complex questions as effectively as interviewers traditionally do. Our experience indicates that interviewer involvement is especially valuable in a tech-driven, self-reported data collection environment. Interviewers verify survey legitimacy, emphasize what needs to be recorded, stress the importance of accuracy, conduct targeted reminding, and help ensure survey completion. Reducing or eliminating this role could negatively impact respondents' participation and data quality, underscoring the need to maintain the level of interviewer support in the HBS.

10.2. Actions for future surveys

For the upcoming Household Budget Survey (HBS) in 2026, we are considering several measures to improve survey performance and data quality:

1. Targeted recruitment: Implement tailored strategies to boost participation, shifting resources to target groups with low response rates, e.g., the elderly and groups with low education level.
2. **Enhanced technology:** Improve scanning technology, editing tools, machine learning models, and back-office systems to support both machine and human data handling. This will aid both respondents in the recording process and in-house processing teams.
3. **Editing process optimization:** Evaluate and enhance editing functionality both in-app to assist respondents and in-house at the institute to improve efficiency and quality of data editing and coding.
4. **Secure login:** Update our secure login to meet current standards, maintaining trust and data protection.
5. **Questionnaire simplification:** Reduce and simplify the questionnaire length for fixed and large expenses to reduce drop out and increase the response burden.
6. **App improvements to secure more use of OCR scanning:** To reduce dropout and improve data quality, further app development is essential. Improvements that will be evaluated are:
 - Digital receipt integration,
 - Improvements of user initiated smart search and algorithms,
 - In-app push notifications to motivate and guide correct recording,
 - Enhanced help
 - Provide a graphical overview of respondent's expenses recorded "to give back to" participants.
7. **Dedicated CATI and support team.** Allowing the CATI team to handle support with a dedicated line and contact point for respondents will enhance service and guidance, providing a better experience that fosters respondents' engagement. Allowing the CATI team to handle support with a dedicated line and contact point for respondents, will enhance the service and guidance of respondents and improve response rates and data quality.
8. **Alternative options for older adults or non-digital respondents:** Explore the need for a non-digital option (e.g., PASI/CATI) for older respondents.

Going forward, our challenge will be balancing user-friendliness with data accuracy. Over time, the app's design and functionality will evolve to improve the self-reporting experience for respondents. Advances in scanning, editing, and machine learning technologies will continue to support data quality. And we can also reduce and simplify the questionnaires and possibly include alternative data collection methods to bridge the gap left by reduced interviewer involvement when using web application. However, we don't believe that replacing interviewer assistance is effective. Our findings show that it is crucial to retain interviewers in some capacity to ensure high response rates and data quality.

References

Berg, N., Holmøy, A. (2024). *Lessons learned from the Norwegian Household Budget Survey 2022*. Statistics Norway paper at Q2024.

Berg, N. & Lund, K. (2023). [*User testing of the new Household. Documentation Report 2023*](#). Statistics Norway publication 2024/19.

Berg, N., & Seferi, G. (2022). *The impact of a smart survey approach on participation and data quality. Tentative results and insights from the Norwegian HBS 2022*. Statistics Norway paper at NSM 2022.

Holmøy, A. & Lillegård, M. (2014). [*Forbruksundersøkelsen 2012. Dokumentasjonsrapport*](#). . Statistics Norway publication 2014/17.

Pettersen, A. M. & Engvik, M. (2022). [*Livskvalitetsundersøkelsen 2022. Dokumentasjonsnotat*](#). Statistic Norway publication 2022/35.

Seferi, G., Holmøy, A., Lillegård, M., & Egge-Hoveid, K. (2024). [*Forbruksundersøkelsen 2022. Dokumentasjonsnotat*](#). Statistic Norway publication 2024/46.

Appendix A: Altinn letters to sub-sample 1) CATI

Altinn letter sent four days before CATI recruitment started



Statistisk sentralbyrå
Statistics Norway

Forbruk «IO-NR»

TIL «NAVN»

Du er invitert til å delta i Forbruksundersøkelsen

Statistisk sentralbyrå (SSB) gjennomfører nå en undersøkelse om husholdningers forbruk av varer og tjenester, og din husholdning er trukket ut til å delta.

Vi kommer snart til å ringe deg

En av våre intervjuere vil i løpet av neste uke ringe deg for å gjennomføre et kort intervju. Etter intervjuet skal husholdningen registrere utgifter til varer og tjenester i en avtalt uke, og fylle ut et webskjema på omtrent 15 minutter. Intervjueren gir deg mer informasjon om hvordan du gjennomfører undersøkelsen.

Hvis du vil avtale tidspunkt for intervju, kan du ta kontakt med oss på telefon eller e-post.

Alle som fullfører undersøkelsen, får et gavekort på 500 kroner. Gavekortet kan brukes i mange butikker over hele landet og i utvalgte nettbutikker.

Svarene dine er viktige

2 000 husholdninger er trukket fra SSBs register til å være med i undersøkelsen. Det er frivillig å delta. Svarene dine blir brukt i forskning og analyse for å sikre at digitale innsamlingsløsninger gir god og pålitelig forbruksstatistikk. Anonyme data fra undersøkelsen inngår i et europeisk forskningsprosjekt.

Du kan lese mer om undersøkelsen på ssb.no/forbruk-svar.

Opplysningene dine er sikre hos oss

- Svarene dine bruker vi kun til forskning og analyse, og vi offentliggjør aldri enkelt svar.
- Frem til 15. juni 2024, kan du trekke deg og be om at svarene blir slettet. Dette gjør du ved å ringe eller sende oss en e-post.
- For at undersøkelsen skal være så kort som mulig, benytter vi opplysninger om deg og din husholdning fra registre SSB har tilgang til. Dette gjelder opplysninger fra Folkeregisteret, opplysninger om utdanning fra skoleeiere og Lånekassen og inntekt fra Skatteetaten og Nav.
- Vi anonymiserer alle opplysninger om deg og husholdningen din innen 1. oktober 2025.

Vi kontakter 10-20 husholdninger etter at undersøkelsen er gjennomført, for å stille noe spørsmål om hvordan det var å delta.

Har du spørsmål?

Kontakt oss på

- e-post: sva@ssb.no
- telefon: 62 88 56 08

Svartjenesten er åpen kl. 09–14 mandag til fredag.

Med vennlig hilsen

Geir Axelsen
administrerende direktør

Altinn letter sent on the start day of the registration

Statistisk sentralbyrå
Statistics Norway

Forbruk {IO-NR}

{Navn}

Din registreringsuke i Forbruksundersøkelsen starter i dag

Takk for at du har sagt ja til å delta i Forbruksundersøkelsen. Som avtalt i startintervjuet skal du nå gjøre følgende:

1. Registrere alle utgifter husholdningen har i en uke fra i dag (mandag til søndag)
2. Fylle ut et webskjema, dette tar omtrent 15 minutter
3. Bekrefte at du er ferdig med registrering og webskjema

Oppgavene gjør du i en webapplikasjon. Nettadressen er: <https://forbruk.ssb.no>

Når du har fullført alle oppgavene, sender vi deg et gavekort på 500 kroner.

Når du åpner applikasjonen første gang, kan den installeres som en app på mobilen din. Innlogging skjer med MinID eller BankID. Når du er innlogget, vil du være det for hele registreringsperioden.

Hvordan registrere utgifter i appen

- Vi anbefaler deg å bruke mobil for registrering, men du kan også bruke nettbrett eller PC.
- Når du logger inn i appen første gang, får du instruksjoner om hva du skal gjøre.
- Du registrerer utgifter ved å ta bilde av kvitteringer fra alle kjøp i perioden.
- Du kan også legge inn kjøp og andre utgifter manuelt. Husk å legge inn en og en vare.
- Du skal registrere alt husholdningen bruker penger på. Husk også utgifter til for eksempel lege, frisør, museumsbesøk, parkering og kino.
- Appen har også et webskjema med spørsmål om utgifter til bolig og kjøp du ikke gjør så ofte. Skjemaet kan fylles ut når som helst i registreringsperioden.

Har du spørsmål?

Kontakt oss på

- e-post: svar@ssb.no
- telefon: 62 88 56 08

Svartjenesten er åpen kl. 09–14 mandag til fredag.

Tusen takk for at du deltar!

Vennlig hilsen
Statistisk sentralbyrå

Appendix B: SMS texts to sub-sample 1) CATI

SMS1 recruitment

Til <<navn>>. Statistisk sentralbyrå (SSB) inviterer deg til å delta i Forbruksundersøkelsen. Vi har sendt deg mer informasjon i Altinn. En av våre intervjuere vil ringe deg en av de nærmeste dagene. Kontakt oss på 62885608 eller svar@ssb.no dersom du vil avtale tidspunkt for intervju. Vennlig hilsen SSB

SMS2 reminder1 recruitment

Hei <<navn>>!
Vi har invitert deg til å delta i Forbruksundersøkelsen. Vi trenger vi svar fra så mange som mulig for å forstå utviklingen. Derfor er dine svar viktige!
Vi ringer deg igjen i løpet av noen dager.
Kontakt oss på 62885608 eller svar@ssb.no dersom du vil avtale tidspunkt for intervju.
Vennlig hilsen SSB

SMS 3 to those recruited - link in SMS and Altinn

Takk for at du har sagt ja til å delta i Forbruksundersøkelsen.
Registreringsuken din starter i dag.
For å registrere utgifter, logger du inn i appen via ID-porten: <https://forbruk.ssb.no> eller via lenken du har mottatt i Altinn i dag.
Husk å ta med alt husholdningen bruker penger på.
Ta kontakt på 62885608 eller svar@ssb.no dersom du har spørsmål.
Vennlig hilsen Statistisk sentralbyrå

SMS 3 to those recruited - link in Altinn

Takk for at du har sagt ja til å delta i Forbruksundersøkelsen.
Registreringsuken din starter i dag.
For å registrere utgifter, logger du inn i appen via lenken du har mottatt i Altinn i dag.
Husk å ta med alt husholdningen bruker penger på.
Ta kontakt på 62885608 eller svar@ssb.no dersom du har spørsmål.
Vennlig hilsen Statistisk sentralbyrå

SMS4 Reminder1 -link in SMS and Altinn

Tusen for at du bidrar til ny Forbruksstatistikk.
Har du spørsmål eller behov for hjelp? Kontakt oss på 62885608 eller svar@ssb.no
Tips: Du kan enkelt registrere alle kvitteringer ved å ta bilde av dem i appen.
Logg inn i appen via <https://forbruk.ssb.no> eller via lenken du har mottatt i Altinn.
Vennlig hilsen SSB

SMS4 Reminder1 -link in SMS and Altinn

Takk for at du bidrar til ny Forbruksstatistikk.

Har du spørsmål eller behov for hjelp? Kontakt oss på 62885608 eller svar@ssb.no

Tips: Du kan enkelt registrere alle kvitteringer ved å ta bilde av dem i appen.

Logg inn i appen via lenken du har mottatt i Altinn.

Vennlig hilsen SSB

SMS5 -Encouragement 1 to those started

Tusen takk for at du bidrar til ny Forbruksstatistikk.

Tips! Ta vare på alle kvitteringer, så slipper du å huske på hva du har kjøpt. Du kan ikke legge inn totalsummer.

Kontakt oss på 62885608 eller svar@ssb.no dersom du lurer på noe. Vennlig hilsen SSB

SMS6 Reminder2 -link in SMS and Altinn

Vi er over halvveis i uken, og det kan være lurt å komme i gang med registreringen.

Ta gjerne kontakt på 62885608 eller svaer@ssb.no dersom du lurer på noe.

Tips: Dersom du registrerer utgifter manuelt, husk å legge inn en og en vare.

Logg inn i appen via <https://forbruk.ssb.no> eller via lenken du har mottatt i Altinn.

Vennlig hilsen SSB

SMS6 Reminder2 -link in Altinn

Vi er over halvveis i uken, og det kan være lurt å komme i gang med registreringen.

Ta gjerne kontakt på 62885608 eller svaer@ssb.no dersom du lurer på noe.

Tips: Dersom du registrerer utgifter manuelt, husk å legge inn en og en vare.

Logg inn i appen via lenken du har mottatt i Altinn.

Vennlig hilsen SSB

SMS7 Encouragement 2 to those started


Din deltakelse i Forbruksundersøkelsen er snart over. Du har gjort en kjempeviktig jobb!

Tips: Skal du kanskje på kafé, i svømmehallen eller på kino i helgen? Husk å registrere slike utgifter også.

Ta kontakt på 62885608 eller svar@ssb.no dersom du har spørsmål. Vennlig hilsen SSB

Appendix C: Altinn letters to sub-sample 2) NO CATI

Altinn letter sent four days before the registration started



Statistisk sentralbyrå
Statistics Norway

Forbruk «IO-NR»

TIL «NAVN»

Du er invitert til å delta i Forbruksundersøkelsen

Statistisk sentralbyrå (SSB) gjennomfører nå en undersøkelse om norske husholdningers forbruk av varer og tjenester, og husholdningen din er trukket ut til å delta.

De som deltar, registrerer alt husholdningen bruker penger på i en avtalt uke. Dette gjøres i en webapplikasjon hvor en tar bilde av kvitteringer og fyller ut et webskjema på omtrent 15 minutter.

Din registreringsuke er: <<kPeriode>>

På startdagen sender vi deg mer informasjon i Altinn om hvordan undersøkelsen gjennomføres. Dersom registreringsuken ikke passer, er det mulig å endre den i appen på startdagen.

Alle som fullført undersøkelsen, får du et gavekort på 500 kroner. Gavekortet kan brukes i mange butikker over hele landet og i utvalgte nettbutikker.

Svarene dine er viktige

2 000 husholdninger er trukket tilfeldig fra SSBs husholdningsregister til å være med i undersøkelsen. Det er frivillig å delta. Svarene dine blir brukt i forskning og analyse for å sikre at digitale innsamlingsløsninger gi god og pålitelig forbruksstatistikk. Anonyme data fra undersøkelsen inngår i et europeisk forskningsprosjekt.

Du kan lese mer om undersøkelsen på ssb.no/forbruk-svar.

Opplysningene dine er sikre hos oss

- Svarene dine bruker vi kun til forskning og analyse, og vi offentliggjør aldri enkelt svar.
- Frem til 15. juni 2024, kan du trekke deg og be om at svarene blir slettet. Dette gjør du ved å ringe eller sende oss en e-post.
- For at undersøkelsen skal være så kort som mulig, benytter vi opplysninger om deg og din husholdning fra registre SSB har tilgang til. Dette gjelder opplysninger fra Folkeregisteret, opplysninger om utdanning fra skoleeiere og Lånekassen og inntekt fra Skatteetaten og Nav.
- Vi anonymiserer alle opplysninger om deg og husholdningen din innen 1. oktober 2025.

Vi kontakter 10-20 husholdninger etter at undersøkelsen er gjennomført, for å stille noe spørsmål om hvordan det var å delta.

Har du spørsmål?

Kontakt oss på

- e-post: svaer@ssb.no
- telefon: 62 88 56 08

Svartjenesten er åpen kl. 09–14 mandag til fredag.

Med vennlig hilsen

Geir Axelsen
administrerende direktør

Altinn letter on the start day of the registration

Statistisk sentralbyrå
Statistics Norway

Forbruk {ioNumber}

{name}

Din registreringsuke i Forbruksundersøkelsen starter i dag

Torsdag i forrige uke, sendte vi deg et brev i Altinn om at din husholdning er trukket ut til å delta i Statistisk sentralbyrås undersøkelse om forbruk av varer og tjenester.

Undersøkelsen består av følgende tre oppgaver:

1. Registrere alle utgifter husholdningen har i uken <<kPeriode>>
2. Fylle ut et webskjema, dette tar omtrent 15 minutter
3. Bekrefte at du er ferdig med registrering og webskjema

Opgavene gjør du i en webapplikasjon. Nettadressen er: <https://forbruk.ssb.no>

Når du har fullført alle oppgavene, sender vi deg et gavekort på 500 kroner.

Når du åpner applikasjonen første gang, kan den installeres som en app på din mobil. Innlogging skjer med MinID eller BankID. Når du er innlogget, vil du være det for hele registreringsperioden.

Dersom registreringsuken ikke passer, har du mulighet til å endre den i appen.

Det er frivillig å delta. Svarene dine blir brukt i forskning og analyse for å sikre at digitale innsamlingsløsninger gi god og pålitelig forbruksstatistikk. Du kan lese mer om undersøkelsen på ssb.no/forbruk-svar.

Hvordan registrere utgifter i appen

- Vi anbefaler deg å bruke mobil for registrering, men du kan også bruke nettbrett eller PC.
- Når du logger inn i appen første gang, får du instruksjoner om hva du skal gjøre.
- Du registrerer utgifter ved å ta bilde av kvitteringer fra alle kjøp i perioden.
- Du kan også legge inn kjøp og andre utgifter manuelt. Husk å legge inn en og en vare.
- Du skal registrere alt husholdningen bruker penger på. Husk også utgifter til for eksempel lege, frisør, museumsbesøk, parkering og kino.
- Appen har også et webskjema med spørsmål om utgifter til bolig og kjøp du ikke gjør så ofte. Skjemaet kan fylles ut når som helst i registreringsperioden.

Har du spørsmål?

Kontakt oss på

- e-post: sva@ssb.no
- telefon: 62 88 56 08

Svartjenesten er åpen kl. 09–14 mandag til fredag.

Tusen takk for at du deltar!

Vennlig hilsen

Statistisk sentralbyrå

Appendix D: SMS texts to sub-sample 2) NO CATI

SMS1 without recruitment

Til <<navn>>. Statistisk sentralbyrå inviterer deg til å delta i Forbruksundersøkelsen. Vi har sendt deg mer informasjon i Altinn.

Ta kontakt på 62885608 eller svar@ssb.no dersom du har spørsmål.

Vennlig hilsen Statistisk sentralbyrå

SMS3 to those not recruited - link in SMS and Altinn

Hei. Statistisk sentralbyrå (SSB) trenger dine svar i Forbruksundersøkelsen.

Deltakere registrerer alt husholdningen bruker penger på i en uke, og **din uke starter i dag**.

Utgifter registreres i forbruksappen. Logg inn via ID-porten på <https://forbruk.ssb.no> eller via lenken du har mottatt i Altinn i dag.

Ta kontakt på 62885608 eller svar@ssb.no for spørsmål, eller se informasjon i Altinn.

Vennlig hilsen SSB

SMS3 to those not recruited - link in Altinn

Hei. Statistisk sentralbyrå (SSB) trenger dine svar i Forbruksundersøkelsen.

Deltakere registrerer alt husholdningen bruker penger på i en uke, og **din uke starter i dag**.

Utgifter registreres i forbruksappen. Logg inn via lenken du har mottatt i Altinn i dag.

Ta kontakt på 62885608 eller svar@ssb.no dersom du har spørsmål, eller se informasjon i Altinn.

Vennlig hilsen SSB

SMS4 Reminder1 - link in SMS and Altinn

Tusen for at du bidrar til ny Forbruksstatistikk.

Har du spørsmål eller behov for hjelp? Kontakt oss på 62885608 eller svar@ssb.no

Tips: Du kan enkelt registrere alle kvitteringer ved å ta bilde av dem i appen.

Logg inn i appen via <https://forbruk.ssb.no> eller via lenken du har mottatt i Altinn.

Vennlig hilsen SSB

SMS4 Reminder1 - link in Altinn

Takk for at du bidrar til ny Forbruksstatistikk.

Har du spørsmål eller behov for hjelp? Kontakt oss på 62885608 eller svar@ssb.no

Tips: Du kan enkelt registrere alle kvitteringer ved å ta bilde av dem i appen.

Logg inn i appen via lenken du har mottatt i Altinn.

Vennlig hilsen SSB

SMS5 -Encouragement 1 to those started

Tusen takk for at du bidrar til ny Forbruksstatistikk.

Tips! Ta vare på alle kvitteringer, så slipper du å huske på hva du har kjøpt. Du kan ikke legge inn totalsummer.

Kontakt oss på 62885608 eller svar@ssb.no dersom du lurer på noe. Vennlig hilsen SSB

SMS6 Reminder2 - Link in SMS and Altinn

Vi er over halvveis i uken, og det kan være lurt å komme i gang med registreringen.
Ta gjerne kontakt på 62885608 eller sva@ssb.no dersom du lurer på noe.
Tips: Passer det ikke denne uken? Du kan flytte registreringsperioden til neste uke i appen.
Logg inn i appen via <https://forbruk.ssb.no> eller via lenken du har mottatt i Altinn.
Vennlig hilsen SSB

SMS6 Reminder2 - Link in Altinn

Vi er over halvveis i uken, og det kan være lurt å komme i gang med registreringen.
Ta gjerne kontakt på 62885608 eller sva@ssb.no dersom du lurer på noe.
Tips: Passer det ikke denne uken? Du kan flytte registreringsperioden til neste uke i appen.
Logg inn i appen via lenken du har mottatt i Altinn.
Vennlig hilsen SSB

SMS7 Encouragement 2 to those started

Din deltakelse i Forbruksundersøkelsen er snart over. Du har gjort en kjempeviktig jobb!
Tips: Skal du kanskje på kafé, i svømmehallen eller på kino i helgen? Husk å registrere slike utgifter også.
Ta kontakt på 62885608 eller sva@ssb.no dersom du har spørsmål. Vennlig hilsen SSB

SMS 9a - Thank you to those completed

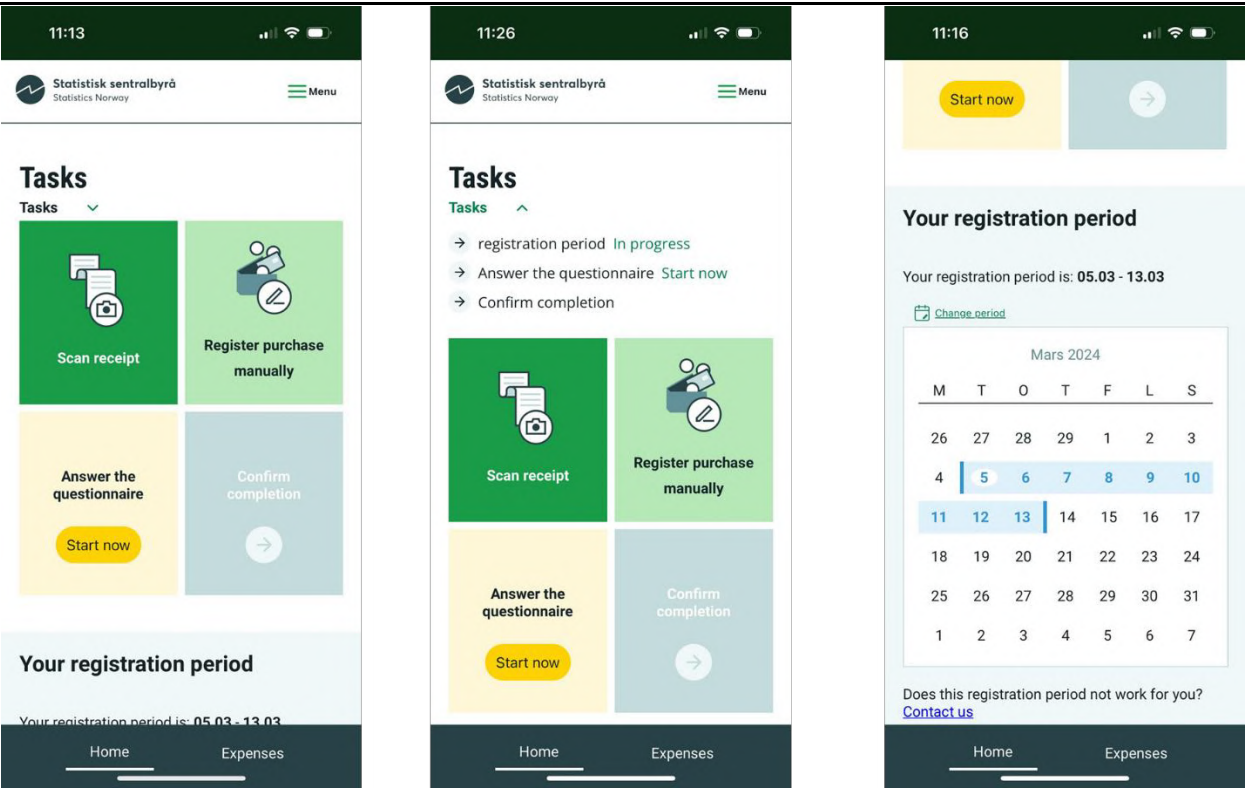
Registreringsperioden på Forbruksundersøkelsen er nå over. SSB takker for din deltakelse. I løpet av en til to uker sender vi deg et gavekort på 500 kroner. Ta kontakt på 62885608 eller sva@ssb.no dersom du har spørsmål. Vennlig hilsen SSB

SMS 9b - Thank you to those started

Registreringsperioden på Forbruksundersøkelsen er nå over. SSB takker for din deltakelse. Dersom du ikke er helt ferdig med registreringen, setter vi pris på om du gjør det innen førstkommande onsdag. Når du er ferdig, er det viktig at du trykker på knappen «Jeg er ferdig med undersøkelsen». En til to uker etter at undersøkelsen er avsluttet, sender vi deg et gavekort på 500 kroner. Ta kontakt på 62885608 eller sva@ssb.no dersom du har spørsmål. Vennlig hilsen SSB

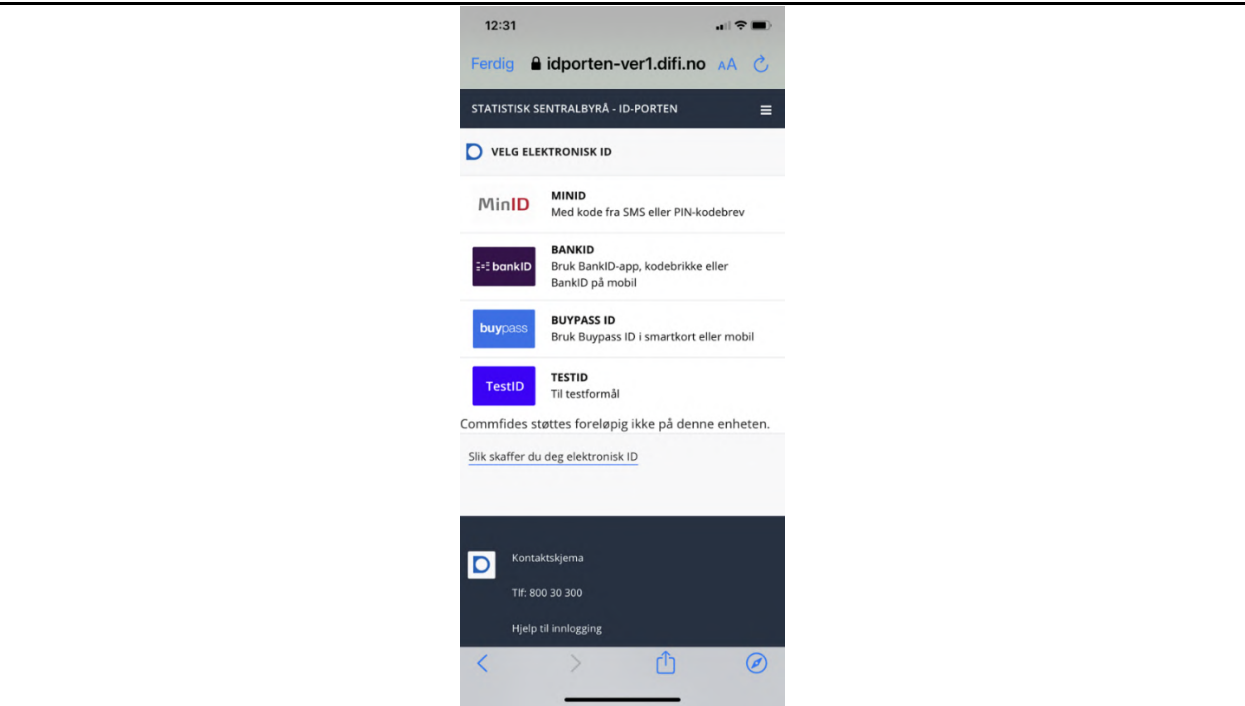
Appendix E: Screenshots from the web application

Figure E.1 Home screen with and without drop-down menus open for task and for calendar



Source: The Household Budget Survey 2022, Statistics Norway

Figure E.2 Login once with the national ID-porten for secure login



Source: The Household Budget Survey 2022, Statistics Norway

Figure E.3 Onboarding or “get started” instructions with consent

12:57

Statistisk sentralbyrå
Statistics Norway

When you participate in the consumption survey, you contribute to important statistics about Norwegians' consumption habits.

1/2

Next

12:57

Statistisk sentralbyrå
Statistics Norway

← Back

In the survey you shall:

- record the entire household consumption for one week
- complete the questionnaire
- confirm that you are done

2/2

Start the survey

12:57

Statistisk sentralbyrå
Statistics Norway

Consents

By registering expenses and answering the questionnaire in this app, you consent that the information can be used for official statistics and research.

What information do we collect?

- Everything you register and scan in the app. Make sure that names or other personal information do not appear on photos of receipts.
- Information about browser and operating system.

Your information is secure with us

- Information from the app is stored separately from birth number and other contact information.
- Receipts you upload are scanned by a third party, but are deleted by them as soon as the scan is performed.
- By 1.10.2025, all information will be anonymised.

It is voluntary to participate

- You can contact us at any time at svar@ssb.no or 62 88 56 08 to request that the information you have provided be deleted.

Statistics Norway follows the privacy rules

- The survey is conducted in accordance with the Statistics Act and the Personal Data Act.

You can read more about privacy at <https://www.ssb.no/omsab/personvern>

I consent

Source: The Household Budget Survey 2022, Statistics Norway

Figure E.4 Receipt scanning

11:13

Statistisk sentralbyrå
Statistics Norway

Tasks

Tasks

Scan receipt

Register purchase manually

Bibliotek

Ta bilde eller video

Velg fil

Your registration period

Your registration period is: 05.03 - 13.03

Home Expenses

11:14

Statistisk sentralbyrå
Statistics Norway

Expenses

Scan receipt

Register purchase manually

Show: ☒ Newest first ☐ Oldest first

01 mar	Kiwi	20,00	
28 feb	REMA	19,90	
	Kiwi	83,00	
14 feb	Nøstet Mitt NØSTET MITT KOLBOTN AS	790,00	
10 des	Meny	469,80	
11 mar	Joker	71,80	

Home Expenses

Source: The Household Budget Survey 2022, Statistics Norway

Figure B.5 Manual registration of purchase

12:20

← Back

Add item

Which product or service did you buy?

Enter a single item, max 30 chars 5/30

Hip wool

Add "Hip wool"

How much did you pay?

Enter precise amount

1,00

Please enter valid amount

12:12

← Back

Add item

Which product or service did you buy?

Enter a single item, max 30 chars 8/30

Hip wool

Eg.: Apples or Hairdresser

How much did you pay?

Enter precise amount

790,00

12:12

← Back

New Purchase

Date

05.03.2024

Shop/place/beneficiary

Nøstet mitt

Item

Amount

Hip wool

790,00

Add item/service

Total

790,00

Save

11:16

Statistisk sentralbyrå

Statistics Norway

Menu

Expenses

Scan receipt

Register purchase manually

Show: Newest first Oldest first

05 mar Nøstet mitt 790,00

01 mar Kiwi 20,00

28 feb REMA 19,90

Kiwi 83,00

14 feb Nøstet Mitt NØSTET MITT 790,00

KOLBOTN AS

10 des Meny 469,80

Home Expenses

Source: The Household Budget Survey 2022, Statistics Norway

Figure E.5 Overview of expenses recorded

11:16

Statistisk sentralbyrå

Statistics Norway

Menu

Expenses

Scan receipt

Register purchase manually

Show: Newest first Oldest first

05 mar Nøstet mitt 790,00

01 mar Kiwi 20,00

28 feb REMA 19,90

Kiwi 83,00

14 feb Nøstet Mitt NØSTET MITT 790,00

KOLBOTN AS

10 des Meny 469,80

Home Expenses

Source: The Household Budget Survey 2022, Statistics Norway

Appendix F: Recruitment interview CATI (Blaise)

Innledning

Innled

Registrer om du skal starte intervjuet eller registrere en overføring, frafall eller avgang.

1. Start intervjuet
2. Overfør til annen intervjuer
3. Frafall
4. Avgang

Her blir det mulig for intervjueren å legge inn standard frafall og avgangsgrunner dersom IO ikke ønsker å delta eller er avgang.

Hvis Innled = 1

Samtykke1

Har du lest e-post eller brev og samtykker til å delta?

1. Ja, jeg har mottatt brevet og samtykker til å delta
2. Ja, jeg har fått referert innholdet i brevet og samtykker i å delta
3. Nei, og jeg samtykker ikke til å delta

Husholdningskartlegging

Start1

Nå kommer noen spørsmål om din husholdning

Nå starter husholdningskartleggingen. Tast <1> for å gå videre.

*Hush

Til husholdningen regner vi alle personer som er fast bosatt i boligen, og som har felles matbudsjett. Personer som er fast bosatt i boligen, men som er borte fra hjemmet, for eksempel på grunn av arbeid, skal regnes med.

Dersom flere i husholdningen

Vi har registrert at følgende ^AntReg personer tilhører husstanden i tillegg til deg.

(Intervjuer leser opp liste på skjermen over husholdningsmedlemmer hentet fra register)

Består husholdningen av de samme personene nå, eller er det noen som skal legges til eller trekkes fra?

1. Husstanden stemmer
2. Personer skal både legges til og trekkes fra
3. Person(er) skal legges til
4. Personer skal trekkes fra

Til aleneboere

EnPers

I følge våre opplysninger bor du alene i husholdningen. Stemmer det?

1. Ja
2. Nei

Hvis Hush= 2 eller 4

AntUt

Hvem er det som ikke tilhører husholdningen?

(Kommer opp liste over husstanden, person(er) som skal tas ut kan krysses av)

Hvis Hush = 2 eller 3

AntNye

Hvor mange personer skal legges til i husholdningen?

For hver person som legges til husholdningen spør vi om navn, fødselsdato, kjønn og familieforhold til IO. For alle andre hentes disse opplysningene fra register.

Leggtil

Dersom en person skal legges til

Nå kommer noen få spørsmål om personen som skal legges til

Dersom flere personer skal legges til

Nå kommer noen få spørsmål om personene som skal legges til. Vi starter med den eldste.

Du legger nå til person nummer [nr.]

Fornavn

Hva er personens fornavn?

Etternavn

Hva er personens etternavn?

FodtDato

Når er [fornavn etternavn] født?

Kjonn

Er [Navn] mann eller kvinne?

1. Mann
2. Kvinne

Slekt (til alle)

Hvilket familieforhold har [navn] til deg?

- 1.
2. Ektefelle
3. Samboer
4. Sønn/datter
5. Stesønn/stedatter
6. Søsken/halvsøsken
7. Stesøsken
8. Foreldre
9. Steforeldre
10. Svigerforeldre
11. Svigersønn/-datter
12. Besteforeldre
13. Barnebarn
14. Annen slektning av IO
15. Annen ikke-slektning

Slektregintro

For alle husholdningsmedlemmer som ikke er nye

Siden vi mangler informasjon om slektsforhold i registeret, vil vi nå spørre om detteSlektreg

Hvilket familieforhold har [Navn] til deg?

- 1.
2. Ektefelle
3. Samboer

4. Sønn/datter
5. Stesønn/stedatter
6. Søsken/halvsøsken
7. Stesøsken
8. Foreldre
9. Steforeldre
10. Svigerforeldre
11. Svigersønn/-datter
12. Besteforeldre
13. Barnebarn
14. Annen slektning av IO
15. Annen ikke-slektning

Avslutt

Husholdningen består nå av

Kommer opp liste over husholdningsmedlemmer

Hvis dette stemmer merker du av for ferdig. Hvis ikke må du gå tilbake og rette opp.

Stemmer dette?

Registreringsperiode

TildFper

Forbruksundersøkelsen gjennomføres ved at husholdningen registrerer sine daglige kjøp og andre utgifter i en webapplikasjon (app) i en uke. I appen er det også et spørreskjema som skal besvares.

For at undersøkelsen skal gi et jevnt forbruk over året, er det svært viktig at [du/dere] registrerer kjøp og utgifter akkurat i den perioden [du/dere] har fått tildelt.

En progressiv webapplikasjon er en vanlig nettside som er laget slik at den gir en app-lignende opplevelse. Appen vil i utgangspunktet fungere både på PC, nettbrett og smarttelefon, på alle nettlesere, on- og offline. Appen for Forbruksundersøkelsen fungerer best på en smarttelefon, med PC kan også benyttes. Nettbrett anbefales ikke. Innlogging skjer via ID-porten.

Perioden [du/dere] skal registrere kjøp og andre utgifter for, er fra [dd.mm-dd.mm]. (Eks. 08.04-14.04)

Er det denne perioden husholdningen fører for?

1. Ja → Epost
2. Nei → NyForerlStart

NyFper

Velg ny registreringsperiode.

Trykk backspace for å velge registreringperiode

Dersom husstanden ikke ønsker å starte førstkommende mandag, får husstanden et valg om å starte registrering en til tre uker frem i tid.

Det kommer da opp en liste med tre aktuelle registreringsperioder.

Kontaktperson

Denne bolken kommer kun dersom mer enn 1 person 18+ i husholdningen

Kontpers

Vi ønsker at en person i husholdningen har hovedansvaret for registrering av kjøp og andre utgifter i appen.

Denne personen vil motta videre informasjon fra oss.

Hvem i husstanden skal ha hovedansvaret for registreringen?

Komme opp liste over husholdningsmedlemmer 18+ som det kan velges fra (er hentet fra kosthuhsoldningsregisteret)

Dersom det velges en kontaktperson som ikke er i oversikten

Kontpersfnrspm

Innlogging i appen skjer via ID-porten, noe som krever at våre systemer har fødselsnummer på alle personer som skal bruke den. Fødselsnumrene blir lagret pseudonymisert/avidentifisert i våre systemer. Personen du har oppgitt som hovedansvarlig for registreringen, er ikke oppført som en del av din husholdning i våre registre.

Vi trenger derfor denne personens fødselsnummer til bruk ved innlogging med ID-porten.

Hva er [Navns] fødselsnummer?**Epost**

Samme dag som registreringsperioden starter, vil du/[navn på kontpers] motta et brev i Altinn med litt mer informasjon om hvordan undersøkelsen skal gjennomføres. Du/ [navn på kontpers] vil også motta en SMS.

Hvis ikke registrert mobilnummer:

Nytelefon

Hvilket telefonnummer kan vi nå deg på?

Appendix G: Web questionnaire

The SSI Recruitment Experiment Survey 2024. Statistics Norway	
Web-questionnaire	
	Content
Deler i skjema	1. About your home
	2. Housing: Ownership and finances
	3. Housing: Electricity and heating
	4. Holiday homes in Norway and abroad
	5. Transport, travel and accommodation
	6. Schooling and education
	7. Infrequent purchases (durable goods)
	9. Other regular expenses
About your home	Hus1
	What type of housing do you live in?
	1. Detached house
	2. Row house, terraced house or semi-detached house
	3. Vertically or horizontally divided building with three or four housing units
	4. Flat
	5. Live in boat, caravan or car
About your home	Bol1
	How many rooms do you have at your disposal in your home?
	Do not include kitchens, bathrooms, hallways, laundry rooms or rooms measuring less than 6 m ² .
	[1-50]
About your home	Bol2
	Approximately, how many square metres does your home measure?
	For basements and attics, only living quarters are to be included.
	:1....9999 m ²
Housing: Ownership and finances	Eier1
	Do you own or rent your home?
	Select 'Rent' if your home is loaned to you by some other means.
	1. Own
	2. Rent
Housing: Ownership and finances	Hvis Eier1 = 1
	Eier2
	Do you own the home as a freeholder or through a housing cooperative?
	1. Freeholder

	2. Housing cooperative (borettslag) or housing stock company (boligaksjeselskap),
Housing: Ownership and finances	Hvis Eier1 = 1
	Laan1
	Do you have a mortgage or loan secured on your home?
	Include all mortgages secured on your home. Do not include shared debt (fellesgjeld).
	1.Yes
	2.No
Housing: Ownership and finances	Hvis Laan1 = 1
	Laan2
	How much remains to be paid on the mortgage/loan?
	[1-99 000 000] NOK
Housing: Ownership and finances	Hvis Laan1=1 (ja)
	LUtg1
	How much do you pay per month for the mortgage/loan secured on your home?
	Enter 0 if you do not pay anything. Do not include payments for shared debt (fellesgjeld)
	Amount _____
Housing: Ownership and finances	Hvis LUtg1>0
	LRent1
	How much of this is interest?
	1. (1-1 000 000) NOK
	2. Don't know
Housing: Ownership and finances	Hvis Laan1=1 (ja)
	LRent2
	What is the current interest rate on your mortgage? Specify this as a percentage.
	If you have mortgages with different interest rates, give the average rate.
	1. [0-9,9] per cent
	2. Don't know
Housing: Ownership and finances	Hvis eier1=2 (leier/låner)
	Husleie1
	Do you pay rent?
	1. Yes
	2. No
Housing: Ownership and finances	Hvis eier1=2 (leier/låner) og husleie1=1 (JA)
	Husleie2
	How much rent do you pay per month?

	[1-99 997] NOK
Housing: Ownership and finances	Hvis eie1=2 (leiere) og husleie1=1 (som betaler husleie)
	Husleie3
	Is electricity included in the rent?
	1. Yes
	2. No
Housing: Ownership and finances	Hvis Eier1=1 og hus1 ≠ 1 (enebolig)
	FUt看1
	Do you pay shared costs (fellesutgifter) to a housing cooperative (borettslag) or through joint ownership (sameie)?
	1. Yes
	2. No
Housing: Ownership and finances	Hvis FUt看1 = Yes
	FUt看2
	How much do you pay in shared costs (fellesutgifter)?
	[1-1000 000] NOK (Per month, quarter, half-year, year)
Housing: Ownership and finances	Til Alle
	Gar1
	Do you own or rent garage space, a carport or parking space?
	1. Own
	2. Rent
	3. No
Housing: Ownership and finances	Hvis gar1=1, 2
	Gar1a
	Do you pay a fixed charge for parking?
	Do not include charges for garage space, carports or parking spaces that are covered in the rent.
	1. Yes
	2. No
Housing: Ownership and finances	Hvis gar1a=1
	Gar1b
	How much do you pay for parking?
	1. [1-100 000] NOK (Select month/quarter/half-year/year)
	2. Don't know
Housing: Electricity and heating	Hvis eier1=1 og hvis eier= 2 og Husleie3=2
	Elut看1

	Is it easier for you to specify your electricity and grid rent costs separately or together?
	1. Separately
	2. Together
	3. The property does not have electricity
Housing: Electricity and heating	Hvis Elutg1 =separat - vise de to første radene.
	Hvis Elutg1=samlet - vise siste rad
	Hvis Elutg1=boligen har ikke strøm – gå tilFyrutg1
	Elutg2
	How much did you pay in electricity and grid rent last month?
	1 Grid rent ... NOK
	2 Electricity ... NOK
	3 Electricity and grid rent together ... NOK
Housing: Electricity and heating	Fyrutg1
	What type of heating do you have in your home?
	If you have a combined wood and bio-oil burner, select both options
	1. Electricity, e.g. panel heater, underfloor heating
	2. Air-to-air heat pump
	3. Other type of heat pump, e.g. air-to-water, underground, water source
	4. Wood burner
	5. Bio-oil burner
	6. Gas heater
	7. Own or shared central heating system
	8. Balanced ventilation with heat recovery
	9. Other, e.g. pellet heating, open fireplace, oil/paraffin/kerosine heater, gas, solar power
Housing: Electricity and heating	Hvis Fyrutg1=7
	Sentralvarme1
	What type of central heating system does your home have?
	1. District heating
	2. Electricity
	3. Wood, woodchips, pellets
	4. Bio-oil
	5. Gas
	6. Heat pump
	7. Other
Housing: Electricity and heating	Dersom Fyrutg1=4,5,6,7 eller 9
	Fyrutg2

	How much do you pay for..
	Selected heating NOK (month/quarter/half-year/year)
	Selected heating NOK (month/quarter/half-year/year)
	Selected heating NOK (month/quarter/half-year/year)
Holiday homes in Norway and abroad	FritidsN1
	Do you own or have the use of a holiday home (cabin or other type of property) in Norway?
	If you both own and have the use of such a property, select all that are applicable.
	1. Yes, I own a holiday home
	2. Yes, I own a holiday home with someone else/others
	3. Yes, I have the use of a holiday home
	4. No
Holiday homes in Norway and abroad	Hvis FritidsN1=2,3
	UtgiftFritid1
	Do you have expenses in connection with the holiday home that you have the use of or own with others?
	1. Yes
	2. No
Holiday homes in Norway and abroad	Hvis UtgiftFritid1=1
	UtgiftFritid2
	How much have you paid in costs in the last 12 months?
	Beløp NOK
Holiday homes in Norway and abroad	Hvis FritidsN1=1
	FritidsN2
	Have you had any of the following expenses for a holiday home in the last 12 months?
	1. Ground rent
	2. Municipal charges
	3. None of the above
Holiday homes in Norway and abroad	Hvis FritidsN2 ikke er lik 3
	FritidsN3
	How much have you paid in costs for ...
	Avkrysset.... Sumfelt NOK (Per month, quarter, half-year, year)
	Avkrysset....Sumfelt NOK (Per month, quarter, half-year, year)
Transport, travel and accommodation	Reiser1
	Have you paid for flights or travel by train, boat, bus or similar for leisure trips in the last 12 months? Do not include package tours where travel and accommodation are both included in the price.

	1. Yes
	2. No
Transport, travel and accommodation	Hvis Reiser1=1
	Reiser2
	How did you travel? Select all that are applicable
	1. Domestic flight
	2. International flight
	3. Boat (ferry, Hurtigruten coastal express etc.)
	4. Train
	5. Bus
	6. Other mode of transport, please specify
Transport, travel and accommodation	For avkryssede
	Reiser3
	How much did you pay for ..
	Selected travelNOK
	Selected travel.....NOK
	Overnatting1
Transport, travel and accommodation	Have you paid for overnight accommodation when travelling in the last 12 months? Do not include package tours where travel and accommodation are both included in the price.
	1. Yes
	2. No
Transport, travel and accommodation	Hvis Overnatting1=1
	Overnatting2
	What type of accommodation did you stay in? (Select all that are applicable.)
	1. Rented holiday home or room in house (Airbnb, Novasol, Finn etc.)
	2. Hotel or similar (e.g. apartment hotel, motel, mountain lodge, bed and breakfast, guest house)
	3. Campsite (caravan, camper van, tent, camping cabin)
	4. School, hall of residence (where this is not covered in tuition fees)
	5. Other rental accommodation (e.g. private boat in a marina, train, company cabin/apartment)
Transport, travel and accommodation	For avkryssede
	Overnatting3
	How much did you pay for...
	Selected accommodationNOK
	Selected accommodationNOK

Schooling and education	Utdanning1
	Have you paid tuition fees for anyone in the household in the last 12 months?
	1. Yes
	2. No
Schooling and education	Hvis Utdanning1=1
	Utdanning2
	Which of the following have you paid tuition fees for?
	1. Private primary or lower secondary school
	2. Upper secondary school
	3. Folk high school
	4. College or university
	5. Other (e.g. adult education, language course etc.)
Schooling and education	Utdanning3
	How much do you pay each month for ...
	If you do not pay per month, try to estimate the monthly cost.
	Avkrysset utdanningNOK
	Avkrysset utdanning.....NOK
	VarigMøbler1
	Which of the following fixtures or items of furniture have you bought or been given in the last 12 months?
	1. Sofa
	2. Garden furniture
	3. Dining table
	4. Bed
	5. Bookcase/bookshelves
	6. Lamps/lighting
	7. Other large items of furniture/furnishings
	8. None of the above
Infrequent purchases (durable goods)	Hvis VarigMøbler1 ikke er 9
	VarigMøbler2
	How much did you pay for ...
	If it was a gift, please estimate the value.
	Selected itemNOK
	Selected
Infrequent purchases (durable goods)	VarigUnderh1
	Which of the following electrical items have you bought or been given in the last 12 months?
	1. Fridge, dishwasher, oven
	2. Washing machine, tumble dryer
	3. Mobile phone

	4. TV
	5. Audio equipment, e.g. speakers, amplifiers
	6. Games console, e.g. Playstation, Xbox etc.
	7. Computer (desktop or laptop), screen, printer
	8. Camera, video camera
	9. Large electrical/motorised tools
	10. Other large or expensive electrical items or entertainment equipment
	11. None of the above
Infrequent purchases (durable goods)	Hvis VarigUnderh1 ikke er 11
	VarigUnderh2
	How much did you pay for...
	If it was a gift, please estimate the value.
	Avkrysset gjenstandNOK
	Avkrysset gjenstand.....NOK
Infrequent purchases (durable goods)	VarigFritid1
	Fritidsutstyr
	Which of the following types of leisure equipment have you bought in the last 12 months?
	1. Skis/ice skates
	2. Hunting/fishing gear
	3. Diving equipment
	4. Tent or other outdoor equipment
	5. Other expensive leisure equipment
	6. None of the above
Infrequent purchases (durable goods)	Hvis VarigFritid1 ikke er 6
	VarigFritid2
	How much did you pay for...
	If it was a gift, please estimate the value.
	Avkrysset gjenstandNOK
	Avkrysset gjenstand.....NOK
Infrequent purchases (durable goods)	VarigKunst1
	Which of the following items of value have you bought or been given in the last 12 months?
	1. Jewellery
	2. Art
	3. Bunad (Norwegian national costume)
	4. Silverware
	5. Musical instrument
	6. Other valuable items
	7. None of the above

Infrequent purchases (durable goods)	Hvis VarigKunst1 ikke er 5
	VarigKunst2
	How much did you pay for...
	If it was a gift, please estimate the value.
	Avkrysset gjenstandNOK
	Avkrysset gjenstand.....NOK
Other regular expences	FasteUtg1
	Which of the following expenses has the household had in the last month?
	Include everyone in the household.
	1. Mobile phone subscription
	2. Mobile phone subscription
	3. Cable TV and streaming services, e.g. Netflix, Disney+ and Spotify
	4. Membership fees for sports associations or gyms
	5. Other membership fees for team activities or associations (marching bands, choirs, Norwegian Trekking Association etc.)
	6. Subscriptions for newspaper, journal or other publication
	7. Child care (early learning centre, out-of-hours school care etc.)
	8. Monthly travel pass/ticket
	9. Student loan or other loan that is not secured on your home
	10. Toll charges
	11. Software subscriptions (Office 365 etc.)
	12. Alarm services
	13. Incurances
	14. None of the above
Other regular expences	FasteUtg2
	In the last month, what are the household's total expenses for
	x1 NOK
	x2 NOK
	x3 NOK
	osv.

Appendix H: Data files

Gross sample

Variables in file *SSI/NorwayGross .csv* (1 973 household):

Variabel name	Description	Values
0 initialDiaryStartDate	Initial Diary start date	
1 diaryStartDate	Actual Diary start date	
2 diaryweek	Diary week	
3 referencePeriod	First week respondents are contacted by interviewers and initial Diary week for those not recruited by interviewers	
4 grp_statusSurvey	Finished, started, Logininfo (sent), Refusal, Not able, Technical, No contact	<ul style="list-style-type: none"> Finished Logininfo Started Not able Refusal Technical No contact
5 grp_statusRecruitment	Recruited, Refusal, Not able, Technical, No contact	<ul style="list-style-type: none"> Recruited Refusal Not able Technical No contact
6 grp_statusDiary	Finished, Started, No contact	<ul style="list-style-type: none"> Finished Started No contact
7 grp_statusQuestionnaire	Finished, Started, No contact	<ul style="list-style-type: none"> Finished Started No contact
8 grp1_age	Age groups, contact person	<ul style="list-style-type: none"> 18-24 years 25-44 years 45-66 years 67-79 years
9 grp_gender	Gender, contact person	<ul style="list-style-type: none"> Men Women
10 grp_householdSize	Household Size	<ul style="list-style-type: none"> 1 person 2 persons 3 persons 4 persons 5 persons or more
11 browserName	Name of browser used for registration	<ul style="list-style-type: none"> Mobile Safari Chrome Samsung Internet Edge Safari Firefox GSA Brave MIUI Browser

Variabel name	Description	Values
12 osName	Type of device (iOS, android, Windows, Mac Os)	<ul style="list-style-type: none"> iOS Android Windows Mac OS
13 Householdtype	Type of household (6 categories, same as when drawing sample)	<ul style="list-style-type: none"> Living alone Couple with small children (youngest child 0-5 years) Couple with older children (youngest child 6-17 years) Couple without resident children Lone parent with children 0-17 years Other type of household
14 grp_countryBackground	Grouped country background, contact person	<ul style="list-style-type: none"> Norway EU/efta, USA, Canada, Australia and New Zealand, incl. UK Other countries Unspecified
15 Centralization	Index for sentralization (for the address where the household lives)	1 is highest, 6 is lowest
16 region	Region in Norway the household lives in	<ul style="list-style-type: none"> Oslo and Viken Inland Agder and South-East Norway Western Norway Trøndelag Northern Norway Unspecified
17 Quartiles	Household income divides in Quartiles (4)	<ul style="list-style-type: none"> 1 = First income quartile 2 = Second income quartile 3 = Third income quartile 4 = Forth income quartile <p>First quartile: Lowest 25 percent of income distribution. Second quartile: Second lowest 25 percent of income distribution. Third quartile: Second highest 25 percent of income distribution. Fourth quartile: Highest 25 percent of income distribution</p>
18 quartileIncome	Household income divided i Quartiles, actual numbers	
19 maritalStatus	Marital status, contact person	<ul style="list-style-type: none"> unmarried married/registered partner widow/widower/surviving partner divorced/divorced partner separated/separated partner
20 grp_educationLevel	Grouped Level of education, contact person	<ul style="list-style-type: none"> Below upper secondary school Upper secondary school Higher education Unspecified

Net sample

Variables in file *SSI/NorwayNett.csv* (446 households):

Variable name	Description	Values
0 initialDiaryStartDate	Initial Diary start date	
1 diaryStartDate	Actual Diary start date	
2 diaryweek	Diary week	
3 referencePeriod	First week respondents are contacted by interviewers and initial Diary week for those not recruited by interviewers	
4 statusSurveyDates	Date survey status is set	
5 statusQuestionnaireDates	Date questionnaire status is set	
6 statusDiaryDates	Date diary status is set	
7 statusRecruitmentDates	Data recruitment status is set	
8 grp_statusSurvey	Possible values. Finished, started, Logininfo (sent), Refusal, Not able, Technical, No contact	<ul style="list-style-type: none"> • Finished • Logininfo • Started • Not able • Refusal • Technical • No contact
9 grp_statusRecruitment	Recruited, Refusal, Not able, Technical, No contact	<ul style="list-style-type: none"> • Recruited • Refusal • Not able • Technical • No contact
10 grp_statusDiary	Finished, Started, No contact	<ul style="list-style-type: none"> • Finished • Started • No contact
11 grp_statusQuestionnaire	Finished, Started, No contact t	<ul style="list-style-type: none"> • Finished • Started • No contact
12 grp1_age	Age groups, contact person	<ul style="list-style-type: none"> • 18-24 years • 25-44 years • 45-66 years • 67-79 years
13 grp2_age	5 year age groups, contact person	<ul style="list-style-type: none"> • 18-24 years • 25-29 years • 30-34 years • 35-39 years • 40-44 years • 45-49 years • 50-54 years • 55-59 years • 60-64 years • 65-69 years • 70-74 years • 75-79 years
14 grp_gender	Gender, contact person	<ul style="list-style-type: none"> • Men • Women

Variabel name	Description	Values
15 grp_householdType	Type of household (6 categories, same as when drawing sample)	<ul style="list-style-type: none"> • Living alone • Couple with small children (youngest child 0-5 years) • Couple with older children (youngest child 6-17 years) • Couple without resident children • Lone parent with children 0-17 years • Other type of household
16 grp_householdSize	Household Size	<ul style="list-style-type: none"> • 1 person • 2 persons • 3 persons • 4 persons • 5 persons or more
17 sum_items	Number of items recorded in a household	
18 itemsManual	Number of items recorded manually in a household	
19 itemsOcr	Number of items recorded with OCR in a household	
20 receiptsManual	Number of receipts recorded manually in a household	
21 receiptsOcr	Number of receipts recorded with OCR in a household	
22 sum_receipts	Number of receipts recorded in a household	
23 browserName	Name of browser used for registration	<ul style="list-style-type: none"> • Mobile Safari • Chrome • Samsung Internet • Edge • Safari • Firefox • GSA • Brave • MIUI Browser
24 osName	Type of device (iOS, android, Windows, Mac Os)	<ul style="list-style-type: none"> • iOS • Android • Windows • Mac OS
25 sentralization	Index for sentralization (for the address where the household lives),	
26 quartile	Household income divides in Quartiles (4)	<ul style="list-style-type: none"> • 1 = First income quartile • 2 = Second income quartile • 3 = Third income quartile • 4 = Forth income quartile <p>First quartile: Lowest 25 percent of income distribution. Second quartile: Second lowest 25 percent of income distribution. Third quartile: Second highest 25 percent of income distribution. Fourth quartile: Highest 25 percent of income distribution</p>

Variabel name	Description	Values
27 decile	Household income divides in Deciles (10)	We suggest decile division: Deciles divide households in Norway into 10 equal-sized groups according to rising income Decile 1 Decile 2 Decile 3 Decile 4 Decile 5 Decile 6 Decile 7 Decile 8 Decile 9 Decile 10
28 quartileIncome	Household income divided in Quartiles, actual numbers	
29 decileIncome	Household income divided in eciles actual numbers	
30 numberOfCalls	Number of calls made to the household	
31 minutesInterview	Length of recruitment interview, minutes	
32 maritalStatus	Marital status, contact person	<ul style="list-style-type: none"> unmarried married/registered partner widow/widower/surviving partner divorced/divorced partner separated/separated partner
33 region	Region in Norway the household lives in	<ul style="list-style-type: none"> Oslo and Viken Inland Agder and South-East Norway Western Norway Trøndelag Northern Norway Unspecified
34 InitialDiaryWeek	Is the households initial Diary week changed or not, (only for CATI groups)	
35 InitialContactPerson	Is the households contact person changed or not (only for CATI groups)	
36 grp_educationLevel	Grouped level of education, contact person changed or not, (only for CATI groups)	<ul style="list-style-type: none"> Below upper secondary school Upper secondary school Higher education Unspecified
37 grp_countryBackground	Grouped country background, contact person	<ul style="list-style-type: none"> Norway EU/efta, USA, Canada, Australia and New Zealand, incl. UK Other countries Unspecified
38 resultCall1	The result of first call attempt	
39 resultCall2	The result of second call attempt	
40 resultCall3	The result of third call attempt	
41 resultCall4	The result of forth call attempt	

Appendix I: Overview qualitative data sources

Table A Overview data sources

1. Post-Q Survey (WP2.1)	n=148	Survey (4 quest.)	No	Participants WP2.1 completed	Evaluation of log on, trust, and user experience
2. Follow-up interviews participants (WP2.1)	n=8	Remote online tests/ Retrospective interview (45 min)	Yes	Participants WP2.1*: - Young adults (18-33) - Seniors (67-71)	Motivation, involvement, accuracy and trust
3. Workshop with the telephone interviewers team (WP2.1)	n=4	Worskhop (60 min)	-	-	Feedback from interivewers

*The participants of the recruitment survey were distributed across four subsamples: With or without telephone recruitment and with a "high" or "low" trust log-in solution.

In addition, we have qualitative insights on usability and attitude towards use of web application from a number of qualitative tests (n=126) during development of web app and field work for the Household Budget Survey 2022.

Documentation reports:

The following qualitative documentation reports from Statistics Norway for the SSI grant is:

1. WP2.1 Post-Q evaluation survey to participants
 - [PostQ_Survey_Questions_Eng_SSB](#)
 - [PostQ_Questionnaire_NO.docx](#)
 - [TopLine_PostQ_Evaluation_Survey.xlsx](#)
2. WP2.1 Follow-up interviewers with participants
 - [Debrief WP2.1 Followup ID Final ENG.pptx](#)
3. WP2.1 Workshop evaluation with telephone interviewer
 - [Summary Interviewers' Experiences WP2.1.docx](#)

NB! Readers outside the project group can get access to the documents by contacting the authors.

Appendix J: Non-response

Tables in this appendix, provides an overview of the response rate and reasons for non-response for sub-sample 1) CATI and 2) No CATI and the four experiments shown in table 2.1. The figures are based on the gross sample, and we have looked at how non-response varies by the contact persons gender, age, and level of education. For the sub-samples, we also look at type of household and region where the household lives.

The tables distinguish between finished and started survey. The former refers to households who have completed all parts of the survey; telephone interview (only sub-sample 1) CATI), web questionnaire, and registration of expenses for a week. In sub-sample 1) CATI, the latter consists of households who have completed the telephone interview but have not started or completed registration or web questionnaire. In sub-sample 2) No CATI, started survey refers to household who have started registration or web questionnaire, but did not finish.

Table J.1 Response rate and reasons for nonresponse by the contact person's gender, age and level of education, type of household and region. Sub-sample 1) CATI

	Finished	Started	Logininfo	Refusal	Not able	No contact	N
All	29,0	1,8	7,0	18,5	3,7	40,0	996
Gender							
Men	26,6	2,9	8,9	22,3	4,7	34,6	448
Women	31,0	0,9	5,5	15,3	2,9	44,3	548
Age							
18-24 years	27,1	0,0	8,3	2,1	4,2	58,3	48
25-44 years	27,9	2,0	8,6	15,9	2,9	42,7	452
45-66 years	31,9	2,0	5,7	21,9	2,5	35,9	401
67-79 years	23,2	1,1	4,2	24,2	12,7	34,7	95
Level of education							
Below upper secondary school	17,4	2,3	9,9	20,9	4,1	45,3	172
Upper secondary school	28,6	1,8	5,7	19,5	3,4	40,9	384
Higher education	36,9	1,6	6,2	16,1	2,4	36,9	384
Unspecified	12,7	1,8	12,7	20,0	14,5	38,2	55
Type of household							
Living alone	27,9	0,5	5,9	17,2	8,3	40,2	204
Couple with small children (youngest child 0-5 years)	31,1	2,4	11,6	17,7	1,8	35,4	164
Couple with older children (youngest child 6-17 years)	29,1	2,3	6,4	19,8	2,9	39,5	172
Couple without resident children	30,3	1,7	4,5	23,0	3,4	37,1	178
Lone parent with children (youngest child 0-17 years)	35,6	1,5	6,8	11,4	3,0	41,7	132
Other types of households	20,5	2,7	7,5	20,5	1,4	47,3	146
Region							
Oslo and Viken	31,0	1,6	7,4	24,7	4,1	31,0	364
Innlandet	30,2	3,8	5,7	30,2	3,8	26,4	53
Agder og Sør-Østlandet	25,2	3,1	7,6	10,7	3,1	50,4	131
Vestlandet	25,2	1,4	5,3	16,3	3,9	47,9	282
Trøndelag	25,2	1,4	5,3	16,3	3,9	47,9	282
Nord-Norge	38,8	1,5	11,9	11,9	3,0	32,8	67

Source: The SSI Recruitment Experiment Survey 2024. Statistics Norway

Table J.2 Response rate and reasons for nonresponse by the contact person's gender, age and level of education, type of household and region. Sub-sample 2) No CATI

	Finished	Started	Non-response	N
All	16,1	4,9	79,0	977
Gender				
Men	15,2	5,3	79,5	434
Women	16,8	4,6	78,6	543
Age				
18-24 years	15,4	7,7	76,9	39
25-44 years	15,1	5,6	79,2	443
45-66 years	17,7	4,2	78,1	407
67-79 years	13,6	3,4	83,0	88
Level of education				
Below upper secondary school	8,6	4,3	87,0	162
Upper secondary school	12,6	4,5	82,9	397
Higher education	23,0	6,2	70,8	370
Unspecified	16,7	0,0	83,3	48
Type of household				
Living alone	16,7	3,9	79,3	203
Couple with small children (youngest child 0-5 years)	13,9	8,2	77,8	158
Couple with older children (youngest child 6-17 years)	17,0	5,1	77,8	176
Couple without resident children	15,2	4,1	80,7	171
Lone parent with children 0-17 years	15,1	7,9	77,0	126
Other type of household	18,2	0,7	81,1	143
Region				
Oslo and Viken	16,3	4,9	78,9	369
Innlandet	15,4	3,1	81,5	65
Agder and Sør-Østlandet	16,2	6,9	76,9	130
Vestlandet	15,4	3,3	81,3	241
Trøndelag	16,8	6,9	76,2	101
Nord-Norge	16,9	5,6	77,5	71

Source: The SSI Recruitment Experiment Survey 2024. Statistics Norway

Table J.3 Response rate and reasons for nonresponse by the contact person's gender, age and level of education, type of household and region. Experimental group 1A) CATI and Low trust

	Finished	Started	Refusal	Not able	No contact	N
All	29,4	7,6	18,4	3,4	41,2	500
Gender						
Men	26,9	10,0	24,2	4,6	34,2	219
Women	31,3	5,7	13,9	2,5	46,6	281
Age						
18-44 years	28,4	8,0	14,4	2,4	46,8	250
45-66 years	33,0	7,7	21,6	2,1	35,6	194
67-79 years	21,4	5,4	25,0	12,5	35,7	56
Level of education						
Below upper secondary education	18,2	9,1	26,3	7,1	39,4	99
Upper secondary education	29,5	6,7	17,1	3,1	43,5	193
Higher education	34,6	7,7	15,9	1,9	39,9	208
Type of household						
Living alone	28,0	6,0	15,0	10,0	41,0	100
Couple with small children (youngest child 0-5 years)	31,7	11,0	18,3	1,2	37,8	82
Couple with older children (youngest child 6-17 years)	31,8	8,2	18,8	1,2	40,0	85
Couple without resident children	26,4	6,6	25,3	3,3	38,5	91
Lone parent with children 0-17 years	40,3	6,0	13,4	0,0	40,3	67
Other type of household	20,0	8,0	18,7	2,7	50,7	75

*Because the group 18-24 years has very few observations, we have combined it with the group 25-44 years.

* Because the group with unspecified education has few observations, we have combined it with the group Below upper secondary education.

Source: The SSI Recruitment Experiment Survey 2024. Statistics Norway

Table J.4 Response rate and reasons for nonresponse by the contact person's gender, age and level of education, type of household and region. Experimental group 1B CATI and High trust

	Finished	Started	Refusal	Not able	No contact	N
All	28,6	10,1	18,5	4,0	38,7	496
Gender						
Men	26,2	13,5	20,5	4,8	34,9	229
Women	30,7	7,1	16,9	3,4	41,9	267
Age						
18-44 years	27,2	12,8	14,8	3,6	41,6	250
45-66 years	30,9	7,7	22,2	2,9	36,2	207
67-79 years	25,6	5,1	23,1	12,8	33,3	39
Level of education						
Below upper secondary education	14,8	15,6	16,4	6,2	46,9	128
Upper secondary education	27,7	8,4	22,0	3,7	38,2	191
Higher education	39,5	7,9	16,4	2,8	33,3	177
Type of household						
Living alone	27,9	6,7	19,2	6,7	39,4	104
Couple with small children (youngest child 0-5 years)	30,5	17,1	17,1	2,4	32,9	82
Couple with older children (youngest child 6-17 years)	26,4	9,2	20,7	4,6	39,1	87
Couple without resident children	34,5	5,7	20,7	3,4	35,6	87
Lone parent with children 0-17 years	30,8	10,8	9,2	6,2	43,1	65
Other type of household	21,1	12,7	22,5	0,0	43,7	71

Source: The SSI Recruitment Experiment Survey 2024. Statistics Norway

Table J.5 Response rate and reasons for nonresponse by the contact person's gender, age and level of education, type of household and region. Experimental group 2A) No CATI and Low trust

	Finished	Started	Non-response	N
All	18,2	5,9	75,9	477
Gender				
Men	16,4	7,0	76,5	213
Women	19,7	4,9	75,4	264
Age				
18-44 years	17,2	7,4	75,4	244
45-66 years	18,6	4,3	77,1	188
67-79 years	22,2	4,4	73,3	45
Level of education				
Below upper secondary education	15,1	5,7	79,2	106
Upper secondary education	13,1	5,1	81,8	198
Higher education	26,0	6,9	67,1	173
Type of household				
Living alone	17,9	6,3	75,8	95
Couple with small children (youngest child 0-5 years)	15,0	12,5	72,5	80
Couple with older children (youngest child 6-17 years)	14,8	3,4	81,8	88
Couple without resident children	23,2	2,4	74,4	82
Lone parent with children 0-17 years	16,1	9,7	74,2	62
Other type of household	22,9	1,4	75,7	70

Source: The SSI Recruitment Experiment Survey 2024. Statistics Norway

Table J.6 Response rate and reasons for nonresponse by the contact person's gender, age and level of education type of household and region. Subgroup 2B) No CATI and High trust

	Finished	Started	Non-response	N
All	14,0	4,0	82,0	500
Gender				
Men	14,0	3,6	82,4	221
Women	14,0	4,3	81,7	279
Age				
18-44 years	13,0	4,2	82,8	238
45-66 years	16,9	4,1	79,0	219
67-79 years	4,7	2,3	93,0	43
Level of education				
Below upper secondary education	5,8	1,0	93,3	104
Upper secondary education	12,1	4,0	83,9	199
Higher education	20,3	5,6	74,1	197
Type of household				
Living alone	15,7	1,9	82,4	108
Couple with small children (youngest child 0-5 years)	12,8	3,8	83,3	78
Couple with older children (youngest child 6-17 years)	19,3	6,8	73,9	88
Couple without resident children	7,9	5,6	86,5	89
Lone parent with children 0-17 years	14,1	6,2	79,7	64
Other type of household	13,7	0,0	86,3	73

Source: The SSI Recruitment Experiment Survey 2024. Statistics Norway

Appendix K: Bias

Tables in this appendix, provides an overview of possible biases for sub-sample 1) CATI and 2) No CATI and the four experiments shown in table 2.1. The tables show the distribution in the gross and the net sample, and the deviation between them, for the following characteristics: contact persons gender, age and level of education and type of household.

Table K.1 Net sample, started, non-response and gross sample by gender, age and education level by contact person and type of household. Percent. Sub-sample 1) CATI

	Gross sample	Started	Non-response	Net-sample	Net - Gross
All	100	100	100	100	100
Gender					
Men	45,0	72,2	45,9	41,2	-3,8
Women	55,0	27,8	54,1	58,8	3,8
Age					
18-24 years	4,8	0,0	5,1	4,5	-0,3
25-44 years	45,4	50,0	46,0	43,6	-1,8
45-66 years	40,3	44,4	38,5	44,3	4,0
67-79 years	9,5	5,6	10,4	7,6	-1,9
Level of education					
Below upper secondary school	17,3	22,2	20,0	10,4	-6,9
Upper secondary school	38,6	38,9	38,8	38,1	-0,5
Higher education	38,7	33,3	34,4	49,1	10,5
Unspecified	5,5	5,6	6,8	2,4	-3,1
Type of household					
Living alone	20,5	5,6	21,2	19,7	-0,8
Couple with small children (youngest child 0-5 years)	17,3	22,2	17,1	17,3	0,0
Couple with older children (youngest child 6-17 years)	17,3	22,2	17,1	17,3	0,0
Couple without resident children	17,9	16,7	17,6	18,7	0,8
Lone parent with children 0-17 years	13,3	11,1	12,0	16,3	3,0
Other type of household	14,7	22,2	16,3	10,4	-4,3

Source: The SSI Recruitment Experiment Survey 2024. Statistics Norway

Table K.2 Net sample, started, non-response and gross sample by gender, age and education level by contact person and type of household. Percent. Sub-sample 2) No CATI

	Gross sample	Started	Non - response	Net sample	Net - Gross
All	100	100	100	100	100
Gender					
Men	44,4	47,9	44,7	42,0	-2,4
Women	55,6	52,1	55,3	58,0	2,4
Age					
18-24 years	4,0	6,2	3,9	3,8	-0,2
25-44 years	45,3	52,1	45,5	42,7	-2,7
45-66 years	41,7	35,4	41,2	45,9	4,2
67-79 years	9,0	6,2	9,5	7,6	-1,4
Level of education					
Below upper secondary school	16,6	14,6	18,3	8,9	-7,7
Upper secondary school	40,6	37,5	42,6	31,8	-8,8
Higher education	37,9	47,9	33,9	54,1	16,3
Unspecified	4,9	0,0	5,2	5,1	0,2
Type of household					
Living alone	20,8	16,7	20,9	21,7	0,9
Couple with small children (youngest child 0-5 years)	16,2	27,1	15,9	14,0	-2,2
Couple with older children (youngest child 6-17 years)	18,0	18,8	17,7	19,1	1,1
Couple without resident children	17,5	14,6	17,9	16,6	-0,9
Lone parent with children 0-17 years	12,9	20,8	12,6	12,1	-0,8
Other type of household	14,6	2,1	15,0	16,6	1,9

Source: The SSI Recruitment Experiment Survey 2024. Statistics Norway

Table K.3 Net sample, started, non-response and gross sample by gender, age and education level by contact person and type of household. Percent. Experimental group 1A) CATI and Low trust

	Gross sample	Started	Non - response	Net sample	Net - Gross
All	100	100	100	100	100
Gender					
Men	43,8	50,0	45,2	40,1	-3,7
Women	56,2	50,0	54,8	59,9	3,7
Age					
18-44	50,0	66,7	50,4	48,3	-1,7
45-66	38,8	33,3	36,9	43,5	4,7
67-79	11,2	0,0	12,7	8,2	-3,0
Level of education					
Below upper secondary education	19,8	16,7	23,1	12,2	-7,6
Upper secondary education	38,6	50,0	38,3	38,8	0,2
Higher education	41,6	33,3	38,6	49,0	7,4
Type of household					
Living alone	20,0	0,0	20,7	19,0	-1,0
Couple with small children (youngest child 0-5 years)	16,4	0,0	16,1	17,7	1,3
Couple with older children (youngest child 6-17 years)	17,0	16,7	16,4	18,4	1,4
Couple without resident children	18,2	33,3	18,7	16,3	-1,9
Lone parent with children 0-17 years	13,4	16,7	11,2	18,4	5,0
Other type of household	15,0	33,3	16,7	10,2	-4,8

Source: The SSI Recruitment Experiment Survey 2024. Statistics Norway

Table K.4 Net sample, started, non-response and gross sample by gender, age and education level by contact person and type of household. Percent. Experimental group 1B) CATI and High trust

	Gross sample	Started	Non - response	Net sample	Net - Gross
All	100	100	100	100	100
Gender					
Men	46,2	83,3	46,5	42,3	-3,9
Women	53,8	16,7	53,5	57,7	3,9
Age					
18-44	50,4	41,7	51,8	47,9	-2,5
45-66	41,7	50,0	40,1	45,1	3,3
67-79	7,9	8,3	8,2	7,0	-0,8
Level of education					
Below upper secondary education	25,8	33,3	30,7	13,4	-12,4
Upper secondary education	38,5	33,3	39,2	37,3	-1,2
Higher education	35,7	33,3	30,1	49,3	13,6
Type of household					
Living alone	21,0	8,3	21,6	20,4	-0,5
Couple with small children (youngest child 0-5 years)	16,5	33,3	15,5	17,6	1,1
Couple with older children (youngest child 6-17 years)	17,5	25,0	17,8	16,2	-1,3
Couple without resident children	17,5	8,3	16,4	21,1	3,6
Lone parent with children 0-17 years	13,1	8,3	12,9	14,1	1,0
Other type of household	14,3	16,7	15,8	10,6	-3,8

Source: The SSI Recruitment Experiment Survey 2024. Statistics Norway

Table K.5 Net sample, started, non-response and gross sample by gender, age and education level by contact person and type of household. Percent. Experimental group 2A) No CATI and Low trust

	Gross sample	Started	Non - response	Net sample	Net - Gross
All	100	100	100	100	100
Gender					
Men	44,7	53,6	45,0	40,2	-4,4
Women	55,3	46,4	55,0	59,8	4,4
Age					
18-44 years	51,2	64,3	50,8	48,3	-2,9
45-66 years	39,4	28,6	40,1	40,2	0,8
67-79 years	9,4	7,1	9,1	11,5	2,1
Level of education					
Below upper secondary school	22,2	21,4	23,2	18,4	-3,8
Upper secondary school	41,5	35,7	44,8	29,9	-11,6
Higher education	36,3	42,9	32,0	51,7	15,5
Type of household					
Living alone	19,9	21,4	19,9	19,5	-0,4
Couple with small children (youngest child 0-5 years)	16,8	35,7	16,0	13,8	-3,0
Couple with older children (youngest child 6-17 years)	18,4	10,7	19,9	14,9	-3,5
Couple without resident children	17,2	7,1	16,9	21,8	4,6
Lone parent with children 0-17 years	13,0	21,4	12,7	11,5	-1,5
Other type of household	14,7	3,6	14,6	18,4	3,7

Source: The SSI Recruitment Experiment Survey 2024. Statistics Norway

Table K.6 Net sample, started, non-response and gross sample by gender, age and education level by contact person and type of household. Percent. Experimental group 2B) No CATI and High trust

	Gross sample	Started	Non - response	Net sample	Net - Gross
All	100	100	100	100	100
Gender					
Men	44,2	40,0	44,4	44,3	0,1
Women	55,8	60,0	55,6	55,7	-0,1
Age					
18-24 years	47,6	50,0	44,4	44,3	-3,3
25-44 years	43,8	45,0	55,6	52,9	9,1
45-66 years	8,6	5,0	44,4	2,9	-5,7
67-79 years					
Level of education					
Below upper secondary school	20,8	5,0	23,7	8,6	-12,2
Upper secondary school	39,8	40,0	40,7	34,3	-5,5
Higher education	39,4	55,0	35,6	57,1	17,7
Unspecified					
Type of household					
Living alone	21,6	10,0	21,7	24,3	2,7
Couple with small children (youngest child 0-5 years)	15,6	15,0	15,9	14,3	-1,2
Couple with older children (youngest child 6-17 years)	17,6	30,0	15,9	24,3	6,7
Couple without resident children	17,8	25,0	18,8	10,0	-7,8
Lone parent with children 0-17 years	12,8	20,0	12,4	12,9	0,1
Other type of household	14,6	0,0	15,4	14,3	-0,3

Source: The SSI Recruitment Experiment Survey 2024. Statistics Norway

Field Report France

Insee

The protocol for the family budget survey consists of two visits to conduct two face-to-face interviews, each lasting approximately one hour. The objective of the survey is to reconstruct the entire budget of a household, so it is essential to complete both questionnaires for the survey to be validated, as well as to fill out a spending diary for a week between the two visits.

Given the burden of this protocol on the responding households, there is historically an attrition rate between the two visits. One of the challenges of the 2024 mode test is to measure this attrition in two distinct protocols:

- A "classic" or "all-paper" protocol, where interviews are conducted face-to-face and the spending diary is provided in paper format.
- A "multimode" protocol where part of the second interview can be completed online, and the spending diary can be filled out on paper or via a smartphone application.

This note presents the response rates and attrition between visits measured during the test for the two protocols, then provides analytical elements of these results in relation to the characteristics of the respondents, and finally proposes a global interpretation of these results to conclude on the protocol to be implemented for the 2026 collection.

1 Measurement of Response Rates and Attrition According to Protocols

Table 1 : Response rates and attrition at different stages of the 2024 experimental BDF survey, by protocol

	Number of Households			Global Response Rate			Attrition Relative to V1		
	Total e	All-Paper	Multimode	Total e	All-Paper	Multimode	Total e	All-Paper	Multimode
In the Sample	2500	831	1669						
V1 Completed	1283	427	856	51 %	51 %	51 %			
Accept V2 in V1	1190	405	785	48 %	49 %	47 %	-7 %	-5 %	-8 %
V1 + V2	1131	391	740	45 %	47 %	44 %	-12 %	-8 %	-14 %
V1 + V2 + diaries filled	1076	386	690	43 %	46 %	41 %	-16 %	-10 %	-19 %

Source: 2024 Experimental BDF Survey

1.1 Response Rate After Visit 1

There are 1283 households for which the first visit was completed, resulting in a 51% response rate for the first visit among the 2500 sampled households.

At this stage, the protocol does not differ for sub-sample 1 (all-paper protocol) and sub-sample 2 (multimode protocol). The response rate is identical between the two sub-samples.

1.2 Global Attrition After Visit 1

Attrition after Visit 1 occurs at different stages:

- **At the end of the first visit**, a variable indicates whether the household accepts the second visit. At this stage, the attrition for the entire test is 3 points: 7% of respondents to the first questionnaire refuse the second visit.
- **At the time of the second visit**, some households that initially accepted did not ultimately accept the appointment for the second visit, increasing attrition. Across the entire test, the response rate decreases by 12%.
- **Considering the presence of a diary for the household**: between the two visits, the household is supposed to have filled out its spending diary. Without it, the survey is incomplete, and the household's responses are not usable.

For this test, it was suggested to the surveyors not to present the entire protocol (2 visits and diary) until the end of the first visit, rather than at the first contact. This likely explains the high attrition rate right after the first visit, which continued to increase between the two visits.

It should also be noted that this attrition is not solely due to households refusing to continue the survey but can also be attributed to diaryistical or technical issues:

- When the first visit occurred towards the end of the test period, the surveyor did not schedule a second visit if it had to take place after the collection period. For the actual survey, it is planned that surveyors will not start first visits in the last week.
- Regarding the diary, follow-up and collection generated some incidents (diary lost in the mail or by the service provider, identifier issues, refer to the collection note?) that did not always allow a diary to be linked to its household, artificially creating attrition.

1.3 Attrition After Visit 1 According to Protocol

A comparative study of this attrition according to each protocol reveals significant differences:

- **At the end of the first visit**, attrition is -5% in the all-paper sample compared to -8% in multimode.
- **At the time of the second visit** a similar contrast is observed: 8% less for the all-paper protocol compared to 14% for the multimode protocol.
- **With the diary** the final attrition is 10% relative to Visit 1 in the classic protocol, compared to -19% for the modernized protocol.

It appears that the multimode protocol led to nearly twice as many households being discouraged from responding to the second visit compared to the all-paper protocol.

1.4 Interpretation and Understanding of Results

The only difference between the two sub-samples being the presentation of a different protocol, it can be hypothesized that the complexity of the protocol presented to the second sub-sample, due to a greater variety of tools made available to households, discouraged some households. This interpretation is reinforced by the fact that a difference in attrition is observed right after the first visit.

This result may seem disappointing and counterintuitive. Indeed, everything possible in the first "all-paper" protocol is also possible in the second protocol. The hope associated with

the multimode protocol of gaining response rate points among certain populations, and thus improving the overall collection rate, is not realized here

Another possible explanation may lie in the difficulty faced by some surveyors who were uncomfortable with the applications and had to present a tool they did not master, thus creating confusion among respondents. This effect may have been even more significant as the instructions given were to push for the use of the application among as many households as possible in the experimental survey.

To better understand the effect on responding households, it is necessary to analyze which respondents were particularly discouraged.

2 Determinants of Attrition Between the Two Visits

To interpret these differences, we analyze the percentage of attrition by sub-sample according to some descriptive variables of the household. The attrition observed at the time of the second visit is preferred over the two other measures of attrition presented in Table 1 (end of the first visit / total attrition including the absence of a diary):

- It is more significant than the attrition at the end of Visit 1 and allows for more contrasted results by household characteristics.
- The absence of a diary may not be the fault of the household.

We present results for age, education level, main employment situation, and frequency of computer use. Note that the sample sizes are sometimes small, particularly in sub-sample 1, and these results should therefore be interpreted with caution.

Table 2 : Effectives and Distribution of Respondents to the First Visit According to Their Characteristics and Sub-Sample

		Sub-Sample 1		Sub-Sample 2	
		respondents V1	% of total	respondents V1	% of total
Age	Under 40	104	24 %	184	22 %
	40 - 50	79	19 %	171	20 %
	50 – 60 ans	104	24 %	201	24 %
	60 and over	140	33 %	299	35 %
Education Level	Middle School	75	18 %	152	18 %
	High school	247	58 %	454	53 %
	Bachelor's Degree	37	9 %	103	12 %
	Master's Degree and above	67	16 %	144	17 %
Employment Status	Unemployed	22	18 %	74	18 %
	Employed	270	58 %	518	53 %
	Inactive	133	9 %	263	12 %
Frequency of Computer Use	Daily	355	83%	741	87%
	Less than daily	72	17%	112	13%

Regardless of the protocol, attrition decreases with age, confirming a classic result of participation rates in surveys. However, it is higher in the multimode protocol regardless of age group: 5 points more for those under 40, 40-50, and 60 and over. The gap is slightly

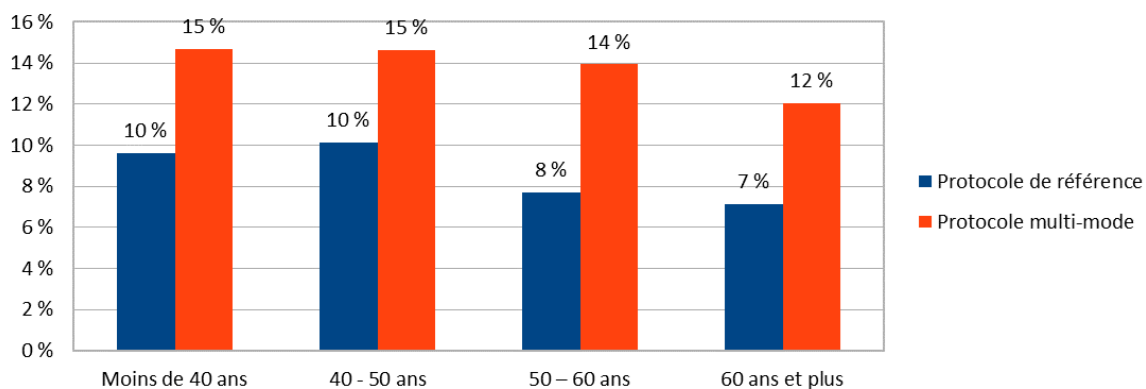
larger for those aged 50-60 (6 points), but age does not seem to be a determining factor in the discouragement for the second visit related to the multimode protocol.

Unemployed respondents have the highest attrition rate between the two visits compared to those who are employed or inactive. They are also more strongly affected by the increase in attrition in the multimode protocol (+10 points). This is the group of respondents with the highest attrition rate among all the characteristics studied (24% in multimode).

Attrition by education level gives more difficult-to-interpret results: it increases from middle school to bachelor's degree level and decreases for respondents with a master's degree and above. The gap between the two protocols is greatest for those with a master's degree and above.

Finally, the frequency of internet use also seems to be correlated with attrition: in the all-paper protocol, it is lower for non-daily internet users (small sample sizes), but it is higher in the multimode protocol, with 15% of these users not responding to the second visit.

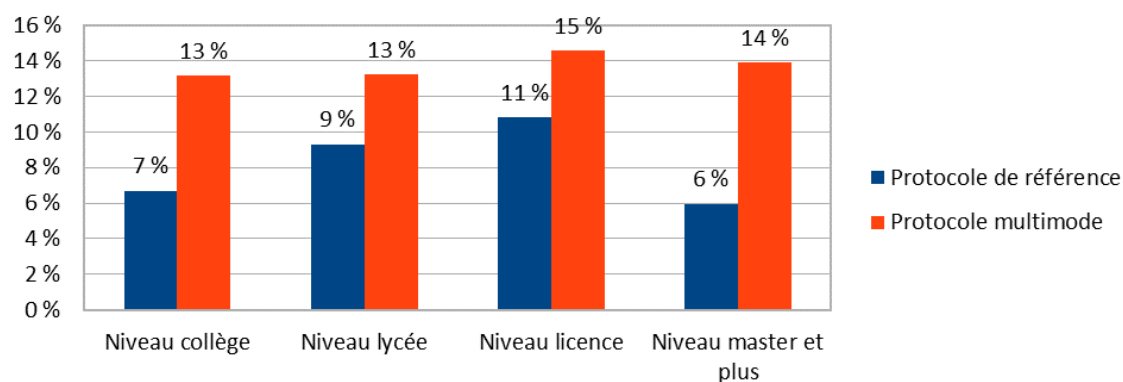
Figure 1 : Attrition Rates by Age and Protocol



Lecture : among people who responded to Visit 1 and are under 40, 10% did not complete the second visit in the reference protocol, and 15% in the multimode protocol.

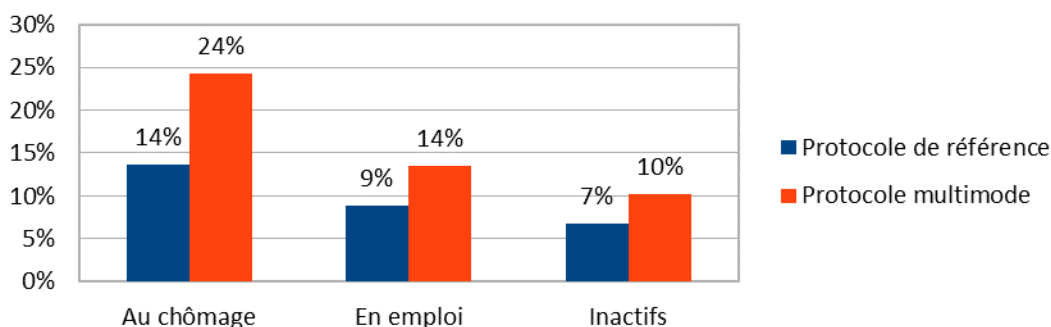
Source: 2024 Experimental BDF Survey

Figure 2 : Attrition Rates by Education Level and Protocol



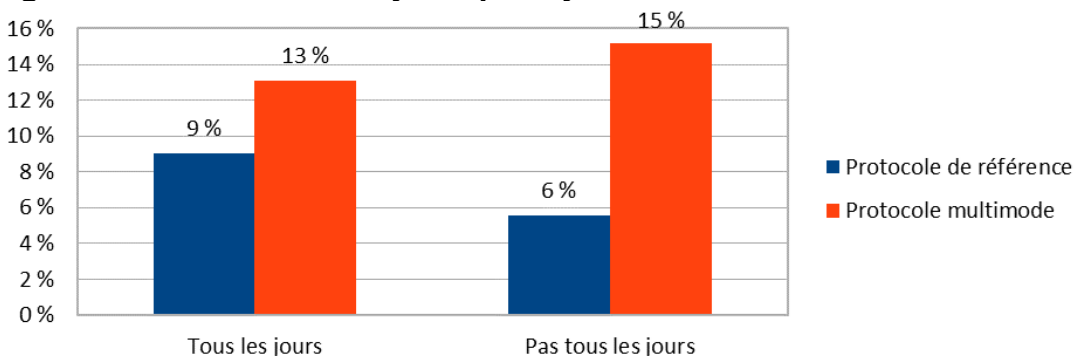
Source: 2024 Experimental BDF Survey

Figure 3 : Attrition Rates by Education Level and Protocol



Source: 2024 Experimental BDF Survey

Figure 4 : Attrition Rates by Frequency of Internet Use and Protocol*



A logistic regression on the determinants of attrition within each sub-sample does not, however, reveal many significant characteristics, except for being unemployed.

3 Consequences for the 2026 Protocol

This test has identified avenues for improving the protocol for the 2026 survey to limit attrition between the two visits.

Option 1: Better Inform Households About the Complete Protocol

Rather than announcing it at the end of Visit 1, the surveyor could present the entire protocol during their first contact with the household.

Benefit: This could potentially reduce the number of refusals for the second visit and thus the time wasted by surveyors in collecting the first visit questionnaire that we cannot exploit. Furthermore, the approach is more ethical, and the surveyor does not feel like they are deceiving the household.

Risk: This could, however, reduce the number of households accepting the first visit and not allow the surveyor to convince a household already well-engaged in the survey to continue.

Option 2: Better Train Surveyors in the Multimode Protocol

The collection tools (questionnaires and application) were finalized shortly before the test, and the accompanying materials for surveyors can be revised. In particular, a video could help surveyors show households how the application works.

Option 3: Better Target the Proposal of New Tools

The new collection tools were introduced for several objectives:

- Improve the quality of the survey (application that directly retrieves the correct data, already coded, and avoids processing losses and errors).
- Reduce the costs of the survey (web questionnaire that reduces the time for the surveyor, application that reduces the need for data entry services for purchased products).
- Make the survey easier for respondents by offering tools more suited to their usage, thus improving quality.
- Gain more respondents, particularly among younger people who often respond less.

With a 42% response rate to the web questionnaire and 42% use of the application, the objectives of improving quality and reducing costs are achieved. When the tools are used, the surveyor's workload is also reduced, but this may have come at the cost of complexity in presenting the protocol and tools, sometimes to households that were not suited to use them.

However, it was not anticipated that these new tools would instead decrease the number of respondents.

It therefore seems necessary to adapt the protocol by limiting the presentation of the tools to households.

Proposed New Protocol:

- The web questionnaire continues to be offered to all respondents (short presentation time, less discriminatory internet use than a smartphone application).
- The application is not offered to everyone; it is filtered based on a question. When it is presented, the paper diary is only provided "in case of technical issues."

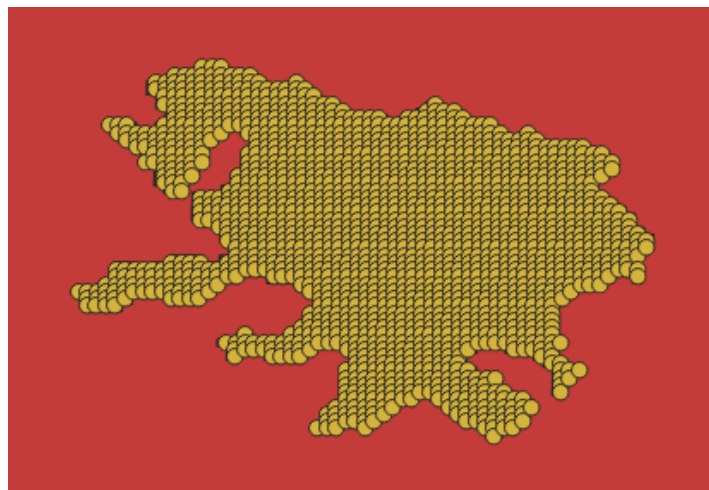
Diagram of the Protocol for the 2025 Test, and if Conclusive, for the 2026 Survey*

Appendix B (Machine learning)

Quality Analysis Report: Comparison between Google Places and Open Street Map (OSM)¹

The primary goal of this explorative analysis was to compare the quality and coverage of Points of Interest (POI) between Google Places (GP) and Open Street Map (OSM). The adopted methodology involved the random selection of a significant number of points (similar to throw arrows) across urban areas with populations exceeding 250,000 inhabitants, in six countries involved in the ESSnet Smart Surveys Implementation (Netherlands, Belgium, Italy, Germany, France, Norway). Urban areas' choice has been adopted to exclude uninhabited areas like seas, lakes, and mountains.

Data collection was limited to a 50-meter radius around each selected point. The points were distributed using a 1kms grid overlay on a GIS map, ensuring uniform coverage of targeted urban areas and appropriate allocation of points across larger regions. The grid adoption guarantees to allocate several points to the larger urban areas, as it happens in the following picture regarding Paris:



In total, 1220 arrows were analyzed. POIs were collected within a 50-meter. At the end of the analysis, Google Places yielded a total of 4563 POIs, while Open Street Map identified 916. Beyond this macro comparison, it's crucial to examine the specific types of POIs collected from both providers.

In OSM, to make the data comparable with those in GP, only POIs with specific tags were considered, specifically:

- amenity (excluding: waste, telephone, recycling)
- office
- shop
- healthcare

¹ Authors of this Appendix are Fabrizio De Fausti, Francesco Piccolo and Marco D. Terribili

- land use
- craft
- tourism
- leisure

The excluded POI Tags in OSM were:

- amenity=waste: Locations for public waste disposal (e.g., trash bins, waste containers)
- amenity=telephone: Public telephones.
- amenity=recycling: Facilities or bins for recycling various materials.
- barrier: Physical obstacles or boundaries such as fences, walls, or gates.
- highway: Roads, streets, and paths for vehicle and pedestrian traffic.
- bus stop: Locations where buses pick up or drop off passengers.
- power: Infrastructure related to electrical power, such as lines, substations, or plants.
- natural: Natural features like trees, rivers, beaches, or cliffs.
- man_made: Artificial structures not covered by other categories, e.g., piers, towers, or storage tanks.
- emergency: Emergency-related features like defibrillators, fire hydrants, or emergency access points.
- addr:housenumber: The specific house or building number in an address.
- addr:city: The city component of an address.
- noexit: Indicates a dead-end or cul-de-sac; no way through at that point.

The excluded POI Types in GP were:

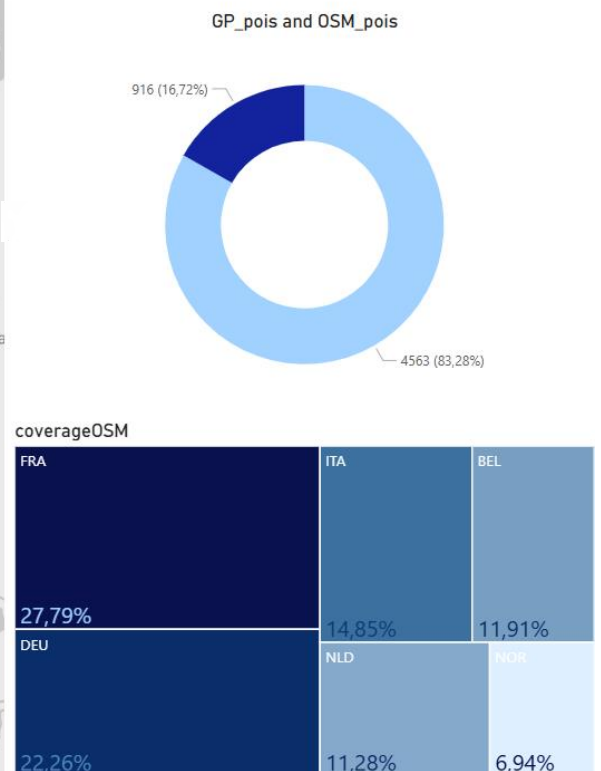
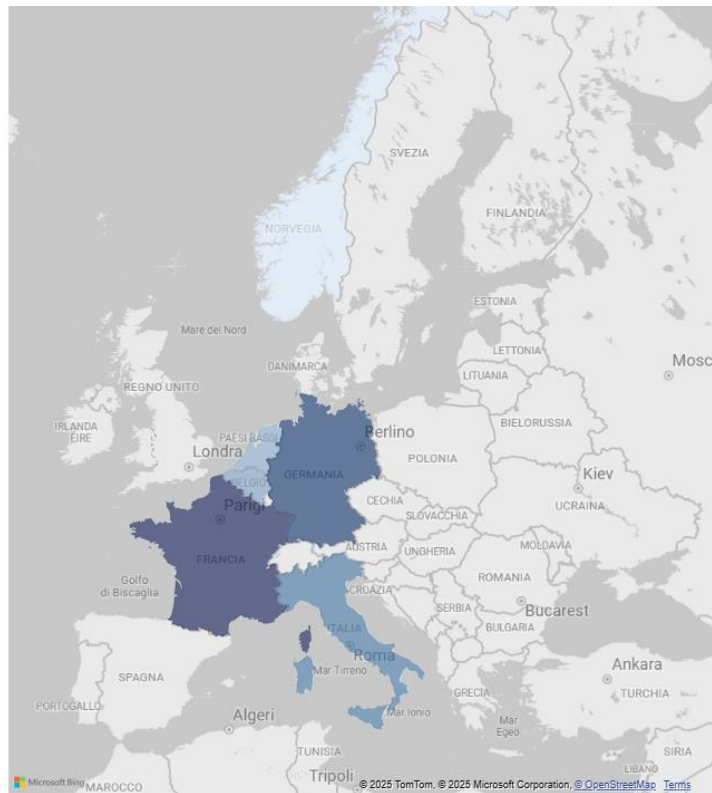
- route: A named road, street, or highway.
- political: A generic type indicating administrative or political boundaries (used with other types like locality or sublocality).
- locality: A city or town-level geographic area.
- neighborhood: A commonly known local area within a city or town.
- sublocality_level_1: A subdivision of a locality, such as a district or borough.
- sublocality: A general term for subdivisions within a city

It was noted that despite the lower number of POIs in OSM, some of these (such as rest benches or public fountains) might not be directly relevant for comparison with Google Places POIs.

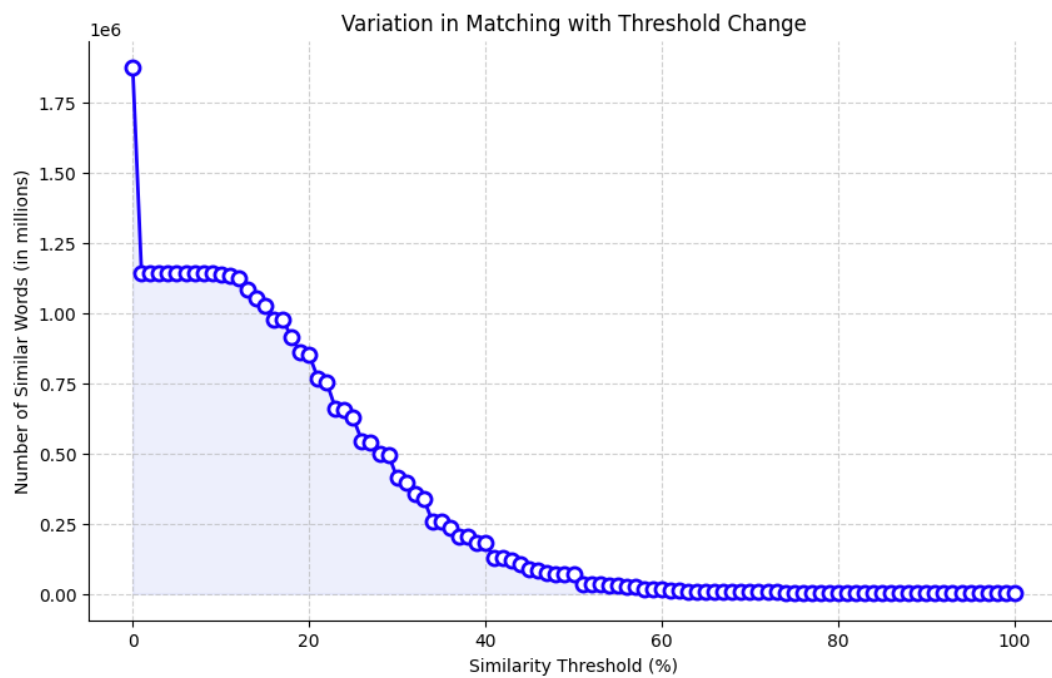
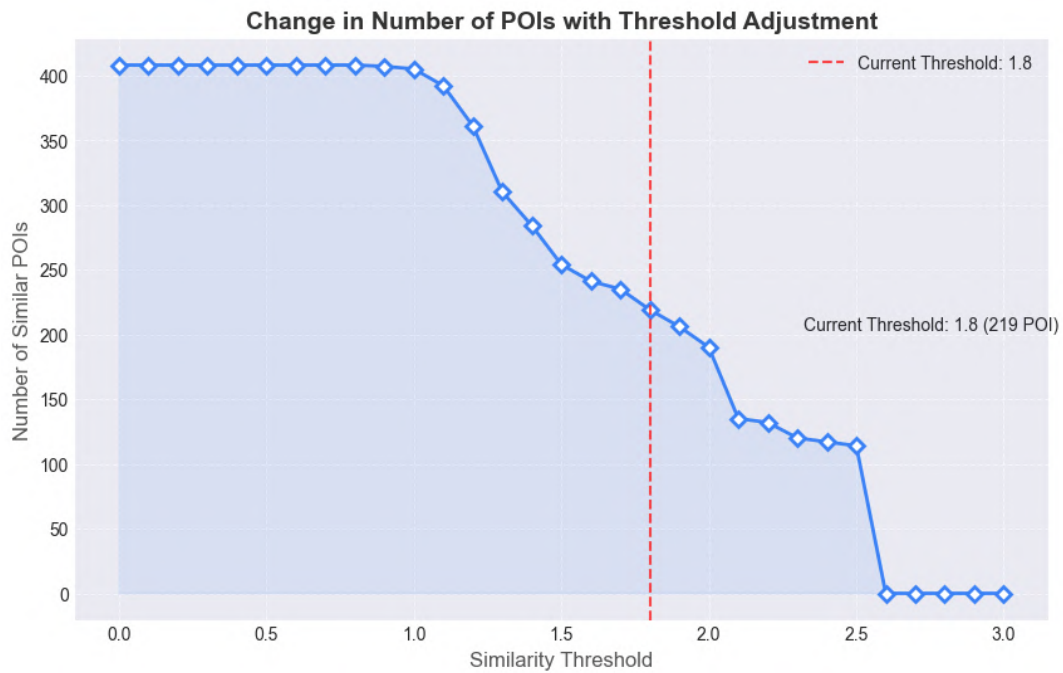
POI coverage varies significantly across countries.

WB_A3	arrows	GP_pois	OSM_pois	countmerged exact	countmerged fuzzy	coverageOSM
DEU	458	1689	376	56	92	22,26%

FRA	297	1119	311	45	50	27,79%
ITA	233	916	136	27	44	14,85%
NLD	129	532	60	18	26	11,28%
BEL	76	235	28	3	5	11,91%
NOR	27	72	5	1	2	6,94%
Total	1220	4563	916	150	219	



An interesting observation is that OSM POIs are not a subset, even partially, of those found in Google Places. Through the record linkage process, only 150 out of 916 POIs could be matched between the two sources, increasing to 219 when relaxing the linkage criteria. The system employs Levenshtein distance-based matching with a 75% similarity threshold as the cut-off point. When a linkage falls below this 75% threshold for either name similarity or category similarity, the POI pair is excluded from consideration. After applying this trimming strategy, the number of POIs considered similar becomes 177. The matching process employs a weighted scoring system where category matches receive greater weight than name matches. This weighting scheme reflects the relative importance of these attributes in determining whether two POIs represent the same physical location. Despite efforts to ensure comparability between OSM tags and GP labels, the analysis highlights significant differences in POI representation between the two platforms. This finding underscores the challenges in integrating location data from different sources, even when they ostensibly cover the same geographic areas.



The results of this analysis, although exploratory and conducted at a macro level, underscore significant quantitative and qualitative differences between Google Places and Open Street Map. The higher density of POIs in Google Places suggests broader coverage and/or a greater propensity of users or businesses to register their data on this platform.

Differences in coverage across the countries involved in the ESSnet project and the unexpectedly low rate of record linkage between the two platforms indicate that, despite some standardization efforts, there is still a significant divergence in POI datasets between Google Places and OSM. This

suggests that combining or comparing different POI sources may be necessary for a comprehensive understanding of POI distribution in urban areas.

Appendix C: Human Computer Interaction and Usability country reports

This Appendix contains the following country reports on small scale usability testing of the Receipt Scanning Microservice (RSM) integration in Household Budget Surveys (HBS) tests, the HBS benchmark tests, the HBS simulation test, and the GeoService Microservice (GSM) integration in Time Use Surveys (TUS) tests and Mobility Survey (MS).*

1. Destatis – Country report on RSM in HBS with MOTUS
2. VUB – Country report on RSM in HBS with MOTYS
3. SSB – Country report on receipt scanning in Household Budget App
4. Insee – Country report on receipt photographing in @HBS App
5. CBS – Country report on receipt scanning simulation in @HBS2.0 App
6. Destatis – Country report on GSM in TUS with MOTUS
7. Istat – Country report on GSM in TUS with MOTUS

**Note that the results from the VUB/Statbel test on GSM in MS with MOTUS have been integrated directly in the final deliverable because of limited time and because of the VUB being the task leader integrating all country findings into one report.*

Results small scale usability test of the scan function

Smart Survey Implementation – WP 2.3

07/2024



Disclaimer: Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Eurostat. Neither the European Union nor the granting authority can be held responsible for them.



Management Summary (1/2)

Background

Smart Surveys are surveys conducted with the help of mobile devices and combine questions for self-report and **smart features for collecting data**. On behalf of *Eurostat* the international project *Smart Survey Implementation (SSI)* is working on the development of different smart features. One of them is a **receipt scanner using optical character recognition** for automatic recording of receipts. The aim is to improve data quality and reduce participant burden.

Methodology

A **usability test** was carried out in **July 2024**. The test object was an app with the receipt scanner as smart feature which was not yet fully trained at the time of the test. **19 participants tested the app** on site at the Federal Statistical Office in the presence of an interviewer.

Research Questions

ACCEPTANCE

Are users willing to use a smart feature to record their receipts?

USABILITY

How easy is the use of the receipt scanner as part of an app?

Management Summary (2/2)

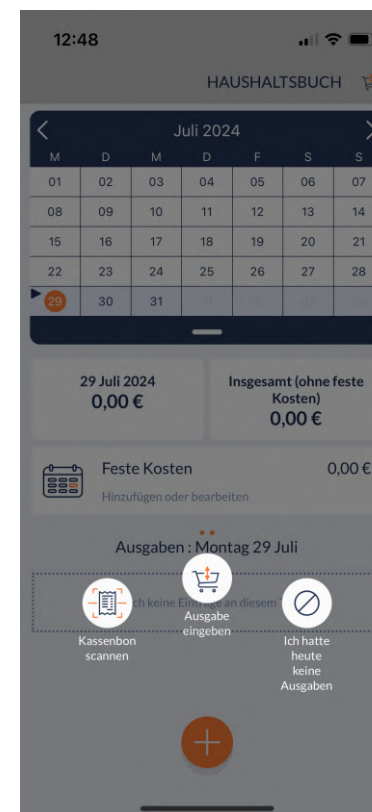
Results

Acceptance – Are users willing to use a smart feature to record their receipts?

The scan function as a smart feature is used and **accepted** by the participants. In general they appreciate the **easy** and **fast** way to record receipts, especially compared to the existing option of the manual input. In Germany a scan function is used by some apps and therefore is perceived as **common**. The app is perceived as **trustworthy** by its professional look and known publisher. It doesn't rise trust or privacy questions by the participants.

Usability – How easy is the use of the receipt scanner as part of an app?

The scan function **can be used** by all participants, but there are some **limitations in terms of user-friendliness**. For example, the term “scan” gets misinterpreted, the click path to the actual scan is long and not always clear, the feedback after scanning takes too long. Poor quality of the scanning results entail a high correction effort. This causes **confusion, dissatisfaction** and sometimes **resignation** among the participants.



Screenshot – Selection options for entering expenses

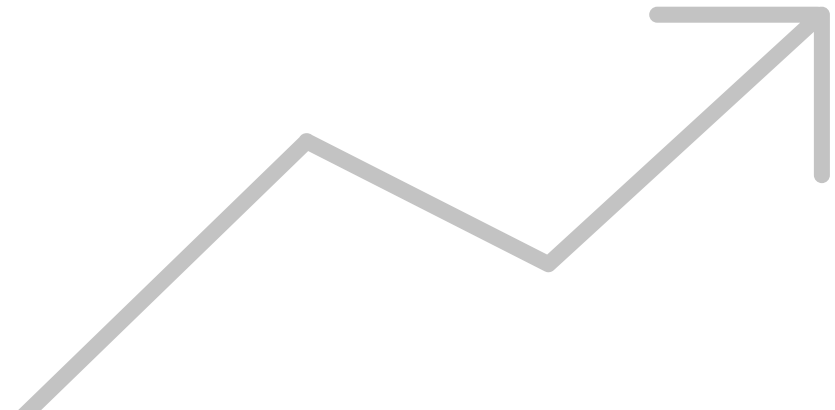


Screenshot – Taking a photo within the app for OCR

Table of Contents

02	Management summary	59	Results in terms of further aspects
05	Conclusion	60	Manual input
12	Background & objectives	71	Overview of expenses
15	Methodology & participants	77	General design
25	Results in terms of scan function	83	Trust & privacy
26	Decision to scan	77	Installation
32	Path to scan	95	Bugs
37	Processing of scan	98	Contact
45	Results of scan		
48	Correction of scan		

Conclusion



Scan function

The **scan function** appears to be an **accepted and even the preferred input option** because it is common, fast and easy. But the **current state of development** is not satisfactory due to a high error-proneness of the scan results.

What needs to be considered?

- » Option to scan a ticket might not be found because the term „scan“ is not associated with „taking a photo“
- » Click path till the point where the photo is taken is long and not always clear
- » Time between taking the photo and receiving positive feedback is too long

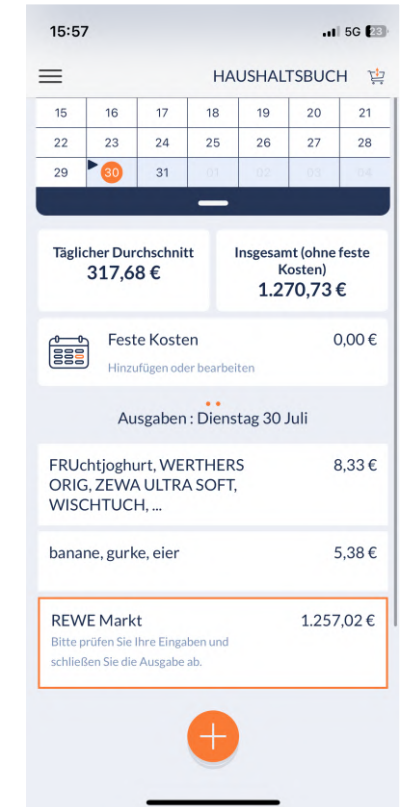


Correction of scan results

There is a high **confidence in technology** and accuracy of the scan results. An incorrectly recognized total sum is immediately noticed, while concrete **indications that draw attention to a necessary correction are being overlooked.**

What needs to be considered?

- » No one likes to correct, and no one accepts such a cumbersome and lengthy correction
- » A sense of duty and correctness make participants correct major errors
- » Visual indications get overlooked easily; textual indications are not concrete enough
- » The necessity of manually adding COICOPs is not understood and not perceived as relevant



Manual input

Especially for **short tickets**, the **manual input** is still attractive. However, the user interface is not explanatory enough and it remains unclear at what **level of detail** the ticket has to be entered.

What needs to be considered?

- » It is not always clear to participants at what point in the manual input process they are
- » Getting started and adding the name and category of an expense item is not smooth
- » The level of detail of the input is not clearly highlighted, e.g. whether the total expense, a single expense item or categories of expense items have to be entered

15:55 5G

NEUE AUSGABE

Ausgabe
banane

Kategorie
Datteln, Feigen und tropische Früchte - frisch oder gekühlt

Anzahl 1 Preis 1,99 €

Ich habe den Artikel zurückgegeben (Reklamation o. ä.) ☐

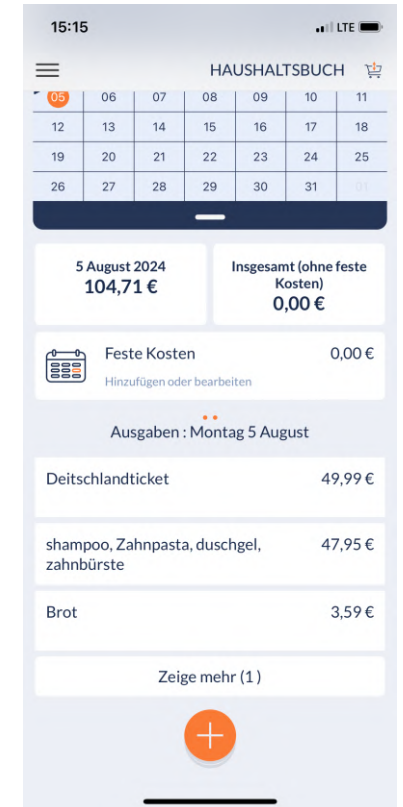
Gesamtpreis: 1,99 € AKTUALISIEREN

Expense diary

The “Haushaltsbuch” (diary) collects all important information and offers a **clear starting point to the input options**. Sometimes, the design of the **user interface should fit user needs better**.

What needs to be considered?

- » Calendar only has few functions and takes up a lot of space
- » Latest tickets are sorted below older ones and therefore are sometimes hidden
- » Not always clear that scrolling is possible and overview of expenses is further below



Trust & privacy

The app is perceived as trustworthy. It has a professional look and a known publisher. Applying the scan function does not raise any data protection issues either.

What needs to be considered?

- » It is not shown at all who is behind the app
- » Data protection declarations are accepted as necessary and are accepted without looking at them
- » No advertising is essential

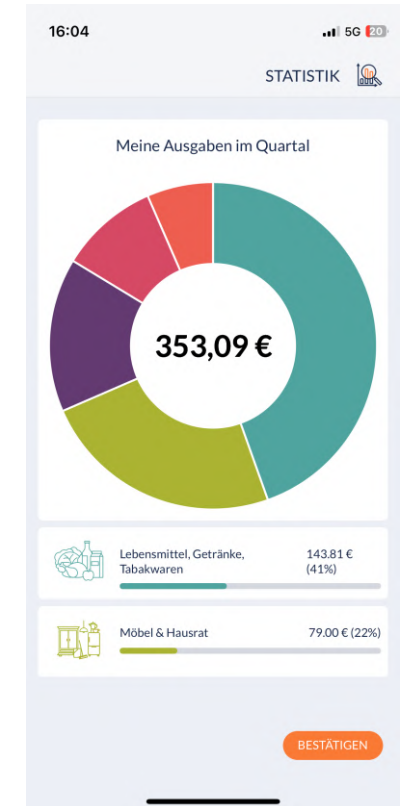


Willingness & motivation to participate

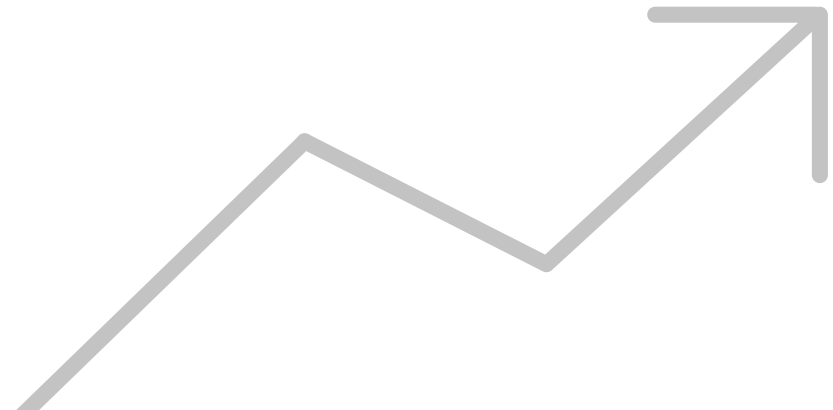
At the current level of technical maturity, the **required effort of adding expenses in everyday life would be too high** for many to participate. However, **getting feedback** on one's own expenses is appreciated.

What needs to be considered?

- » The diagram could be enriched with more information
- » Technical maturity and ease of use is key to keep participants on board and thus for the success of the studies and quality of the data

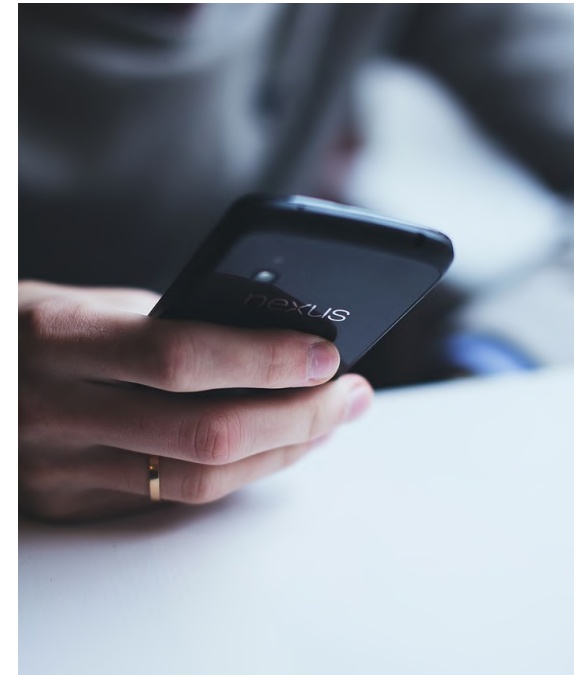


Background & objectives



Smart Survey Implementation (SSI) Project

- » Was announced and is mainly financed by the **European Union**
- » Smart Survey = surveys that combine **smart sensor data and traditional survey questionnaire**
- » Main goals (besides others) are to **develop and test shared smart microservices** for statistical production systems and optimize push-to-smart recruitment and motivation strategies
- » Its project consortium includes **11 project partners**
- » Runs from May 2023 – April 2025
- » Is organised in 5 workpackages

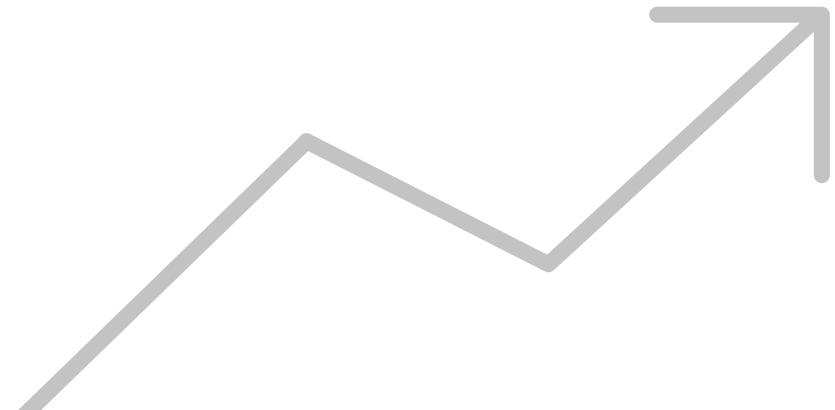


Symbolic image:
<https://pixabay.com/de/photos/smartphone-technologie-handy-lg-1281632/>

Usability test within SSI

- » Goal of the SSI project is the development of a microservice which can be integrated in apps to **scan receipts with OCR**, read its expense items and add them to a expense diary
- » This microservice should in second step also be able to **automatically categorise all expense items to the correct COICOP** product classification
- » It is planned to use this microservice for the **European Household Budget Survey** (HBS)
- » Main goal of this usability test was to test a first version of the OCR microservice

Methodology & participants



Qualitative usability interviews

- » 20 individual interviews planned, **19 conducted**
- » Carried out **inhouse** in Federal Statistical Office (Wiesbaden, Germany)
- » Moderated by an interviewer and supported by another person in the room next door
 - » Role of moderator and supporter alternating between four different people
 - » Supporter responsible for recording technology and logging
- » Time frame: 90 minutes
- » Incentive: 40,00 EUR
- » Survey period: 16th to 26th of July 2024

Test set up

- » Interviewer and participant sit opposite to each other at a table
- » Smartphone used by the participant is connected with the monitor in terms of better traceability for the interviewer and recording
- » Recording of image via monitor and sound via microphone in the middle of the table

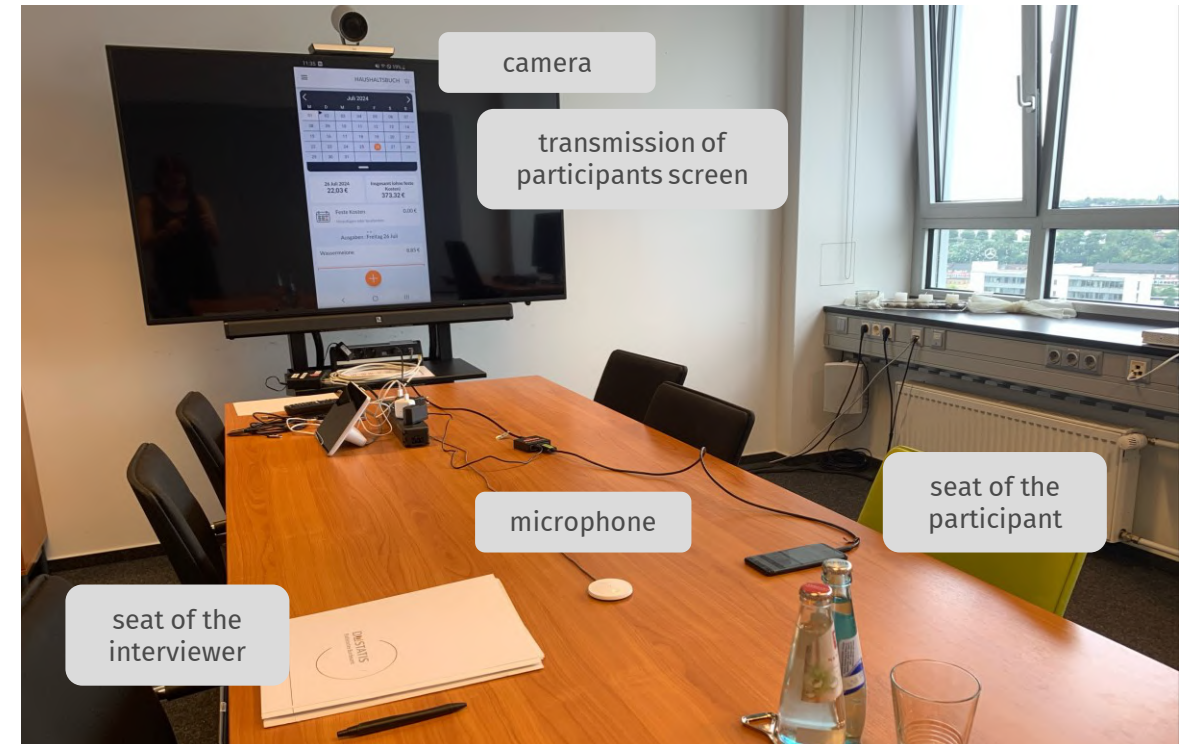
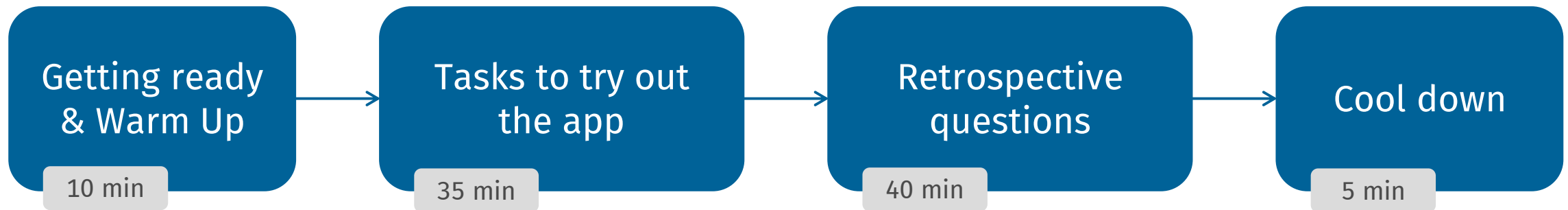
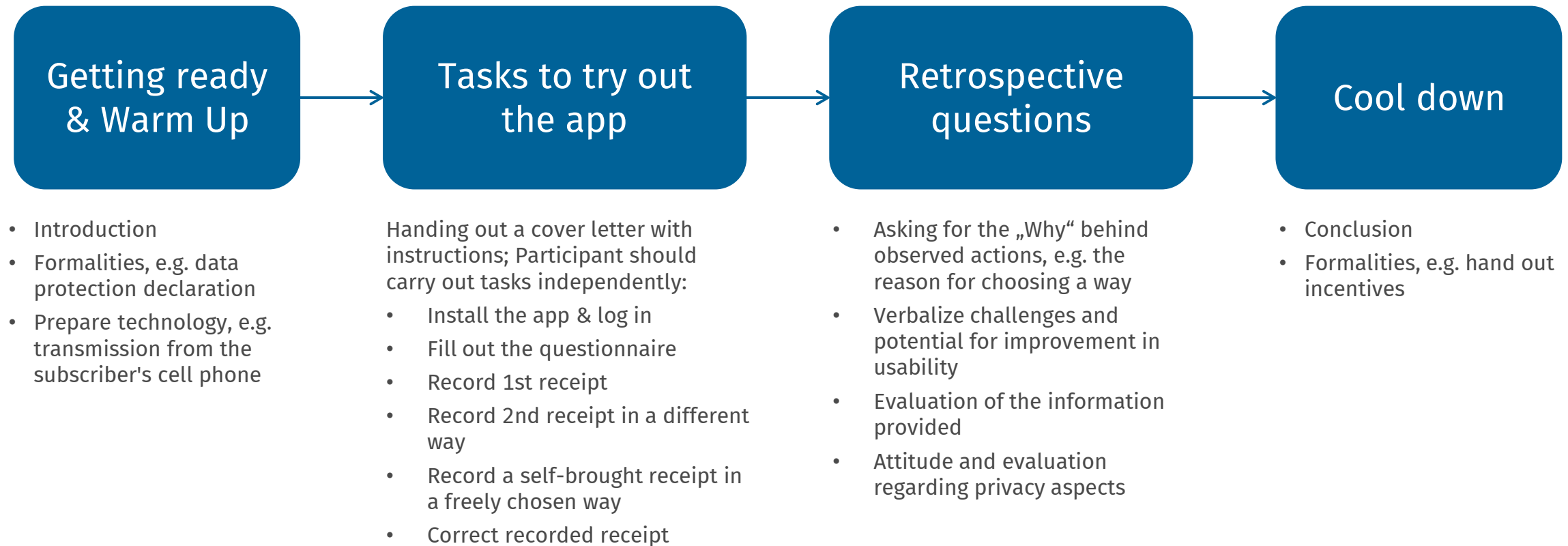


Photo of the test setup

Procedure of the interviews



Procedure of the interviews in detail



Aquisition of participants

Two different paths for acquisition

- » Recruitment from **own existing pool**
- » Recruitment via **advertisement at *Kleinanzeigen*** (second hand sales platform)

Overview of participants (1/2)

Participant	Gender	Age	Education*
TP01	Male	66	Higher
TP02	Female	68	Medium
TP03	Female	19	Student
TP04	Female	56	Medium
TP05	Male	36	Medium
TP06	Male	30	Higher
TP07	Male	60	Higher
TP08	Female	28	Medium
TP09	Female	42	Medium
TP10	Male	18	Student

Participant	Gender	Age	Education*
TP11	Male	69	Medium
TP12	Female	33	Medium
TP13	Female	66	Medium
TP14	Male	60	Medium
TP15	Female	59	Medium
TP16	Male	36	Higher
TP18	Female	32	Medium
TP19	Female	21	Student
TP20	Male	20	Student

*Classification of the educational level
Higher = University degree
Medium = Graduation, vocational training
Lower = 9 years of school, no graduation
Student = Education not finished yet

Overview of participants (2/2)

Participant	Gender	Age	Education*	Version**	Phone***
TP01	Male	66	Higher	2	Personal
TP02	Female	68	Medium	2	Destatis
TP03	Female	19	Student	2	Personal
TP04	Female	56	Medium	1	Destatis
TP05	Male	36	Medium	1	Personal
TP06	Male	30	Higher	1	Destatis
TP07	Male	60	Higher	1	Destatis
TP08	Female	28	Medium	1	Personal
TP09	Female	42	Medium	1	Personal
TP10	Male	18	Student	1	Personal

Participant	Gender	Age	Education*	Version**	Phone***
TP11	Male	69	Medium	2	Personal
TP12	Female	33	Medium	2	Personal
TP13	Female	66	Medium	2	Destatis
TP14	Male	60	Medium	2	Destatis
TP15	Female	59	Medium	2	Destatis
TP16	Male	36	Higher	2	Personal
TP18	Female	32	Medium	2	Personal
TP19	Female	21	Student	2	Destatis
TP20	Male	20	Student	2	Destatis

*Classification of the educational level
 Higher = University degree
 Medium = Graduation, vocational training
 Lower = 9 years of school, no graduation
 Student = Education not finished yet

**Version of the App
 During the test, fields in the manual input were removed and the results of the scan function improved through training data.

*** Used phone in the test
 When the phone of a participant was not compatible with our system for recording or was not charged sufficiently, we offered a Samsung or iPhone from Destatis

Distribution of acquisition criteria

Age distribution

as targeted

Age range	<26	26-49	50-65	>65
Number of participants	4	7	4	4

Gender distribution

slightly more female participants

Gender	Male	Female
Number of participants	9	10

Education distribution

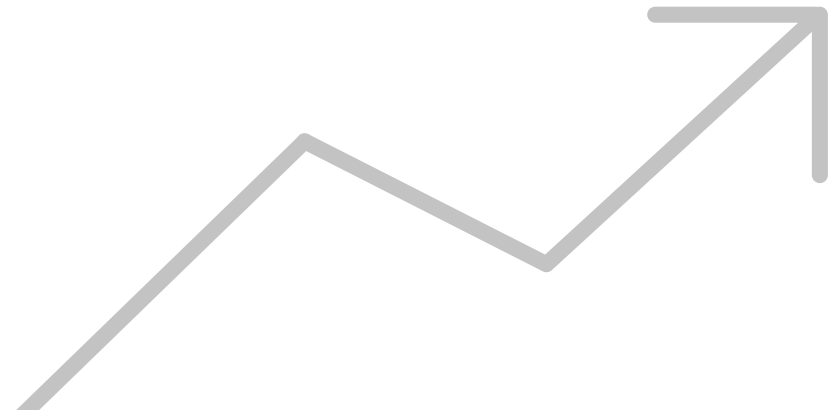
no target set

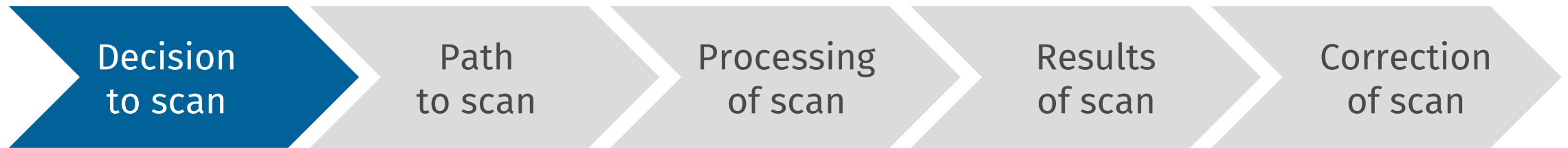
Educational level	Student	Low	Medium	Higher
Number of participants	4	0	11	4

Reading and interpretation notes

- » Qualitative Research considers small samples and therefore we do not provide results in the form of percentages. Instead we use the terms **occasionally/few, sometimes/some** and **often/many** to give an orientation about the frequency of observed behaviors and evaluations.
- » Qualitative Research does not only show results, it also **asks for the “Why”** underlying a result and therefore **gives a better understanding** of the processes leading to a result.
- » Quotations are used for a more visual presentation of the results. The abbreviation of a test person (e.g. TP01) is used to indicate who made a statement.

Results in terms of scan function





Click path to scan a ticket

Choosing the scan option

Choosing „Image“

Choosing „Take Picture“

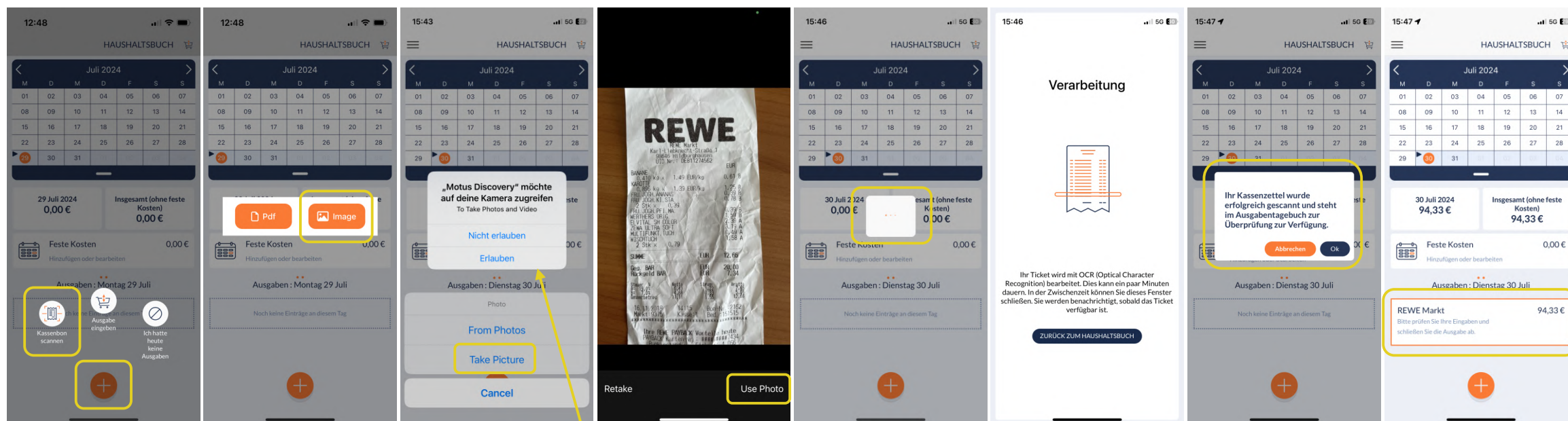
Taking a picture

Loading animation

Processing in progress

Processing completed

Scanned ticket visible



In the first attempt, the respondent is asked to allow access to the camera and the photo library

Preferred option is scan (1/2)

» Theoretically, all participants **would prefer the scan option** to the manual input – but only theoretically, as the current state of development is not satisfactory

- » Handling is (potentially) faster
- » Handling is easier (if result is correct)
- » Participants are not willing to type in all the information
- » Participants are used to scan functions

"I'm used to scan things from other applications, which are then processed." (TP05, m36)

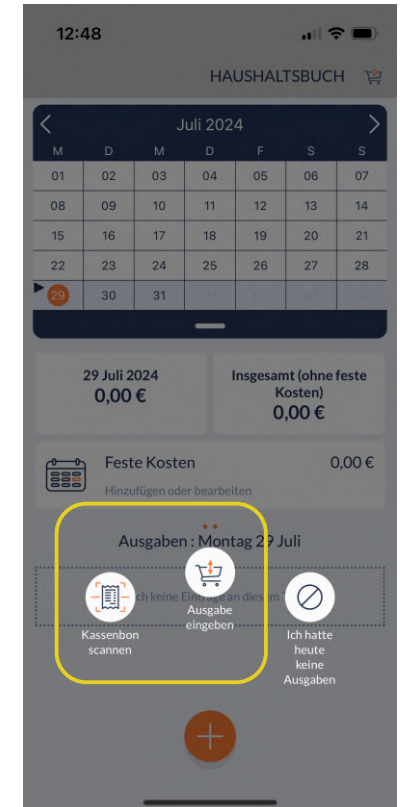
*"I'm happier with scanning because **you don't have to enter so much.**" (TP19, f21)*

*"In any case, **scanning is better** if it doesn't take forever and doesn't make too many mistakes." analogous quote, (TP01, m66)*

"I'm lazy." (TP06, m30)

Preferred option is scan (2/2)

- » Many criticize the **current state of development** of the scan function
 - » Many errors in scan result compared to the original ticket
 - » Resulting in a high effort to correct → See Chapter „Correction of scan“
- » **Manual input** is still attractive **in few situations**
 - » Sometimes, when the ticket is short (one to three items)
 - » Occasionally, when the ticket is hand written



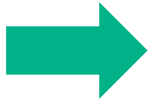
Scan not always considered an option in first attempt

- » In the first attempt, **half of the participants used the scan** option and the other half used the manual input
- » Reasons **why the scan was not used** in first attempt
 - » Personal inhibition threshold, missing experience
 - » Misinterpretation of the term „scan“ – participants assume that a barcode or QR-code is necessary

*“Because I’m used to it (manual input). Because I’m old. Because I **don’t have that much experience** with scanning.” (TP13, f66)*

*“Because I **didn’t realise** that you can scan without a barcode.” (TP07, m60)*

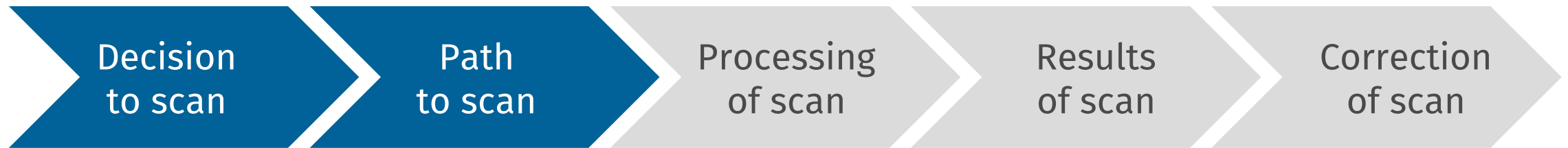
Our recommendation



Do not use the term “Kassenbon scannen”* but “Kassenbon fotografieren”**
→ This gives users a clear understanding of the option

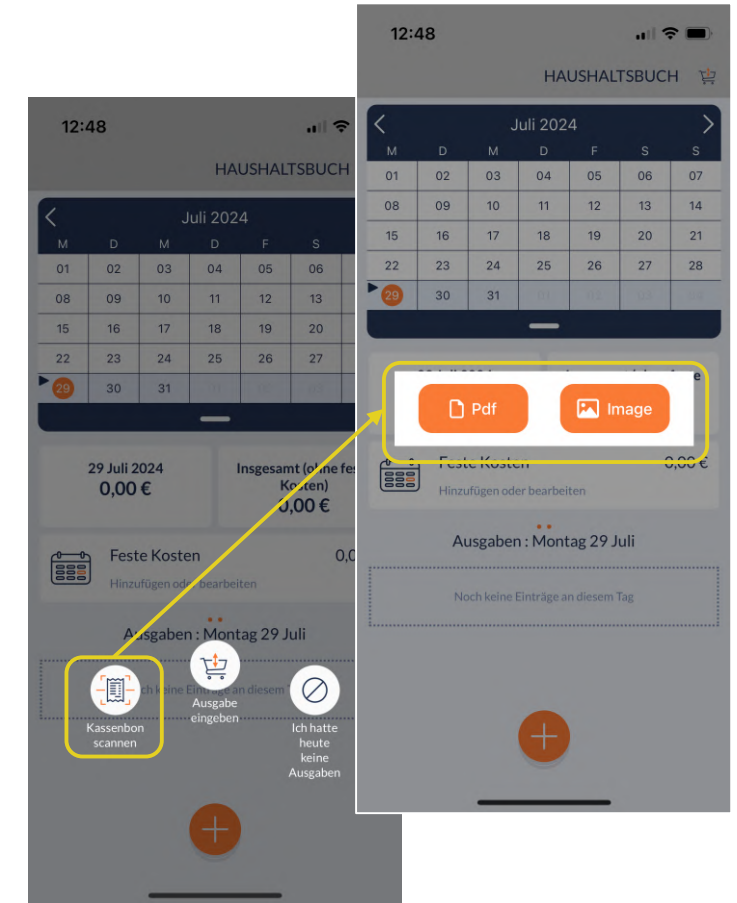
* Scan ticket

** Take picture of ticket



The path to take a picture contains hurdles

- » Some participants have problems with the **English language** and would prefer German terms
- » For some participants the two options **<Pdf>** and **<Image>** are not clearly distinguished
 - » Some don't understand the term **<Pdf>** at all
 - » Few don't understand the term **<Pdf>** in this context; occasionally they want to use a **<Pdf>** and did not understand the option anyway
 - » Few only want a direct way to take a picture



Taking pictures is a learned behaviour (1/2)

- » Many participants decided to **take a picture and do it in a very confident way**, although there were no explanations

Observed behaviours and thoughts from few participants

- » Smoothing out crumpled ticket
- » Not taking a picture of the whole ticket (e.g. only products, no sum)
- » Uncertainty in handling of long tickets (1)
- » Steep angle when taking picture makes ticket conical (1)
- » Covering important information with hand (2)
- » Uncertainty regarding influence of blue ticket paper (3)



Taking pictures is a learned behaviour (2/2)

- » Few participants **would like to receive an indication while scanning**, e.g. a frame around the ticket, which turns green when it was well placed
- » Few participants prefer **choosing pictures from the gallery**
 - » Participants want to take pictures with the regular camera during the day, so that they do not have to take the tickets home
 - » Participants want to upload all collected tickets at once

"Then I can just put them all in a jiffy." (TP19, f21)

Our recommendations

1

Separation of the option „Kassenbon scannen“* into **two single options**: „Kassenbon fotografieren“** and „Kassenbon hochladen“***

→ This allows users to understand the options better

→ This allows a shorter click path

2

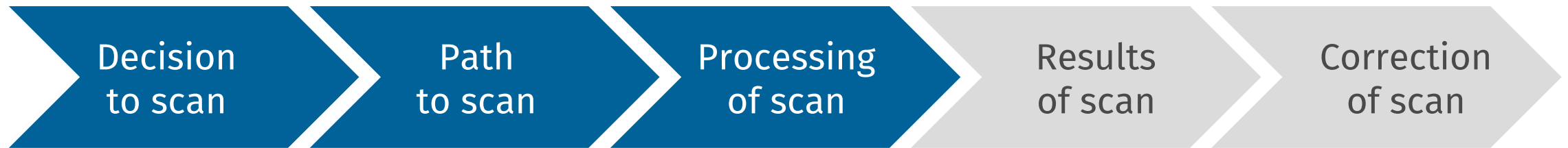
Short explanation of what is important for taking a picture before using the option for the first time

→ Gives users orientation on how to make the ticket „readable“ for the system

* Scan ticket

** Take picture of ticket

*** Upload ticket



After taking the picture, participants were left alone

» Directly after submitting the picture, the **participants do not perceive a feedback**

- » Background: Android users did not receive any feedback, Apple users were shown a loading animation (three dots)
- » Many participants think **something went wrong**
- » A few participants think they **personally did something wrong**
- » A few Apple users think something went wrong, because the upload **takes so long** (although they are shown the loading animation)

*"Oops, something went wrong with the scan."
(TP01, m66)*

*"Nothing appears."
(TP04, f56)*

Processing time takes often longer than expected

- » According to half of the participants, the time between taking a picture and being shown the result is **too long**
- » Participants are **used to faster applications**

*"You are used to this fast pace,
that things go relatively quick."
(TP15, f59)*

*"At home I would
impatiently click back."
(TP11, m69)*

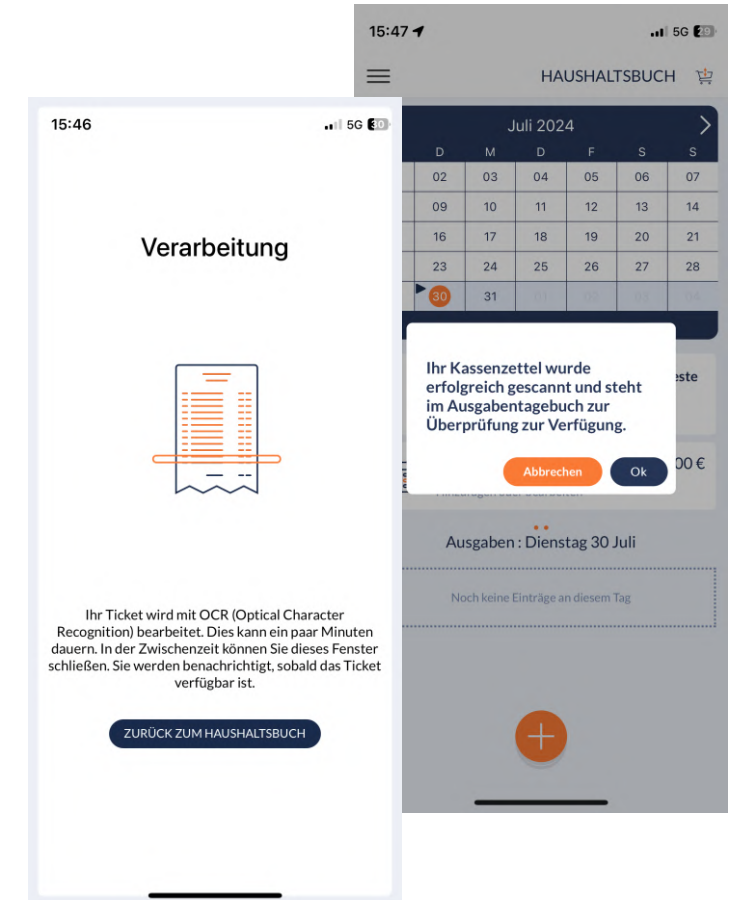
*"If I wanted to do this every
day – no – then I would
delete the app." (TP19, f21)*

Infoscreens about processing are perceived as helpful (1/3)

- » For some the **infoscreens bring relief** that scan went fine
- » Often **described as helpful** because they explain what happens in the background

*"Now it is doing something. Now something could happen."
(TP15, f59)*

"I think it's very good, you know that something is happening and that it works." (TP18, f32)



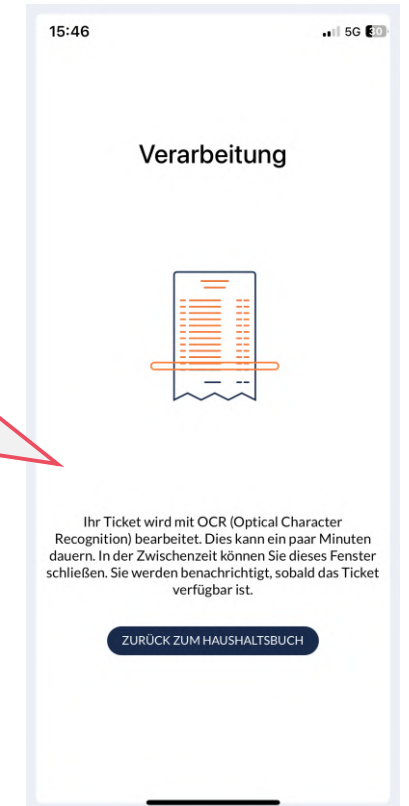
Infoscreens about processing are perceived as helpful (2/3)

» Few participants point out **negative details** regarding the notification

- » Text too long; few would not read it
- » Text written too technically
- » Misinterpreted as succesfully scanned ticket
- » Insecurity about closing the window (through misinterpretation of the animation as scan progress indicator)
- » „takes few minutes“ discourages participants

“At the first sentence - I don't feel like reading the rest.” (TP10, m18)

» For few participants the animation is a **positive detail**



Infoscreens about processing are perceived as helpful (3/3)

- » Notification about successfully scanned ticket is often an **occasion to look at the results of the scan**, some even interrupt their current task
- » Occasionally the first infoscreen is still visible in the background when second infoscreen appears and therefore participants are not sure if the process is actually completed
 - » Closing the second infoscreen does not automatically close the first infoscreen



Our Recommendations (1/2)

- 1 **Inform immediately** after submitting a photo (without any delay, without showing three dots as symbol for processing)
→ Gives users immediate reassurance
- 2 If possible, significantly **shorten the processing time** of a ticket
→ Reduces uncertainty during the waiting period
- 3 **Adjust the text** of the first information about processing, e.g. „Ihr Kassenzettel wird verarbeitet. Sie können das Fenster schließen.“*
→ Makes the content more understandable for users

* Your receipt is being processed. You can close the window.

Our Recommendations (2/2)

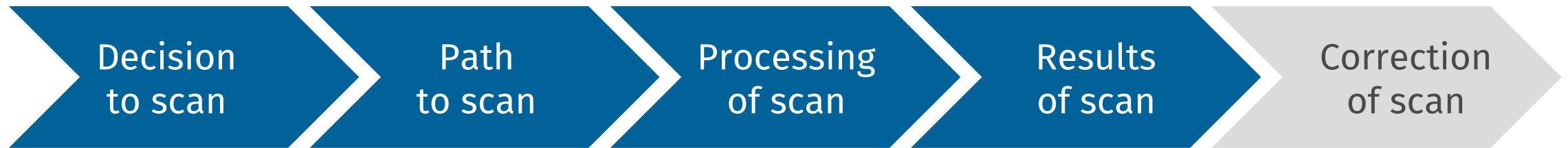
4 Ensure that the first info screen disappears when the second one appears

→ Creates clarity

5 Adapt the button of the second notification, show only one button e.g. <x> or <schließen>* for closing the notification

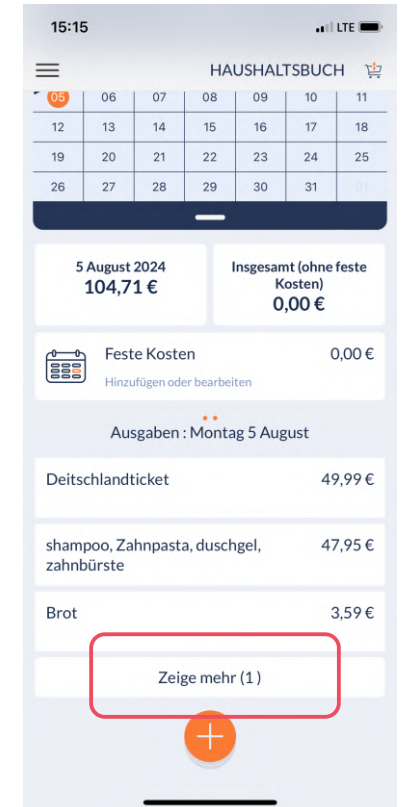
→ Gives users a more common option

* Close



Some can't find the result of the scan

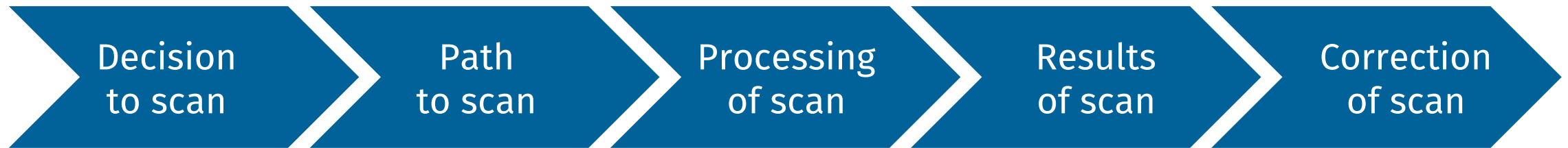
- » While some directly notice the displayed result of the scan, others can not find it
- » Reasons why few participants can not find results
 - » Result hidden behind the button <show more>
 - » In case of wrong totals, few did not identify the result as their corresponding ticket



Our Recommendations

- 1 Show latest result in first place
→ Feed logic makes the presentation similar to other apps
- 2 Show all results at one page and do without the button „Zeige mehr“*
→ Reduces click path and shows all expenses directly
- 3 Show the calender in the minimized version as default
→ Creates more space for the essential things

* Show more



Click path to correct a scanned ticket

Scanned ticket
(tentative)Pen for
correctionHint to correct
expense itemsHint to add
categoryCorrected
informationCorrected
expense itemsCorrected ticket
(committed)

The sequence of screenshots illustrates the correction process:

- Scanned ticket (tentative):** Shows a calendar for August 2024 with the 20th selected. The total for August 20, 2024, is 303,86 €. The item 'ALDI SUD' is highlighted with an orange frame.
- Pen for correction:** The same calendar view, but the 'ALDI SUD' item is now highlighted with a red line.
- Hint to correct expense items:** A list of items with their prices: 781819 Champ. Braun 400g (189,00 €), 55,00 €, 0,95 €, 41453 Creme a1a Cuisine (0,89 €), and 783855 Zucchini (58,00 €). A red line is under the first item.
- Hint to add category:** A form to add a category. The category 'Anderes Gemüse - frisch oder gekühlt' is selected. The price is 189,00 €.
- Corrected information:** The same form as the previous step, but the category is now 'Anderes Gemüse - frisch oder gekühlt'.
- Corrected expense items:** A list of items with their prices: 781819 Champ. Braun 400g (189,00 €), Lauchzwiebeln (0,55 €), Morzarella (0,95 €), 41453 Creme la Cuisine (0,89 €), Zucchini (0,58 €), and beutel (0,01 €). A red line is under the first item.
- Corrected ticket (committed):** The final calendar view showing the corrected total for August 20, 2024, as 191,98 €.

Red line disappears

Red text disappears

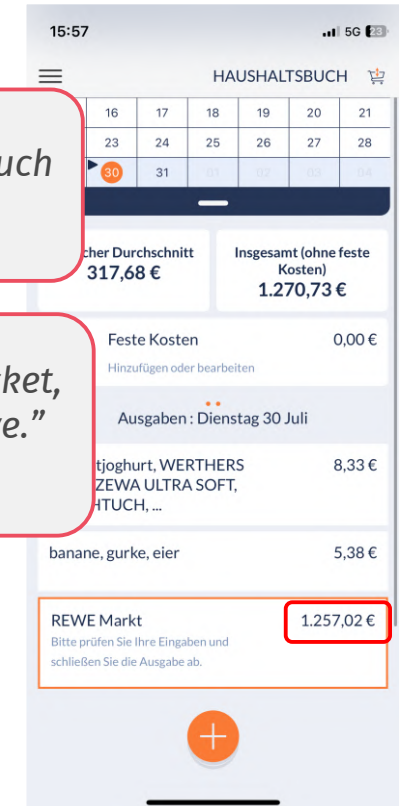
Orange frame and grey text disappears

Many realize that tickets were detected incorrectly

- » Many determine whether a receipt was detected correctly **based on the displayed total sum**
 - » Recurring error: detected sum is way too high (e.g. 4.568 EUR) or no sum was shown (0 EUR)
- » Many participants realize incorrect scan results, however they **did not use the correction as a matter of course**
 - » Many, but not all participants correct errors in the scan results independently
 - » Few delete the entire scan result and try scanning again for better results

"But I didn't spend that much money." (TP19, f21)

"That's an expensive ticket, it wasn't that expensive." (TP16, m36)



Reasons why participants correct results

» Create accuracy

- » Some want a correct result (for themselves) in their diary
- » Few see themselves as an orderly person

» Experience or learned behaviour

- » Few state that they always check receipts
- » Few believe text recognition results always need to be checked

*"[I correct] because I'm a neat person. After all, this should have a purpose. It's no use if there's false information in there that you can't do anything with."
(TP14, m60)*

Reasons why participants do NOT correct results

» Conversely, if the total of a receipt is **displayed correctly**, some see **no reason to correct**

"For large purchases, it would be very exhausting to correct everything." (TP18, f32)

» Correction takes **too much effort** & is considered **too time-consuming**

"I would throw away receipts after scanning and no longer check them. I would trust the app to get the results right." (TP12, f33)

» Many stated they would not want to correct all products

» Some criticize the overall correction effort

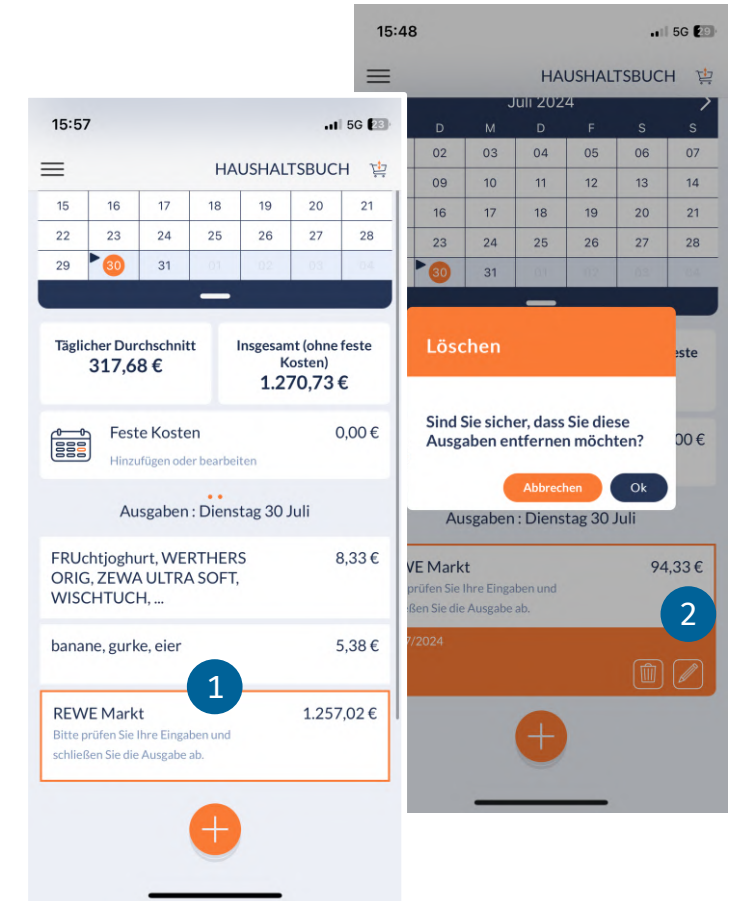
» Some criticize the effort to add a category to each expense item to receive a correct ticket, the relevance of the category is not clear to the participants

"I actually thought that the app would do the categorization automatically. It's a bit tiring." (TP10, m18)

» **Visual hints** are rarely an impetus to correct (orange frame, grey & red hint texts) → [see next charts](#)

Correction notes are too inconspicuous (1/2)

- » Many do **not perceive a clear difference** between the initial scan result, which still needs to be corrected (tentative), and the corrected result (committed) ⁽¹⁾
 - » Half of the participants **misinterpret the orange frame** as a symbolization for scanned tickets or simply as a visual element
 - » Few **misinterpret the grey text** as a general hint and not as a concrete call to action
- » However, for many participants the path to correction is easy to understand because the „pen“ **symbol is learned** ⁽²⁾



Correction notes are too inconspicuous (2/2)

- » Some correction notes within a ticket are unclear
 - » Only few correct because of the red hint text and line (1)
 - » Few are confused why the **red text is still shown** even though the name and price is corrected (1)
 - » Some **don't recognize the red line** under the categorie as a sign for correction (2)
- » Therefore it is not surprising that some participants **want** to have **more obvious and concrete indications**

"It was not clear to me what information was missing. The item and price are there ..." (TP12, f33)

The image shows two overlapping screenshots of a mobile application interface, likely for a shopping cart or ticket management system. The interface is in German. The top screenshot (left) shows a list of items with a red line under the category 'BANANE' and a red hint text 'Bitte geben Sie zusätzliche Informationen'. The bottom screenshot (right) shows a similar list with a red line under the category 'HTUCH' and a red hint text 'Bitte geben Sie zusätzliche Informationen'. Both screenshots have a red circle with a number (1 or 2) indicating the correction note. The interface includes fields for 'Datum der Ausgabe?' (Date of issue), 'Was war Ihre Ausgabe?' (What was your issue?), and 'Preis' (Price). The total amount is shown as 'Gesamtbetrag: 94,33 €'.

Correction of expense items (1/3)

» Some correct either all products or at least some products of a scanned receipt, while only few correct no products at all (1)

» Many correct some information within the detailed product view, although most participants do not correct all information per expense independently (2)

! Please note that this insight only applies to our test situation and is limited. In a pretest setting, participants might tend to correct more than in a real survey, especially considering that participants criticized the high correction effort.

The image displays two screenshots of a mobile application interface for correcting expense items. The left screenshot shows a list of items with a '1' highlighting the 'BANANE' entry. The right screenshot shows a detailed view of a product with a '2' highlighting the category selection.

Left Screenshot (15:48):

- Header: NEUE AUSGABE
- Section: Datum der Ausgabe? Dienstag, 30 Juli 2024
- Section: Was war Ihre Ausgabe?
- Item 1: BANANE, 0,00 € (highlighted with a blue circle '1'). Below it: Bitte geben Sie zusätzliche Informationen.
- Item 2: (empty), 0,00 €. Below it: Bitte geben Sie zusätzliche Informationen.
- Item 3: KAROTTE, 0,00 €. Below it: Bitte geben Sie zusätzliche Informationen.
- Item 4: (empty), 0,00 €. Below it: Bitte geben Sie zusätzliche Informationen.
- Item 5: FRU.JOGH.ANANAS, 0,39 €.
- Item 6: FRU.JOGH.ANANAS, 0,39 €.
- Bottom: Gesamtbetrag: 94,33 €. Button: SPEICHERN.

Right Screenshot (15:55):

- Header: NEUE AUSGABE
- Section: Kategorie (highlighted with a blue circle '2'). Below it: eln, Feigen und tropische Früchte - frisch gekühlt.
- Section: Menge: 1, Preis: 1,99 €.
- Section: Haben den Artikel zurückgegeben (Reklamation o.ä.) (empty circle).
- Bottom: Nettopreis: 1,99 €. Button: AKTUALISIEREN.

Correction of expense items (2/3)

- » Many **correct the price**, although some do not correct minor deviations in price ⁽¹⁾
 - » Few attribute deviations to rounding errors caused by the app when **rounding prices** of individual items
 - » Few name a **lack of motivation** for (time-consuming) correction of only minor deviations
- » Many **correct the product name** ⁽²⁾, only few do not correct minor errors, e.g. “\$” instead of “%”

„When it is only about one euro difference, I probably wouldn't look through 20 positions“ (TP09, f42)

15:55 5G

NEUE AUSGABE

Ausgabe
banane

Kategorie
Datteln, Feigen und tropische Früchte - frisch oder gekühlt

Anzahl 1 Preis 1,99 €

Gesamtpreis: 1,99 € AKTUALISIEREN

Correction of expense items (3/3)

» Many do not **correct missing categories** (3)

- » Many criticize the insufficient indication that a category needs to be supplemented
- » Some do not consider the category as relevant
- » Some consider it too much work, too detailed
- » Few mention that the app should fill this in itself

"It goes into too much detail, which is not relevant." (TP15, f59)

"It might be annoying that it's red, but I would accept it and say: no time, no desire to do it." (TP09, f42)

» Some do not **correct the quantity** (4)

- » Some only focus on price (& name)
- » Few do not immediately get that the field „price“ refers to the single item price

15:55 5G

NEUE AUSGABE

Ausgabe
banane

3 Kategorie
Datteln, Feigen und tropische Früchte - frisch oder gekühlt

4 Anzahl 1 Preis 1,99 €

Ich habe den Artikel zurückgegeben (Reklamation o. ä.)

Gesamtpreis: 1,99 € AKTUALISIEREN

Our Recommendations

Give users more obvious and concrete indications on what information is missing

1

Use appropriate **signal colours** instead of orange frame
e.g. red frame (correction needed) and a green frame (correction done)

2

Explicitly **name what is missing**

i.e. the grey/red text should point out what information needs to be corrected

3

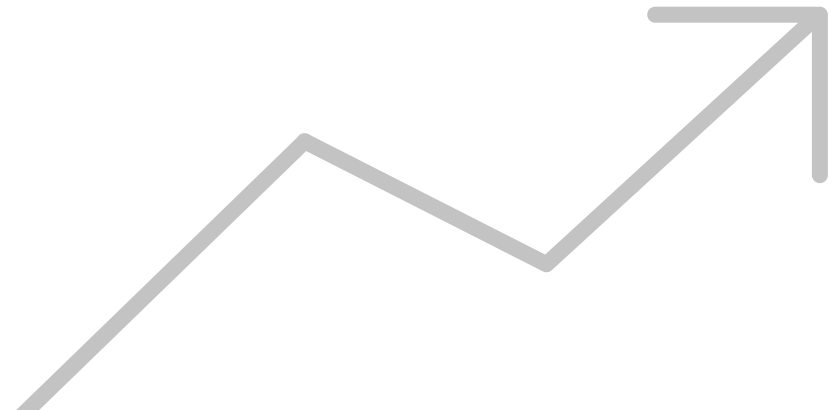
Using symbols as a quick reference

e.g. <!/> for <correction needed>, <√> for <correction done>

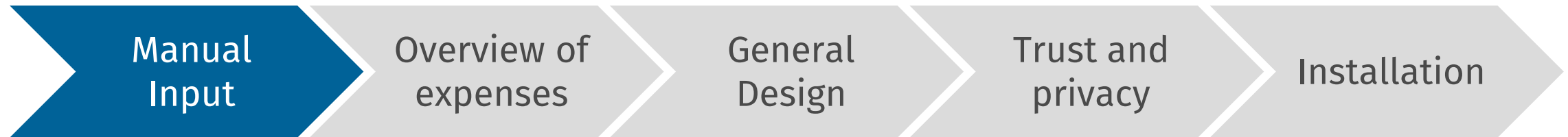
4

Automatic assignment of COICOP to every single expense item or no assignment of COICOP at all

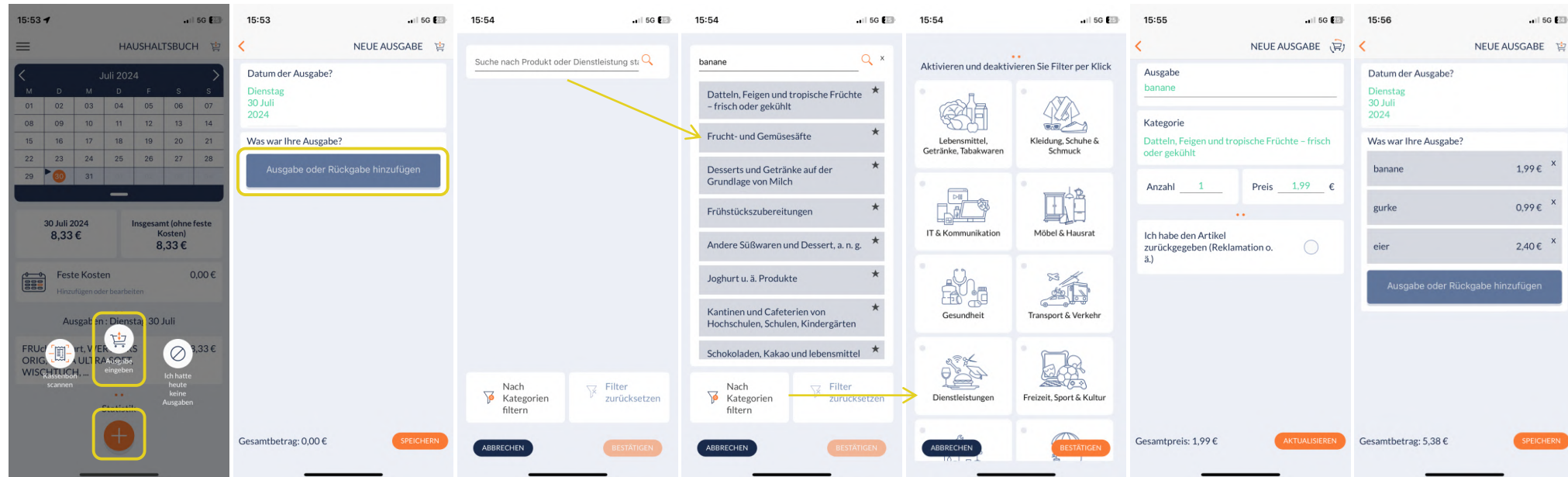
Results in terms of further aspects



Manual input

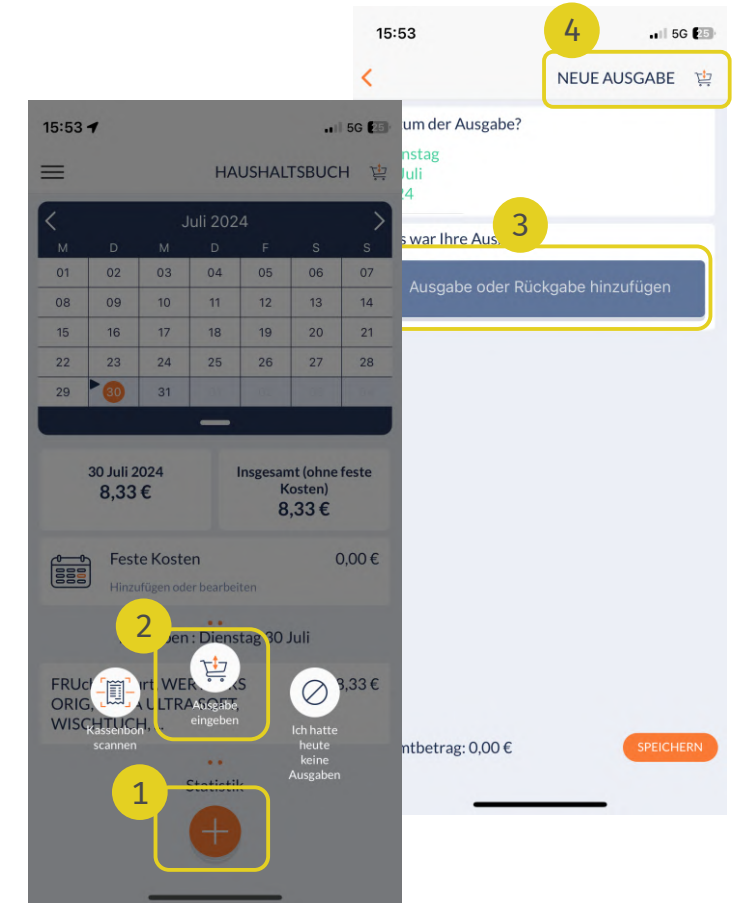


Click path for manual input

Choose the
manual inputAdd multiple
expense itemsEnter
productAssign suitable
categoryOpt. prefilter
categoriesEnter detail
informationExpense items of
a purchase

The path to manual input contains hurdles

- » Many **immediately understand** the path to manual input
 - » For some the <+> is not a learned symbol, so they try out different ways (e.g. tap on <fixed costs>) (1)
 - » After tapping on the <+>, all understand which option to choose to open the manual input (2)
- » Some seem to be confused and do not understand **how to add an expense** (3)
 - » Some do not perceive the blue button as a start to add expense items, especially as it is not designed clearly as a button (3)
 - » Few tap on „Neue Ausgabe“ because it is more obvious than the blue button which seems to be inactive (4)



Unclear that expense screen is referring to whole ticket

- » Users **do not know that a digital twin** of the physical ticket **has to be created**
- » No indication that this interface is on the **level of a single ticket**
 - » Indications such as “Neue Ausgabe”* or <Ausgabe hinzufügen>** are not clear
- » Leads to multiple follow up problems – in the first step it is not clear **what has to be added** via the button
 - » Some add the whole sum of a ticket (e.g. use name „groceries“), few create own categories within a ticket and add these (e.g. use names „groceries“, „drinks“)
 - » Few create a new expense for each product

15:53 5G

NEUE AUSGABE

Datum der Ausgabe?

Dienstag
30 Juli
2024

Was war Ihre Ausgabe?

Ausgabe oder Rückgabe hinzufügen

“I would only record the total because the effort would be too high. Total would be enough for me.” (TP12, f33)

Search is not expected when input is asked

» Many do not know **how to proceed on this page**

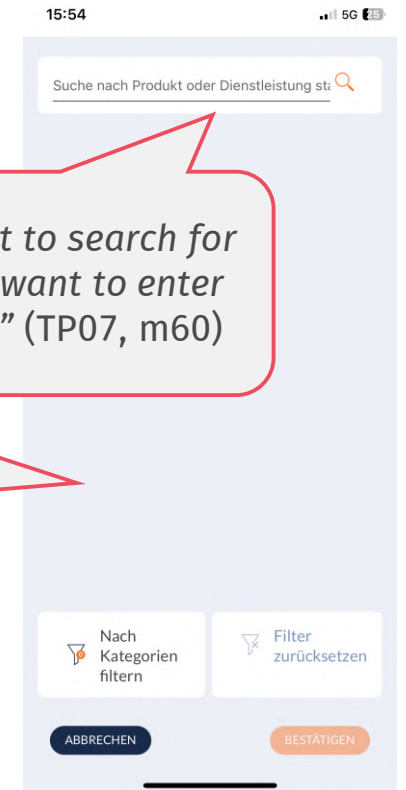
- » Searching is not intuitive when trying to add a product
- » Some tap on <Nach Kategorien filtern>* first (and each time), because they do not understand that it is a prefilter for the search function and rather believe it to be a menu to choose the actual category from – benefit of the filter is not understood

» Many do not know **how to use the search bar**

- » Many are unsure with what degree of detail to add expense items ((1) name of the market, (2) name of the single product or (3) product category)
- » Some do not read the text and few do not understand the instruction
- » Few do not see the search bar at all
- » Few tap on the magnifier without entering text

"I don't want to search for anything, I want to enter something." (TP07, m60)

"I can't get any further now." (TP04, f56)



Some difficulties when selecting type of expense (1/2)

» Few do **not understand the concept** of assigning categories

» Some **struggle choosing** the right category

» Too many categories

» Similar-sounding categories

» Inappropriate categories: insecure about which category is correct

"I feel a little overwhelmed. Are all these supposed to be suggestions?" (analogous quote of TP10, m18)

"That's quite a lot of choice. Do I have to be so precise?" (TP09, f42)

» Automatically shown suggestions are **shown with a delay**, therefore not perceived as automatic search, try to confirm by tapping enter or magnifier



Some difficulties when selecting type of expense (2/2)

» Occasionally perceived problems

- » No category is shown, insecurity how to deal in this case
- » Few enter the name of a market or product category and do not get appropriate suggestions, choose wrong categories

» This leads to the following consequences

- » Some think about it for a long time
- » Few fear putting in wrong category with regard to official statistics
- » Some choose a wrong category
- » Many perceive choosing the type of expense as a high burden (see correction function)



Only minor problems with input mask

- » Many participants are easily able to fill in data into the input mask
- » Field „Anzahl“* causes problems ⁽¹⁾
 - » Few do not immediately get that the field „price“ refers to the single item price and later realize that the price is multiplied by the quantity
 - » Few do not put in any data
 - » Few want to enter a decimal (e.g. weight: “0,8 kg”)
- » Field „Preis“** causes difficulties ⁽²⁾
 - » Some forget to enter the comma
 - » Subsequent error if the entire purchase is recorded as one entry

“I would enter everything as one sum because it is all food.” (TP11, m69)

15:55 5G

NEUE AUSGABE

Ausgabe
banane

Kategorie
Datteln, Feigen und tropische Früchte frisch
oder getrocknet

1 2

Anzahl 1 Preis 1,99 €

Ich habe den Artikel zurückgegeben (Reklamation o. ä.)

Gesamtpreis: 1,99 € AKTUALISIEREN

Keyboard blocks <Speichern> button

- » Many do not find the <Speichern>* button because it is hidden by the keyboard – some even become insecure
 - » Few try to go back by using the arrow (top left corner) which is intended for leaving
 - » Some tap on <Neue Ausgabe>** in the top right corner and believe it is a button
 - » Depending on the smartphones model the keyboard also shows the options <Öffnen>*** or <Weiter>***, which few participants try to use
- » Tapping outside of the keyboard is not perceived as an option to close it and to get to the button <Speichern>

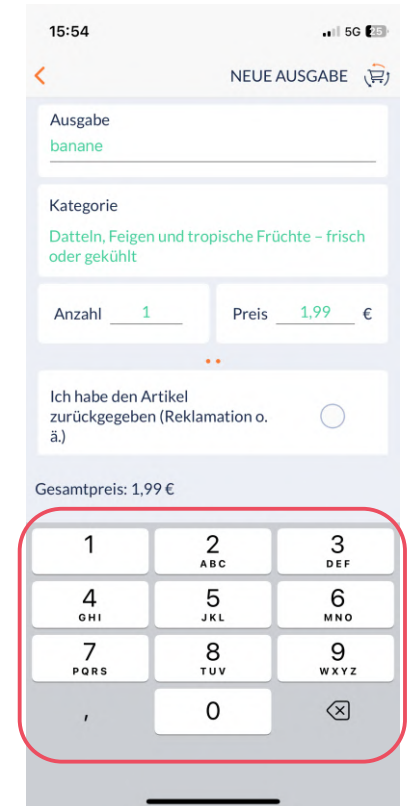
*Save

**Add new expense

***Open or next

“Then I have to press somewhere. If things go badly, you press on a field and delete it by mistake.” (TP09, f42)

“How can I get out of here now?” (TP12, f33)



New expense item: problems with item details

- » The first seven test persons were asked for more comprehensive information regarding the product*
- » Here, problems arised:
 - » The difference between “Anzahl” (quantity), “Menge” (amount) and “mehrere Artikel” (several articles) is unclear to some - It is unclear how to deal with a package containing several pieces (e.g. eggs) ⁽¹⁾
 - » Some enter the total number of items on the ticket
 - » Geschäftliche Ausgabe (business expenses) - If you move the slider to 100 % private, 'business' should be deactivated
- » Payment
 - » Some are surprised that cash payment cannot be selected in addition to card payment

08:40

NEUE AUSGABE

Ausgabe
Wassermelone

Kategorie
Anderes Obst - frisch oder gekühlt

Anzahl 3 Preis 2,95 €

Details zu Lebensmitteln, Getränken, Tabakwaren (Bitte während der Detailstichprobe angeben)

Umfasst die Ausgabe mehrere Artikel?

☒ Nein, es ist 1 Artikel

☐ Ja, _____ Artikel

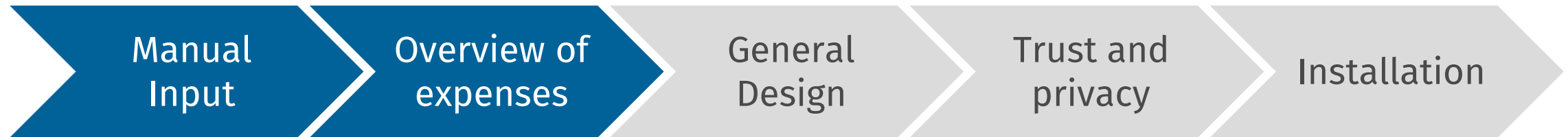
Menge 0 Gemessen in.. kg

Rabatt 0,00 €

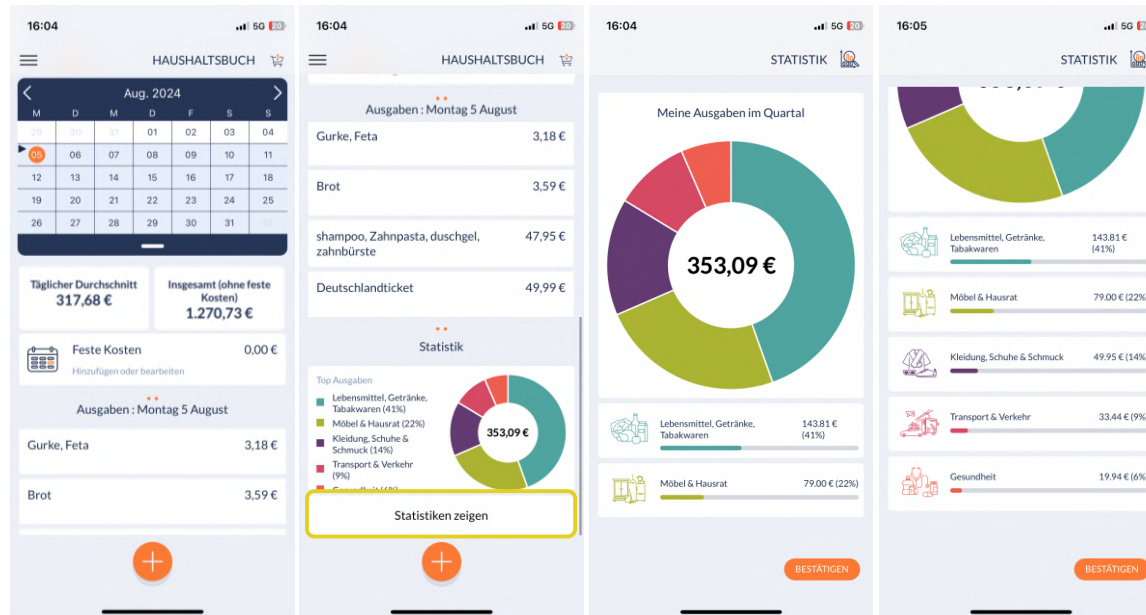
Gesamtpreis: 8,85 € AKTUALISIEREN

Our Recommendations

- 1** → Give the expense page a heading or have it assigned, change wording of „Ausgabe“ to „Einkauf“
→ Makes clear on which level respondents are on (expense or expense item) and thus avoids consequential errors
- 2** → Adapt text and design of the blue button to add expense item; e.g. (1) label button with „Add product“ only or (2) add an orange plus symbol for adding
→ To make clear to the users that it is a tapable button and what happens behind the button
- 3** → Field for entering the product name should look like an input mask and not like a search mask, e.g. (1) remove magnifying glass and filter, (2) place input field more prominently, (3) change text to „Produkt eingeben“, (4) implement automatic allocation of COICOP, (until possible: leave out magnifier symbol and filters)
→ Include only those elements that respondents need



Click path to overview of expenses

Scrolling down in
„Haushaltsbuch“First
overviewDetailed
overviewCategories in
Detail

Some hurdles on the path to the expenses overview

- » Some participants find the overview in first attempt, others need more time or they only find it at the tip of the interviewer

Hurdles mentioned occasionally

- » Placement below single expenses not expected (when no expenses are recorded, then overview easier to see)
- » Not aware that scrolling is possible
- » Total costs (box below calendar) interpreted as overview, more details expected by clicking on
- » Expected in burger menu



Getting feedback on expenses is important

» For many the overview of expenses is important

Reasons for appreciation

- » Reward for the effort, getting something back
- » Creates a clear overview on input
- » Personal feedback on expenses, for self control

"If I'm going to make the effort, then I'm going to get something out of it." (TP02, f68)

"Without an overview, it wouldn't be very clear if you only had receipts." (TP20, m20)

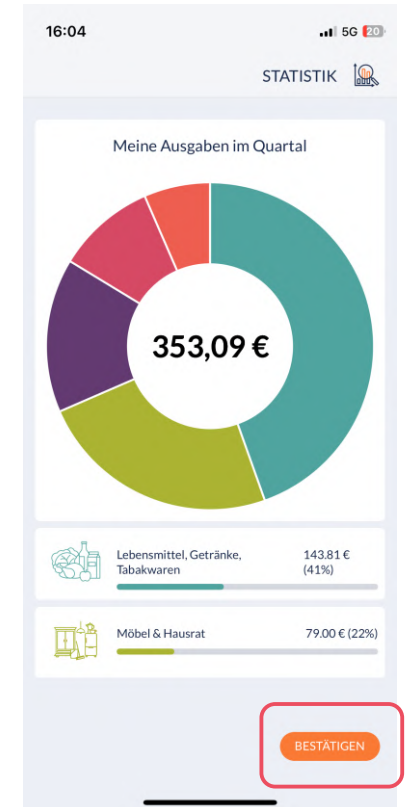
"What can I learn from the overview over my spendings?" – cited analogous after (TP14, m60)

Participants like the presentation

- » On the whole **many positive comments**, although there are few suggestions for improvement

Smaller suggestions for improvement

- » Showing the expenses in € by mouseover
- » Unclear why „confirm“ is required via button



Our Recommendations

- 1 Design button „Statistik anzeigen“* in the clear look of a button
→ Creates a stringent design and thus better orientation for users
- 2 Rename the button „Bestätigen“** in the detail view to „Schließen“***
or use a <x> as sign for closing the window
→ Gives users a clear understanding what the button is for
- 3 Adding more content to the pie chart or reduce the size of the pie chart
→ Provides users with more information in a clearer way

* Show statistics

** Confirm

*** Close



Remarks on design of diary (1/2)

- » Many participants did not use/discover the option to minimize the calendar – **Smaller visualization of the calendar** perceived as better ⁽¹⁾
- » Few participants would like a **visual indication in the calendar** on the days where expenses are recorded
- » For few participants the **reference period** of total expenses is not clear (the field “Insgesamt”*) ⁽²⁾
- » Two orange dots interpreted as a possibility to swipe horizontally (same as in expense item screen) ⁽³⁾
- » Some participants have to search around for some time to find the button to add expenses (it seems not to correlate with the age of participants) ⁽⁴⁾

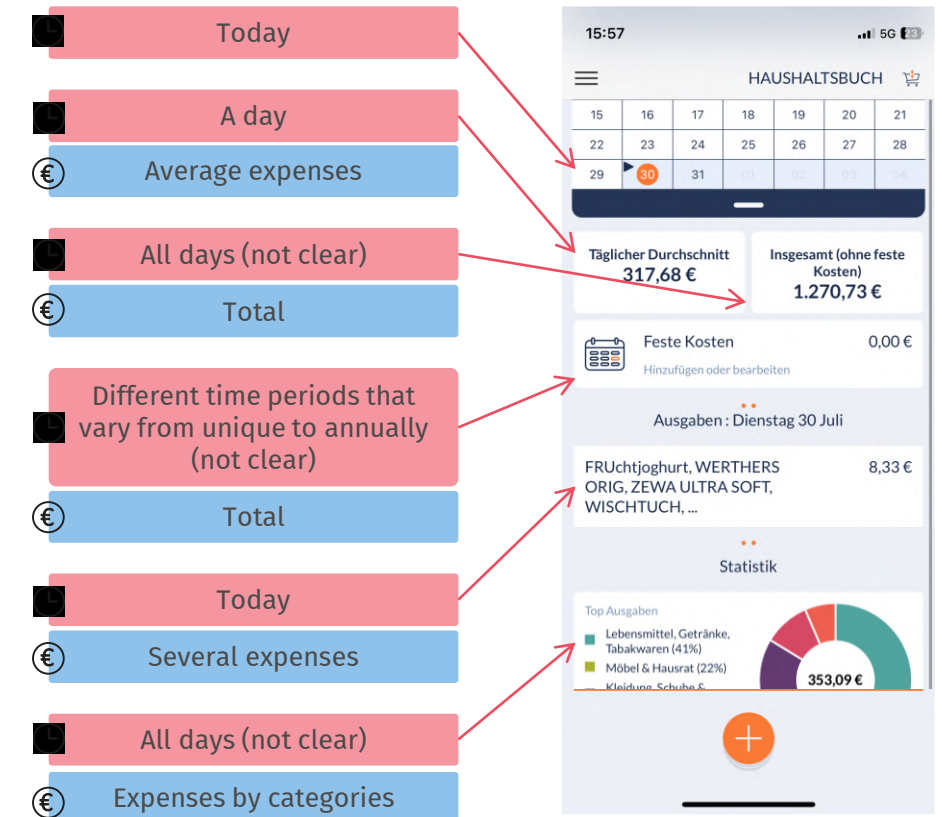


* Total (without fixed expenses)

Remarks on design of diary (2/2)

Not explicit mentioned by participants:

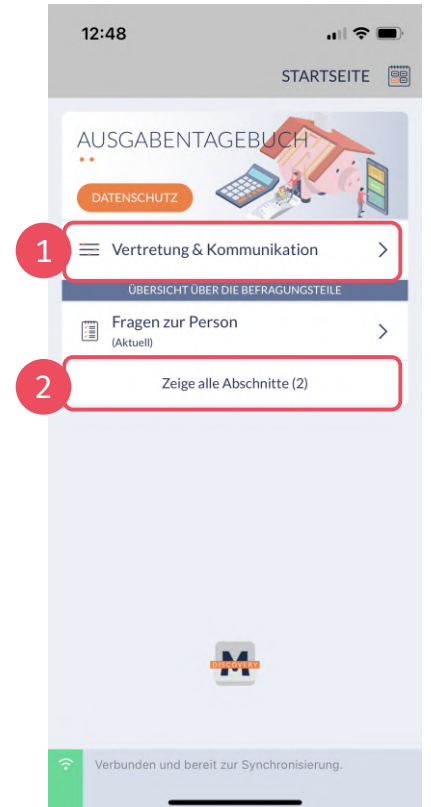
- » Page shows expenses in numerous relations in regards to (1) period of time and (2) interpretation – therefore the page might appear overloaded with information



Occasionally mentioned remarks on design of landing page

- » Landing page does **not have many functions**
- » Menu item „**Vertretung und Kommunikation**“* is **confusing** due to the name (not self-explanatory) and the position (expected under settings) ⁽¹⁾
- » **Hiding additional menu items** is found to be unnecessary when the page contains space to display all of them directly ⁽²⁾

* Representation and communication



General aspects

- » In general usage of the app, it is very **often necessary to close, commit, save, update** something, in order to proceed
- » In combination with the high degree of detail in the data collection, this makes for a **tiring user experience**

Our Recommendations

1

Revision of the overview page

→ To make the important things more clear

2

Reduce the high variance in temporal and expenditure-related references at the overview page

→ To reduce complexity



Participants trust the app

» All participants find the app trustworthy

Sometimes supported by

- » „Statistisches Bundesamt“ as publisher of the app
- » Professional appearance of interface
- » No advertising

Occasionally mentioned it could further be strengthened by

- » Adding a logo of „Statistisches Bundesamt“
- » A better fitting name (tested under the name „Motus Discovery“)

! These learnings are in line with the results of our focus group discussions, previously conducted within the SSI project (11/2023)

Participants are willing to share information

- » For participants it makes **no difference if they scan their own ticket or a ticket from someone else** (here: the ticket that was given to them in the test)
- » Many participants would **enter all expenses in the app**
 - » Only few participants would leave out special expenses (in order to not falsify the result) or
 - » Few would leave out sensitive expenses (theoretically brothel)
 - » Some do not seem to be aware that the data is processed by the National Statistical Institute



The path from the invitation letter into the app

Invitation letter
with instructions

Download from
app store

Info:
language

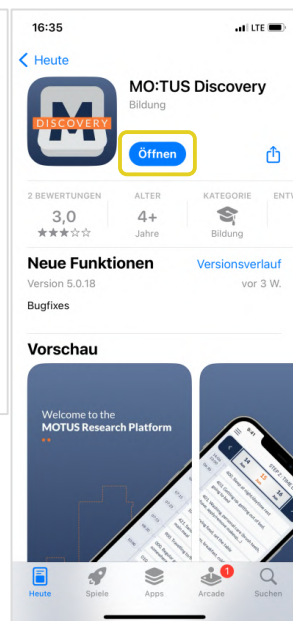
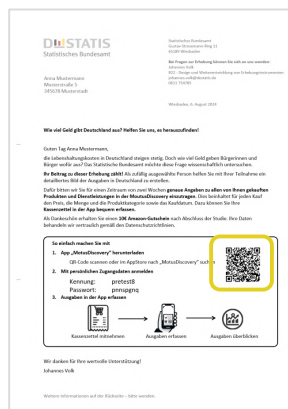
Info:
functionality

Info:
datasecurity

Info:
push notification

Info:
Results

Log In



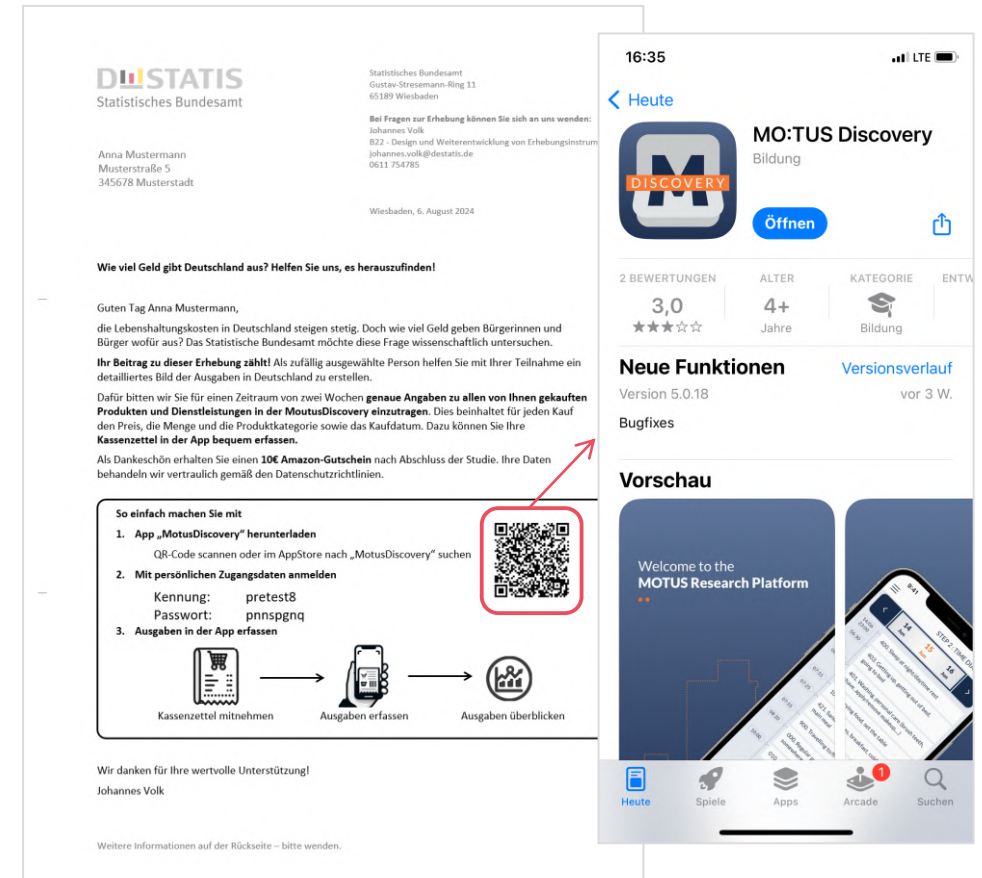
Path into the app is familiar (1/2)

- » All participants are able to easily download the app from the app store, it is something they are used to do
- » The participants **read** the invitation letter with **different degrees of intensity**
 - » Many participants only read the first page, don't recognize the printed back
 - » Some participants skip/skim the text and go directly to the box explaining the three steps
 - » Unclear how many days the expenses have to be reported



Path into the app is familiar (2/2)

- » Almost all participants appreciate the QR-Code as a direct link to the app in the app store
- » Shortens the way to the app – no typing and no search necessary
- » More comfortable, easy
- » Contemporary



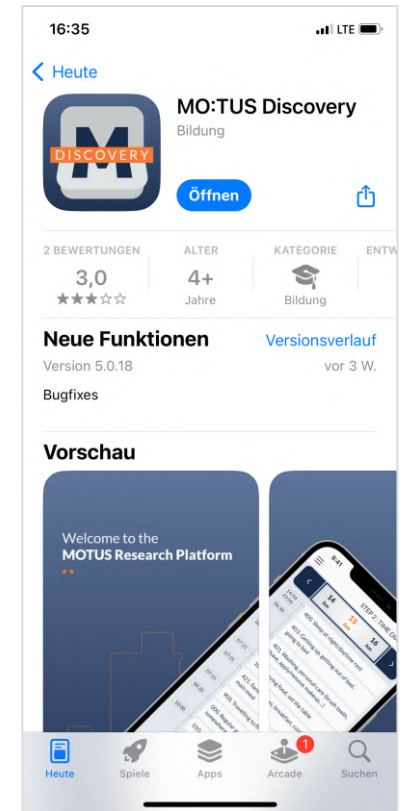
Information in app store receives little attention

- » Most participants download the app directly **without checking the given information** (e.g. reviews, infoscreens)

What catches the eye of few participants?

- » Size of the app (quite large)
- » Name of the app (not suitable and unexpected)

Notice: Low attention to information **may be related to the test situation** (e.g. might create trust, questioning instructions might be uncomfortable)



Infoscreens as an introduction are familiar

- » Some participants describe infoscreens as **normal** introduction without showing big emotions
- » Half of the participants **go fast through the screens**
 - » Some skipped them directly
 - » Some seemed to read only the headlines
- » **Good content, brief and to the point**, no major criticisms

*"Infoscreens are available for all apps, it would make a shady impression if pages did not come."
(TP09, f42)*

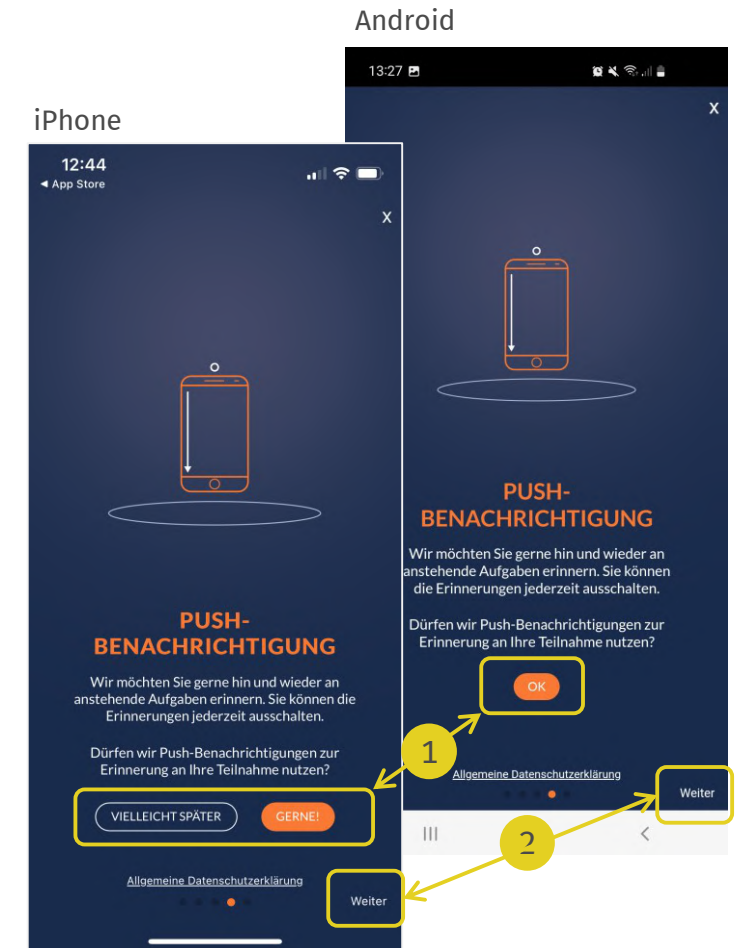
"Rather helpful, but I know that from so many apps. I just skip it. I find my way around that way too." (TP10, m18)

"I think it's good that it's held generally at the beginning [and you're] not bombarded at the start." (TP13, f66)

Minor comments on infoscreens

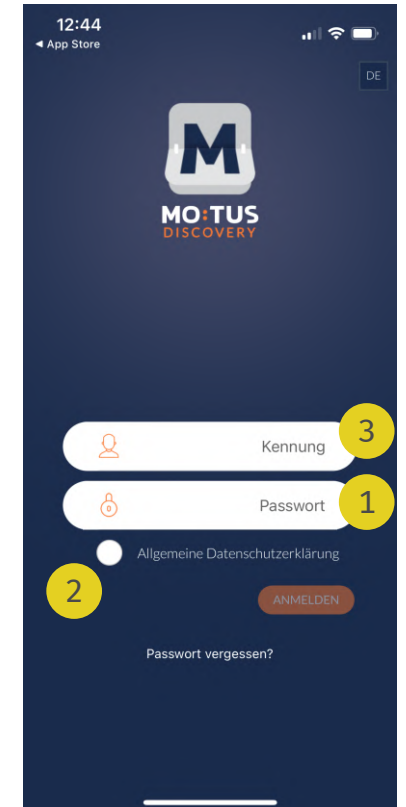
» Occasionally mentioned aspects

- » Push notifications: not clear if the notifications will be allowed when using the button <weiter> (1)
(The user interface differs between Android and iOS in the test situation)
- » Results: the mentioned connection between input of expenses and political decisions is not clear
- » Button <weiter> does not look like a button, is small and in the corner (2)
- » Light font on dark background bothersome
- » Switch from dark background in introduction to light background in the app itself is perceived as inconvenient



Login is easy to handle

- » All participants **found it easy** to login, even though half of them **needed two approaches**
- » Frequent notes by the participants
 - » Some want an option to show the password (to make input and correction easier) (1)
 - » Some did not tick the box for data protection agreement (it seemed too inconspicuous, for few a slider is more common) (2)
 - » Few wondered about the term „Kennung“ („Anmeldename“ is more common) (3)



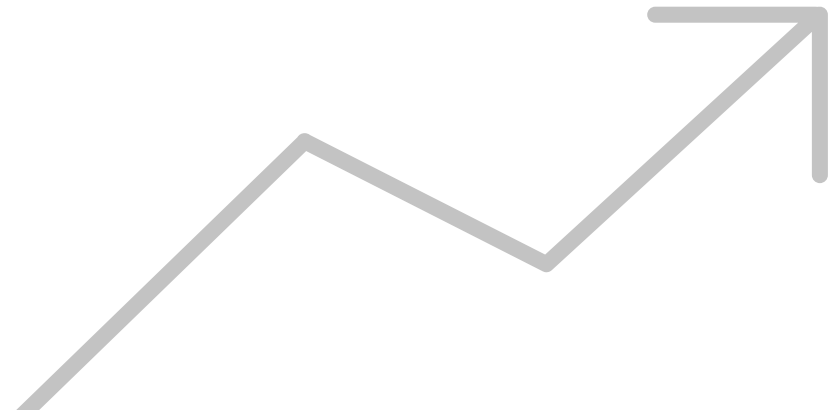
Our Recommendations

1

Add the function to show the password at the login

→ To make entering the password easier

Bugs



Bugs (1/2)

- » Introduction screens: font is regular on all screens, except for screen 1, which has a thin font. Fonts should be consistent
- » Differences between android & iOS:
 - » Only under iOS, after uploading a scan, the „three-dot-animation“ is displayed. It is missing under Android.
 - » On the infoscreen about the push notifications, under iOS, the options „gerne“ + „später“ + „weiter“ are shown. Under android, only „ok“ + „weiter“ are shown, which leaves unclear whether participants agreed or not.

Bugs (2/2)

- »» Diary page does not refresh/update automatically:
 - »» Amount of deleted tickets is still shown/included in the total amount field on the upper right corner below the calendar (in dairy)
 - »» Results of scanned tickets are not immediately shown in diary, except participants confirm the notification that ticket is successfully scanned
 - »» Expense overview (pie chart) not updated automatically, after adding new expenses



SMALL-SCALE USABILITY TEST OF OCR IN MOTUS

A THINKING ALOUD EXPERIMENT



BRUSSELS INSTITUTE
FOR SOCIAL AND
POPULATION STUDIES
Research Unit
Tempus Omnia Revelat (TOR)

MAIN CONCLUSIONS

A COMMON MISTAKE THAT PEOPLE MAKE WHEN TRYING TO DESIGN SOMETHING COMPLETELY FOOLPROOF IS TO UNDERESTIMATE THE INGENUITY OF COMPLETE FOOLS

- DOUGLAS ADAMS



01 ONE SIZE DOES NOT FIT ALL BUT SOME GENERAL IMPROVEMENTS ARE POSSIBLE

01 ONE SIZE DOES NOT FIT ALL BUT SOME GENERAL IMPROVEMENTS ARE POSSIBLE

MAIN FINDINGS

- ❧ UI judgement ranged from “very professional” to “old fashioned”
- ❧ Confusion arose when wordings in notifications did not match wordings of buttons
 - ❧ e.g. “expenditures” in Dutch was often interpreted as “total cost”
 - ❧ e.g. “update” vs “save”
- ❧ UI interface layout caused some confusion
 - ❧ e.g. editing option fell off the screen
 - ❧ e.g. not knowing that one can scroll
 - ❧ e.g. tapping “monthly expenditures” to add ticket

POINTS OF ATTENTION

- ❧ Communication and wordings must be very clear and understandable
- ❧ UI layout (1)
 - ❧ Hard coded language must be consistent
 - ❧ Colouring of buttons must be consistent (e.g. “follow the orange trail” – quote from TP)
- ❧ UI layout (2)
 - ❧ On small screens calendar takes up too much space
 - ❧ “Monthly expenditures” is very dominant
 - ❧ New tickets could appear on top
 - ❧ Make clear that scrolling is an option

02 SCAN OPTION IS PREFERRED METHOD, BUT EXPECTATIONS ARE EXTREMELY HIGH

02 SCAN OPTION IS PREFERRED METHOD, BUT EXPECTATIONS ARE EXTREMELY HIGH

MAIN FINDINGS: SPEED

- ⌘ Expect very high processing speed
 - ⌘ Notification reads that processing might take up several minutes
 - ⌘ Randomly timing shows tickets are processed between 1 and 1.5 minutes
 - ⌘ Majority of respondents shows physical signs of impatience after 20-30 seconds
- ⌘ Majority of respondents wanted to continue scanning other tickets

POINTS OF ATTENTION

- ⌘ Processing speed is crucial
- ⌘ Communication is crucial
 - ⌘ Maybe show a "tickets being processed: #"

02 SCAN OPTION IS PREFERRED METHOD, BUT EXPECTATIONS ARE EXTREMELY HIGH

MAIN FINDINGS: IMAGE

- ⌘ Click path is too elaborate and confusing
 - ⌘ '+' button
 - ⌘ Use receipt
 - ⌘ Take photo
- ⌘ Cropping image is confusing for some respondents
 - ⌘ Younger respondents do not crop (e.g., "if it's a good OCR, it will know")
 - ⌘ Older respondents think they need to crop, but have trouble
 - ⌘ Majority of respondents think they did not crop correctly if scanned ticket is incorrect

POINTS OF ATTENTION

- ⌘ Reduce click path
 - ⌘ + and then 📷
- ⌘ Communication is crucial
 - ⌘ Leave out cropping option if not needed

02 SCAN OPTION IS PREFERRED METHOD, BUT EXPECTATIONS ARE EXTREMELY HIGH

MAIN FINDINGS: QUALITY

- ⌘ High confidence in OCR and several tickets come are presented correctly
- ⌘ Response to incorrect results is mixed
 - ⌘ If close match to ticket, majority of respondents cannot be bothered to correct
 - ⌘ If not not close match, majority of respondents think they did something wrong with scanning
 - ⌘ Some respondents correct straight away, others need to be prompted by interviewer

POINTS OF ATTENTION

- ⌘ Training data is very important
- ⌘ Communication is crucial
 - ⌘ Is there a way to prompt respondents to check the scanned results against the ticket?

02 SCAN OPTION IS PREFERRED METHOD, BUT EXPECTATIONS ARE EXTREMELY HIGH

MAIN FINDINGS: CORRECTING

- ✂ Correcting tickets is tedious
 - ✂ The need for assigning COICOPs is not understood and hard to complete
 - ✂ Currently 'discount' is extracted as a product which adds the discount to the ticket
 - ✂ The majority of respondents think they can correct the discount by making the amount negative, which cannot be done

POINTS OF ATTENTION

- ✂ Training data is very important
- ✂ For scanned tickets
 - ✂ A classification algorithm is expedient
 - ✂ The extended context options should not be showed even when toggled on because OCR does not extract that information
- ✂ Dealing with discounts is tricky
 - ✂ Allowing adding an overall discount to a ticket is easy for respondents
 - ✂ But it might encourage respondents to just add discount to reach the correct amount

Note: for this experiment the expenditure diary showed all options. These options can be toggled on or off in the back office of the app.

03 ONLINE SMART SURVEY DOES NOT MEAN NO INTERVIEWER NEEDED

03 ONLINE SMART SURVEY DOES NOT MEAN NO INTERVIEWER NEEDED

MAIN FINDINGS

- ⌘ When stuck and asked about how they would normally proceed, the majority of respondents said **they would give up**.
- ⌘ However, through trial and error – and some help for the oldest respondent – all respondents **came to grip with the registration flow**.

POINTS OF ATTENTION

The **learning curve is very steep**. An interviewer could take up the role of instructor/case manager by:

- ⌘ explaining how the app works
- ⌘ providing a help line
- ⌘ following up on progress

04 CLASSIFICATION IS DIFFERENT FOR RESPONDENTS AND STATISTICIANS

04 CLASSIFICATION IS DIFFERENT FOR RESPONDENTS AND STATISTICIANS

MAIN FINDINGS

- ⌘ The majority of respondents types in the exact product name which hardly ever matches relevantly with the COICOP list
 - ⌘ e.g. “dinosaur milk chocolate biscuits”
- ⌘ The description of categories at the detailed level is long and thus takes up quite a lot of screen per category
- ⌘ The lack of a classification algorithm implies that when respondents correct a scanned ticket, they need to add a COICOP

POINTS OF ATTENTION

- The **COICOP classification does not make sense to respondents.**
- ⌘ The wordings of the COICOP classification at the lowest level could be revised
- ⌘ The process of classifying expenditures needs to be made easier e.g. by using classification algorithms
- ⌘ COICOP classifications might not be the most insightful classification to provide feedback

05 INCREASING RELIABILITY OF DATA BY ADDING MORE CONTEXT INCREASES RESPONDENT BURDEN

05 INCREASING RELIABILITY OF DATA BY ADDING MORE CONTEXT INCREASES RESPONDENT BURDEN

MAIN FINDINGS

- ⌘ Respondents generally fall into two categories
 - ⌘ “That’ll do” are the ones that provide as little information as possible to reach the correct amount
 - ⌘ “Accurate” are the ones that try to do everything right
- ⌘ The “accurate” respondents still have trouble understanding some of the detailed context questions
 - ⌘ ‘multiple’ items is not understood
 - ⌘ the UI does not clearly show that the unit can be changed

POINTS OF ATTENTION

- ⌘ An **extended expenditure diary is more burdensome** because
 - ⌘ scanning function does not pick up extra details
 - ⌘ there are too many fields to be completed
 - ⌘ not all fields apply to all expenditures which is confusing
- ⌘ More **contextualized data might be acquired** by
 - ⌘ using dedicated (panel) respondents
 - ⌘ using focus days/weeks
 - ⌘ offering ‘expert’ mode

Note: for this experiment the expenditure diary showed all options. These options can be toggled on or off in the back office of the app.

METHODOLOGY & PARTICIPANTS

METHODOLOGY & PARTICIPANTS

RESEARCH TEAM

Supervisor	Theun Pieter van Tienoven
Test coordinator	Petrus te Braak
Interviewers	Petrus te Braak
	Francisca Mullens
	Nelly Geerts
	Jens Verhoeven
Analysis	Theun Pieter van Tienoven

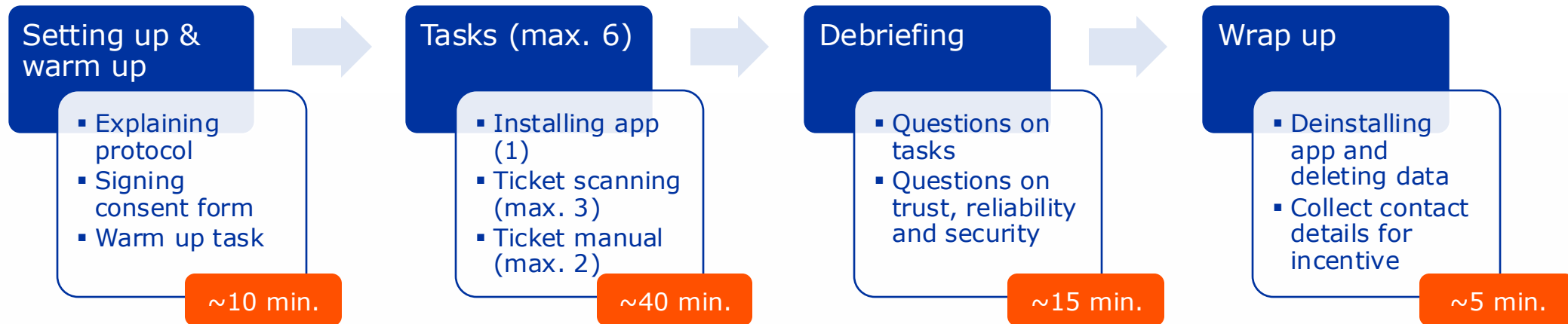
METHODOLOGY & PARTICIPANTS

SET UP

- 20 test persons (TP) planned, **18 conducted**
- Voluntary sampling in interviewers' network
- TPs used their own smartphone
- Time frame: 60-90 minutes
- Incentive: 20.00€ gift voucher from online store
- Survey period: May-June 2024

METHODOLOGY & PARTICIPANTS

PROCEDURE



METHODOLOGY & PARTICIPANTS

TASK LIST

Task #	Receipt type	Usability focus	Rationale	Observer focus
Task 1	-	Warm-up/primer	Practice think-aloud	Actively stimulate verbalizations of thoughts
Task 2	-	Install app and log-in	Prepare/access	
Task 3	A1. Short and clear	Choose approach	Start easy	
Task 4	A2. Short and clear	Force inversed approach	Make sure both approaches are tested	
Task 5	B1. Medium in need of editing	Forced scanning approach	Continue with more complex task/receipt, that we know needs editing.	Does the participant notice errors from scanning and does the participant edit errors?
Task 5-bis		Forced editing	A chance to edit errors.	
Task 6	C1. Personal	Choose approach	Continue with personal receipt. Which approach they choose now that we know that they are aware of both approaches.	Are considerations about sensitivity and personal information more prominent when sharing their own receipt? Do they notice errors from scanning? Do they edit errors?
Task 6-bis		Forced editing	A chance to edit errors.	

METHODOLOGY & PARTICIPANTS

TASK LIST *continued*

Optional, depending on time.

If used 'manual' method for task 6.

Task 7	C2. Personal	Force scanning	Continue with personal receipt. Which approach they choose now that we know that they are aware of both approaches.	Are considerations about sensitivity and personal information more prominent when sharing their own receipt? Do they notice errors from scanning? Do they edit errors?
Task 7-bis		Forced editing	A chance to edit errors.	

If used 'scanning method for task 6.

Task 8	B2. Long in need of editing	Forced scanning approach	Continue with more complex task/receipt, that we know needs editing.	Does the participant notice errors from scanning and does the participant edit errors?
Task 8-bis		Forced editing	A chance to edit errors.	
Platform specific				
Task 9	All receipts	Find and check Expense Overview	Navigation and usability	

METHODOLOGY & PARTICIPANTS

PARTICIPANTS

#	Sex	Age	Education	Phone
TP1	M	63	Secondary	Iphone 13
TP2	M	26	Master	Xiaomi Redmi Note 12S
TP3	F	42	Master	Sony Xperia 10V
TP4	M	62	PhD	Samsung A54
TP5	F	30	Master	Iphone SE
TP6	F	82	Lower	Samsung Galaxy A22
TP7	F	31	Master	Pixe l6
TP8	F	24	Master	iPhone 11
TP9	M	28	PhD	Samsung Galaxy A14

#	Sex	Age	Education	Phone
TP10	M	27	Bachelor	Pixel 6a
TP11	M	53	Lower	iPhone 11
TP12	F	52	Lower	Pixel 4a
TP13	M	25	Higher	iPhone 13
TP14	F	25	Bachelor	iPhone 12
TP15	M	25	PhD	iPhone 6
TP16	F	65	Secondary	iPhone SE 2020
TP17	F	29	PhD	iPhone X
TP18	M	32	Master	Samsung Galaxy A55

DETAILED FINDINGS

TASK

INSTALLING THE APP AND LOGGING IN USING CREDENTIALS IN INVITATION LETTER

INSTALLING THE APP

TASKFLOW

Three options to install the app

Download in
app store

1. By scanning the QR-code.

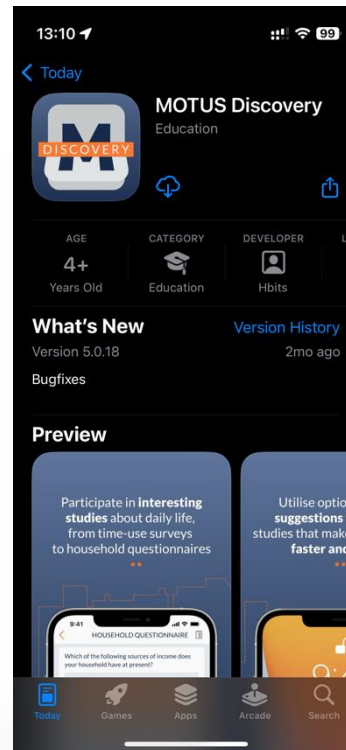


App Store



Play Store

2. By visiting the App Store of Google Play store and search for MOTUS Discovery.
3. By visiting <https://demo1-frontoffice.motusbuilder.io> and following the instructions.



INSTALLING THE APP

FINDINGS

⌘ All but one respondent use the QR code because:

- ⌘ it is fast
- ⌘ it is familiar
- ⌘ it prevents installing the wrong app

⌘ All but one respondent do not review app in store

- ⌘ too much time
- ⌘ never check review
- ⌘ if app is allowed in store, it must be trustworthy

Potential bias

Few respondents mention that reviews are not relevant because the app is part of a study that is conducted by a university.

ONBOARDING

TASKFLOW

Allow
notifications

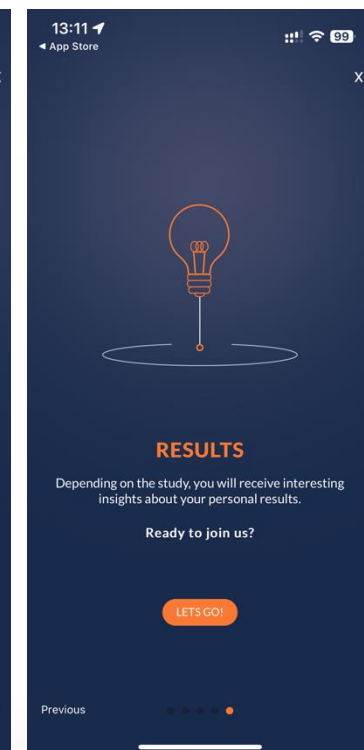
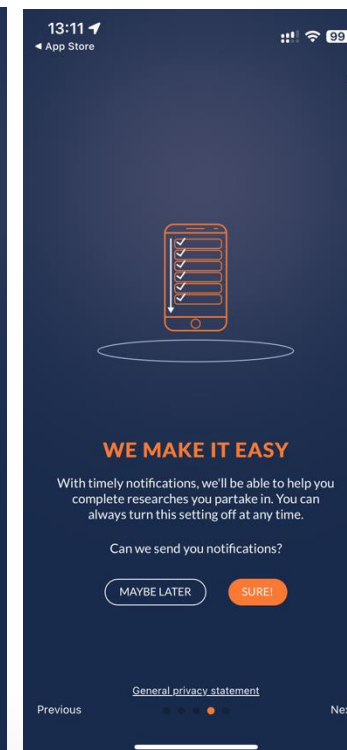
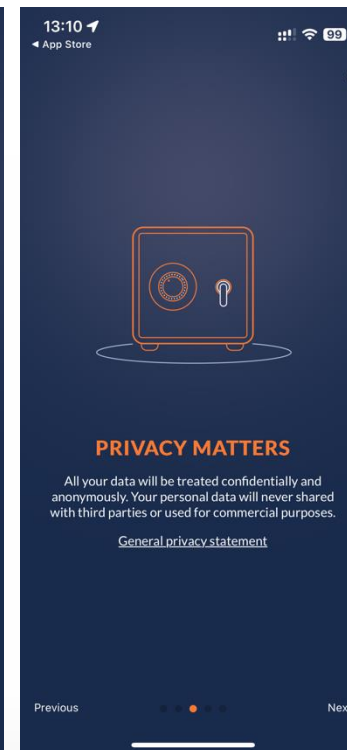
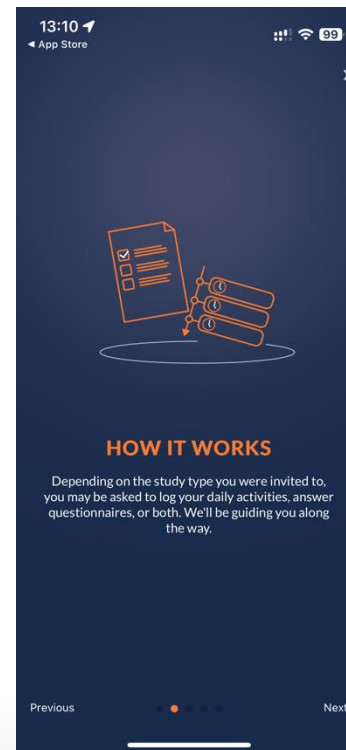
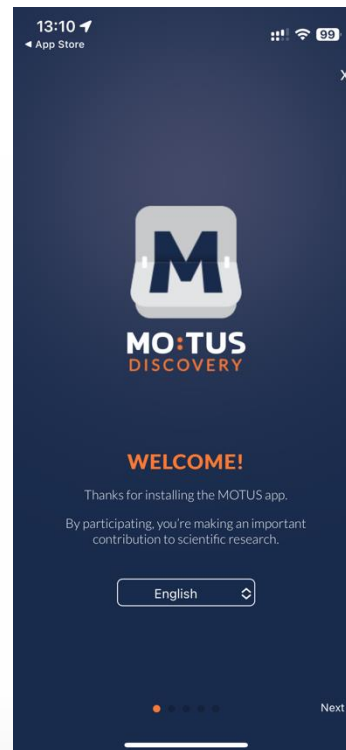
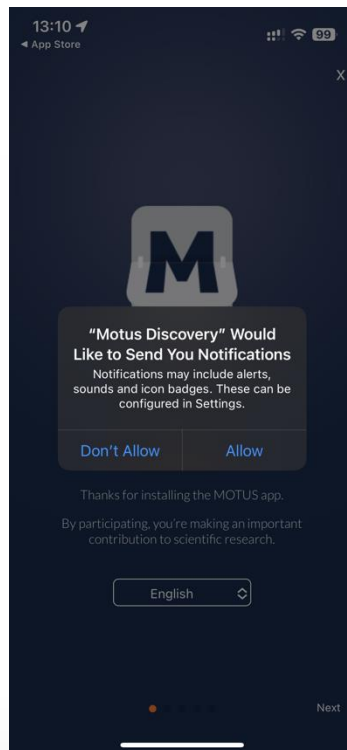
Info on
language

Info on
functionality

Info on
privacy

Info on
notifications

Info on
results



ONBOARDING

FINDINGS

- ⌘ The majority of the respondents does not allow MOTUS to send notifications
- ⌘ The majority of the respondents does not read the onboarding screens
 - ⌘ too much text and takes too much time
 - ⌘ chose to install the app so information is irrelevant
 - ⌘ those who did read the onboarding screens considered them relevant
- ⌘ None of the respondents read the privacy policy and several respondents tried to logon without accepting it

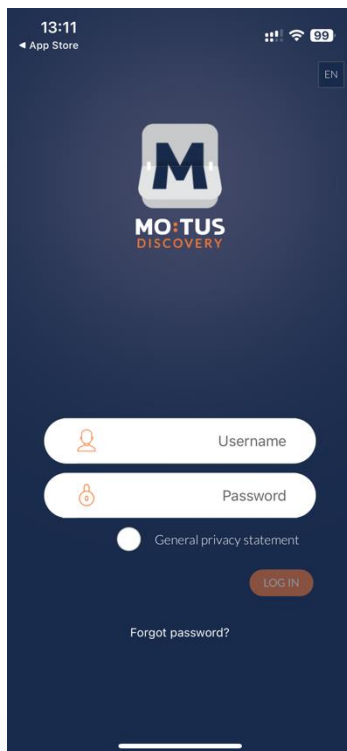
Privacy policy

Accepting the privacy policy is a formality and for some even an annoyance.

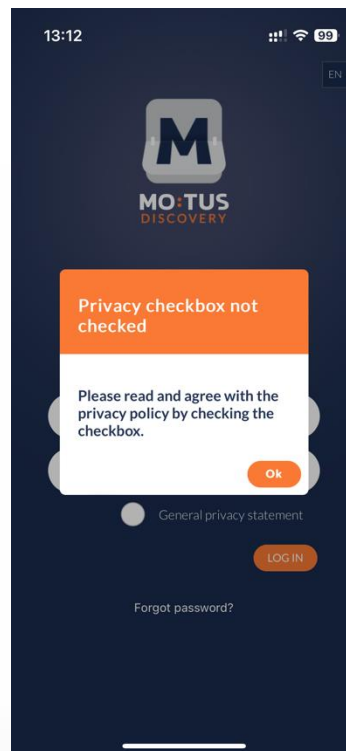
"It will probably be necessary to make the app work, right?" – TP14

LOGGING IN TASKFLOW

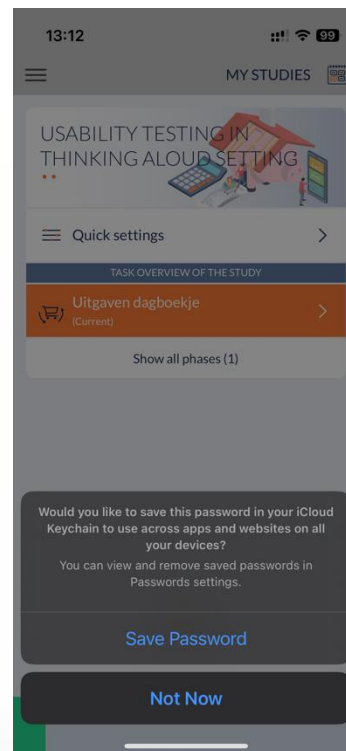
Enter
credentials



Accept
privacy policy



Save
password



ONBOARDING

FINDINGS

- ⌘ The majority of the respondents had no trouble copying the credentials from the letter to the app
 - ⌘ one older respondent started entering an email address
 - ⌘ one older respondent remarked that it would be good to use an eye-icon to check the password
- ⌘ Long and complex randomly generated usernames and passwords are not recommended
- ⌘ None of the respondents saved the password

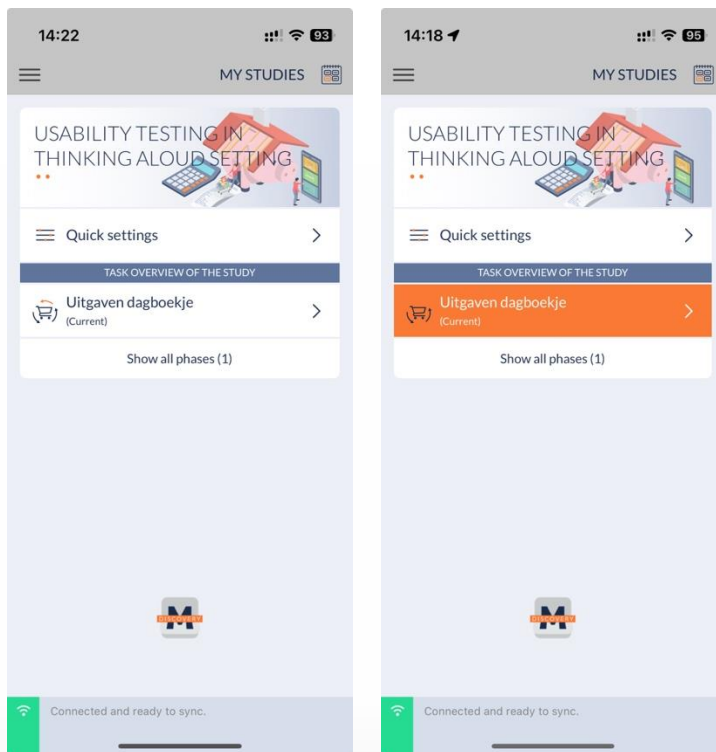
Credentials

This experiment used simple credentials. The username was 'test' followed by three numbers. The password consisted of six lowercase letters.

MY STUDIES

TASKFLOW

View home screen



Note: the expense diary turns orange when past the finishing date of the registration week

MY STUDIES

FINDINGS

- 🔗 The majority felt that the home screen was self-explanatory

Orange works

Due to a delay in testing, a few tests were run after the end of the registration week. On the 'my studies' screen, the expense diary task colours orange. Once explained, one respondent noted that an orange diary gave a better signal of what needed to be done/what to click on.

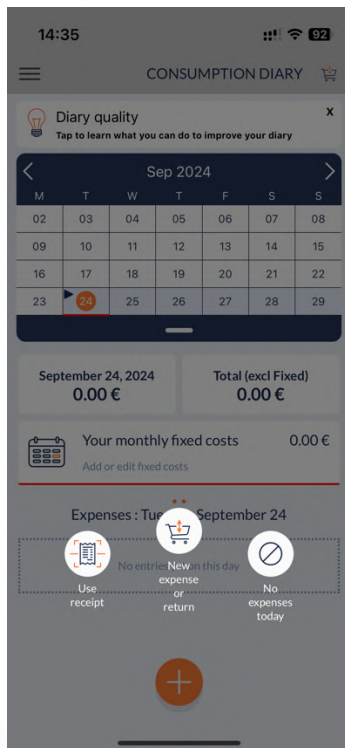
TASK

SCANNING A TICKET

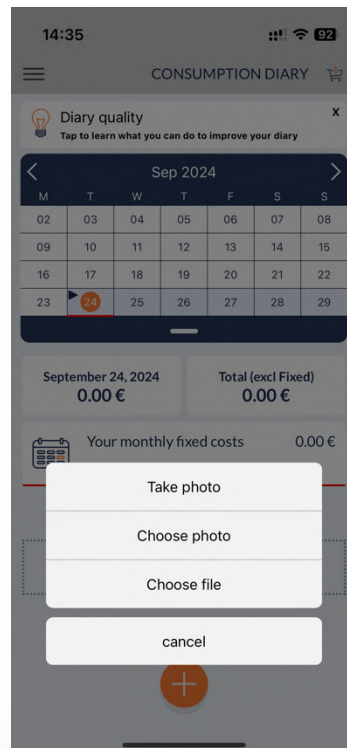
SCANNING A TICKET

TASKFLOW

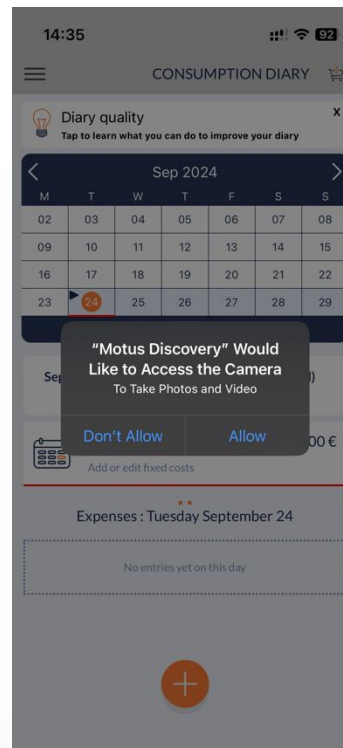
Click on '+'
and 'use ticket'



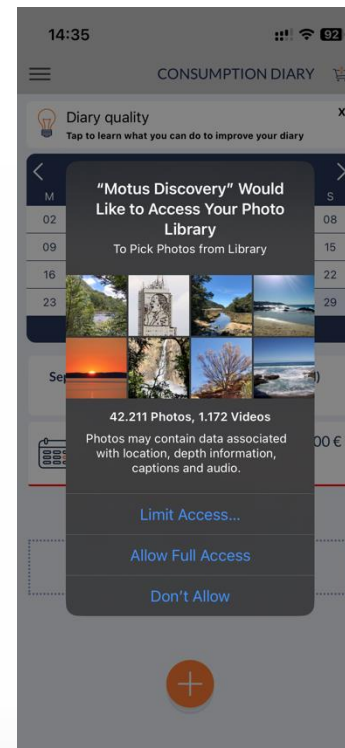
Click on
'take photo'



Click on
'take photo'



Manage access
to photos



(Re)take and
use photo



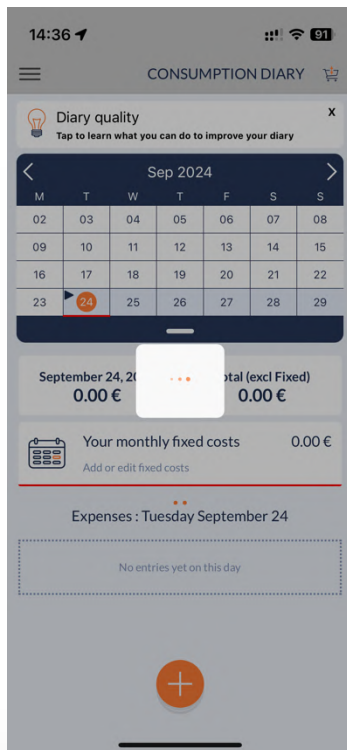
Crop photo

Note: version 5.0.20 does not show the cropping option anymore

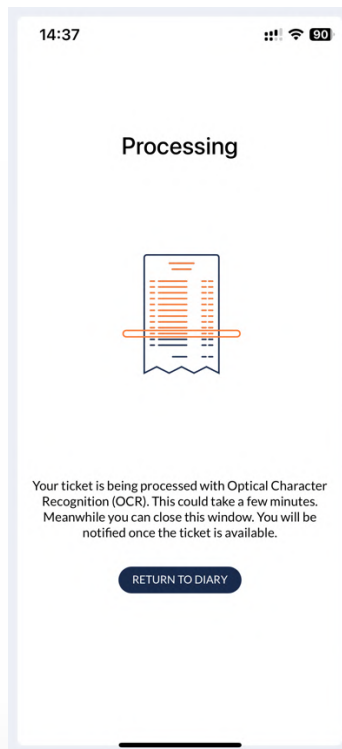
SCANNING A TICKET

TASKFLOW

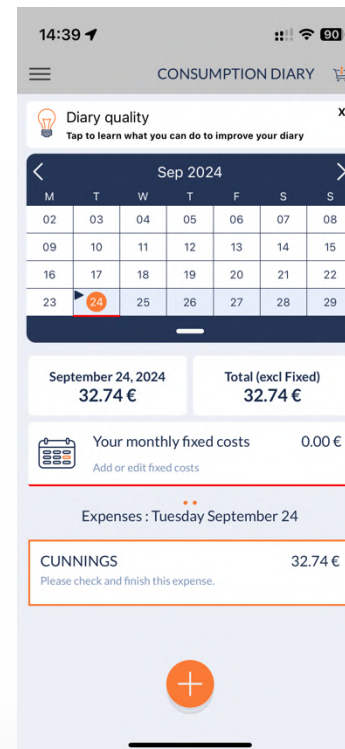
Wait for
information



Processing in
progress



Processing
complete



Scanned
ticket visible

SCANNING A TICKET

FINDINGS: GENERAL

- ⌘ All respondents prefer the scan option
 - ⌘ It is fast, familiar, and considered easier
- ⌘ Minor comments on scan option
 - ⌘ A two-click path is preferred (+ and then 📷)
 - ⌘ Concise wordings 'scan' vs 'photo' vs 'image'
 - ⌘ Cropping is unnecessarily complicated because only older respondents try to crop, yet they are least experienced in doing so
- ⌘ Expectations exceed current state of development
 - ⌘ Processing time is considered too long
 - ⌘ Too many errors in scan results

SCANNING A TICKET

FINDINGS: PROCESSING SPEED

- ⌘ The majority of respondents read the notification saying that processing might take several minutes.
- ⌘ Timed processing speed varies on average between 1 and 2 minutes with the occasional outlier longer than that.
- ⌘ The majority of respondents considered this too long and show signs of impatience and doubt after 20 to 30 seconds.
 - ⌘ Most respondents doubted the quality of the image taken and redid the scan resulting in multiple tickets
 - ⌘ Some started adding the ticket manually
 - ⌘ Some respondents wanted to keep scanning other tickets to keep going
 - ⌘ Other respondents mentioned they would give up if not for the sake of the current study



Several minutes

Albeit correctly communicated that processing takes several minutes, for the majority of respondents this took too long. Additional communication might be appropriate.

SCANNING A TICKET

FINDINGS: CORRECTNESS

- ⚡ The majority of the scanned tickets were incorrectly yet only slightly off. For most of the tickets, the cause of error was easily identifiable.

Nutri-boost discount on Delhaize ticket

- ⚡ Delhaize gives Nutri-boost discount on products with Nutri-score A or B
- ⚡ The discount is most often shown in a line directly under the product as a product with expense of 0€ and discount equal to the Nutri-boost discount

However, at least two problems are identified

- ⚡ If the discount is not picked up correctly, respondents cannot change this, because they get the error notification that the discount cannot exceed the product price (i.e. 0€)
- ⚡ The total discount is also summarized at the bottom of the ticket and often included as a product line with discount added as an expenditure

Training data

Training data is of utmost importance for the microservice to work.

Delhaize Boondael
Avenue du Bois de la Cambre 1
1050 Bruxelles
Tel: 02/672.89.48
TVA: BE0805034870

17/06/24 16:22 PR P-6 C-12 T-6075

0451061295421
1L OATLY HAVER 3,15
739437661650
NUTRI-BOOST -10% -0,31
1L OATLY HAVER 3,15
739437661650
NUTRI-BOOST -10% -0,32
375G OR FLAKES CHO 1,79
540011369936

Product line picked up correctly

Nutri-boost discount not picked up correctly

1L OATLY HAVER

NUTRI-BOOST -10\$

Name of expense item

NUTRI-BOOST -10\$

Type of expense item

tentative

Return or expense item

return item

Price	Discount	Quantity	Price
0.00 €	0.00 €	1	0.00 €

Keep item private

discard item from household budget

375G OR FLAKES CHO

10:03 NEW EXPENSE OR RETURN

NUTRI-BOOST -10\$

Type of expense
Free addition / I don't know

Quantity 1 Price 0,00 €

Invalid input

Total price of expense cannot be lower than discount. Change your expense, please.

Discount in Euro? 0,32 €

Is this a return item?

Total price: --- ADD

10:03 NEW EXPENSE OR RETURN

NUTRI-BOOST -10\$

Type of expense
Free addition / I don't know

Quantity 1 Price 0,00 €

Additional information (optional)

Does your expense consist of multiple items?
No, it's 1 item

Measuring unit of the item(s) 0 Unit kg

Discount in Euro? 0,32 €

Is this a return item?

Total price: --- ADD

Correcting the Nutri-boost discount yields an error message



DUPLICATE RECEIPT

Delhaize Boechout
Alexander Franckstraat 6
2530 Boechout
Tel: 03/454.32.97
BTW: BE 1000 790 669

29/05/24 14:45 PR P-1 C-56 T-3695

6X1L DDL MELK HALF 6,85

NUTRI-BOOST -10% -0,68

150G MELK KAMEL 2,69

AVERB BR NOT &ROZ 3,09

NUTRI-BOOST -10% -0,31

GELE PERZIK 500G 2,29

NUTRI-BOOST -10% -0,23

TACCHINO ARROSTO 5,69

360G DLL BOTERWAF 4,49

ODEND LIGHT BEL 6 5,59

BROCCOLI 1,49

NUTRI-BOOST -10% -0,15

TOTAAL 32,18

KORTING NUTRI-BOOST -1,37

TOTAAL 30,81

POI: 01937010
KLANTTICKET

DELHAIZE

Nutri-boost discount milk

Nutri-boost discount on bread

Nutri-boost discount on peaches

Nutri-boost discount on broccoli

Total Nutri-boost discount on ticket...

...included as expenditure

KORTING NUTRI-BOOST

Name of expense item

KORTING NUTRI-BOOST

Type of expense item

tentative

Return or expense item

☐ return item

☐ Keep item private

☐ discard item from household budget

Price	Discount	Quantity	Price
-------	----------	----------	-------

1.37		1	1.37
------	--	---	------

€	0.00 €		€
---	--------	--	---

SCANNING A TICKET

FINDINGS: CORRECTNESS *continued*

- ⚡ Invoices are challenging for the current state of development of the OCR microservice, although the cause of error was easily identifiable.

Three digits on Colruyt invoice

- ⚡ Invoices from the Colruyt provide prices with three decimals, which the OCR interpreted as thousands
- ⚡ These invoices also give prices excluding VAT, which is included at the end of the ticket, but not picked up by the OCR.

Invoices

Invoices are a specific type of ticket that often:

- ⚡ contains personal data
- ⚡ gives unit prices excl. VAT

ALGEMENE VERKOOPSVORWAARDEN:
zie keerzijde

De gegevens op het kassicket
worden door het Instituut
Quality Control op hun juistheid
en volledigheid gecontroleerd.

colruyt laagste prijzen

Colruyt Food Retail N.V.
Edingensesteenweg 196 - 1500 Halle
Tel. 02 345 2345 - www.colruyt.be
RPR Brussel
BTW-BE0716 663 615 - IBAN BE72 2930 2544 8916 SWIFT GEBABEBB

Tech ergens een lagere prijs gezien? Bel ons: 02 345 2345.
We passen onze prijs aan en betalen u het verschil terug.

276.2 L

Een smakelijke BBQ voor de laagste prijzen

Vind inspiratie, recepten, tips en heerlijke ingrediënten
voor al je barbecues op
colruyt.be/nl/barbecue.

Zaterdag 29/06/2024 11:05 Tel. winkel : 038 257178 Openingsuren: Uw kassabediende :
CADI WILRIJK Tel. beenhouwerij : 038 898576 Ma - Do : 08.30 - 20.00 Emma
JAARSTRAAT 34 Vr : 08.30 - 21.00
2610 WILRIJK Z : 08.30 - 20.00

Factuur

#	Art.Nr	Benaming	Leeggoed	Hoev.	Eenheidsprijs EXC. BTW	Bedrag EXC. BTW
A	11727	BONI BIO Koffie gemalen Peru FT 250g		4	3,953	15,811
A	7140	LIPTON Green Tea Intense Mint 20st		2	1,877	3,755
A	11879	LIPTON Black Tea Forest Fruits 20st		2	2,16	4,321
A	22256	LIPTON Mint infusion 20st		2	3,085	6,17
A	22257	LIPTON Infusion Camomile 20st		2	3,009	6,019
A	22258	LIPTON Infusion Rosehip 20st		2	2,16	4,321
		-----Dranken				
C	21729	BREF badkamerreiniger cit.spray 750ml		2	3,05	6,099
C	14074	BONI ECO ruitenspray 750ml		2	1,975	3,95
C	17032	NIVEA deo Fresh Natural 200ml		2	3,099	6,198
C	1311	NATURE BOX cd avocado/argan 385ml		4	3,826	15,306
		Professional Plus korting	3100722	30%		-5,56
		-----Non Food				
		Totaal bonnen Professional Plus kaart: € 5,56				
		BTWCode A C Totaal				
		BTW % 6,00% 21,00%				
		Dranken 40,40				40,40
		Non Food 31,55				31,55
		Excl BTW 40,40 31,55				71,95
		BTW 2,42 6,63				9,05
		Totaal korting leverancier (buiten BTW): € 5,56				
FOOD INC € 42,82			TOTAAL GOEDEREN € 71,95		TOT.LEEGGOED € 0,00	TE BETALEN € 75,44
207 231 043.00.07 3524			EUROCARD/MASTERCARD		€ 75,44	
FR9012 660066 - 98312016			Totaal betaald		€ 75,44	
			Teruggave		€ 0,00	

Personal data blacked out

3-digit prices are recognized as thousands

Total sum of 66.39€ (not shown on receipt) is calculated without VAT in microservice

Respondent needs to add VAT manually

Name of expense item

btw

Type of expense item

0000001 - Free addition / I don't know

Return or expense item

☐ return item

Price Discount Quantity Price

9.05 € 0.00 € 1 9.05 €

Keep item private

☐ discard item from household budget

SCANNING A TICKET

FINDINGS: CORRECTNESS *continued*

- ↪ The length or quality of the tickets partially determines the correctness of the results.

Wrinkled/folded tickets

- ↪ Wrinkled or folded tickets are especially hard to be processed correctly when folds go over the product prices

Long tickets

- ↪ Long tickets are harder to 'scan' because respondents physically take more distance from the ticket
- ↪ None of the respondents used the zoom function on the camera to capture a long ticket. It is therefore not known what this means for the quality of the ticket and the correctness of the processing

Communication

Additional communication could focus on:

- ↪ smoothing out tickets
- ↪ how to take image of long ticket



Example of drummed-up ticket that required substantial editing

SCANNING A TICKET

FINDINGS: CORRECTNESS *continued*

↪ The layout of tickets varies a lot.

Unit price vs. Total price

↪ Some tickets present unit price in a first column and total price (based on the unit/amount bought) in the second column. The OCR currently seems to prioritize the first column over the second column.

Language of tickets

↪ Albeit not part of this round of testing, it is worth noticing that some countries have more than one official language and tickets might either be presented in one of those languages or might be bi-lingual.

Training data

Training data is of utmost importance for the microservice to work.

LE GOUT

LE GOUT SRL
AVENUE DES AZALEES 40 - 1030 BRUXELLES
TVA: BE 0632.875.906

ADDITION

Date : 06.07.24 09:57 T: EMP.1
Garçon: LE GOUT

Qt	Description	PxUnit	Total C
1	BAGUETTE A L'ANCIEN		2,50 C
2	CROISSANT	1,60	3,20 C
1	JUS DE POMME AUBEL		3,90 C
322g	COMTE 24 MOIS	49,90	16,07 C

Total : 25,67

	Taux	Base	TVA	TTC
C	6%	24,22	1,45	25,67
CARTE				-25,67

TOTAL 25,67 EUR

** MERCI - A BIENTOT **

Unit price of *croissants*

Unit price of cheese

BAGUETTE A L&ANCIEN

CROISSANT

Name of expense item
CROISSANT

Type of expense item
tentative

Return or expense item	Price	Discount	Quantity	Price
return item	1.60 €	0.00 €	1	1.60 €

Keep item private
discard item from household budget

JUS DE POMME AUBEL

322gCOMTE 24 MOIS

Name of expense item
322gCOMTE 24 MOIS

Type of expense item
tentative

Return or expense item	Price	Discount	Quantity	Price
return item	49.90 €	0.00 €	1	49.90 €

Keep item private
discard item from household budget

TASK EDITING AND COMMITTING DATA

EDITING AND COMMITTING DATA TASKFLOW

Click on ticket

Click on 'pencil'

Check fields + warnings

Tap on product to complete

Repeat until no warnings

09:13

CONSUMPTION DIARY

Diary quality
Tap to learn what you can do to improve your diary

Apr 2025

07 08 09 10 11 12 13

April 10, 2025
96.30 €

Total (excl Fixed)
161.29 €

Your monthly fixed costs
0.00 €

Add or edit fixed costs

Expenses: Thursday April 10, 2025

ZUNNINGS 96.30 €
Please check and finish this expense.

+

09:13

CONSUMPTION DIARY

Diary quality
Tap to learn what you can do to improve your diary

Apr 2025

07 08 09 10 11 12 13

April 10, 2025
96.30 €

Total (excl Fixed)
161.29 €

Your monthly fixed costs
0.00 €

Add or edit fixed costs

Expenses: Thursday April 10, 2025

ZUNNINGS 96.30 €
Please check and finish this expense.

04/10/2025

+

09:15

NEW EXPENSE OR RETURN

store's country?
Australia

Where did you make the expense?
BUNNINGS
Speciality store

What was your expense?

4PK ULTIMATE LITH E304199000 15.83 € x

ENR CR1632 LITHIUM BATTERY 2PK 13.37 € x

20X15X5.5CM723300018 D 11.00 € x

420G FAST GRAB 7.25 € x

0328109 MINI ROLLER FRAME PRORENOVATOR 0.00 € x

Total amount: 47.45 €

SAVE

Checkout discount, coupons, empty bottle

09:16

NEW EXPENSE OR RETURN

Name of expense
paint

Type of expense
Consumer goods for drawing/painting purposes

Quantity 1 Price 5.55 €

Additional information (optional)

Does your expense consist of multiple items?
☒ No, it's 1 item

Measuring unit of the item(s) 0 Unit kg

Discount in Euro? 0.00 €

Is this a return item?

Total price: 5.55 €

UPDATE

09:17

NEW EXPENSE OR RETURN

When did you make the expense?
Thursday April 10 2025

Country? (If online expense: which is the store's country?)
Australia

Where did you make the expense?
BUNNINGS
Speciality store

What was your expense?

4PK ULTIMATE LITH E304199000 15.83 € x

ENR CR1632 LITHIUM BATTERY 2PK 13.37 € x

20X15X5.5CM723300018 D 11.00 € x

420G FAST GRAB 7.25 € x

paint 5.55 € x

Total amount: 53.00 €

SAVE

Checkout discount, coupons, empty bottle

SCANNING A TICKET

TASKFLOW

Click on
'update'

09:17

NEW EXPENSE OR RETURN

When did you make the expense?
Thursday April 10 2025

Country? (If online expense: which is the store's country?)
Australia

Where did you make the expense?
BUNNINGS
Speciality store

What was your expense?

4PK ULTIMATE LITH E304199000	15.83 €	X
ENR CR1632 LITHIUM BATTERY 2PK	13.37 €	X
20X15X5.5CM723300018 D	11.00 €	X
420G FAST GRAB	7.25 €	X
paint	5.55 €	X

Total amount: 53.00 €

SAVE

Checkout discount, coupons, empty bottle

Confirm all is
correct

09:18

NEW EXPENSE OR RETURN

When did you make the expense?
Thursday April 10 2025

Country? (If online expense: which is the store's country?)
Australia

Where did you make the expense?
BUNNINGS
Speciality store

What was your expense?

4PK ULTIMATE LITH E304199000	15.83 €	X
ENR CR1632 LITHIUM BATTERY 2PK	13.37 €	X
20X15X5.5CM723300018 D	11.00 €	X
420G FAST GRAB	7.25 €	X
paint	5.55 €	X

Total amount: 53.00 €

SAVE

Checkout discount, coupons, empty bottle

Did you check all expenses and can this ticket be considered as completed?

No Yes

Ticket is
committed

09:18

CONSUMPTION DIARY

Diary quality
Tap to learn what you can do to improve your diary

Apr 2025

M	T	W	T	F	S	S
07	08	09	10	11	12	13

April 10, 2025 53.00 €

Total (excl Fixed) 117.99 €

Your monthly fixed costs 0.00 €
Add or edit fixed costs

Expenses : Thursday April 10, 2025

BUNNINGS
Speciality store 53.00 €

04/10/2025

Insights

+

EDITING AND COMMITTING DATA

GENERAL FINDINGS

- ⌘ Editing the data can be confusing
 - ⌘ Discount causes issues when processed as product line
 - ⌘ Add context – **especially classification** – is unclear
- ⌘ Respondents check the total amount
 - ⌘ Often 'close enough' will do for them
- ⌘ Not always clear that data are tentative
 - ⌘ When respondents consider that the total amount is correct or will do, they consider their task done and do not commit the data

Tentative vs committed

The reasoning behind tentative and committed data from a privacy perspective and to give users control over their data is not well understood by respondents. Many see committing data as an extra – unnecessary – action. Communicating/explaining this is therefore of great importance.

EDITING AND COMMITTING DATA

EDITING DATA

- ⌘ Discount processed as a product line causes issues
 - ⌘ If the discount is incorrectly processed as 0 EUR and respondents try to edit the discount product line, they will get an error message that the discount cannot be greater than the product.
 - ⌘ Respondents then must delete the discount product line and 'add' the discount to the correct product.

Training data

Training data is of utmost importance for the microservice to work.

- ⌘ Editing or assigning a classification is not understood
 - ⌘ Respondents do not follow the logic of scientific classifications such as COICOP
 - ⌘ This is partially due to not knowing the level of detail they need to enter in the product search field to retrieve the correct category suggestion (see section on manual entry)

- ⌘ An ML algorithm to classify products could be useful
- ⌘ The question should be asked if classification is a task for respondents

Delhaize Boondael
Avenue du Bois de la Cambre 1
1050 Bruxelles
Tel: 02/672.89.48
TVA: BE0805034870

17/06/24 16:22 PR P-6 C-12 T-6075

0451061295421
1L OATLY HAVER 3,15
739437661650
NUTRI-BOOST -10% -0,31
1L OATLY HAVER 3,15
739437661650
NUTRI-BOOST -10% -0,32
375G OR FLAKES CHO 1,79
540011369936

Product line picked up correctly
Nutri-boost discount not picked up correctly

10:03 NEW EXPENSE OR RETURN

NUTRI-BOOST -10%

Type of expense
Free addition / I don't know

Quantity 1 Price 0.00 €

Invalid input

Total price of expense cannot be lower than discount. Change your expense, please.

Discount in Euro? -0.32 €

Is this a return item?

Total price: ---

10:03 NEW EXPENSE OR RETURN

NUTRI-BOOST -10%

Type of expense
Free addition / I don't know

Quantity 1 Price 0.00 €

Additional information (optional)

Does your expense consist of multiple items?
☒ No, it's 1 item
☐ Yes, _____ items

Measuring unit of the item(s) Unit
kg

Discount in Euro? -0.32 €

Is this a return item?

Total price: ---

Correcting the Nutri-boost discount yields an error message

1L OATLY HAVER

NUTRI-BOOST -10%

Name of expense item
NUTRI-BOOST -10%

Type of expense item
tentative

Return or expense item	Price	Discount	Quantity	Price
return item	0.00 €	0.00 €	1	0.00 €

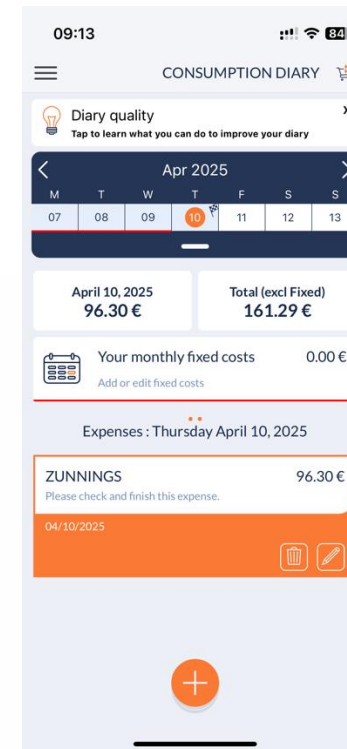
Keep item private
☐ discard item from household budget

375G OR FLAKES CHO

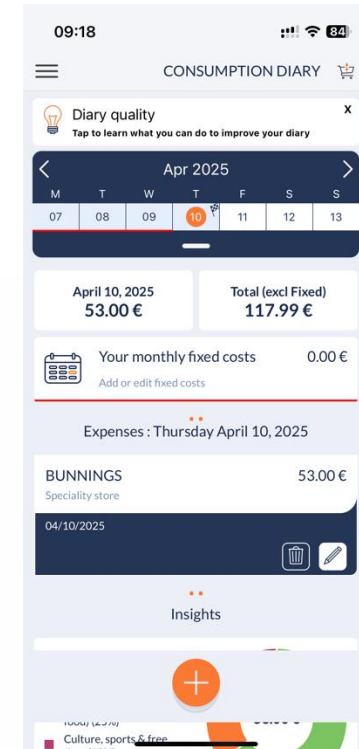
EDITING AND COMMITTING DATA

COMMITTING DATA

- ❖ Respondents do not read that they need to commit data
 - ❖ Confusion arises when the total amount is correct but in the expense-block it still reads that editing needs to be done
 - ❖ The colour coding – from tentative = orange to committed = blue – makes sense but only after it has been done/seen for the first time (i.e., it is not expected)



Tentative



Committed

TASK MANUAL ENTRY

MANUAL ENTRY

TASKFLOW

Click on '+' and
'use ticket'

14:35

CONSUMPTION DIARY

Diary quality
Tap to learn what you can do to improve your diary

Sep 2024

M	T	W	T	F	S	S
02	03	04	05	06	07	08
09	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29

September 24, 2024
0.00 €

Total (excl Fixed)
0.00 €

Your monthly fixed costs
0.00 €

Expenses : Tuesday, September 24

No entries New in this day
expense or return

No expenses today

Use receipt

+

Click on 'add an
expense or
return'

11:33

NEW EXPENSE OR RETURN

When did you make the expense?
Tuesday October 1 2024

Country? (If online expense: which is the store's country?)
Belgium

Where did you make the expense?
Delhaize
Supermarket / Hypermarket

What was your expense?

Add an expense or return

Total amount: 0.00 €

SAVE

Type 'product'
and 'select
category'

11:29

bread

Other bakery products made from bread dough, e.g. breadcrumbs, wafers made from bread dough, baking peas

Rye and mixed bread (without wholemeal bread)

Bread and wholemeal bread (incl. Pumpernickel)

White bread (without bread for toasting)

Nougat-based bread spread

Crisp bread

Bread without further designation

Cheese, processed, e.g. ready-made

no filters

Reset filters

CANCEL

CONFIRM

Provide details
about expense

11:29

NEW EXPENSE OR RETURN

Name of expense
bread

Type of expense
Bread and wholemeal bread (incl. Pumpernickel)

Quantity 1 Price 2.95 €

Additional information (optional)

Does your expense consist of multiple items?
☒ No, it's 1 item
☐ Yes, _____ items

Measuring unit of the item(s) Unit
0 kg

Discount in Euro? 0.21 €

Is this a return item? ☐

Total price: 2.95 € - 0.21 € = 2.74 €

ADD

Save expenses

11:32

NEW EXPENSE OR RETURN

When did you make the expense?
Tuesday October 1 2024

Country? (If online expense: which is the store's country?)
Belgium

Where did you make the expense?
Delhaize
Supermarket / Hypermarket

What was your expense?
bread 2.95 €
-0.21 €

Add an expense or return

Checkout discount, coupons, empty bottle refund?

Professional expense? ☐ Paid with creditcard? ☐

Total amount: 2.74 €

SAVE

Find expenses in
diary

11:32

CONSUMPTION DIARY

Diary quality
Tap to learn what you can do to improve your diary

Oct 2024

M	T	W	T	F	S	S
01	02	03	04	05	06	07
08	09	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	31	01	02	03	04

October 1, 2024
2.74 €

Total (excl Fixed)
35.48 €

Your monthly fixed costs
0.00 €

Expenses : Tuesday October 1, 2024

Delhaize
Supermarket / Hypermarket 2.74 €

+

MANUAL ENTRY

FINDINGS: GENERAL

❧ Unclear what is expected

- ❧ The level of detail that is required to register expenses is not always clear

❧ Bottom up approach

- ❧ Logic from scientific point of view but not (necessarily) the user point of view
- ❧ Downsides: loss of detail in case respondents only enter total amount and loss of control mechanism (i.e., sum of expenses matches total)

❧ Context confusing*

- ❧ Not all context is relevant or well understood (e.g. does bread needs to be added as quantity or grams)

❧ 'Done' button on numpad is missing

- ❧ To return to the diary field from the numpad, respondents need to tap outside the numpad to make it disappear. This is not a common action and causes confusion

Communication and in-app assistance can be an important element to manage respondents' expectations and responsibilities.

*Note that context questions can be toggled on/off in the back-office

Household budget diary

Tijddagboek BTUS24

COICOP CLASSIFICATION

Select the classification you want to use

OPTIONAL FIELDS

Entry overview

- ☐ Hide country question
- ☐ Hide where did you make this expense question
- ☐ Hide professional expense question
- ☐ Hide paid with credit card question

Expense details screen

- ☐ Hide quantity question
- ☐ Hide consist of multiple items question
- ☐ Hide measuring unit of the expense question
- ☐ Hide discount question

Setting context
questions in the back
office

MANUAL ENTRY

FINDINGS: ENTERING PRODUCT OR SERVICE

For the majority of respondents, it is unclear what they should enter in the search field.

⌘ Confusion about what to enter.

- ⌘ Several respondents interpret the Dutch translation of 'expenses' as the *total of their ticket*
- ⌘ As a result, they type *groceries* in the product/service search bar
- ⌘ One respondent typed the *total amount* of the ticket in the product/service search bar
- ⌘ There is a *delay in the search results* that is in some cases long enough for respondents to doubt their entry in the product/service search bar

⌘ Confusion about the level of detail to enter.

- ⌘ Without any examples or suggestions, it is often unclear what *level of detail* to enter in the product/service search bar.

"What do I need to enter? Food, fruit, bananas, or bananas from South America?" – TP

- ⌘ As a results, almost all respondents the *exact product name*, which does not yield sensible suggestions from the underlying COICOP classification list.

MANUAL ENTRY

FINDINGS: COMPLETION TIME

- ⌘ The learning curve is very steep
 - ⌘ Some respondents *needed help* from the interviewer to continue with the task
 - ⌘ The majority of respondents mentioned they *would not participate/continue* if not for the sake of this experiment
- ⌘ Even without the additional context, completing a ticket takes too much time
 - ⌘ The 'nutri-boost' discount on lots of products creates additional handling per product entry and respondents wonder why they cannot *enter a total discount*
 - ⌘ The *additional context is confusion* because not every product allows for the additional context to be completed
 - ⌘ All respondents raise *concerns about longer tickets*, which they agree are much more common for weekly grocery shopping

MANUAL ENTRY

FINDINGS: UI

- ⌘ Terminology/word usage leads to confusion among many respondents
 - ⌘ Some respondents *needed help* from the interviewer to continue with the task
 - ⌘ The majority of respondents mentioned they *would not participate/continue* if not for the sake of this experiment
- ⌘ Even without the additional context, completing a ticket takes too much time
 - ⌘ The 'nutri-boost' discount on lots of products creates additional handling per product entry and respondents wonder why they cannot *enter a total discount*
 - ⌘ The *additional context is confusion* because not every product allows for the additional context to be completed
 - ⌘ All respondents raise *concerns about longer tickets*, which they agree are much more common for weekly grocery shopping

TASK QUESTIONS ON TRUST, PRIVACY AND SECURITY

TRUS, PRIVACY AND SECURITY

GENERAL FINDINGS

⌘ There is no mistrust

- ⌘ Professional *design* and absence of advertisements enhance trust
- ⌘ *Trust in the institutions* issuing the app / conducting the survey enhances trust

⌘ Respondents have no problem entering expenses

- ⌘ Some respondents consider *leaving out sensitive expenses* (e.g. jewellery store, pharmacy)
- ⌘ Using researcher-issued ticket or personal ticket made no difference

⌘ Respondents are aware that receipts are stored on a server but have *no concerns about data privacy and data security*

BUG REPORT

BUG REPORT



A respondent scanned a ticket and after returning to the expense diary started adding another ticket manually. When the notification appeared that the processed ticket was ready, the respondent lost all the manually inputted data.



T.P.VAN.TIENOVEN@VUB.BE



BRUSSELS INSTITUTE
FOR SOCIAL AND
POPULATION STUDIES
Research Unit
Tempus Omnia Revelat (TOR)

SSI/WP2.3 Small Scale Tests/OCR - Report from Norway (V2)

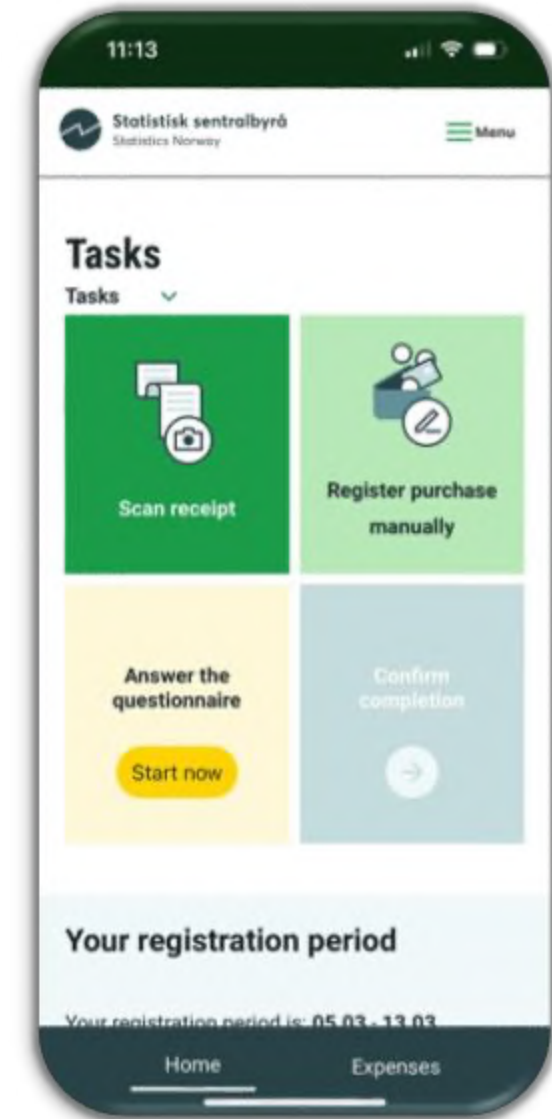
KARIANNE LUND AND NINA BERG, STATISTICS NORWAY, 22.08.24



Statistisk sentralbyrå
Statistics Norway

Content

1. Test design and selection of test respondents
2. Analyses:
 - General usability
 - Usability 1. Survey communication
 - Usability 2. Interpretations
 - Usability 3. Functionality - scanning and manual
 - Usability 4. Missing functionality
 - Usability 5. Functionality – other (important) issues
 - Engagement
 - Trust
 - Accuracy
 - Privacy
 - Target groups - special topics?
3. Experiences with test design
4. Strengths and limitations
5. Attachment



Test design

- Remote online tests, shared mobile screen
- 1 hour +/-
- Recruited thorough SSB's panel (visitors to ssb.no)
- Fieldwork week 22-23, 2024
- Incentives used
- All but one used their smart phone in the tests
- Moderated and recorded by two SSB methodologists

(See attachment for Test protocol (1), Interview guide (2), & Illustration of SSB's HBS 2022 web app (4) and Analysis method (5))



Selection of test respondents (N=11)

- Young

- 3 men
- Age 18, 24 and 28
- Overall very good technical skills

- Adults

- 4 women
- Age 50, 53, 53, and 60
- Overall good technical skills

- Seniors

- 4 men
- Age 62, 70, 74 and 76
- Overall good technical skills



Key characteristics of test respondents

Young

Male, 18, part of parents split HHs

Male, 24, technology optimist, strong desire to automate data collection and reduce manual efforts

Male, 28, emails SSB after the test to make sure its authentic

Adult

Female, 50, not Norwegian mother tongue

Female, 60, former social researcher, pt. in long term sick leave

Female, 53, advisor municipal, experienced & efficient digital user

Female, 53, independent interior designer, teenage son lives at home, experienced & efficient digital user

Seniors

Male, 70, (former) business entrepreneur (also IT), pro technical development

Male, 74, former advisor within property, struggled w technology, but not app, not that engaged in survey

Male 76, former museum conservator, have written surveys, critical to response burden and accuracy, used an advanced computer

Male, 61, former managing director in finance, will not use person ID nr for unimportant services (=surveys), sceptical to processed food

General usability

- Availability of/access to all receipts is a major problem. User testing indicates that respondents are not willing to invest extra efforts to search for, or recapture in detail what they bought – the result being missing information or very rough estimates with scarce details (at best)
- User testing suggest that some respondents would prefer and expect solutions for upload of digital expenses, both from memberships apps, debit cards and bills. This could contribute to lower the response burden and assist detailed reporting
- User testing indicate that when respondents review their scanned expenses, summary information is used actively including store name, date, total amount, but also number of products as this is summary information provided both in the app and on many receipts - hence easy to compare
- Respondents generally find manual registration burdensome, and it requires a high cognitive engagement compared to registration by scanning



Usability 1. Survey communication

- User testing clearly indicates that the test respondents find the first impression of the survey to be in line with what they would expect considering SSB as the surveyor. It is underlined that the design is tidy, easy to understand, minimalistic in a good way.
- The surveyor being SSB is the one main factor that contributes to the respondents feeling of trust to the survey. Several respondents explain that the sender being SSB makes them lower their guards and assume that they are in good hands. Being invited to the survey by SSB makes respondents less cautious to read introduction information thoroughly because they assume an invitation from SSB is serious.
- Login through Bank-ID is the second most contributing factor to respondents feeling of trust to the survey. User testing clearly indicates that the login requirement through Bank-ID adds to the perceived level of trust to the survey. It is expected as a requirement today for public services.
- Onboarding is short and sharp, but user tests indicate that respondents hurry through this in order to get to the task. Popup-tips are not well integrated – the few who reads it don't report any problems understanding the text, but struggle to see the connection/relevance to the task?
- User testing clearly indicates that the visual design of the login page where they can choose to continue in browser or store the app on their phones, does not communicate clearly that the «save as homescreen» is what SSB consider to be the preferred option, neither does it reveal how to practically proceed if choosing this option. Most respondents chooses to continue in the browser. This could be because there is only one action button at the bottom of the page? If we wanted people to save the solution to the home screen, there would have to be a button for that. So, two buttons at the bottom. "Continue to save on home screen" or "Continue in browser«. This is an example of a design that do not fully support what we want to communicate.
- User testing suggest that few respondents choose to navigate to the product list when a scanning is complete. At first, not everyone realize that there are more details collected from the receipt than what they can see on the overview screen that only shows date, store name and total amount. This design does not seem to support that we encourage respondents to control - and if necessary, edit – the registered expenditures. If we expect respondents to control and edit their scanned receipts, we need to communicate this clearly. One way that may contribute would be to send the respondents to the product list from the scan instead of the overview page. No prompt during scanning process to check and edit, TP expect system to make alerts.



Usability 2. Interpretations

«Take picture»

- User testing clearly indicates that the terminology "take a picture of the receipt" and "scanning" (progress bar) could have been reconsidered. Text while waiting for the scan could include more information about what is going on, for example "Please wait while we'll get the information from the receipt etc", "Wait a moment while we save your expenses" or something like that? Surely there are many other apps that we can be inspired by? Not necessarily intuitive that «take photo» means scanning. Some think that «take picture» means «take picture», and need to scroll down to see that it is scanning. Terminology "take picture" and "scanning" is not consistent throughout app.

«Save» and «add item»

- Clear evidence is found in the user tests, that respondents struggle with both identifying and interpreting the functionality in buttons «save» and «add item» buttons.

«Store name»

- User testing suggest that how the store name should be entered is interpreted quite differently, some enter only the store name, some include the location, some include the numerical code from the receipt, some enter themselves at the end because they think they are the payee.

«Most recent»

- User testing indicates that «most recent» can be interpreted as most recent *scanning/registration*, hence the order buys are presented does not necessarily make sense to all.

Pencil symbol

- We are made aware of, in user tests, that the symbol is used twice on the same screen but does not appear to mean the same thing. Inconsistent design?

Kilo products

- User testing reminds us that kilo products *may* be registered by respondents by calculating and adding the price per kilo instead of the amount from the receipts. There is however no reason pt to assume that this is a common misunderstanding.



Usability 3. Functionality - scanning and manual

Scanning

- All respondents choose to scan when given the option to scan receipt or register manually. User tests indicate that the reasons for this can be that the placement of the scanning option at the top left on the screen (reading direction) makes it the first element respondents see (hence perceives as “the first and best”), some say it's designed in a brighter, more compelling colour and yet some think that it sounds like a “cooler” way to register, compared to “register manually” (“manually” may be associated with “manual labour” which generally has connotations as burdensome). Most respondents realize that two methods are available.
- All respondents explain that they prefer scanning, first and foremost because its more time efficient. User tests clearly indicates that the scanning option reduces the response burden both in terms of how much time it takes to register expenditures, but also in terms of how respondents experience the cognitive effort required from them in order to provide a proper response. Respondents' response process are seriously simplified by scanning, compared to manual registration.
- User tests indicate that respondents tend to feel less obligated to tend to the correctness of – and be responsible for – details in expenditures that are “automatically” registered “by the system” (perceived to be = SSB), as opposed to expenditures that are manually registered.
- The waiting time for scanning is substantial. A long waiting time while the system completes the scanning may have more than one effect. User testing indicates that some respondents find the wait surprisingly long. A minority say they put up with it, as the app shows a process window. A long waiting time obviously contributes to increase the response burden, hence also increase the risk for dropouts from the survey.

Manual registration

- User testing clearly indicates that manual registration of expenditures is the least desired option. Some even say they would “never do this” if they were not “forced to do it”. This could mean that expenditures that are not documented by a paper receipt, will not be registered at all. This would have implications for the survey's coverage.
- When registering manually many errors occur.
- Some respondents say they assume that less details are needed in manual registration. They leave out detailed product specification, f ex type of milk (type of Biola). (They would not be willing to put in the required effort).
- Some respondents seem to feel in more control of the data when registering manually. This could imply a perceived reduced need to review and edit the data when a registration is complete.
- Respondents control of manually registered data may also give more opportunities to manipulate the data from the receipts as it is easier to knowingly overlook certain products that one does not want to share with the surveyor due to sensitivity reasons.



Usability 4. Missing functionality

Deleting a single product

- User testing indicates that respondents may find it helpful if a functionality was added for deleting a single product from the product list. This may seem risky from the surveyor's point of view, due to a higher risk of respondents deleting products that might make them look bad, typically alcohol, candy. Some state that it would be easier to delete an item that was wrongly scanned and rather add a new item manually.

«Sum» or «Total»

- User testing also suggest that respondents may benefit from more «sum» or «total» functions than in today's solution. This is mentioned on the overview page, to sum up all expenditures by date, but also on the «edit product» page.



Usability 5. Functionality – other (important) issues

Save/add item

- User test clearly indicates that respondents struggle to understand the logic of buttons for «save» and «add item». The confusion covers both the terminology used but also the placement of the buttons on the page. Several respondents explicitly prefer that the buttons we want them to use is placed in the reading direction. They find it unnatural to fill in information from left to right, from top to bottom, and then have to return to the top of the page for the action button they need in order to proceed. As a result of this, some edits information and find it frustrating that edits are not saved. Not everyone realize that they need to save both product item and receipt. This could result in respondents lack of motivation to review and edit.

Search words

- Search words may assist respondents if used actively but seems like test persons does not embrace the opportunity to reduce response burden. Can also confuse when the word typed does not match the available search word, it can give a feeling that one did something wrong.

Error messages

- Error messages are triggered often and for no reason at all. Even if none of the TR's comment on it, it may increase the experienced response burden and may also be considered disturbing visual elements with no apparent function. It does not seem logic that red printed error messages are shown on the screen before respondents has had a chance to avoid it by performing as intended. It contradicts the perspective several respondents have of the visual design of the app as easy to understand and minimalistic.



Engagement (1)

Motivation

- The already high response burden in the survey may make engagement harder than in any other survey. User testing indicates that to engage respondents in for example manual registration when needed, and reviewing and editing information, they need really convincing arguments for why they should invest the required effort.

Manual registration

- From user tests we have learned that respondents use the expenditures registered from scanning as a template for how the manual registration should be done. Hence, the detail level may be perceived as unreasonably high. It seems unbelievable that the equivalent level of detail is doable on manual registration.

Review and edit

- User testing clearly indicates that respondents rarely review, and even more rarely edit their registrations. Every action that requires extra efforts from the respondent, such as reviewing and editing scanned expenditures, (as well as collecting receipts from other HH-members, search for electronic receipts, control second sources etc.) risk not being done at all, or not being done very properly. However, willingness to check for errors may increase with the size of the amount – as small sums is perceived to not have enough impact.

Reading instructions

- User testing indicates that respondents reads very little of introduction info. NB: This may be a test effect – some read more because they are under observation, some read less because it's only a test and someone is waiting for them to continue.



Engagement (2)

Scan or manual

- User tests clearly indicate that respondents choose to scan receipts rather than register manually, when given the option.

Scanning

- Scanning encourages little engagement, partly because respondents land on the overview page after scanning, and there is no prompting to check. Most respondents also explain that they expect that little engagement is required when scanning – «have faith in the system», it's not their responsibility (not only an attitude, but a sincere expectation).

Search words

- Few (see and?) engage in using the search words actively, very few highlight this as a help and some even find it confusing, especially when there is a mismatch between the product name on receipt and available search word.

Trust – accuracy (1)

- Generally, respondents seem to perceive scanning as more accurate than manual registration, and they comment that it has more details.
- User testing clearly indicates that respondents may be overly confident in SSB and “the systems” ability to quality assure the accuracy of the data, especially when scanning – some review, but editing is rare. Most do not perceive checking the scanning as their responsibility.
- The general perception of the respondents seem to be that the respondents are more responsible for the accuracy when registering manually.
- Several suboptimal scans reduce the trust to accuracy, also difficulties to save edited information.
- App’s ability to scan, not their own memory and ability to report detailed, is considered when evaluating accuracy.



Trust – accuracy (2)

- It is pointed out that correct data requires a series of events, hence many risks for reduced accuracy – incl. motivated respondents that have access to all of their (and HH mbrs) expenditures and have the ability to register correctly etc..
- User testing indicates that respondents may be inclined to register big single product purchases manually to “be certain” that big purchases are correctly registered.
- Accuracy seem to be of little concern when the total amount is low, because it is thought to have little impact on the statistics.
- Concern for spelling accuracy in product names seem to be of respondents very least concern.
- User testing suggests that respondents may expect accuracy to be better if it was possible to upload data from purchases directly from web portals such as membership pages etc..



Trust – privacy (1)

- Respondent's perception of the survey being trustworthy seems to be mainly due to SSB as sender, but also highly supported by the Bank-ID login, a professional looking app with that seems to be a tidy and to the point survey communication.
- When prompted, respondents say they are normally careful about opening survey links and checking sources. They tend to feel embarrassed not showing such behavior in the test.
- User testing suggest that respondents may try to avoid being judged or wrongly “categorized” for what they buy. Products such as alcohol, candy, energy drinks are mentioned, as well as products from sex shops – the first category may be more relevant than the last (hypothetical – no one said they had a receipt from a sex shop).



Trust – privacy (2)

- Expenses nor receipts are generally not perceived as sensitive – it is information that can be shared
- Receipts rarely contain personal information (card numbers are not complete), some respondents briefly check this during the test (before/during the test) to be sure of this.
- Respondents have solid trust in SSB's ability to handle and store information according to privacy law.
- The topic of privacy is difficult and complex to understand, respondents seem uncomfortable or unfamiliar with discussing this in the tests.
- User testing clearly indicates that adjusting the receipts out of privacy concerns, such as cropping etc., are not considerations made.
- That SSB use webapp and scanning in their data collection does not affect perception of trust in SSB.



Target groups – special topics?

- 18-45 years
 - Issues with paper receipts & response burden.
- 46-59 years
 - Most willing to meet surveys requirements.
- 60 years+
 - Struggle more with technology in general and opening the survey (not how to use the app).
 - Most will use mobile if it is recommended, but preference for larger screens and large fonts grow with age.



Experiences with test design

- Too many technical things that could go wrong
- Digital receipts from us did not work very well – made test difficult, TPs involvement lower?
- Choice of receipts not good enough for the test – scanning results were inconsistent
- More engagement in personal receipts
- Difficult to uncover trust issues among (positive) test persons
- Decision on trust was made in connection with decision on participation – long before BankID and choice of registration method
- Analysis setup with full transcripts and notes of observations seem more demanding than worthwhile?
- How to structure data from interviews most efficiently? The Excel format ignores the flow and development of the conversation



Strengths and limitations

- Test persons who sign up for tests like this are often more likely to be pro-SSB, have higher level of trust towards government etc.
- Qualitative testing gives insight but does not provide estimates on how widespread experiences or events are
- Most receipts used in tests are provided by SSB, this may affect how respondents solved the tasks
- Only one gender covered in each age group
- Analyses may be affected by former, related testing on the same survey (pro and con)

Attachments

Attachment list

1. Test protocol
2. Interview guide
3. Receipts used in test (clear and long receipts)
4. Illustration of SSB's HBS web app:
 - a) Description
 - b) Video screen of app's workflow
 - c) Screenshots
 - d) Workflow description in Excel
5. Analysis method and summary per test person

1. Test protocol

- See test protocol/[Protocol SSI small scales usability tests](#)
- Main purpose to gain insight in the following topics for smart surveys:
 1. Citizen's involvement and engagement in smart surveys
 2. Acquisition, processing and combining of data
 3. Trustworthiness and privacy safeguards
- Test methods:
 1. Observation and Think Aloud (TA) method while test person solves test tasks
 2. Debrief or retrospective interview after TA test



2) Interview guide

- Structure:

- I. Observation and TA of usability test of recruitment and trust

- 1. Warm-up
 - 2. Installation and onboarding instructions
 - 3. Short and clear receipt – choose approach (A1)
 - 4. Short and clear receipt – forced inversed approach (A2)
 - 5. Medium/long receipt – scanning of receipt in need of editing (B1)
 - If did not find, check, and edit: Force editing
 - 6. Personal receipt – choose approach (C1)
 - If not scan: Force scanning
 - If did not find, check, and edit: Force editing

If time:

- 7. Find and check all receipts registered in Expense Overview
 - If not scan: Force scanning

- II. Debrief/retrospective after all tasks are completed

- III. Trust

- Accuracy and credibility
 - Trust/Privacy and data security

- Interview guide:

- Norwegian: [SSI_WP2_3_Guide OCR_Norwegian](#)
 - English: [WP2_3_Interview guide_Trust_SSB/English](#)

English interview guide

SSB

Interview Guide – SSI/WP2.1

Usability test of recruitment and trust

Bold black = Read out
BLUE CAP = Instructions to moderator
Blue = Possible probe questions

INTRO (5 min)

Hello! First, I'd like to say that we're appreciate that you want to participate in this test of Statistics Norway's (SSB) Household Budget Survey app. SSB use this survey to produce statistics on spending habits in Norway.

My name is ... and I will be leading this interview.

The interview will take approximately an hour.

After the interview, we will send you a € 30 gift certificate that can be used in a wide range of stores.

The purpose of this interview is to get an understanding of what works well and what doesn't. We've prepared some tasks and questions for our conversation today that we would like to hear your thoughts on. Note that I have not been involved in making the app we will test, so please don't hesitate to make critical comments. In fact, we are specifically interested in any issues or obstacles you may see, as we aim to learn from problems and improve the app. Therefore, we encourage you to share your honest opinions with us.

ENGAGE TP

Please tell me, have you participated in any user tests like this earlier? How was that?

IF "NO":

How about surveys, do you participate in surveys from time to time?

Can you remember any one of them? Please tell me about the most recent one, what was it about?

CONCENT

To make it easier for me to pay full attention to the test, I'd like to record it. That allows me to review it later to control what was actually said and done. These recordings are only accessible to a selected few and will be treated with confidentiality. We will not quote names or use images from them, and they'll be deleted after three months.

Is it okay for us to record?

SSB

Do you have any questions, or shall we proceed with the test?

[START RECORDING](#)

INSTRUCTIONS (5 min)

In today's test we have some tasks for you to solve in our Household Budget Survey app. These are realistic tasks that survey respondents could encounter in the app.

I will give you the tasks one by one. I'd like you to complete each task in the app without my interference or assistance as we're aiming for as realistic an experience as possible. I wouldn't be there to help if you were doing this at home either. Note that we are testing the usability of the app, not you. I would like to know what you are thinking and what considerations you are making to complete each task. Please comment on buttons you're unsure about, words that are difficult to understand, or anything else you notice in the app. I would appreciate it if you [speak](#) your thoughts out loud and explain as much as possible of what you see and do, and I'll observe and listen to what you say.

TASKS (25 min)

TASK 1: WARM-UP

[READ OUT TASK 1](#)

Before we continue, let's do a short practice task, just to get accustomed to the practice of thinking aloud. Please use any website to find the price for a flight ticket for one adult from Rome to London on the 1st of May. Please share your thoughts with me as you progress.
(Probe if necessary to aid test respondents think-aloud.)

INTRODUCTION TO TASKS

I have about six tasks for you today.

After each task let me know when you are done, and I will give you the next task.

Please take all the time you need to complete each task and "think aloud" while you work.

We will go through all tasks together and discuss them in greater detail when all the tasks are completed.

Any questions before we continue?

Okay, then let's start.

[READ AND HAND OUT TASK 2.](#)

Interview guide cont.:

SSB

TASK 2: INSTALLATION AND ONBOARDING INSTRUCTIONS

Please, open app, (start screen sharing,) login, follow instructions in the app, and go to start screen.

<http://consumption.ssb.no/>

OBSERVE

- How did downloading and login go? Does TP encounter any problems? Anything that surprises TP?
- Does TP read onboarding instructions?
- Do TP stop at consent? Read it? Object?

IF TP ASKS FOR ASSISTANCE:

Say Please, do your best to solve the task without my help, as I would not be here if this was a real survey. Wait for TP's initiative.

Assist only if necessary to move test forward.

Unless the only other option for you (considering the obstacle you experience now) is to drop out of the test, I urge you to try once more before I help you. Maybe with a different approach this time? If another try does not get you anywhere, I will of course help. Ok?

Give TP instructions on what to click on to solve first step of the task.

Wait for TP to proceed further.

If needed, ask what they think is necessary to do to proceed to next step/complete the task. Remind them to think aloud.)

DEBRIEF/RETRO BEFORE TASKS CONTINUES

Before I give you the next task, how did the login go? Why was it difficult or easy to log in? Explain?

What did you think about the log in method (Altinn/sms with hyperlink)?

Do you have any thoughts on the information you received in the app?

READ AND HAND OUT TASK 3 WITH RECEIPT A1.

TASK 3: RECEIPT A1 – SHORT & CLEAR – CHOOSE APPROACH

Please enter receipt marked "A1".

SSB

REMIND TP TO "THINK ALOUD" AND EXPLAIN WHAT THEY SEE ON THE SCREEN.

OBSERVE

- Does TP choose manual or scanning? Does TP encounter any problems? Anything that surprises TP?
- Does TP crop image?
- Does TP consider to hide/delete personal information? Do they save ticket/data?
- Does TP look at the data saved/the overview? Do they check and edit?

READ AND HAND OUT TASK 4 WITH RECEIPT A2.

TASK 4: RECEIPT A2 – SHORT & CLEAR RECEIPT – FORCED INVERSED METHOD

Please enter receipt marked "A2" [manually/by scanning] this time.

REMIND TP TO "THINK ALOUD" AND EXPLAIN WHAT THEY SEE ON THE SCREEN.

OBSERVE

- Do TP encounter any problems? Anything that surprises them?
- What method do they seem to prefer?
- Do they look at the data saved/the overview? Do they check and edit?

READ AND HAND OUT TASK 5 WITH RECEIPT B.

TASK 5: RECEIPT B1 – MEDIUM RECEIPT/SCANNING/NEED EDITING

Please scan receipt marked "B1".

REMIND TP TO "THINK ALOUD" AND EXPLAIN WHAT THEY SEE ON THE SCREEN.

OBSERVE

- Does TP comment on registration method? Notice any preferences or problems with the method?
- Does TP look at the data saved/the Overview?
- Does TP check and edit?
- Does TP notice any problems?

Interview guide cont.:

SSB

IF TP DID NOT FIND, CHECK, EDIT RECEIPT:

READ AND HAND OUT TASK 5-bis (receipt B1)

TASK 5-bis: FORCED EDITING

Please proceed to make sure the entry is correctly registered.

REMINd TP TO "THINK ALOUD" AND EXPLAIN WHAT THEY SEE ON THE SCREEN.

OBSERVE

- o Does TP find and look at the saved data?
- o Does TP notice any problems with the scanning?
- o What kind of editing do they say is need/do they do?
- o If none, what are the considerations?

READ AND HAND OUT TASK 6.

TASK 6: PERSONAL RECEIPT C1 – CHOOSE APPROACH

Please register the receipt you brought with you (Receipt "C1").

REMINd TP TO "THINK ALOUD" AND EXPLAIN WHAT THEY SEE ON THE SCREEN.

OBSERVE

- o Does TP comment on considerations about privacy and sharing personal data?

TASK 6-bis: FORCED EDITING/CHANCE TO EDIT

Please proceed to make sure the recording is correctly registered.

REMINd TP TO "THINK ALOUD" AND EXPLAIN WHAT THEY SEE ON THE SCREEN.

OBSERVE

- o Does TP find and look at the saved data?
- o Does TP notice any problems with the recording?
- o What kind of editing do they say is need/do they do?
- o If none, do they express any considerations?

IF TIME

If used manual method for task 6.

SSB

READ AND HAND OUT TASK 7.

TASK 7: PERSONAL RECEIPT C2 – FORCED SCANNING APPROACH

Please scan one of the receipt you brought with you (Receipt "C2").

REMINd TP TO "THINK ALOUD" AND EXPLAIN WHAT THEY SEE ON THE SCREEN.

OBSERVE

- o Does TP comment on considerations about privacy and sharing personal data?

TASK 7-bis: FORCED EDITING

Please proceed to make sure the scan is correctly registered.

REMINd TP TO "THINK ALOUD" AND EXPLAIN WHAT THEY SEE ON THE SCREEN.

OBSERVE

- o Does TP find and look at the saved data?
- o Does TP notice any problems with the scanning?
- o What kind of editing do they say is need/do they do?
- o If none, do they express any considerations?

If used scanning method for task 6.

READ AND HAND OUT TASK 8.

TASK 8: LONG IN NEED OF EDITING RECEIPT B2 – FORCED SCANNING APPROACH

Please scan the receipt you brought with you (Receipt "B2").

REMINd TP TO "THINK ALOUD" AND EXPLAIN WHAT THEY SEE ON THE SCREEN.

OBSERVE

- o Does TP comment on considerations about privacy and sharing personal data?

TASK 8-bis: FORCED EDITING

Please proceed to make sure the scan is correctly registered.

REMINd TP TO "THINK ALOUD" AND EXPLAIN WHAT THEY SEE ON THE SCREEN.

Interview guide cont.:

SSB

OBSERVE

- Does TP find and look at the saved data?
- Does TP notice any problems with the scanning?
- What kind of editing do they say is need/do they do?
- If none, do they express any considerations?

READ AND HAND OUT TASK 9.

TASK 9: FIND AND CHECK EXPENSE OVERVIEW

Please try to get an overview of all the expenses you have entered.

REMIND TP TO "THINK ALOUD" AND EXPLAIN WHAT THEY SEE ON THE SCREEN.

OBSERVE

- Does TP find and look at the saved data/Overview?
- Do they do any checking and edit?
- What kind of editing do they say is need/do they?
- If none, what are the considerations?
- Do they see that there also is a chart with overview of their registrations?

SSB

DEBRIEF AFTER ALL TASKS ARE COMPLETED (25 min)

RETROSPECTIVE

GO THROUGH EACH TASK AND ASK FOLLOW-UP QUESTIONS BASED ON OBSERVATIONS.

Let's go through each task.

TASK 2: INSTALLATION AND ONBOARDING INSTRUCTIONS

First, what was your first impression of the app when you first logged in? Please, tell me what made you think that.

Did the first impression make you trust the app or not?

PROBING:

- Have you logged in to an app like this before?
- What did you think of the two-step login with ID-porten/Bank-ID? Are you familiar with it? Do you trust it is secure? Probe for examples of apps that use it.
- How much or how little of the information did you read? What do think about the amount of information? Too scarce or too elaborate? What did you think about the consent required?

Interview guide cont.:

SSB

TASK 3 and 4: RECEIPT A1 – SHORT & CLEAR – CHOOSE APPROACH & FORCED INVERSED

How did you experience entering the first ticket I gave you?

What was easy and what was difficult with adding receipt A1? What did you like? What did you not like?

Please explain your thinking before you started registering expenses from the first ticket – what were your considerations at that point?

How did you decide which entry method to use?

Did you realize that there were two different methods to record receipts? How easy/hard was it to choose? Why?

What did you find most unexpected as you entered these expenses?

If relevant:

Please describe how you proceeded to prepare the receipt before scanning it.

Please share your thoughts. Did you consider to prepare or edit the receipt before scanning? Why did you try/not try to crop or edit the picture of the receipt? What was the reason for this, and what type of information was this about?

Which entry method did you prefer? Why? What are the arguments for and against using either manual entering or scanning? What kind of receipts would be better to scan, and what kind would be better to register manually, in your point of view?

What do you think happens with the information from the receipt when you say “save”?

Would this thought have any effect on your decision to scan or not receipts?

TASK 5 and 5bis: RECEIPT B1 – MEDIUM RECEIPT/SCANNING & NEED OF EDITING

How did you experience scanning the second ticket I gave you go?

What was easy and what was difficult with task 5/receipt B1?

What did you like? What did you not like? What surprised you at this point?

If errors were checked/corrected:

When and how did you notice that the scan did not read the receipt correctly?

What made you decide to edit the errors you noticed? What was easy and what was difficult regarding the correction of the results?

If errors were not corrected:

What did you expect to have to do after scanning the receipt?

SSB

TASK 6 and 6bis: PERSONAL RECEIPT C1 – CHOOSE APPROACH & CHECK/EDITING

What did you think of entering the ticket you brought with you? What was easy and what was difficult with task 6/receipt C1? What did you like? What did you not like? Anything unexpected or peculiar?

Why did you choose scan/manual registration?

IF MANUALLY:

Did your choice to manually enter your purchases have anything to do with the receipt being a personal receipt? How?

ALL:

How do you feel about entering your own personal receipt in this app? Did you make any considerations about personal or private information on the receipts?

IF TIME

If used manual method:

TASK 7 and 7bis: RECEIPT PERSONAL C2 – FORCED SCANNING & EDITING

How did you experience scanning a personal ticket?

What was easy and what was difficult with task 7/receipt C2?

What did you like? What did you not like? What surprised you at this point?

If errors were checked/corrected:

When and how did you notice that the scan did not read the receipt correctly?

What made you decide to edit the errors you noticed? What was easy and what was difficult regarding the correction of the results?

If errors were not corrected:

What did you expect to have to do after scanning the receipt?

If used scanning method:

TASK 8 and 8bis: RECEIPT B2 – LONG RECEIPT/FORCED SCANNING & EDITING

How did you experience scanning the fourth ticket I gave you go?

What was easy and what was difficult with task 8/receipt B2?

What did you like? What did you not like? What surprised you at this point?

If errors were corrected:

When and how did you notice that the scan did not read the receipt correctly?

What made you decide to edit the errors you noticed? What was easy and what was difficult regarding the correction of the results?

If errors were not checked/corrected:

What did you expect to have to do after scanning the receipt?

TASK 9: FIND AND CHECK EXPENSES

How did you find the overview of expenses?

What was easy and what was difficult with finding and checking the overview of expenses?

What did you like? What did you not like? Anything unexpected or peculiar?

What do you think of the overview of expenses can be used for?

Did you also see the overview chart?

IF SO:

Please share your thoughts about this chart? What did you like? What did you not like?

Anything unexpected or peculiar?

What do you think the chart can be used for?

TRUST

Overall, what contributed to your trust or distrust of the app you tested today? What makes you say that?

How does the option to scan and upload receipts affects your trust of the app?

To what extent do you think the data collected in the app will be correct or reliable in a real survey? What makes you say that? What is the reason?

How does the option to scan and upload receipts affects your opinion of the correctness or reliability of the data collected through the app?

And how about manually entry? How correct or reliable do you think manual registration will be? And which of the methods will in your opinion catch expense data most correctly?

How about personal information, to what extent do you think the app processes personal data in a secure and private manner or not?

How does the option to scan and upload tickets affects your judgement of the security and privacy guarantees of the app?

Which factors in your opinion reduce the trustworthiness of the data or statistic to be produced by this app?

Do you have any additional comments, concerns or suggestions you would like to share?

Background Questions and Conclusion

MAKE NOTES:

- Gender
- Age
- Household size (1/2/3+)
- Children under 18 years (Yes/No)
- Employment status (Employed/Unemployed/Disability pension/Retired)
- Native language
- Perception of digital skills (High/Low) (should be revealed by test, or ask if TP uses computers at work/Use many apps on their phone/Consider themselves digital proficient)

Thank you for participating!

We will send a € 30 gift voucher (We have address/email for delivery).

3. Receipts used in test

Task 3: Clear receipt (Kiwi (A1)):

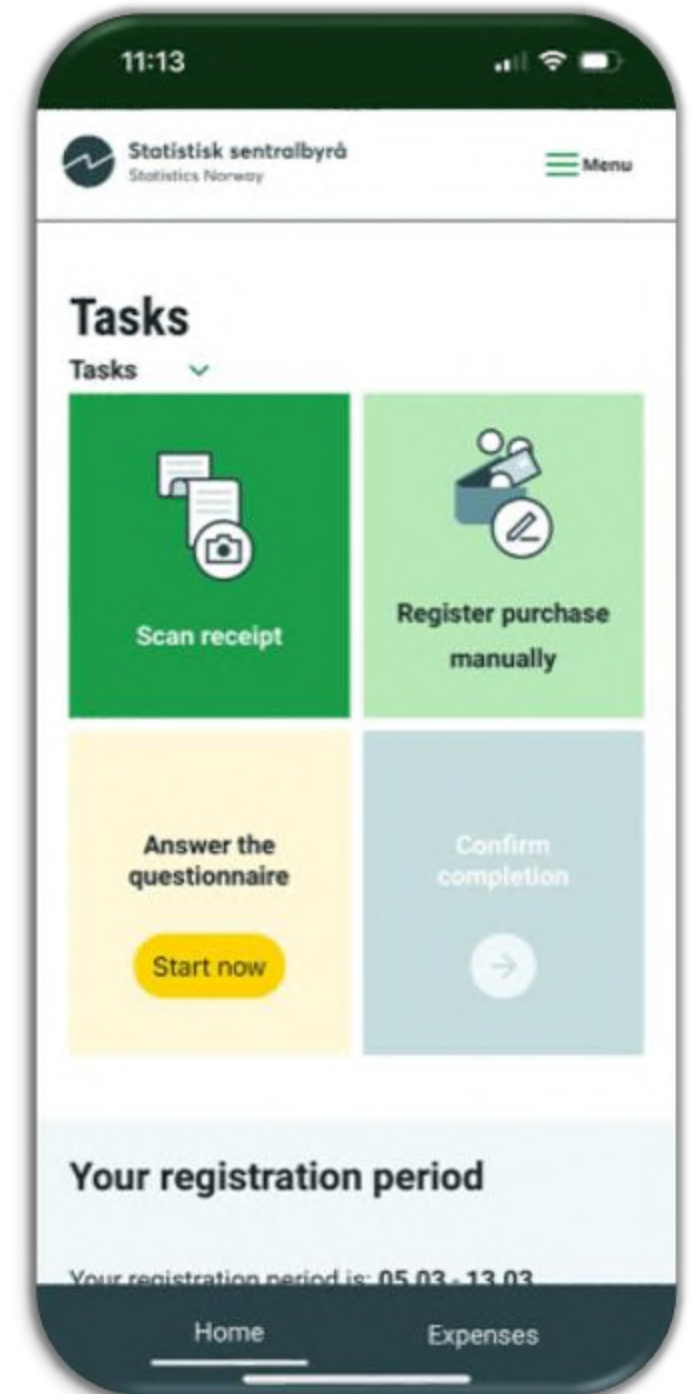


Task 5: Long receipt (Rema1000 (B1)):



4a) SSB's HBS web application

- Opens from a weblink in any internet browser
- Independent of platform and device
- Secure login with a two-step authenticator
- Registration options: 1) Scan receipt or Register purchase manually
- Editing options – but no prompting/requirement
- Usability in focus



4b) Video screens of app's workflow

1. Login:

- [Test login & onboarding](#)

2. Clear receipt:

- [Receipt \(clear\) Rema1000 \(groceries\)](#)
- [Scanning Rema1000 \(clear\) video](#)
- [Manual reg McDonalds video](#)

3. Receipt need editing due to incorrect # of prod items or due to discount:

- [Receipt \(edit\) Kiwi \(groceries\)](#)
- [Scanning Kiwi \(edit\) video](#)
- [Editing Kiwi \(edit\) video](#)

4. Find Expense overview

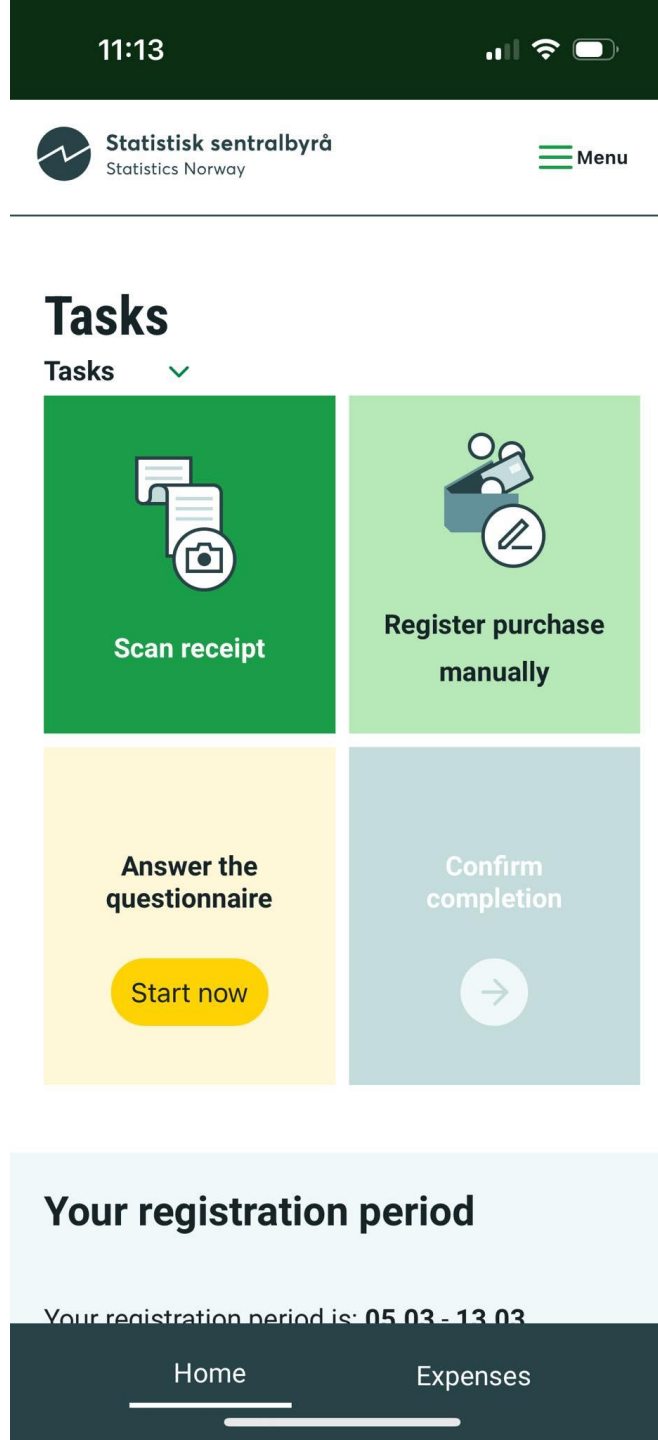
- [Expense overview screen](#)
- [Find Expense overview](#)

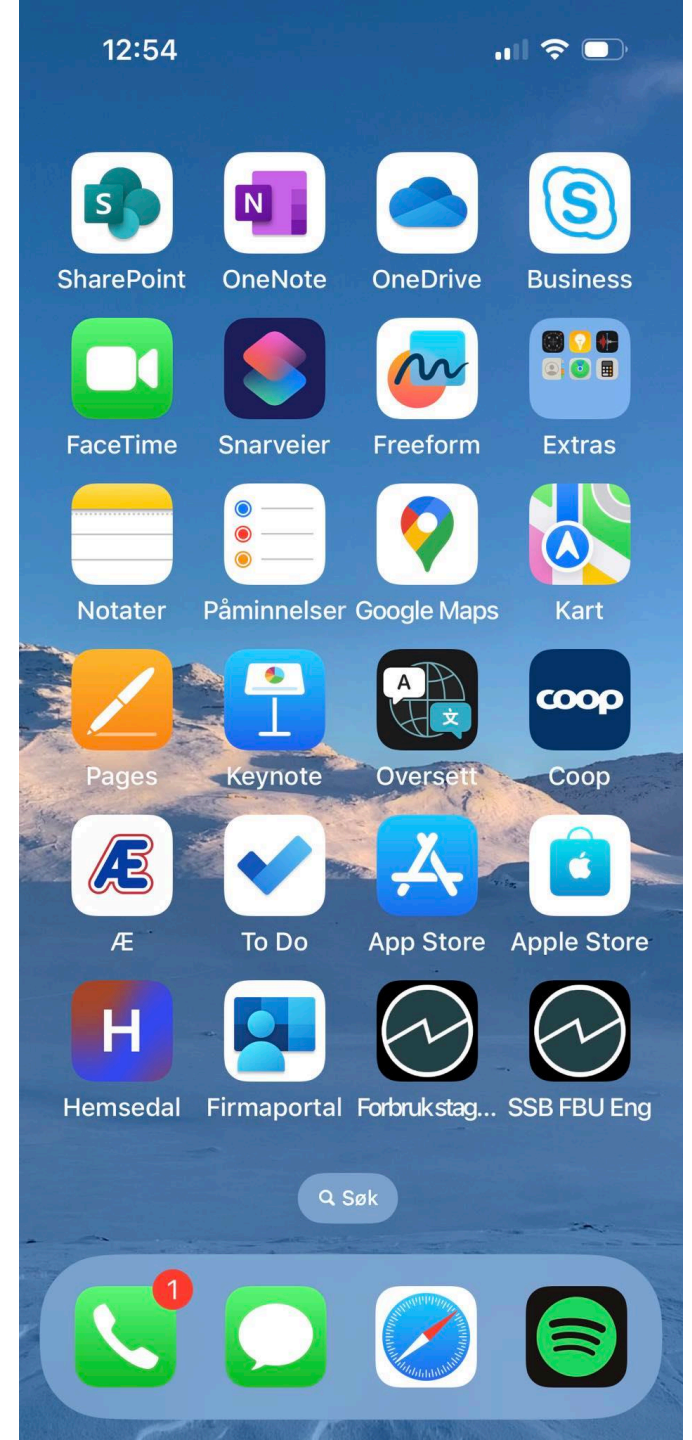
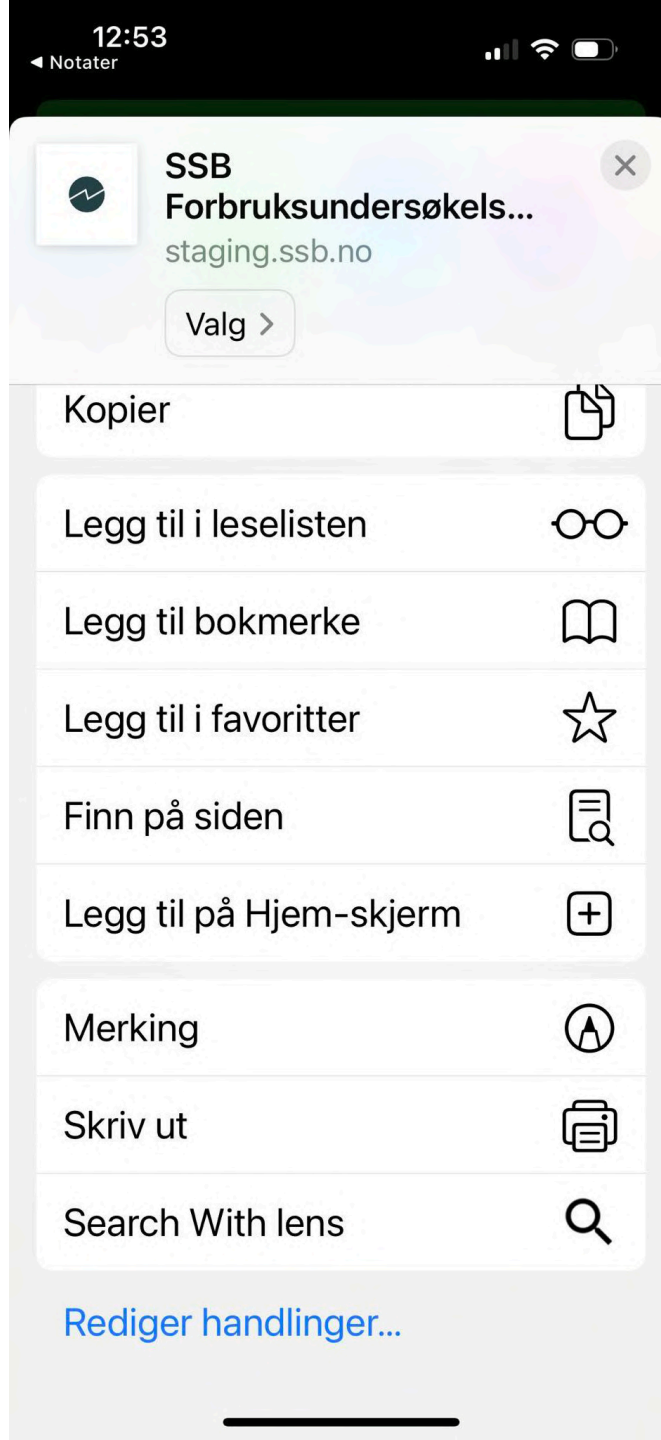
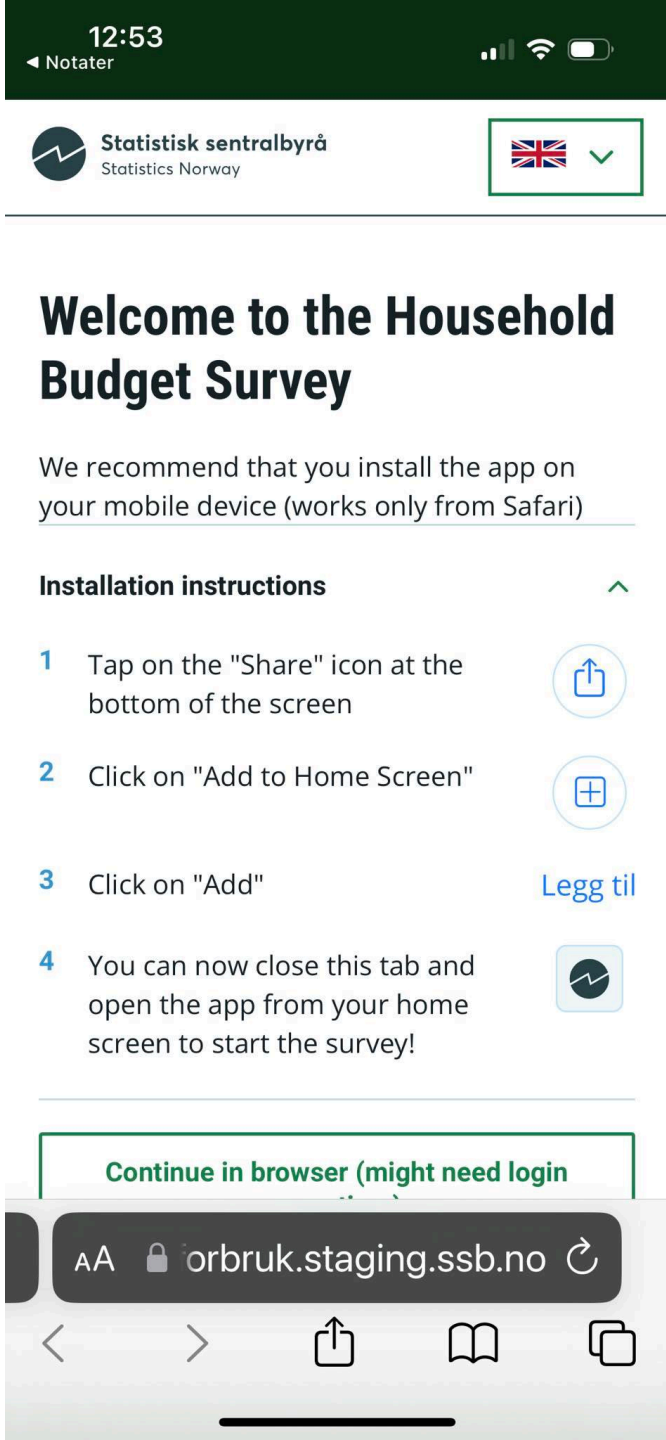


4c) Screen shots SSB web app

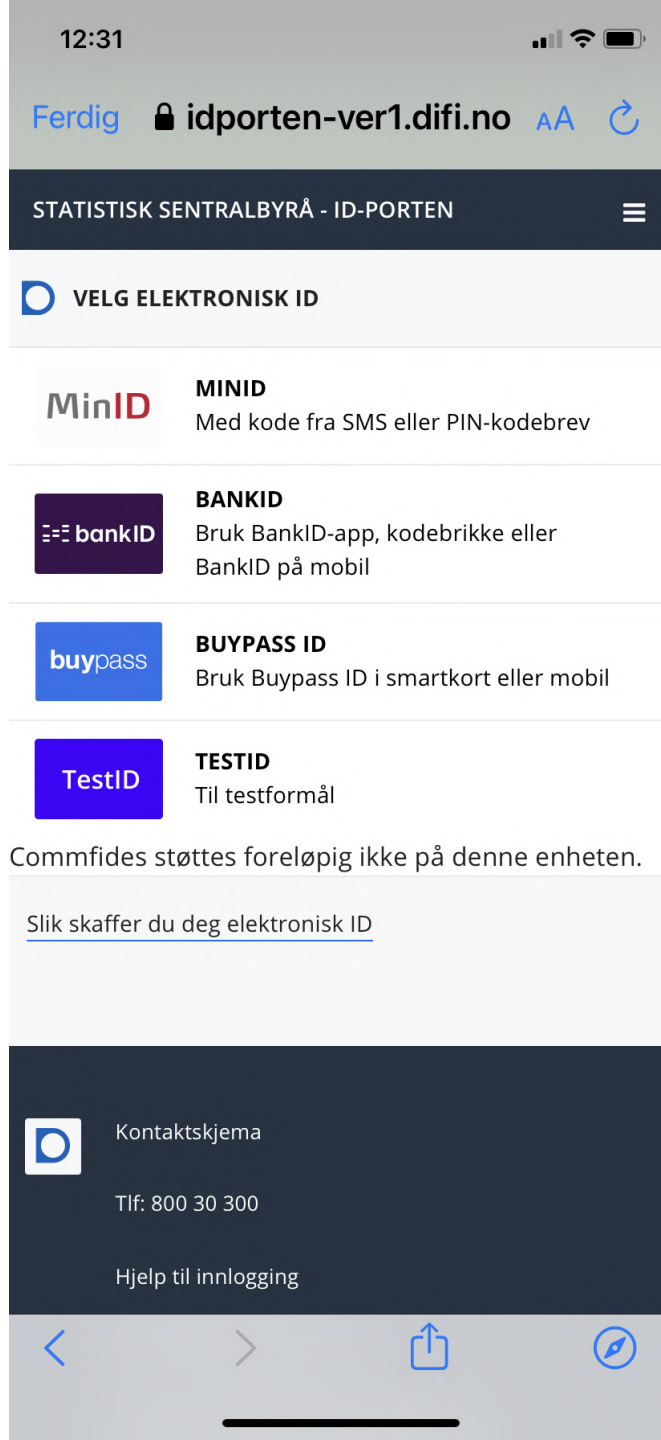


Start screen





BankID



12:57

Statistisk sentralbyrå
Statistics Norway

When you participate in
the consumption survey,
you contribute to
important statistics about
Norwegians' consumption
habits.

1/2

Next

12:57

Statistisk sentralbyrå
Statistics Norway[← Back](#)

In the survey you shall:

- record the entire household consumption for one week
- complete the questionnaire
- confirm that you are done

2/2

[Start the survey](#)

12:57

Statistisk sentralbyrå
Statistics Norway[Menu](#)

Consents

By registering expenses and answering the questionnaire in this app, you consent that the information can be used for official statistics and research.

What information do we collect?

- Everything you register and scan in the app. Make sure that names or other personal information do not appear on photos of receipts.
- Information about browser and operating system.

Your information is secure with us

- Information from the app is stored separately from birth number and other contact information.
- Receipts you upload are scanned by a third party, but are deleted by them as soon as the scan is performed.
- By 1.10.2025, all information will be anonymised.

It is voluntary to participate

- You can contact us at any time at svar@ssb.no or 62 88 56 08 to request that the information you have provided be deleted.

Statistics Norway follows the privacy rules

- The survey is conducted in accordance with the Statistics Act and the Personal Data Act.

You can read more about privacy at
<https://www.ssb.no/omssb/personvern>

[I consent](#)



Consents

By registering expenses and answering the questionnaire in this app, you consent that the information can be used for official statistics and research.

What information do we collect?

- Everything you register and scan in the app. Make sure that names or other personal information do not appear on photos of receipts.
- Information about browser and operating system.

Your information is secure with us

- Information from the app is stored separately from birth number and other contact information.
- Receipts you upload are scanned by a third party, but are deleted by them as soon as the scan is performed.
- By 1.10.2025, all information will be anonymised.

It is voluntary to participate

- You can contact us at any time at svar@ssb.no or 62 88 56 08 to request that the information you have provided be deleted.

Statistics Norway follows the privacy rules

- The survey is conducted in accordance with the Statistics Act and the Personal Data Act.

You can read more about privacy at <https://www.ssb.no/omssb/personvern>

I consent

Consents

By registering expenses and answering the questionnaire in this app, you consent that the information can be used for official statistics and research.

What information do we collect?

- Everything you register and scan in the app. Make sure that names or other personal information do not appear on photos of receipts.
- Information about browser and operating system.

Your information is secure with us

- Information from the app is stored separately from birth number and other contact information.
- Receipts you upload are scanned by a third party, but are deleted by them as soon as the scan is performed.
- By 1.10.2025, all information will be anonymised.

It is voluntary to participate

- You can contact us at any time at svar@ssb.no or 62 88 56 08 to request that the information you have provided be deleted.

Statistics Norway follows the privacy rules

- The survey is conducted in accordance with the Statistics Act and the Personal Data Act.

You can read more about privacy at <https://www.ssb.no/omssb/personvern>

I consent



12:58



Scan receipt

You can easily register a purchase by taking a picture of or uploading a receipt.



1/4

Next

Your registration period is: 05.03 - 13.03

[Change period](#)

Mars 2024

M	T	O	T	F	L	S
26	27	28	29	1	2	3

Home

Expenses

12:58



Register manually

If you are missing a receipt, you can register it manually.



2/4

Next

Your registration period

Your registration period is: 05.03 - 13.03

[Change period](#)

Mars 2024

M	T	O	T	F	L	S
26	27	28	29	1	2	3

Home

Expenses

12:58



Expenses

All the expenses you register are in the "Expenses" tab at the bottom of the screen.



3/4

Next

[Change period](#)

Mars 2024

M	T	O	T	F	L	S
26	27	28	29	1	2	3

Home

Expenses

12:58



Bills and fixed expenses

Most fixed expenses are registered in the questionnaire, but you can enter other bills you have during the registration period in the app (register manually button)

4/4

Great!

Your registration period is: 05.03 - 13.03

[Change period](#)

Mars 2024


M	T	O	T	F	L	S
26	27	28	29	1	2	3


Home

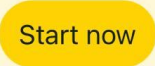
Expenses


Tasks

Tasks 


Scan receipt


Register purchase manually

Answer the questionnaire


Confirm completion


Your registration period

Your registration period is: **05.03 - 13.03**


Home


Expenses

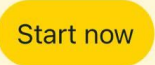
Tasks


Tasks 

- registration period **In progress**
- Answer the questionnaire **Start now**
- Confirm completion


Scan receipt


Register purchase manually

Answer the questionnaire


Confirm completion


Home

Expenses

Start now



Your registration period

Your registration period is: **05.03 - 13.03**

 [Change period](#)

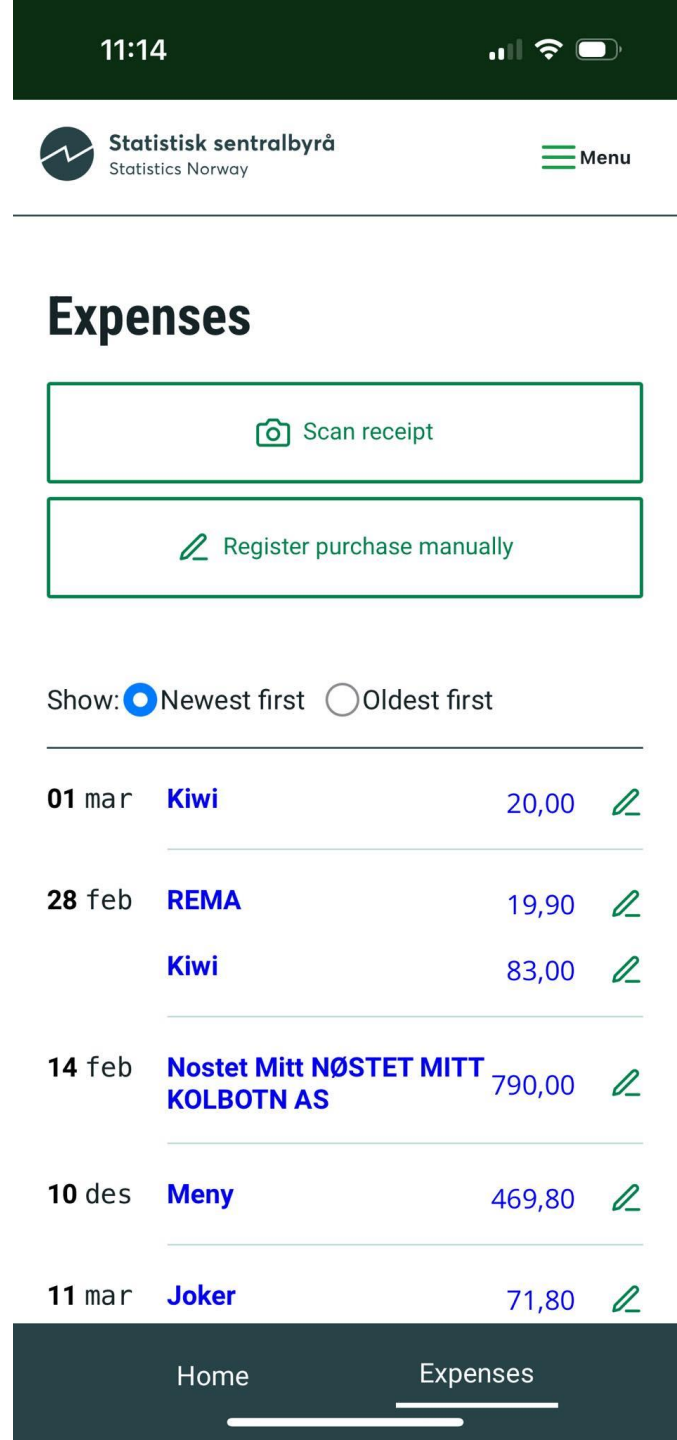
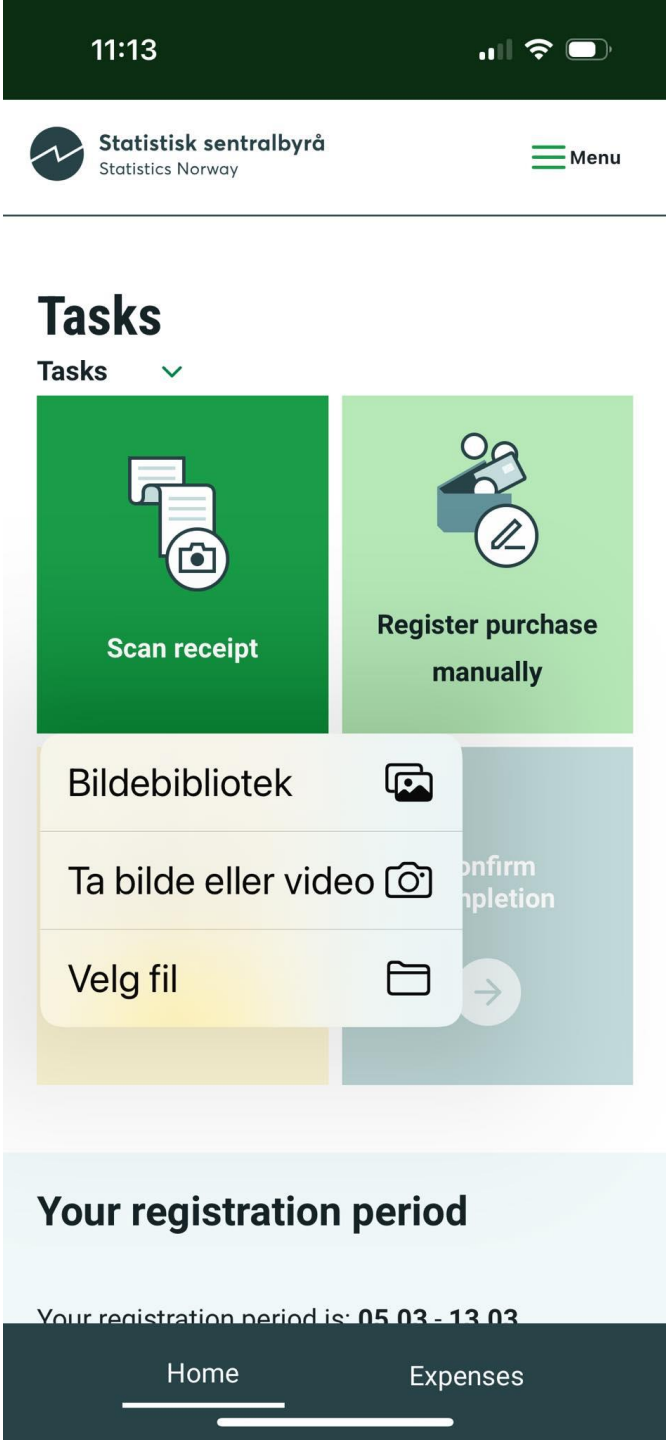
Mars 2024						
M	T	O	T	F	L	S
26	27	28	29	1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	31
1	2	3	4	5	6	7

Does this registration period not work for you?

[Contact us](#)

Home

Expenses



11:15

← Back

New Purchase

Date

05.03.2024

Shop/place/beneficiary

Item	Amount
⊕ Add item/service	
Total	0,00

Save

12:09

Date

05.03.2024

Shop/place/beneficiary

Nøste mitt

Item	Amount
⊕ Add item/service	
Total	0,00

Add at least 1 line item

Autoutfyll kontakt

Ferdig

q w e r t y u i o p å

a s d f g h j k l ø æ

↵ z x c v b n m ⌫

123 mellomrom retur

😊

12:11

← Back

Add item

Which product or service did you buy?

Enter a single item, max 30 chars0/30

ⓘ

Eg.: Apples or Hairdresser

How much did you pay?

Enter precise amount

1,00

12:20

← Back

Add item

Which product or service did you buy?

Enter a single item, max 30 chars5/30

Hip w

Add "Hip w"

How much did you pay?

Enter precise amount

1,00

Please enter valid amount

^v

Ferdig

qwertyuiopå

asdfghjkløæ

↵

zxcvbnm

✖

123

mellomrom

retur

😊

12:12

← Back

Add item

Which product or service did you buy?

Enter a single item, max 30 chars8/30

Hip wool

ⓘ Eg.: Apples or Hairdresser

How much did you pay?

Enter precise amount

790,00

^v

Ferdig

12ABCDEF

4GHI5JKL6MNO

7PQRS8TUV9WXYZ

,0✖

12:12

← Back

New Purchase

Date

05.03.2024

Shop/place/beneficiary

Nøste mitt

Item	Amount
Hip wool	790,00
⊕ Add item/service	
Total	790,00

Save

11:16

📈

Statistisk sentralbyrå
Statistics Norway

Menu

Expenses

📷 Scan receipt

📝 Register purchase manually

Show:

Newest first

Oldest first

05 mar

Nøstet mitt

790,00

📌

01 mar

Kiwi

20,00

📌

28 feb

REMA

19,90

📌

Kiwi

83,00

📌

14 feb

Nostet Mitt NØSTET MITT
KOLBOTN AS

790,00

📌

10 des

Meny

469,80

📌

Home


Expenses

11:15

← Back

Edit purchase

Delete



Scanned receipt

Date

14.02.2024

Shop/place/beneficiary

Nostet MittNØSTET MITT KOLBOTN AS

Item	Quantity	Amount
10 Hip Wool	1	790,00
<div>⊕ Add item/service</div>		
Total	2 varer	790,00

Save

12:35

← Back

Edit purchase

Delete

Date

05.03.2024

Shop/place/beneficiary

Nøste mitt

Item	Amount
Hip wool	790,00
<div>⊕ Add item/service</div>	
Total	790,00

Save

12:13

← Back

Edit purchase

Delete

Are you sure you will delete the whole purchase ?
Nøstet mitt ?
(this operation cannot be reversed)

Cancel

Delete

Item	Amount
10 Hip Wool	790,00
<div>⊕ Add item/service</div>	
Total	790,00

11:16

Statistisk sentralbyrå
Statistics Norway

Menu

Expenses

Scan receipt

Register purchase manually

Show: ☒ Newest first ☐ Oldest first

05 mar	Nøstet mitt	790,00	
01 mar	Kiwi	20,00	
28 feb	REMA	19,90	
	Kiwi	83,00	
14 feb	Nøstet Mitt NØSTET MITT KOLBOTN AS	790,00	
10 des	Meny	469,80	

Home

Expenses

Statistisk sentralbyrå
Statistics Norway

[← Back](#) English

Questionnaire

The questionnaire outlines many common fixed expenses and it may be helpful to have some information available in advance.

We ask, among other things:

- Mortgages and interest rates
- Insurances
- Renovation expenditures within the last 12 months
- Expenditure on holiday homes
- Travels

The questionnaire takes approximately 20 minutes to complete. You can continue later if you cancel along the way.

Completing the questionnaire is one of the mandatory tasks to finish the survey

[Start questionnaire](#)

← Back

English ✓

About your home

What type of housing do you live in?

☒ Detached house

Row house, terraced house or semi-detached house

Vertically or horizontally divided building with three or four housing units

Flat

Live in boat, caravan or car

→ Next

← Back

English ✓

Housing: Ownership and finances

Do you have a mortgage or loan secured on your home?

☒ Yes

No

Include all mortgages secured on your home. Do not include shared debt (fellesgjeld).

Previous

→ Next

Housing: Ownership and finances

How much remains to be paid on the mortgage/loan?

1 000 000 kr

Ferdig

1

2

3

ABC

DEF

4

GHI

5

JKL

6

MNO

7

PQRS

8

TUV

9

WXYZ

0

✕

4d) Workflow description in Excel

(See: [WP2.3-WorkFlowApp-SSB](#))

4d) Workflow description in Excel

- See: [WP2.3-WorkFlowApp-SSB](#)



5. Analysis method and summary



Data and analysis methods

- Video recording and transcripts of all tests
- Observations incorporated (actions/verbalizations separated) into transcripts
- Data chart with observations, verbal quotes, and description of how each TP solved tasks
- Data analysed both vertically per test and horizontally per research question
- Summary of main findings for each TP
- Summary of findings structured per research question

Structure data chart

1) Observations and verbatims

- Use of Teams
- Use of bank-ID for secure login
- Login to the web app
- Usability of the web app
- Use of manual entries vs scanning
- Preferred registration method
- Error checking
- Personal receipt

2) Debrief/retrospect

- First impression
- Experience with registration methods / Preferred method
- Attitudes towards error checking and editing
- Thoughts on privacy
- Trust in the sender, the survey, the app, and scanned data
- Attitude towards data accuracy and reliability
- Secure data storage and privacy

Summary of main findings per test person

- See [WP2.3 Summary per test Norway](#)



Thanks for your attention

Your contact person

Johannes Volk

johannes.volk@destatis.de

+49 611 75 4785

+49 160 92434256

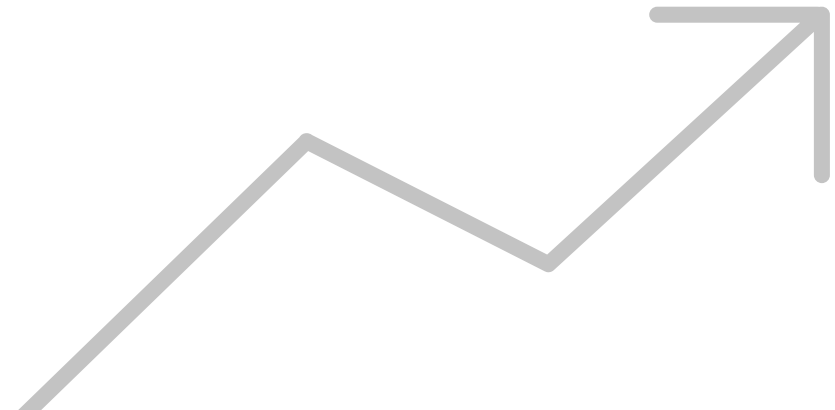
The team behind the survey

Johannes Volk

Lasse Häufglöckner

Anja Sommer

Sarah Schöffling





REPUBLIC OF SLOVENIA
STATISTICAL OFFICE RS



Statistisk sentralbyrå
Statistics Norway



Universiteit
Utrecht



UNIVERSITÄT
MANNHEIM



Report CBS test scanning function @HBS app

v.1.0 20-2-2025

Deirdre Giesen, Daphne Debije, Janelle van den Heuvel, Jasmina Saitovic & Marina Veldhuizen

Smart Survey Implementation

Grant Agreement Number: 101119594 (2023-NL-SSI)

*Disclaimer: Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Eurostat.
Neither the European Union nor the granting authority can be held responsible for them.*

French HBS app usability tests

SSI
06/02/2025



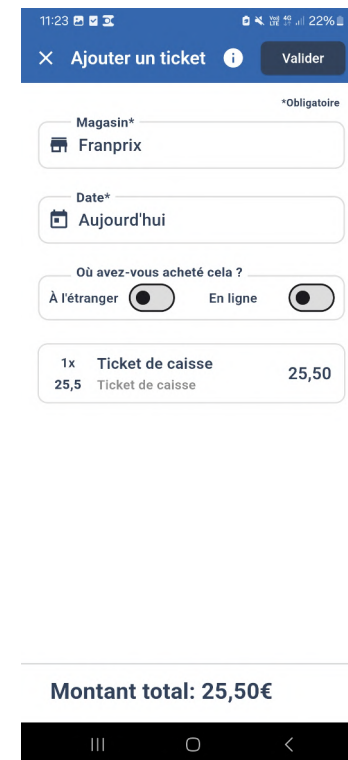
- 1 FRENCH HBS APP**
- 2 DESIGN OF THE TEST**
- 3 TASK BY TASK OBSERVATIONS**
- 4 DEBRIEFING QUESTIONS**
- 5 MAIN FINDINGS**

01

FRENCH HBS APP

BASED ON CBS @HBS APP

- Rebuilt to be integrated in Insee SI from the open code
- Modifications of the app :
 - Conception choice : limit the respondent burden
 - OCR and classification in back office
 - Only photo check : presence of written characters
 - The respondent fill in the total amount of the receipt manually
 - Implementation of Open Id connect to meet Insee authentication standards



02

DESIGN OF THE TEST

REMINDER OF THE SCENARIO

- Task 1 : warm up (select a route to get to the Eiffel Tower)
- Task 2 : download and log in (+ first debriefing questions)
- Task 3 : simple ticket, choice of the input method
- Task 4 : simple ticket, other method
- Task 5 : medium ticket (+ check if the data is correct)
- Task 6 : personal ticket
- Optional tasks
 - Task 7 & 8 : longer tickets
 - Task 9 : overview of all expenses
- Debriefing questions

USE OF NORWAY INTERVIEW GUIDE

TASK 2 : DOWLOAND AND LOG IN

- A document was provided to participants with :
 - instructions on how to download the application
 - ID and password for access.
- Same model as the document actually provided to the respondents for the survey



Carnet numérique de l'enquête
Budget de famille 2024

● Comment accéder à l'application ?

1) Téléchargez l'application mobile Enquête Budget de famille à l'adresse suivante

iPhone		Android	
https://apps.apple.com/fr/app/enqu%C3%AAt-budget-de-famille/id6471562020		https://play.google.com/store/apps/details?id=fr.insee.householdbudgetsurvey	

Ou tapez « enquête budget insee » dans le store

2) Installez et lancez l'application

3) Cliquez sur connexion

3) Saisissez votre identifiant et votre mot de passe individuels
Respectez minuscules, majuscules et caractères spéciaux

Identifiant	Mot de passe
XXXXX	&&&&&9999

TESTS ORGANIZATION:

- Took place from the 12 to the 21 of november 2024
- In person at Insee's headquarters near Paris
- One moderator and one observer for each session
- audiovisual recording of participant hands and phone

RECRUITEMENT OF PARTICIPANTS :

- Through a private platform (tandemz)
- Selection based on :
 - Willingness to attend the session at Insee's office
 - Possession of a recent smartphone (≤ 4 y.o.) to avoid compatibility issues with the app
 - Agreement on the privacy policy
- Financial reward

15 PARTICIPANTS

- gender / phone / self-assessment of digital proficiency

Less than 25 y.o.	25 – 50 y.o.	50 – 65 y.o.	More than 65 y.o
Man Iphone advanced	Woman Iphone intermediate	Woman Iphone intermediate	Woman Android basic
Man Iphone intermediate	Woman Android intermediate	Woman Android advanced	Man Android basic
Man Iphone advanced	Woman Android intermediate	Man Iphone advanced	Man Iphone basic
	Woman Android advanced		
	Man Android advanced		
	Man Android advanced		

03

TASK BY TASK OBSERVATIONS

THE INSTALLATION AND LOGIN PROCESS WERE DEEMED EASY AND RELIABLE BY ALMOST ALL PARTICIPANTS.

« The app is good, quite self-explanatory. You just have to follow along—it's well designed »

« Simple, minimalistic, and it guides me step by step. »

« It remains easy and accessible to everyone. Installing an app is not difficult—everything has an app now (banking, paying rent). The information is sufficient, explaining what to do and where to go. »

« You have to take the time to do it. For me, it's simple, but for my parents, it's a bit long and complicated. However, I understand the security aspect of the survey. »

MINOR DIFFICULTIES WERE ENCOUNTERED BY SOME PARTICIPANTS.

- A participant was surprised by the outcome of using the QR code:
"When the QR code is available, I will use it. I was just surprised—I didn't expect that I had to download the app. Otherwise, I would have gone directly to the Play Store. I thought the link would take me straight to the login/sign-up page."
- Two participants were unable to use the first password provided to connect themselves into the app but succeeded with a second one
The provided passwords contained special characters, sometimes difficult to read
- One participant (65+ y.o) experienced real difficulties because they searched for the app on Google instead of going to the Play Store. The moderator had to guide them step by step. However, they did not blame the app itself:
"It went fine, it was my mistake."

9 USERS CHOSE THE CAMERA, AND 6 CHOSE TO ENTER THE EXPENSE MANUALLY.

THE CHOICE BETWEEN THE TWO METHODS SEEMS CLEAR, WITH THE TWO ICONS FOR THE CAMERA AND KEYBOARD.

ONE PERSON MENTIONED THAT THE ICONS ARE A BIT SMALL, ESPECIALLY THE KEYBOARD.



FUNCTIONALITY AND FEATURES

- Adding Store and City: One user suggested adding the ability to include the city along with the store name, which was a feature later implemented in the updated version of the app
- Photo Scanning Expectation: Several users were surprised that the photo didn't automatically scan the total amount, assuming that was part of the photo function. They expected it to auto-fill, but the app instead prompted them to manually enter the amount.

"It's a shame I have to enter the total amount; I thought it would be automatic."

INTERFACE FEEDBACK:

- Icons & Navigation: The cropping function was well-received, but one user struggled with understanding the validation process. After some confusion, they realized that the small check symbol was used to confirm and validate the receipt. This took some time to understand.
- Back Button Usage: One participant mentioned returning to previous screens to verify if the data had been correctly entered, showing a need for clearer navigation cues or confirmation of actions.

ALL PARTICIPANTS ARE ABLE TO FILL IN THERE EXPENSES MANUALLY

- One participant chose not to complete the process, because it was too time consuming, but he understood how it should be done

"ADD AN EXPENSE" BUTTON LACKS OF CLARITY:

- Several participants were confused by the "Add an Expense" button, thinking it meant adding the total of the ticket rather than individual items.

"When I click on 'Add an Expense,' I thought I was supposed to enter the total amount of the ticket, not list everything I bought. I thought the app should allow me to scan and add the ticket total."

"I had no idea what to search for. I tried 'groceries,' but that didn't work. Then, I tried 'lettuce,' and it was there. But, this felt tedious, entering every item individually."

"Find a product or service? Uh, I have no idea, I'll write 'groceries.' There's no 'groceries.' This is a joke... well then, what did I buy? It would seem crazy to write 'lettuce'... oh, it's there, so I'll do it one by one. I thought we were supposed to enter the total amount, but here, it's really one by one."

- Participants had less information about the app than survey respondents (lack of manual)
- The term "expense" was replaced by "item" in the latest version of the app.

TWO DIFFERENT ATTITUDES WHEN THE EXACT PRODUCT NAME IS NOT FOUND IN THE LIST:

- Some participants find the closest category

"I'm going to add grated emmental, I find the item 'emmental, cheese,' so I click on it and enter the price."

- Others add the product to the existing list:

"Sweet cream under pressure: I couldn't find it in the list, so I click on 'Not found.'"

SINCE THERE IS NO OCR IN APP, TASKS 5, 7 & 8 WERE QUITE SIMILAR IN OUR TEST DESIGN

- Since receipts are longer, photo input is systematically chosen.
- Once the task is already done for the first time, subsequent expense entries are smooth and quick.

"I go ahead and do the same thing as before. I take the photo, the first steps are the same."

"It's a matter of habit, after that it's easy."

- Several methods to input long ticket

« This time there are a lot, but it fits in the photo, I'm going to crop everything below the total amount. »

"I'll take a photo. Earlier they said we could take it in 2 or 3 parts (...) Ok, I have the two photos merged »

Participants are able to check the expenses

"Not explained, so I'll try it on my own. Two options, date or expense tab, I go through the tab, I find my photo, so it's very convenient."

"I prefer going to expenses, and there I see all my receipts. I click on the last receipt, the little arrow. I have access to the receipt photo, the amount, the location, and I can modify, duplicate, or delete. It's correct."

"Maybe in the expenses section. In the expenses tab, you can see the expenses. I click on it, and the receipt shows up along with the amount."

NEED TO PREPARE THE RECEIPT BEFORE TAKING THE PHOTO

"Can I still control it when I take the photo? It's a very long receipt, I'm trying to flatten it."

"I'm trying to smooth out the receipt as best I can because it's going to take me a while to type everything, so I prefer to do it with a photo. I go to the camera."

"I flatten the receipt that was folded in four."

NO WORRIES ABOUT SHARING PERSONAL DATA AT THIS POINT

TWO SCREENS ALLOW TO GET AN OVERVIEW OF EXPENSES :

- « expenses » : list of all expenses
- « overview » : charts presenting expenses by date or store
- All participants were able to access one of these two screens

DESIGN ISSUE :

- The order in which expenses appear is not considered optimal: expenses should be sorted according to the date the expense was made, rather than the date the expense was entered (correction made following these comments).

04 DEBRIEFING QUESTIONS

Receipt scanning is judged more time efficient :

"I prefer the photo, it's quick. It's a shame it doesn't scan automatically, but it's not a big deal. It took me just a few seconds, but that's it."

"Both methods are easy, but the photo one is so much more convenient and faster: no need to enter the products manually."

"If it's the weekly shopping, it takes too much time to enter it manually, I'm trying to save time, it's not a time saver."

"If there are a lot of items, the method is to take a photo."

"If you have two or three items, the manual method can be fine, but otherwise, having the photo and being able to see it seems simpler."

"I prefer the photo method because you can see it directly without having to enter everything."

– Receipt scanning prevents lost of information :

"When we get home, we take a photo before losing it, we archive it."

– One discordant participant :

"I think it's easier to do it manually, it's much faster than using the photo."

PHOTO IS GENERALLY CONSIDERED MORE RELIABLE

- **Proof that the expense really took place :**

"The receipt is the real proof of purchase. Scanning it ensures accuracy of the amounts."

"The photo is proof that the purchase really happened"

"The photo provides assurance that it's a real purchase."

- **Better quality, reduction of errors :**

"It might be more intrusive, but from the Insee's point of view, it's definitely more reliable and of better quality."

"Receipts in photos are genuine purchases, so the margin of error is low. I think if we enter manually, we might forget some lines, but the photo shows all the details. »

"For me, the camera is essential. If we type them in, we can make mistakes! The camera is more reliable. "

ONE PARTICIPANT DISAGREES :

"The most reliable is manual entry, but the photo is still good."

"Trust? I'm not there yet, but if I download an expense tracking app, I already trust it."

- **Simplicity of the design generates trust...**

"First impression: the home screen looks good. I landed directly on the calendar (of course, there was the tutorial first). It immediately showed me the days I need to complete and where to enter my data."

"It looks like a serious application."

"The first screen is clean and easy to read. »

"Clear, concise, and well-explained instructions."

- **... Insee reputation as well**

"It's an Insee app, and they are well known for statistics, so I trust it."

"It says Insee, which is a well-known and reputable statistics institute."

"I trust the app because it's from Insee with the logo, and I know the institution."

"Since it's Insee, I tend to trust it—I know the organization by name."

Simplicity of the design, comparison with banking apps

"It looks like a very serious app, it resembles a banking app; it's very organized, very clear; we have the calendar to help navigate; it's both complete and minimal."

"Overall, it reminds me of the banking app, to see the categories where you spend the most."

Frugality of the information requested

« I can trust the app, not particularly intrusive, we're just asking for a receipt. »

« no personal data »

Insee reputation

"Trust in the app, the design is good, it's Insee, it's well-known, it's not a new organization that was just created."

"it's still Insee, well-known in statistics, we often hear about Insee in the news, I would fully trust this app, the data won't be going all over the place"

"The name Insee itself inspires trust, knowing that it's generated by them, a large organization, a trusted brand."

Respondants are not convinced at their individual level..

"The app won't be directly useful to me in terms of my expenses, it doesn't add anything to the management of my budget."

"The app itself wouldn't be useful to me."

... but understand the utility in the context of a survey

"I guess, in an anonymized way, this helps you collect data for statistics."

"There's a lot of work behind it, the data goes to Insee, which establishes the average expenses based on households, family composition, and where we live."

OVERALL POSITIVE JUDGMENT ON RELIABILITY:

"The data is correct and reliable, yes."

"Of course, it's accurate, we played along. I don't see the point in lying or entering fake products."

"It's me who declares them, so it's reliable."

POSSIBLE BIAS: INDIVIDUAL LEVEL

- Difficulty in using the app:

"In my case, it's not reliable: I couldn't manage to enter an expense."

- Risk of forgetting/mistakes:

"Quite reliable, but some expenses might be missed, especially routine ones. I'm afraid I won't remember all my expenses."

"Users need to remember to enter their purchases into the app every time they make a purchase."

"Source of error: if we enter the wrong amounts."

"Forgetting expenses could reduce the reliability of the survey."

"People need to be motivated to accurately describe all their expense"

- Some expenses are personal, and users are more reluctant to share them in the app:

"I wouldn't necessarily enter all my expenses into the app, there are certain purchases I make that I wouldn't want to share with the app, like a piece of jewelry for my wife."

ANOTHER BIAS: LACK OF REPRESENTATIVENESS OF EXPENSES

"It's a question of data volume. If only Parisians use the app, the data won't be reliable. The usage rate should be spread across France."

"It all depends on how representative it is. If a large portion of the population installs it, it will reflect real-life spending."

"Will everyone play along? That's the issue."

Participants don't seem to have major concerns about the app's security. When asked, they tend to have a positive assumption:

« In terms of privacy security: Insee, in principle, we can trust you. However, if it's fully anonymized beforehand, it's fine. »

One respondent negatively views the lack of information on the subject:

"The app is very simple, but nothing indicated that it's secure, so for me, it's not secure."

PARTICIPANTS ARE GENERALLY NOT VERY CONCERNED ABOUT HOW THEIR PERSONAL DATA IS HANDLED

- General trust in public and private data collection privacy :

"I categorically refuse to think we are being tracked, surveilled, or tattooed, no... it's a small extra, I'm already in databases, like my health insurance card."

- Trust that Insee will handle the data properly, even if they didn't get explicit information about it :

"Regarding the receipt, I trust you on the anonymization, there are technical means to do it."

"Anonymization is guaranteed, not worried about data, if I see the usefulness."

"I don't know how the app handles personal informations, I haven't checked the settings. Does it share info, partnerships, or anything? I don't think the receipts are used for anything else by the app."

"I trust INSEE to handle this confidentially."

- Participants don't think they share personal data :

"I haven't looked, but I imagine it's secure and private. It's just a receipt, I'm not giving a credit card number."

"I don't really see what needs to be processed. I think it's secure. Receipts don't impact me unless my name is on them."

- participants recruitment via a private online platform may introduce some bias on this topic

04

MAIN FINDINGS

USABILITY OF THE APP :

- Participants are able to complete the tasks.
- Once each method is performed for the first time, entering subsequent expenses becomes easy.

POSITIVE FEEDBACK FROM PARTICIPANTS :

- about the design
- about functionalities

INSEE IS A TRUSTED BRAND:

- When installing the app
- For submitting consumption data.

THE VAST MAJORITY OF PARTICIPANTS FAVOR RECEIPT SCANNING METHOD

- It's judged faster...
- And more reliable (proof of purchase)

SEVERAL PARTICIPANTS EXPECTED THE INFORMATION FROM THE PHOTOS TO BE PROCESSED AUTOMATICALLY.

- However, after this initial disappointment, they consider the method to be quick and use it

THE THREE PARTICIPANTS OVER THE AGE OF 65 WERE THE ONES WHO EXPERIENCED THE MOST DIFFICULTY AT EACH STEP OF THE TEST.

- Despite the tutorials, user needs to guess what he's expected to in several steps.
- This is a common characteristic of apps, which follow standard design principles, where navigation relies on the user's ability to explore screens and features.
- This process requires trial and error :
 - For those who are comfortable and familiar with such processes, these trials/errors are quick and seamlessly integrated into their workflow.
 - For less experienced users, each trial/error takes time and undermines the participant's confidence in their ability to complete the tasks.

THOSE PARTICIPANTS MIGHT NOT HAVE ABLE TO USE PROPERLY THE APP IN REAL SURVEY CONDITIONS

- In contrast, participants under the age of 25 completed tasks significantly faster than others.
- This raises the question of whether age should be considered as a criterion for offering the app instead of paper diary, as it might be a good proxy for the ability to use it effectively.

THE SESSIONS HIGHLIGHTED SOME DESIGN FLAWS IN THE APP, WHICH LED TO IMPLEMENT EVOLUTIONS :

- « expense » button in manual input screen
=> changed to « item »
- Automatic launch of tutorials at each new screen confused several respondents, including those the most at ease with the use of the application
=> automatic launch was disabled (except for the very first screen the user faces when he opens the app)
- Online / abroad categories created confusion
=> the text of this screen zone was edited

Retrouvez-nous sur

insee.fr



Simon Moreau (simon.moreau@insee.fr)

HBS

DSDS / DRCVM / Division conditions de vie

Summary main findings

A prototype of the @HBS app was tested in the lab with 20 test respondents (Rs) with different ages, educational levels, digital skills and language skills. The focus of test was on the receipt scanning function of the app, which was simulated but felt realistic for Rs.

- Most Rs appreciate the app and find it rather easy to use.
- Most Rs appreciate the scanning function and feel this reduces response burden and increases data quality (compared to manual entry).
- But for Rs with low digital skills and difficulty reading Dutch, using the app without assistance is very difficult or impossible, with assistance most of them can learn.
- All Rs experience a steep learning curve when using the app.
- Several small and some big issues were identified that need to be resolved before the app can be implemented. Most importantly, Rs do not understand what is asked of them nor the purpose of it (response task not clear, Rs think app is for them to gain insight in their own spending).
- Rs do not show or say (when asked) that they are concerned with privacy or trust issues of either the app or the study, they also do not think scanning or manual data entry matters for privacy and trust.
- When asked, most Rs express concerns about the quality of the data collected with the app due to expected low and selective response (too much work and too difficult for some people).
- Several Rs think scanning will improve overall data quality as the raw data entered will be of higher quality, and scanning reduces burden and provides CBS with more information (i.e. assume scanning will give CBS an option to analyze pictures of receipts).

Summary main conclusions

- An app with a receipt scan function seems a highly valuable instrument for collecting data for HBS (less burden, more quality) for Rs with at least some basic level of digital and language skills.
- For others Rs personal assistance and/or other modes of data collection are needed.
- Speed of scanning and the output quality of scanned data seem very important for user experience (and thus response and data quality).
- Explanation of the response tasks should be thoroughly improved both in communication materials and within the app.

Outline

Background & goals of the test

Tested materials & test design

Letter and leaflet, Installing the app, Login, Expectations of the app

Onboarding

User interface and user experience

Scanning or manual entry: preferences and experiences

Entering expenses

- Manual entry
- Scanning receipts
- Editing scanned information
- Learning curve

Trust, privacy and security

Other findings

Background

Overall goal SSI project: Develop, implement and demonstrate the concept of smart surveys.

Smart surveys = surveys that combine self-report questions with smart features collected via sensor-enabled devices such as smartphones, wearables, and other devices, aiming to enhance data quality, reduce burden on participants, and provide more timely and granular data.

Goals of this test

- Test usability of the @HBS app with a focus on the scan function
- Understand if and how scan function affects trust in the app and trust in the quality of the data collected.

Tested materials

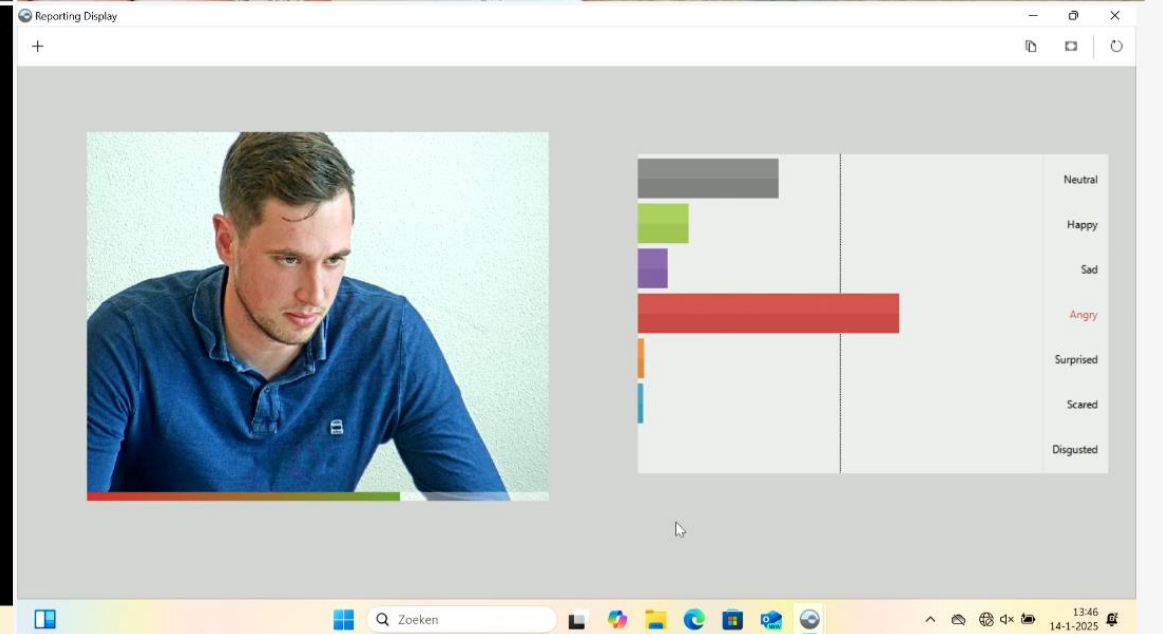
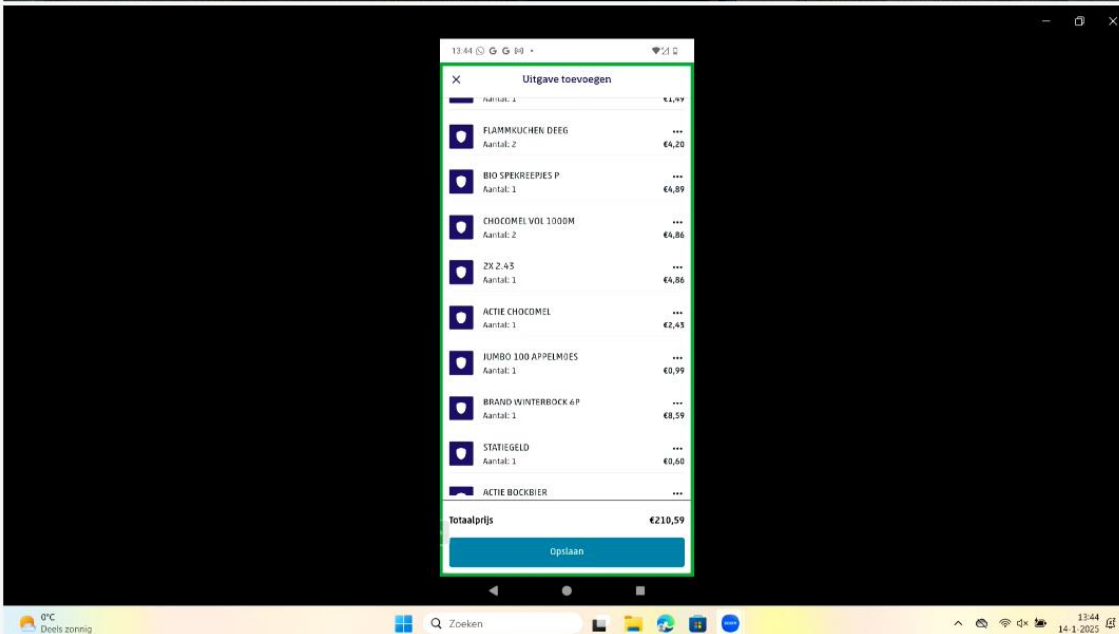
Letter & Leaflet

Prototype of @HBS app on lab phone

- ✓ login screen
- ✗ general and fixed costs questionnaires (shown but not implemented)
- ✓ start questions about vacation period
- ✓ calendar and data collection period
- ✓ manual entry of expenses
- ✓ (simulation of) scanning function receipts
- ✗ lookup tables for stores and expenditures
- ✗ classification of expenditures
- ✗ upload of digital receipts (not shown and not implemented)
- ✓ overview of all entered expenses

Test design

- Qualitative, small scale “cogability” testing (mix of UX test and cognitive interview).
- Focus on “think aloud” during tasks.
- Test protocol coordinated with other similar tests in SSI project.
- In lab at CBS .
- 1 test leader (interviewer) in test room, 1 (or more) observers in observation room.
 - Interviews and observations done by test team (4 methodologist, 1 UX expert)
- Recording with 3 PTZ cameras: 1 on face (with Facereader software, but this software not used for analysis), 1 on hands/phone, 1 overview and screen capture of phone via screen sharing in Zoom (see example view during test on next slide).
- Test duration about 1 hour.
- Testing period 14-29 January 2025



Test protocol

- Introduction, informed consent
- Test task (looking up a route) to practice think aloud & get used to test phone
- Envelop with letter and brochure -> what would you do with this?
- How would you download app (scan code or search in store)?
- Login
- Entry of first (short) test receipt:
 - what would you do
 - ask to enter manually
- Scan of second (short) test receipt (simulation, 5 second, no errors)
- Scan of third (medium) test receipt (simulation, 20 seconds, some errors) + correct errors
- Manual entry of personal receipt
- Scan of fourth (long) test receipt (simulation, 50 seconds, some errors) + correct errors
- General impression of app
- Evaluation of tasks and preferences for scanning or manual entry
- Trust in app
- Trust in data collected with the app
- Final ideas and comments

Recruitment

- Recruitment targets coordinated with other similar tests in SSI project
- Recruitment via list of previously recruited test persons (16), a contact at a local school for secondary vocational education (1), family and friends (2), CBS intern (1).
- 30 euro gift certificate for participation

Analysis

- Individual report of each test in template; draft by observer, review by interviewer
- Tasks scored on scale of
 - 1=without help and without problems
 - 2=without help but with some problems/errors
 - 3=without help but with major problems/errors
 - 4=task performed with help
 - 5=could not perform task with help either
- All reports read by all members of test team
- Summary of findings prepared per topic by one team member, discussed by test team in several iterations.

Test respondents

13 men, 7 women

5 low, 8 medium, 7 high education

(*4 students)

7 qualified by interviewer as having low digital skills

R0 intern for pre-pretest

R13 was planned but had to cancel

Education= highest completed or current (if student)

Lower= primary education, first three years of senior general secondary education (HAVO) and pre-university secondary education (VWO), prevocational secondary education (VMBO) and lower secondary vocational training and assistant's training (MBO-1).

Medium= upper secondary education (HAVO/VWO), basic vocational training (MBO-2), vocational training (MBO-3), and middle management and specialist education (MBO-4)

Higher= Higher professional education / university education.

Participant	Gender	Education	Age	Low digital skills	Dutch language skills
R0	Male	Higher*	24	no	native
R1	Male	Higher	56	no	native
R2	Male	Higher	42	no	native
R3	Female	Higher*	23	no	native
R4	Male	Medium	73	yes	native
R5	Female	Medium	37	no	native
R6	Male	Lower	54	yes	non-native, beginner
R7	Male	Medium	75	no	native
R8	Female	Medium*	20	no	native
R9	Male	Medium	19	no	native
R10	Male	Higher*	25	no	native
R11	Female	Medium	59	yes	native
R12	Male	Lower	47	yes	non-native, beginner
R14	Male	Medium	53	no	native
R15	Female	Higher	29	no	native
R16	Female	Medium	28	no	native
R17	Male	Higher	67	no	native
R18	Male	Lower	54	yes	native, low literacy
R19	Female	Lower	65	yes	native, low literacy
R20	Male	Lower	58	yes	native, low literacy



Centraal Bureau voor de Statistiek
Postbus 4481 | 6401 CZ Heerlen
www.cbs.nl

Dhr. Janssen
Abc-sstraat 123
1234AB
Nederland

correspondentienummer 0558947
onderwerp CBS Uitgaven Onderzoek
datum 29 januari 2025

Beste

Hoeveel kunnen Nederlanders per maand kopen? Kunnen bepaalde groepen steeds meer uitgeven en andere juist minder? En wordt het leven in Nederland duurder of goedkoper? Om antwoord te krijgen op deze belangrijke vragen, voert het Centraal Bureau voor de Statistiek (CBS) het Uitgaven Onderzoek uit. Doet u mee?

Wat krijgt u van ons?

Als u deelneemt en het onderzoek afrondt, ontvangt u na afloop een **VVV Cadeaukaart van 20 euro**.

Wat vragen wij van u?

Voor dit onderzoek vraagt het CBS een aantal huishoudens om tijdens **een periode van twee weken** de dagelijkse uitgaven bij te houden in een app. Uw huishouden is daar één van. Het maakt niet uit of u veel of weinig uitgaven hebt. U vertegenwoordigt veel andere huishoudens in Nederland. Uw deelname is belangrijk.

Hoe doet u mee?

Meedoen kan alleen met een smartphone. Het CBS heeft hiervoor de app **CBS Uitgaven** ontwikkeld. Download de app op uw smartphone en voer de uitgaven van het hele huishouden in. Het is ook mogelijk om de app te downloaden op meerdere smartphones. Zo kan elk lid van het huishouden de eigen uitgaven invoeren. De ingevoerde gegevens worden automatisch samengevoegd.



1. Hebt u een **Android** toestel? Download de app CBS Uitgaven vanuit de Play Store. Hebt u een **Apple** toestel? Download de app CBS Uitgaven vanuit de App Store. Scan de QR-code om de app direct te downloaden.

2. Open de app en log in met uw persoonlijke gebruikersnaam en wachtwoord:

Gebruikersnaam: XXXXX
Wachtwoord: XXXXX

Den Haag
Henri Faasdreef 312 | 2492 JP Den Haag
Postbus 24500 | 2490 HA Den Haag
+31 70 337 38 00

Heerlen
CBS-weg 11 | 6412 EX Heerlen
Postbus 4481 | 6401 CZ Heerlen
+31 45 570 60 00

Bonaire
Blv. Gobernador N. Debrot 67,
Unit 9 | Kralendijk | Bonaire
+599 717 86 76



Vervolgblad 1 behorende bij brief van 10 januari 2025

Gebruik de app direct na het installeren. Houd vanaf komende maandag in de app twee weken de uitgaven van uw huishouden bij. Het gebruik van de app is heel eenvoudig. Op de webpagina www.cbs.nl/uitgaven en in de bijgevoegde folder vindt u meer informatie.

Over een paar dagen krijgt u mogelijk bezoek van een CBS-medewerker. Mocht u nog vragen hebben, dan kunt u deze stellen.

Uw gegevens zijn veilig

Bij al onze onderzoeken zijn uw gegevens veilig. Onderaan deze brief leest u daar meer over.

Heeft u nu al vragen?

Bel ons gerust op (045) 570 64 00. Wij zijn bereikbaar van maandag tot en met vrijdag tussen 9.00 en 17.00 uur. U kunt ook naar contactcenter@cbs.nl mailen. Of kijk op www.cbs.nl/uitgaven voor meer informatie.

Ik dank u alvast hartelijk voor uw tijd en medewerking.

Met vriendelijke groet,

Joost Huurman
Hoofddirecteur Dataservices, Research en Innovatie

Uw gegevens zijn bij al onze onderzoeken veilig. Dit is een verplichting van het CBS die in de wet is bepaald. Het CBS heeft verschillende maatregelen getroffen om uw gegevens te beschermen. Zo zorgt de wet ervoor dat uw gegevens alleen voor statistische doelen worden gebruikt. Geen enkele instelling kan toegang eisen tot de gegevens die het CBS verzamelt. In de statistische informatie van het CBS zijn persoonlijke gegevens nooit te herkennen.

Het CBS verzamelt zelf gegevens maar krijgt ook veel bestanden van andere instellingen. Die informatie voegen we samen. Het CBS maakt statistieken over de Nederlandse samenleving met die informatie. Zo werken we zo zuinig mogelijk.

Het CBS gebruikt uw informatie uitsluitend voor statistische doeleinden. De gegevens worden nooit aan derden verstrekt. Uw anonimiteit is strikt gewaarborgd.

Voor wat er feitelijk gebeurt



Uitgaven Onderzoek 2024

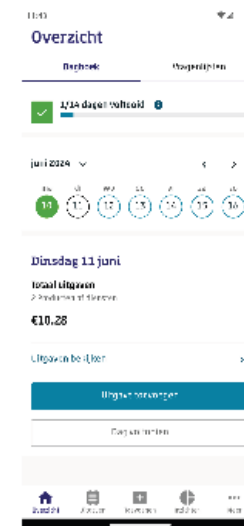
Uw antwoord telt!

Het CBS Uitgaven Onderzoek is een belangrijk onderzoek. Door dit onderzoek kan het CBS zien hoeveel duurder het leven in Nederland wordt. Ook wordt duidelijk welke groepen mensen genoeg of te weinig geld hebben om te kopen wat ze nodig hebben. Deze informatie is belangrijk voor het maken van beleid voor Nederland.

Het CBS nodigt uw huishouden uit om mee te doen aan het onderzoek. Kunnen we op u rekenen?

Wat vragen we van uw huishouden?

Wij vragen uw huishouden (dit zijn u en alle mensen die bij u wonen op dit adres) om twee weken alle privé uitgaven in te vullen. Privé uitgaven zijn bijvoorbeeld uitgaven in de supermarkt, in een webwinkel, in de horeca of bij de kapper. Vaste lasten (zoals huur of verzekeringen) hoeft u niet in te vullen. Daarnaast vragen wij u om een korte vragenlijst in te vullen. Download hiervoor de app CBS Uitgaven op uw smartphone en log in met de gegevens uit de brief. In de brief vindt u meer informatie over het installeren van de app.



Hoe vult u de uitgaven in?
U vult de uitgaven in door op de + (Toevoegen) te tikken in de app. Kies dan voor 'Handmatig invoeren' of 'bonnen scannen'. Scan de bon als u bijvoorbeeld veel boodschappen heeft gedaan. Alle producten en prijzen verschijnen dan in één keer in de app.

Zijn alle uitgaven van een dag ingevoerd?
Tik in het overzicht op 'Dag voltooien'. Doe dit ook als uw huishouden een dag niets heeft gekocht. Dat is ook belangrijke informatie voor ons onderzoek.

Wat krijgt u van ons?
Als uw huishouden twee weken deelneemt en het onderzoek afrondt, ontvangt u als dank een VVV Cadeau kaart van 20 euro.

Hoeveel tijd kost het u om mee te doen?
Meedoen kost weinig tijd. Doordat u kassabonnen kunt scannen, bent u maar een paar minuten per dag kwijt. Helpt u ons mee?

Wilt u meer informatie over het onderzoek of over het gebruik van de app?
Kijk dan in de app of op www.cbs.nl/uitgaven.

Findings

- All Rs said they would open the envelop if they received it at home
 - Letters being addressed to them -> not “the inhabitants of X”
 - Envelop has the logo of CBS.
- 14 Rs saw both letter and leaflet, 6 only saw letter, of which 3 only read front page letter
- 4 Rs say they would require help from family, social workers or CBS for understanding the letter or to download the app (all Rs with low digital skills)
- One respondent thought he should call CBS to make an appointment (possibly triggered by the sentence that an interview may visit, but not discussed with R) but than checked again and understood he was asked to download an app
- Only one respondent when reading the letter stressed the sentence on the household design and the possible visit from interviewer, most respondents only seem to read materials superficially

Letter

Not systematically evaluated, but when discussed:

- Most found letter clear and complete
 - Understood topic is expenditures due to examples
 - Said they did not miss anything
- One R confused about incentive but others did recognize it
- One comment about less visible informed consent statement
- Several spontaneous remarks that gaining insight would motivate them to participate in the study

Leaflet

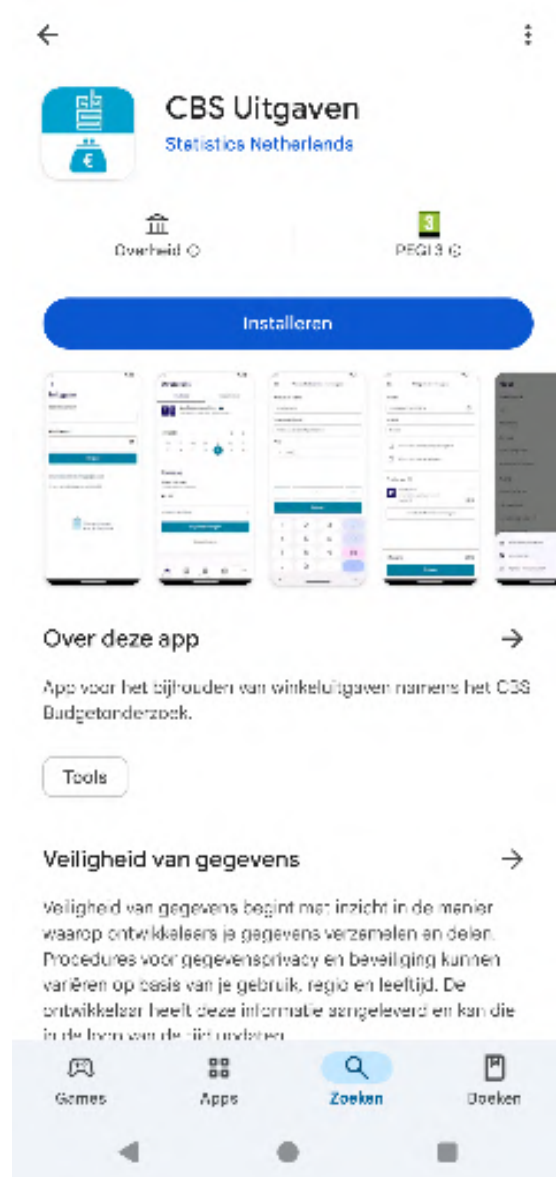
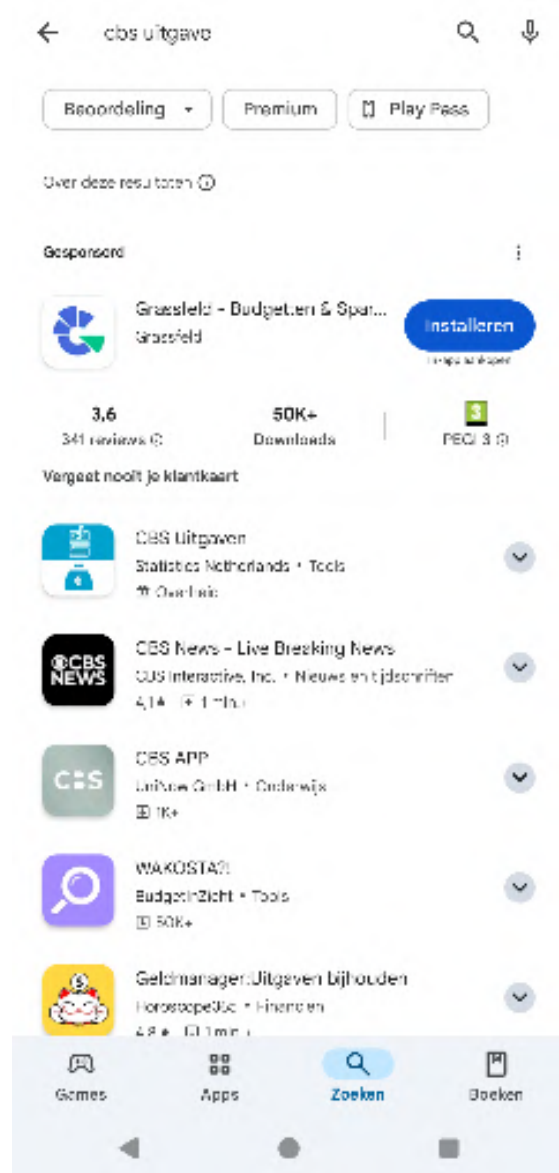
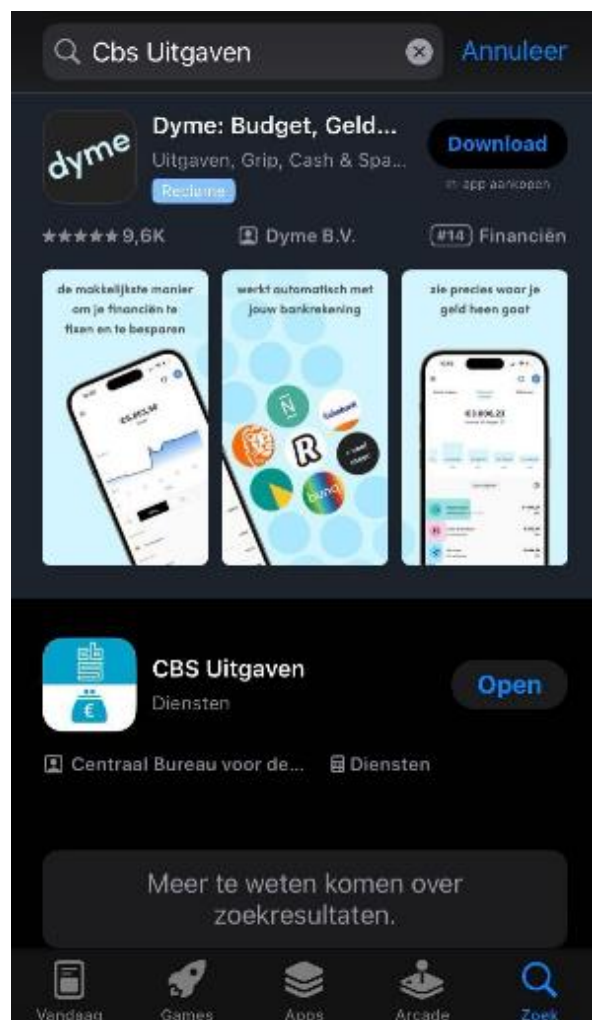
Not systematically evaluated, but when discussed:

- Most found it insightful
- Some remarks on not accessible font size and design
- Some remarks that leaflet was not sufficient and should be more like a manual
- One R thought picture on leaflet did not match study
- One R expected results from previous HBS study

Understanding the response task

Most respondents do not understand the response task at all when they download the app, and many of them still do not even after they have tried the app!

- Many think the app is intended for themselves to gain insight in spendings, not clear that this is a data collection instrument.
- Not clear that there are several tasks: the general and fixed costs questionnaires and the diary (but may be partly due to prototype).
- Not clear that expenditures from all household members should be entered
- Not clear that there is a two-week data collection period
- Not clear which expenditures should be entered
- Not clear which level of detail is needed when entering expenditures



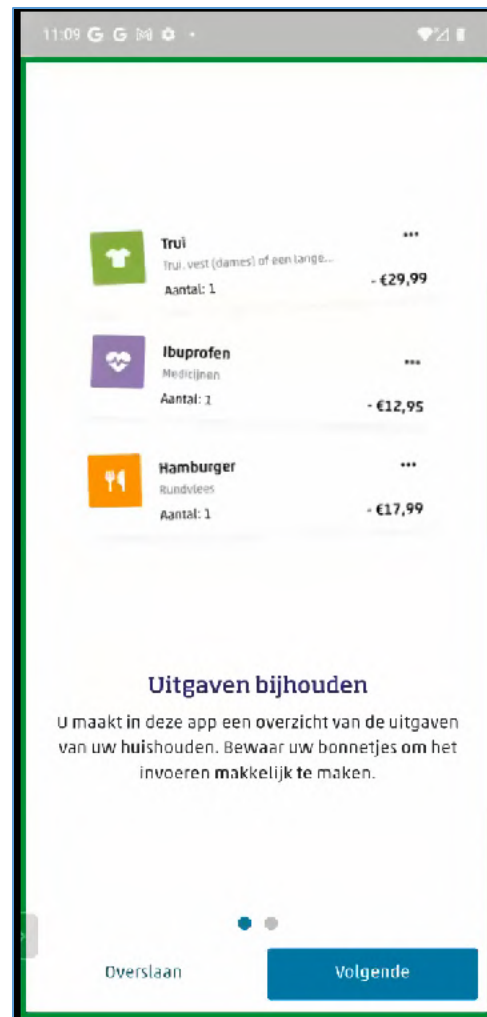
Installing the app

For the first 10 interviews installing the app was only discussed theoretically (“how would you do that?”), for the second 10 interviews respondents were asked to install the app on their private phone.

- QR code: 9 respondents
 - Used camera or QR-code scanner app
- App/play store: 5 respondents
 - As stated in the letter -> “CBS uitgaven”
 - Recognized name and logo. Some verified manufacturer
- Other: Going to the cbs.nl website (1)
- Not able: 5 respondents (all with low digital skills)

Considerations for QR code or manual search

- Use QR code: Convenience
- Not use QR code: recent QR code hacking attacks (1), did not see (2), bad quality of camera (1), does not know how to use QR code (1)
- 1 R with low digital skills a bit reluctant but succeeded with encouragement from interviewer.



Login

Note: passwords used were inadvertently based on security level for business surveys and thus more complicated than usual for CBS household surveys (e.g. use of special characters and upper and lower case)

- Most Rs could login easily, found login credentials in letter and used the “eye” to verify
 - Error in prototype -> most respondents reacted upon
- 3 Rs had problems logging in and used their own e-mail addresses or pressed on “password forgotten”
 - We also see this in error reports of our standard log in page
- 2 Rs could not login by themselves
- All 5 Rs with major problems with login had low digital skills.

Expectations of the app

- Most assumed the purpose of the app was to provide an overview of their spending or to increase awareness of their financial habits.
- None of the respondents mentioned the study as the app's primary purpose.
- Many expected to receive feedback or a summary of their data.
- For some, initial perceptions remained unchanged even after completing the test

May be partly due do to prototype as general questionnaire and fixed costs questionnaire were not implemented.

However, an app may also be more seen as tool for own use and not for data collection by survey institute.

Conclusions and recommendations letter and leaflet, downloading app and login (1/2)

Letter and leaflet do not provide enough information on the response task

Recommendations

- Do not burden letter with more information. Instead try to highlight information on observation period etc. by using bold font.
- Information in leaflet should provide more info on response task. It should be more like a manual for the app itself (info on what and how to add expenses). Use very concrete examples to illustrate.

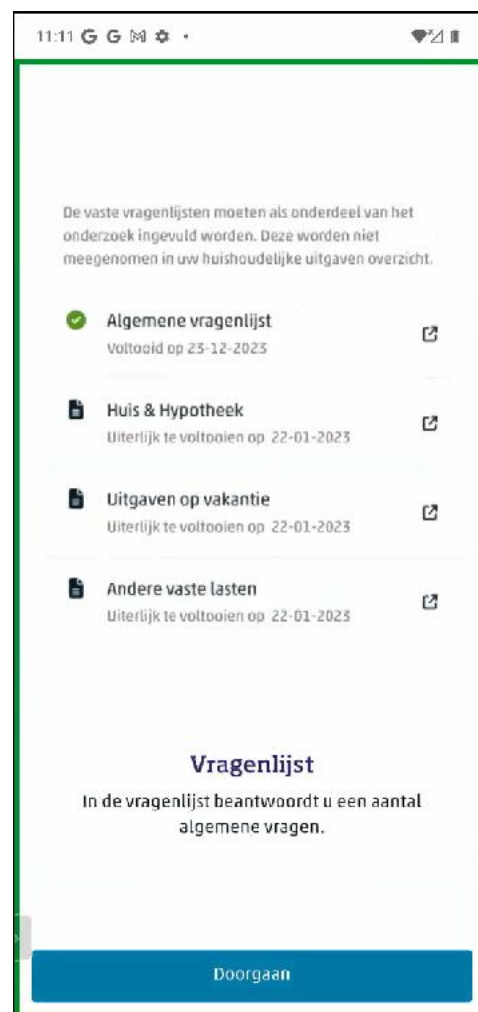
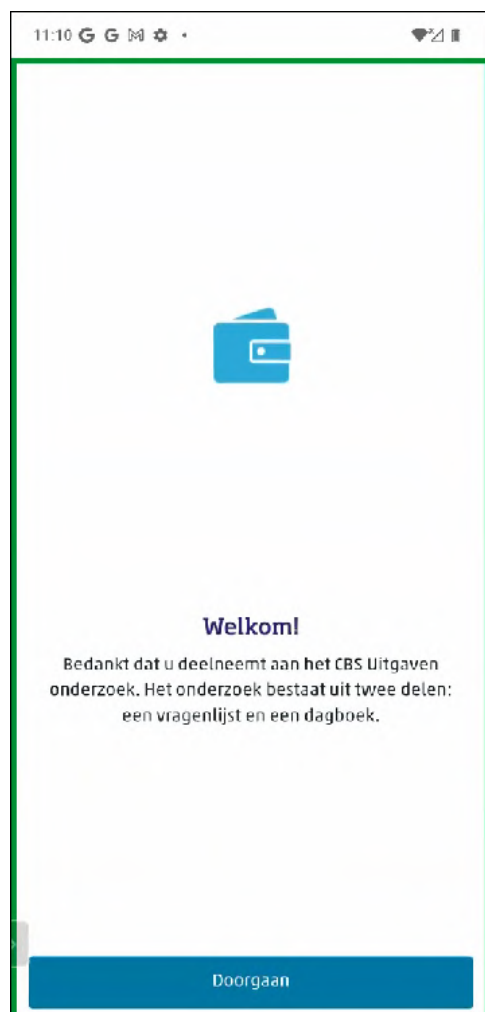
Also more info on response task needed in app (will be discussed later). As some people hardly read the letter and leaflet and others are less likely to read information on the screen, crucial information should be offered both in the app and in the letter and leaflet.

Conclusions and recommendations letter and leaflet, downloading app and login (2/2)

Instructions on downloading and login can be easily found and used for people with some digital skills.

- Important to offer both options (QR-code and app/playstore)
 - Respondents seem to have a clear preference
 - App seems locatable and recognized sufficiently in the app
 - Consider to mention QR option first to nudge Rs to this most efficient option.
- Important to offer personal help and/or other data collection modes to also reach people with lower digital skills.

Onboarding



11:11 G G M ⚙ • 📶 🔋

Uw invulperiode

U kunt meedoen aan het onderzoek van maandag 20 januari 2025 tot en met zondag 2 februari 2025

Is iemand uit uw huishouden deze periode 7 nachten of meer op vakantie?

☐ Ja

☐ Nee

Doorgaan

11:11 G G M ⚙ • 📶 🔋

Uw invulperiode

U kunt meedoen aan het onderzoek van maandag 20 januari 2025 tot en met zondag 2 februari 2025

Is iemand uit uw huishouden deze periode 7 nachten of meer op vakantie?

☒ Ja

☐ Nee

Omdat uitgaven bijhouden op vakantie lastig kan zijn, willen we u vragen om voor een andere invulperiode te kiezen. Selecteer hieronder de eerstvolgende invulperiode waarin niemand uit uw huishouden meer dan 7 nachten op vakantie is.

☐ Maandag 27 januari 2025 tot en met zondag 9 februari 2025


☐ Maandag 3 februari 2025 tot en met zondag 16 februari 2025

☐ Maandag 10 februari 2025 tot en met zondag 23 februari 2025

☐ Maandag 17 februari 2025 tot en met zondag 2 maart 2025

Doorgaan

11:11 G G M ⚙ • 📶 🔋



Uw dagboek start op maandag 20 januari 2025!

Doorgaan

Onboarding

- Rs with moderate to high digital skills navigated the onboarding process smoothly, using swipe gestures.
- Rs with lower digital skills encountered difficulties navigating the onboarding process: didn't recognize the onboarding and assumed they needed to take action.
 - They were often more meticulous in reading the onboarding texts
- Used terminology as “expenses” was misunderstood and had different meanings to users.
- Even some individuals with moderate to high digital skills who stated the onboarding was clear where uncertain about the next tasks.
- Different time periods in both calendars in onboarding and calendar in real study is confusing/ irritating for some Rs
- Question on vacation is not clearly presented (font too small)

User interface and user experience (1/2)

- The design was described as professional
 - Which contributed to the confidence in the app and in CBS
- Functionalities were perceived as intuitive and easy to use
- When asked about overall opinion, almost all respondents in general positive about the app (when also pointing out improvements needed)
- Onboarding, completing the vacation questions, the use of the calendar, the marking of the reporting period in the calendar, how to add an expense and navigation via the menu at the bottom of the page were all intuitive for most respondents
- But some users with lower digital skills had difficulty locating the button for adding expenses

User interface and user experience (2/2)

- Many had difficulty locating a back button or returning to a previous screen (back link does not work as expected; Rs expect to go back within the app but they leave the app when choosing back button).
- Most users did not check the added amount in the overview
- A few experienced challenges with the fonts (color, size and font)
- App did not seem to generate strong engagement or long-term motivation.
 - When asked if they would participate in the @HBS study, almost all say they would but....
 - Several also stated they did not think they would keep track of all expenses for two weeks
 - Several stated the output as presented in current was not relevant to them (but some also find current list of expenses interesting)

Onboarding: conclusions and recommendations

Onboarding does not have a clear goal and fails as a supportive tool.

Recommendations:

- State the goal of the use of the app clearer before the onboarding (letter and leaflet)
- Relocate welcome screen before login.
- Simpler use of language + important terms like 'expenses' should be explained
- Enhance comprehension and provide additional support in navigating the application, example of using images, animations or videos
- Show logo of CBS and link to research project more prominent in the app (is now hidden behind the i – button) to keep Rs aware of the purpose of using the app; explore if reference to CBS / study can be communicated more prominently in other elements of the app.
- Help improve transparency and user understanding (by incorporating missing details about the purpose of the survey and referencing CBS)
- Use same example calendars in onboarding and if possible choose a month that is obviously different from reporting period.

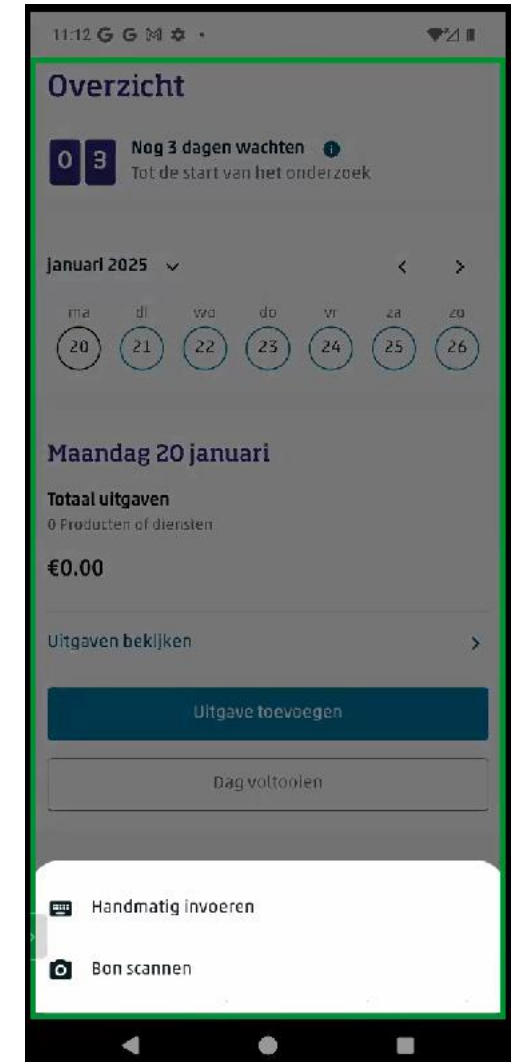
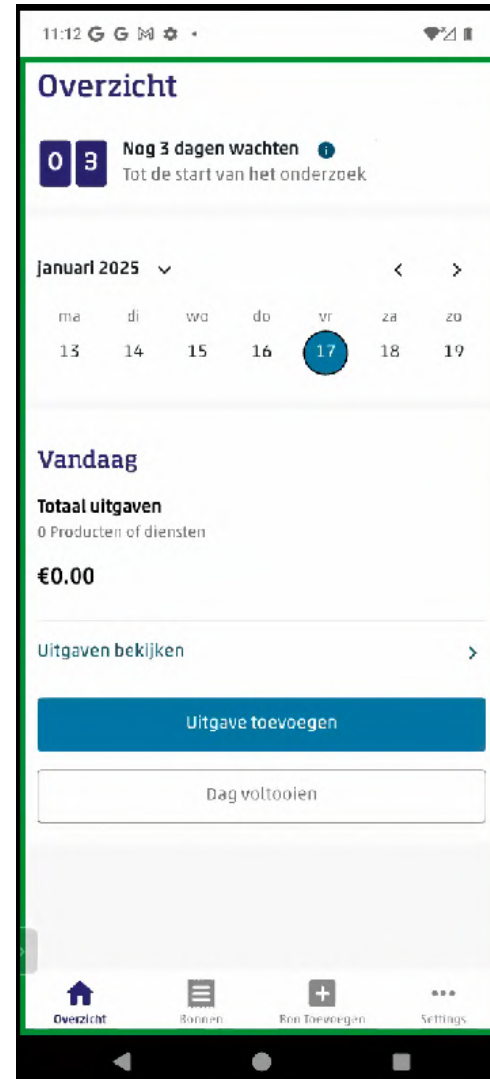
UI and UX: conclusions

- Key features clearly presented and accessible.
- Respondents appreciate the look and feel of the app.
- However, use of the app is challenging for users with low digital skills
 - Finding the important 'add expenses' button in the overview screen challenging for this group
- A potential lack of engagement or motivation was observed with respondents expecting but missing personal value or benefit of using the app.

UI and UX: recommendations

- Explore alternative placements or design for the *Add Expenses* button (enhancing visibility and accessibility)
- Reconsider the design of back-navigation and close buttons—consider adding text labels to indicate where they lead
- Change font of vacation question to make it clear this is a question, make it as large or separate it from the introduction text on the reporting period.
- Incorporate more motivational or engaging features. This could involve integrating elements that provide personal financial insights (possibly after completing the reporting period if there are concerns about affecting their spending behaviour with feedback).
- Check consistency of wording in the app (e.g. use of ‘products’ and ‘products or services’).

Entering expenses



Scanning or manual entry: first preferences

When the respondents were asked how they would prefer to enter the first (short) short receipt before having tried either option:

- 15 Rs prefer scanning receipts, because they perceive it to be faster and easier than entering them manually
- 3 Rs prefer manual entry
- no preference was recorded for the remaining two.

Preferences after experiencing both options

- Most prefer scanning (13 Rs)
- Some prefer a mix: manual for short receipts and scanning for long receipts (4 Rs)
- 2 prefer manual entry
- For 1 R preference not discussed in interview

Considerations for scanning or manual entry

Pro scanning

- Quicker and less effort
- Provides CBS with additional data

Pro manual

- Long processing time for scan with errors and lots of editing not worth the time, quicker to enter manually
- Easier than correcting scanning errors

Manual data entry flow

11:12 G G M S +

✕ Uitgave toevoegen

Datum

20 januari 2025

Winkel

Bijv. Mediamarkt

☐ Dit is een buitenlandse uitgave

☐ Dit is een online uitgave

Producten (0)

Voeg alle producten die u heeft gekocht apart van elkaar toe aan de uitgave.

+ Product of dienst toevoegen

Totaalprijs €0,00

Opslaan

11:12 G G M S +

< Winkel

Zoek een winkel

Zoek naar een winkel zoals Albert Heijn, Mediamarkt of Etos

q w e r t y u i o p
a s d f g h j k l
z x c v b n m

?123 , ' ~ . ✓

11:12 G G M S +

< Winkel

Zoek een winkel

Jumbo

? Jumbo
Onbekend

q w e r t y u i o p
a s d f g h j k l
z x c v b n m

?123 , ' ~ . ✓

11:12 G G M S +

✕ Uitgave toevoegen

Datum

20 januari 2025

Winkel

Jumbo

☐ Dit is een buitenlandse uitgave

☐ Dit is een online uitgave

Producten (0)

Voeg alle producten die u heeft gekocht apart van elkaar toe aan de uitgave.

+ Product of dienst toevoegen

Totaalprijs €0,00

Opslaan

11:12 G G M S +

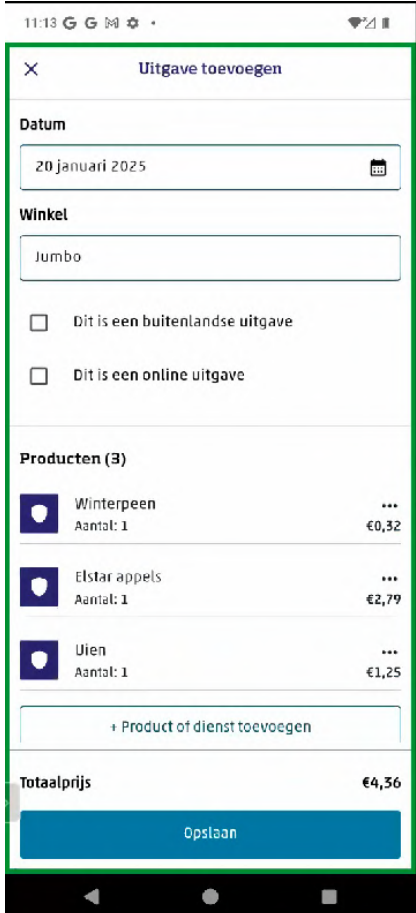
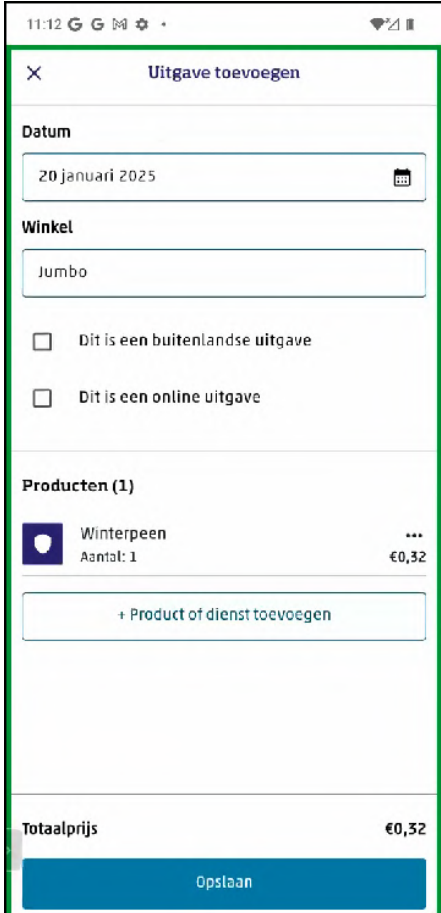
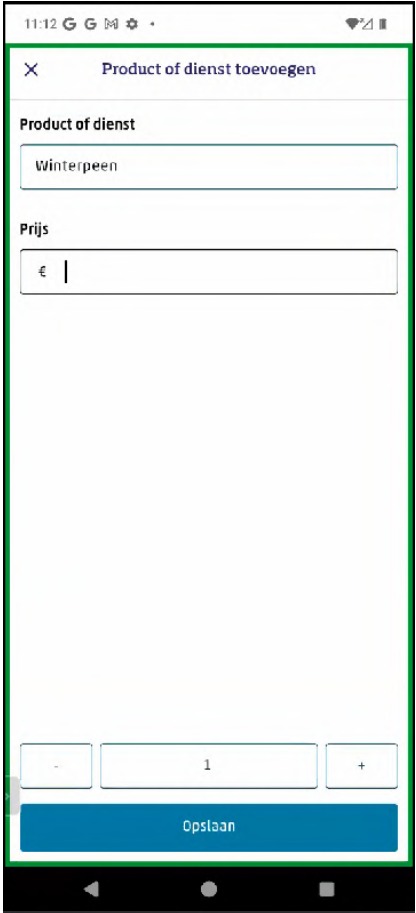
< Product of dienst toevoegen

Zoek een product of dienst

Zoek naar een product of dienst. Het is niet nodig om karaktereigenschappen toe te voegen aan het product (zoals damestrui of zwarte trui).

q w e r t y u i o p
a s d f g h j k l
z x c v b n m

?123 , ' ~ . ✓



Manual entry: experiences

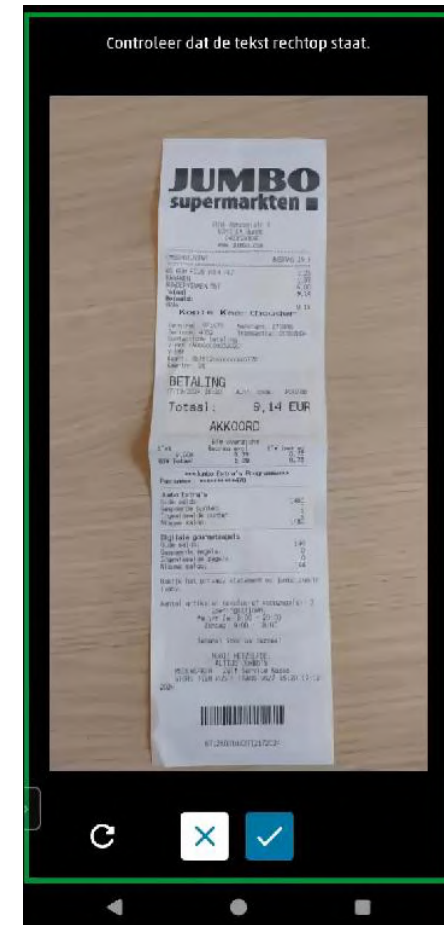
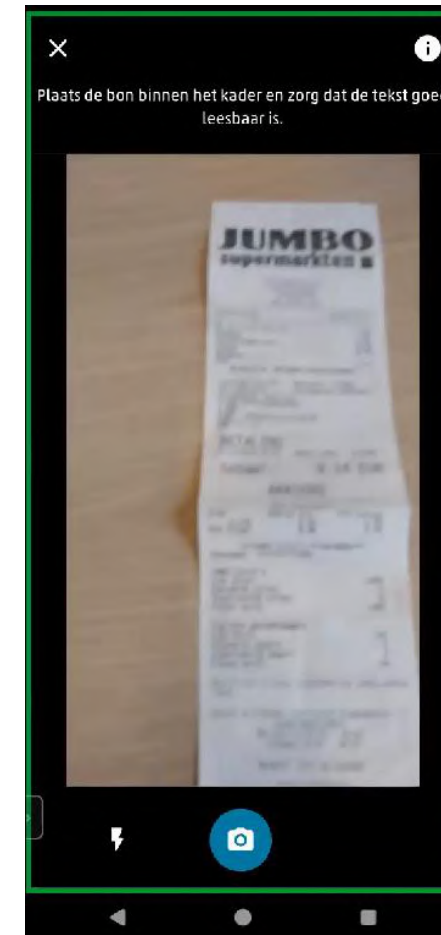
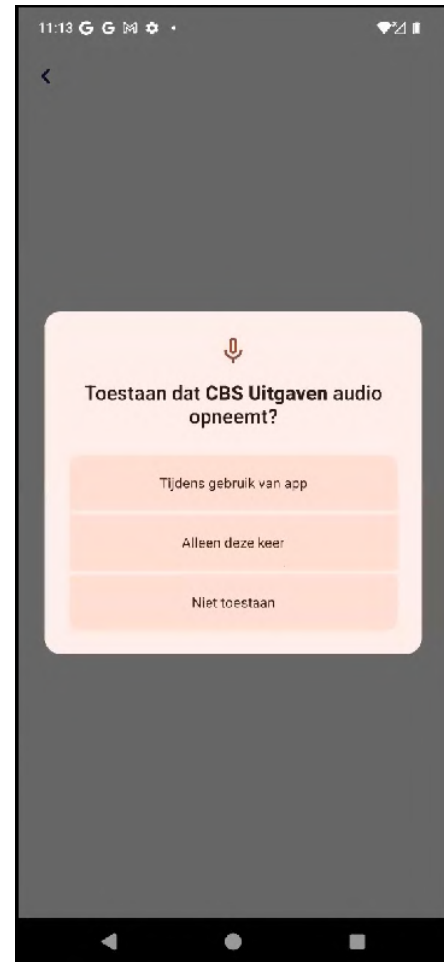
	Short test receipt	Personal receipt
No issues	10	11
Minor issues	2	2
Major issues but no help needed	2	2
With help	4	2
Unable to complete task, even with help	2	0
Task not tested	0	3

Many Rs use total as check for their data entry

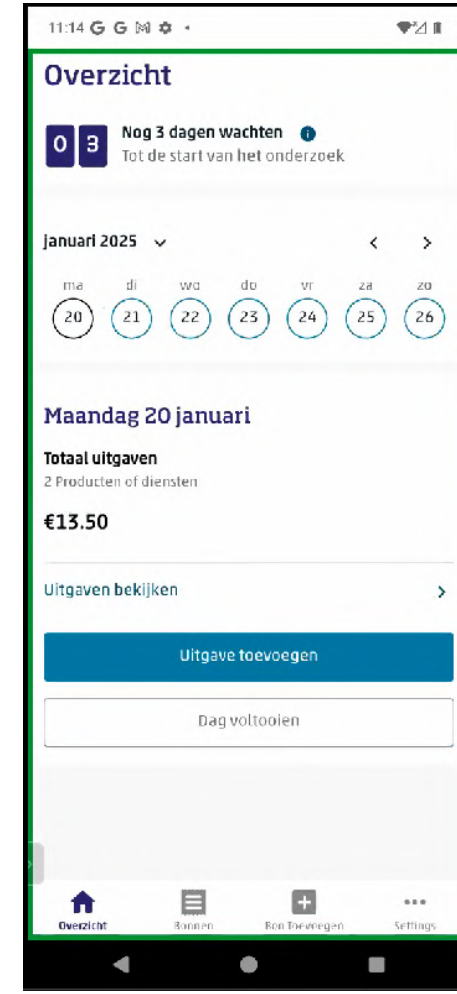
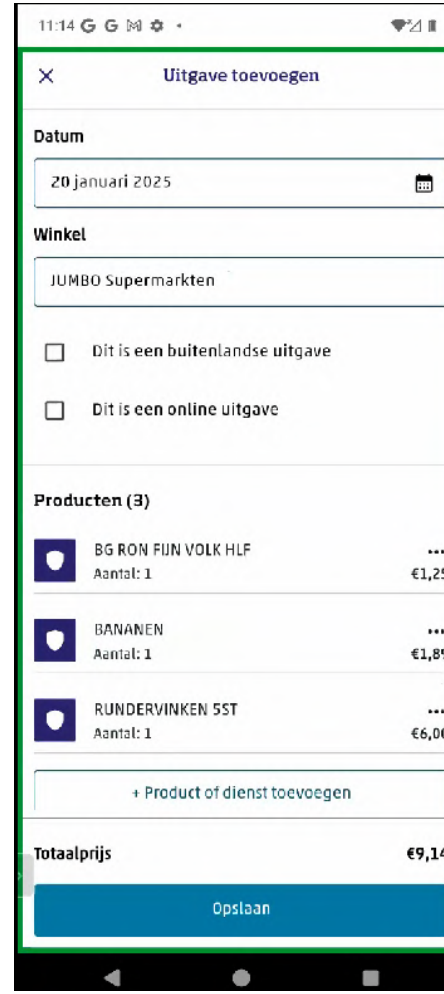
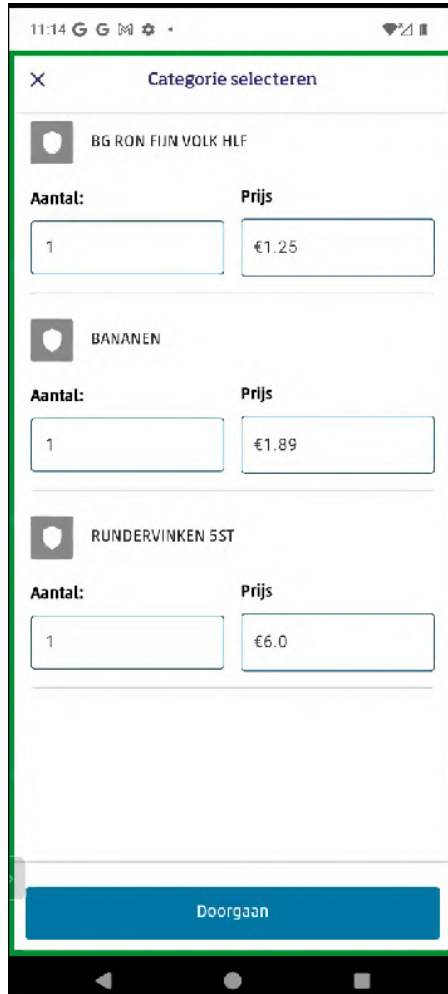
Short test receipt: : of the 8 Rs with major issuer or who needed help, 7 had low digital skills

For personal receipt: of the 4 Rs with major issues of who needed help 3 had low digital skills

Scanning flow (1/2)



Scanning flow (2/2)



Scanning experiences

	Short (without errors)	Medium (with errors)	Long (with errors)
No issues	10	10	7
Minor issues	5	7	8
Major issues but no help needed	3	1	2
With help	2	1	1
Unable to complete task, even with help	0	0	0
Task not tested	0	1	2

Many Rs use total as check for their data entry

Short receipt: : of the 5 Rs with major issuer or who needed help, 4 had low digital skills

Medium receipt: of the 2 Rs with major issues or who needed help, 1 had low digital skills

Long receipt: of the 3 Rs with major issues or who needed help, 2 had low digital skills

If issues, issues found both with manual entry and scanning tasks

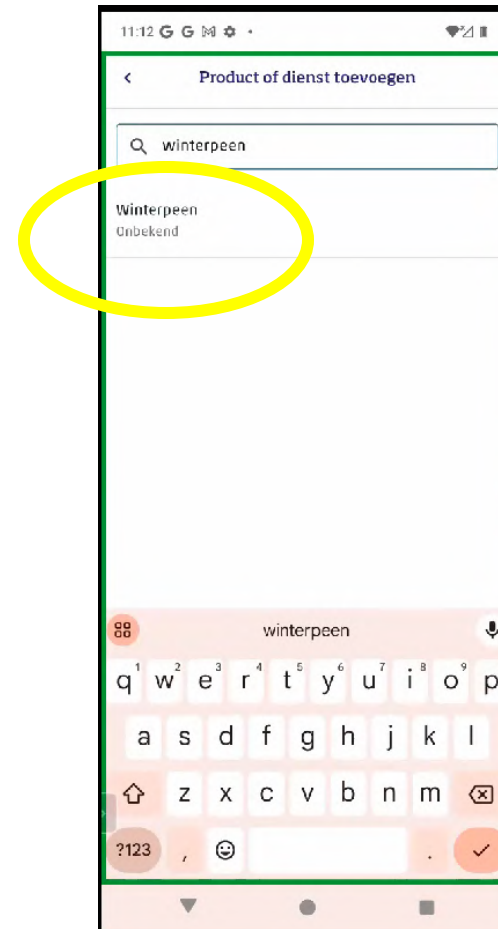
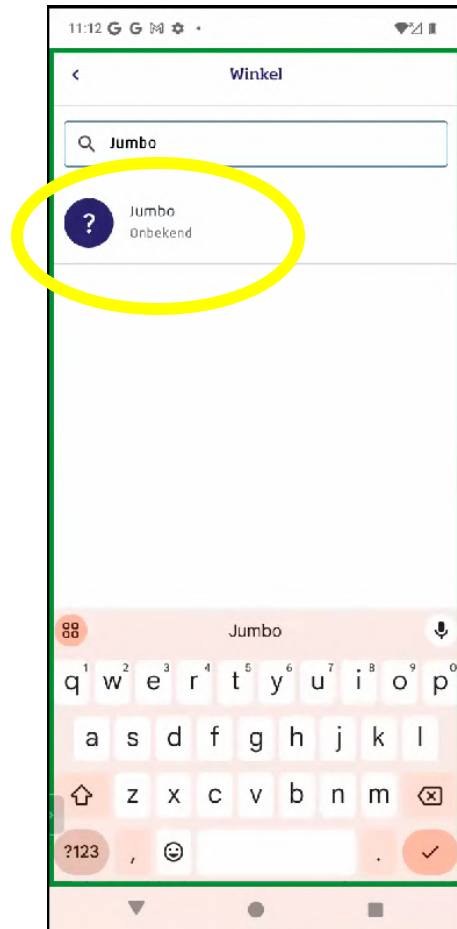
- Check boxes for “foreign expense” and “online expense” confusing for many
 - 7 Rs think they must choose one option, 1 expresses doubt if these can be left blank
 - 3 Rs find the term foreign expense unclear
 - 1 R suggest adding an option for “Dutch store” to the store info
- Not clear if and how to report discounts, deposits and refunds
 - Several strategies seen for discounts: ignoring (not reporting), calculate paid price either by heart (partly making errors in doing so) or with calculator, entering the discount with minus sign, entering a subtraction formula in the price input.
 - Two Rs feel that deposits and stamp savings should not be entered as they will be refunded.
- Not all receipts contain recognizable / individual product names. Even with own recent receipt R could not reproduce which price belonged to which product. These types of receipts are typical for smaller local shops and market stalls.
- Notation of decimals (both , and . allowed in the app).
- 1 R notices that “add an expense” assumes there is already something entered. Better may be “enter expense” or ask for each day “did you have expenses for this day?”

If issues: issues specific for entering manually (1/2)

- Product details not entered correctly (e.g. “groceries” in stead of item descriptions) or uncertainty about required level of detail.
 - If/how to enter quantities (e.g. how to enter 1 package with 5 sausages, kg or number of carrots).
 - Some Rs type in products exactly as they are on the receipt)maybe putting in more detail than needed) others use briefer descriptions
 - 1 R would like to scan the barcode on a product to enter the product details
- Many Rs show or say that they expect search and autofill option when entering stores and products (not yet implemented in tested prototype).
- Some Rs also expect that once a store is added it will not be “unknown” anymore.
- Not immediately clear and some really struggle with selecting the Storename or Product to add it (see example next slide), may be due to lack of search function in prototype.

How to select store/item not intuitive (but may be due to prototype without search function)

After typing store/item
Rs must select
store/item by tapping
on highlighted text



If issues: issues specific for entering manually (2/2)

- 1 R entered personal receipt from Lidl app on their own phone, this would not work easily if the @HBS app was also on their phone.
- 1 R (with low digital skills and limited Dutch knowledge) says the task requires concentration, especially with the numbers
- 1 R (with low digital skills and limited Dutch knowledge) also enters the total of the receipt. Sees that the total of the receipt is not correct but does not understand why.
- 1 R (with low digital skills and low literacy) could not figure out how to add a product, stopped task frustrated.
- 1 R is annoyed that with each product that is entered both the checkmark at the bottom of the keyboard (to indicate entering text is ready) and the save button (in the app) have to be pressed.

If issues, issues specific for scanning (1/2)

- Several Rs find it strange that for accessing camera also permission for audio access is required.
- 1 R thinks scanning is done outside the app and goes to QR app
- 2 Rs thought the phone would automatically start scanning without tapping the camera icon
- Many Rs make pictures of the receipt without all corners visible (thus ignoring the instruction that is shown each time a picture is made)
- 7 Rs try to scan the barcode on the receipt in stead of taking a picture of the receipt.
 - When informed that they should get a notification that the entire receipt should be scanned some still keep scanning the barcode (but may be because scanning the bar code also leads to scanned data in the simulation).
- Several Rs express doubts about what part of the receipt should be in the picture
- Several Rs express uncertainty if the picture is readable, they are unable to check this by zooming in on the picture.
- Taking a picture of the large receipts is challenging for some, some stand up to be able to take the picture.
- 1 R continues holding the phone in the same position after the picture is taken, thinking that moving may interrupt the scanning process

If issues, issues specific for scanning (2/2)

- Many Rs became impatient with the long processing time, especially for the long receipt. Sighs and irritation were observed as well as some participants outright stating it.
 - Several said or showed that they would exit the scanning process if it takes too long.
 - They noted that the process takes longer for longer receipts and mentioned they might try taking a new photo instead.
 - 1 R suggests users should be informed about the longer waiting time for processing, to prevent them from stopping midway.
- 5 Rs do not notice there are errors in scanned data
- 1 R says more instruction for scanning is needed for less digital skilled and non-native Dutch users, explains that non-native users will benefit from this as they often use a translator app on their phone
- 1 R who shops a lot in Germany states it would be preferable that also receipts from outside the Netherlands can be scanned. In border regions it is quite common to shop abroad.

When noticing errors in scanned information

- Response to noticing errors mixed:
 - Some would try to make a new picture
 - Many would try to adjust the errors, but not necessarily all errors and always (depending on if they have the time or how important they think the errors are)
 - A few say they would not adjust errors
 - 1 R would enter the receipt manually
 - 1 R thought the paper receipt contained errors as it did not match the scan
 - 1 R states that if he would have to correct all errors in the long receipt, he would not fill it in

Editing scanned information

	Medium (with errors)	Long (with errors)
No issues	12	8
Minor issues	1	1
Major issues but no help needed	1	3
With help	4	3
Unable to complete task, even with help	1	1
Task not tested or results not clear	1	4

Most Rs were succesful in correcting errors by adjusting prices, editing product names, adding a missing product or removing a wrong item.

Issues while correcting errors

- The two editing screens with different functionalities are very confusing and not functional
 - Total price is not visible on the first screen
 - Items can only be added in second screen
 - Not sure where they should make corrections (can they still change things after pressing 'proceed' on the first screen?)
 - Second screen seems most intuitive (and providing overview with total amount), using the three dots to edit works intuitive as well.
 - One R wants to swipe to delete items
- 1 R tried to remove a product by setting the number of product to zero, but product remained on the list (without an amount).
- 1 R tried editing a name which was not possible due to bug in simulation
- 1 R needed step by step instructions

Learning curve entering expenses

- Significant learning curve, with participants becoming much faster after adding their first expense.
- For manual entry most Rs became increasingly more efficient when manually adding products over time. Some Rs also outright stated that after they had done or seen it once they would be (more) successful doing it again.
- Scanning behavior changes over the tasks. At first many Rs tried scanning only the barcode at the final task only 1 R did. At the final task many Rs were trying to only scan the products and prices instead of the whole receipt. Indicating they changed their perception of what was needed for the scanning process based on what they saw as feedback from the scanning (and not based on the instructions shown with each scan).
- The first time errors occur respondents are searching and thinking of ways to correct the problem. After correcting an issue the first time they are much faster doing it again. Even respondents who needed step by step instructions the first time were often able to do it on their own the next time.

Entering expenses conclusions

- Entering expenses is easy for Rs with some digital skills and learnable for most Rs with low digital skills (if helped).
- Rs appreciate the scanning function very much.
- But many improvements needed to ensure response and data quality.

Entering expenses recommendations (1/2)

- Enable entering data on multiple devices, preferably also on a PC; for some easier when entering a lot of data manually and easier when having to copy data from a (store or banking) app on their phone.
- Reconsider the wording of “add expenses” (“uitgave toevoegen” in Dutch). Enter expenses (“uitgave invoeren”) may be better.
- Redesign the “foreign expenses” and “online” expenses to make clear these are only to be checked if applicable. Add an instruction stating this.
 - Consider the use of “expenditure in foreign store” to make clearer what is intended (one can do expenditures at home in shops abroad).
- Consider using “take a picture” in stead of “scan a receipt” as this may be a more correct description of the task and prevent scanning the barcode.
- Use in the instruction of the scanning a more recognizable version of a receipt (or picture).
- Enable zooming in on the picture taken to check it.
- Preferably start scanning receipt automatically once “scan receipt” is selected (let camera detect receipt).
- Give immediate feedback in the app about quality of picture of receipt, instructions are overlooked.
- If checking the scanned and/or manually entered data is important this should be made an explicit step in the data entry process; make totals of receipt easily available to check this (but be aware that sometimes totals do not have to match if expenses for business or other household are included on receipt).

Entering expenses recommendations (2/2)

- Communicate clearly if pictures of receipts are stored and used (to be transparent and – if we do not store the receipts – make clear we cannot correct errors in the scanned data).
- Skip the first edit screen after scanning
- Improve the processing time when processing pictures, if not possible explicitly inform Rs about this (to prevent frustration and help them plan their data entry).
- Provide clear instructions on the response task, including details on discounts and refunds. Use these also in examples shown.
- Search list for stores and expenses are needed to both make data entry easier and to show the desired level of detail.
- Only accept comma notations for decimals (according to what is customary in Dutch).
- Add the option to upload digital receipts and to add pictures from camera roll.
- Add the option to scan receipts from neighbour countries.

Trust and privacy

- Personal concerns about privacy hardly mentioned spontaneously and if so mostly in positive context (“I trust CBS”, “a letter is more trustworthy than an email”.)
- Nobody mentions trust considerations when asked about their impression of the app and only one person when asked if they would participate in this study (“ I would participate in this study, I know CBS has developed this app carefully”) .
- If asked directly: all say they trust the app
- But... when talking about data quality several mention that others may not participate because they may have issues with trust and privacy

Issues with trust and privacy

- A few mention that the permission for audio access needed for scanning strange
- A few mention they would not like to report certain expenses (e.g. medical expenses, mortgage)
- One R has a general concern for safety scanning QR code
- One R would not report details of small businesses where things are bought (“it is not up to me to report what they sold, I do not know how they keep their books”).
- One R would not like to share data on expenses with parents in household

Factors enhancing trust (or not caring)

- Trust in CBS/government/laws and regulations
- Data on groceries not sensitive
- Use of a letter (as opposed to email)
- Option to contact CBS by phone to check authenticity
- App looks professional/serious/developed carefully
- No personal info required for login (e.g. email or social security number)
- Required permissions can be restricted to when using the app.
- Pictures of receipts not stored in the app
- “In this day and age everybody is watching what you are doing anyway”.

Does mode of data entry matter for trust and privacy?

Scanning or manual data entry not relevant for trust and privacy

- But permission for audio access when scanning strange
- One respondent mentions there may be sensitive data on a receipt but trusts on protection of these data because of regulations

Trust in data quality / quality of the statistics based on the app

When asked, many mention concerns about the data quality:

- Selective/reduced response (due to trust/privacy concerns, language problems, not being able to use app and/or finding it too much effort)
- Data entry prone to error (both quality issues with scanning and making errors with manual entry mentioned spontaneously)

Does mode of data entry matter for data quality?

- Several Rs think scanning is better:
 - Better quality
 - Less effort -> higher likelihood people will do this
 - More control of data for CBS
- One R thinks manual entry is better as scanning has errors and people may not make the effort to correct errors
- One R appreciates both modes of entry are possible, as not everybody will be able to scan receipts.
- One R mentions that upload of digital receipts should be added as people will forget to collect paper receipts.

Conclusions and recommendations Trust

- For our test respondents trust and privacy no big issue, *but this may be because of selective group and setting of the test.*
- Approach strategy seems to enhance trust: letter, name of institute, possibility to contact institute, use of anonymous login details, professional look & feel of the app
- Permission for audio access for scanning should be skipped or if that is not possible explained
- To be considered
 - Option to use Face ID and/or other type of (optional) protection for accessing app on phone (mentioned by one R in the context of usability).
 - Option to not share data within household

Features desired by Rs

- More insight in own spendings
- Option to zoom in on picture taken to check quality
- Uploading digital receipts
- Synchronize app with digital receipts in apps from stores
- Synchronize app with app with digital shopping lists
- Scanning a barcode on a product to enter product details
- Being able to share the app with others who may be interested in it
- Option not to show scanning instruction every time

Other findings & ideas (1/2)

- Term “private” expenses was wrongly interpreted as personal expenses of R (vs common of expenses of R and partner). -> reconsider the term “private”, maybe better “do not report business expenses”.
- Regarding the data collection period 1 R commented that she would postpone the groceries she would usually do in the weekend so that they would fall in the reporting period. -> stress in instructions we want to collect data about how they would usually spend in the data collection period.
- One R stated that interviewers could be used to motivate respondents and aid in privacy concerns or other questions -> interviewers may be helpful to reach certain groups
- One R answers when asked if he would participate “I think so, also because I don't spend a whole lot in a week or in a fortnight, so it's not that much extra work. The fact that it can be done by phone and is an app would be more likely to persuade me to take part than having to do it behind a computer every time or having to do it separately. I am on my phone more than enough so then that is also a small effort. -> indicates risk of selective response by people who do not spend too much and indicates importance of offering the option to participate using a phone
- One R says the term diary is not logical/clear -> reconsider using “diary”

Other findings & ideas (2/2)

- Some Rs spontaneously share how they would go about with the timing of filling in the expenses; some say they would do it immediately in the store, others would do it in the evening and one would do it weekly. -> to improve data quality, consider nudging Rs to frequent (and at least daily) reporting
- One R was not sure if it was possible to use the app on a tablet, which would have been the preferred device -> consider showing in the leaflet that the app works on a tablet, for those who prefer tablets it may be also better for data quality and response burden if they use the tablet with a larger keyboard.

Strenghts and limitations of this test

- 😊 In-depth observation and evaluation of response flow
- 😊 In-depth observation and evaluation specific tasks
- 😊 Diverse test group
- 😊 Simulation felt real to test respondents
- 😞 Artificial setting (in the lab, test phone)
- 😞 Test respondents may be positively biased towards CBS
- 😞 Test with prototype
- 😞 Only diary part of response task for @HBS tested

Retrouvez-nous sur

insee.fr



Simon Moreau (simon.moreau@insee.fr)

HBS

DSDS / DRCVM / Division conditions de vie

Usability test: Geo-assisted entry function

Smart Survey Implementation – WP 2.3

02/2025



Disclaimer: Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or Eurostat. Neither the European Union nor the granting authority can be held responsible for them.

Table of contents

03	Management summary	39	User assessment of the geo-assisted entry
06	Most important findings	45	Usability of the geo-assisted entry
18	Background & objectives	63	Trust & privacy
21	Methodology & participants	71	Selecting activities
31	Results in detail	78	General issues
33	First impressions	83	Technical issues & bugs
35	User behaviour: entry creation	87	Contact
37	Point in time of entry creation		
38	Preferred entry option		

Management summary ^(1/3)

Background

Smart surveys are surveys conducted with the help of mobile devices which combine questions for self-reporting and smart features for collecting data. On behalf of *Eurostat* the international project *Smart Survey Implementation (SSI)* is working on the development of different smart features. One of them is a microservice which collects location data to support entry creation in time-use diaries. The aim is to reduce response burden and improve data quality by providing locations and times of participants whereabouts as suggestions in a newly developed geo-assisted entry function.

Methodology

A usability test was carried out in December 2024. The test object was an app with a prototype of the geo-assisted entry as a smart feature which was not yet fully developed at the time of the test. 14 participants tested the app on site at the Federal Statistical Office in the presence of an interviewer. Moreover, a focus group with 5 participants tested and discussed the new functionality.

Research questions

BENEFIT

Do users see a benefit in the geo-assisted entry?

USABILITY

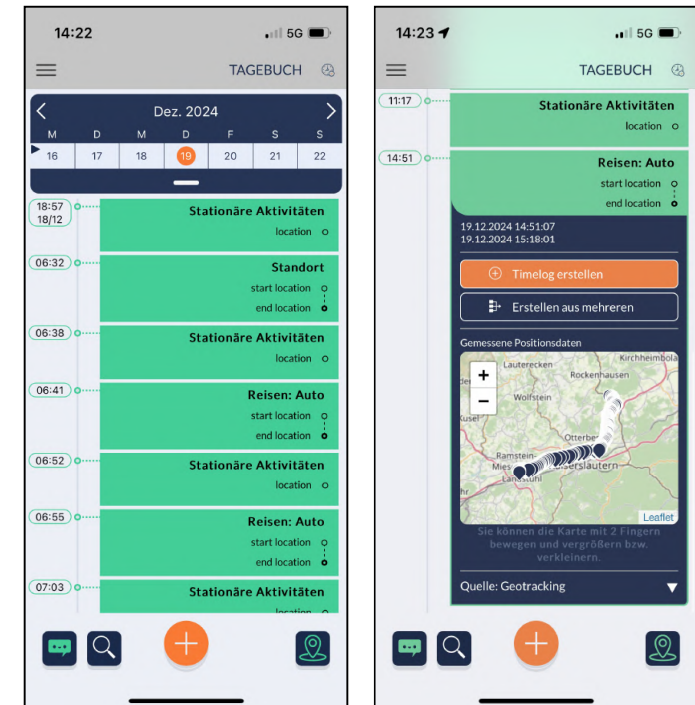
How easy is the use of the geo-assisted entry as part of a diary app?

Management summary ^(2/3)

Results benefit

- » The general benefit of the geo-assisted entry depends on users' entry behaviour and their activities:
 - » If activities are entered some time after they took place the geo-assisted entry offers a thought support function.
 - » Users who entered their activities immediately after they took place did not need a reminder. Moreover, as geo-based suggestions were often only available with some delays, the benefit of the new entry mode could not be noticed by them.
 - » If successive activities take place at the same location, adding entries with the geo-assisted function is more complicated than using the manual entry.

Current design

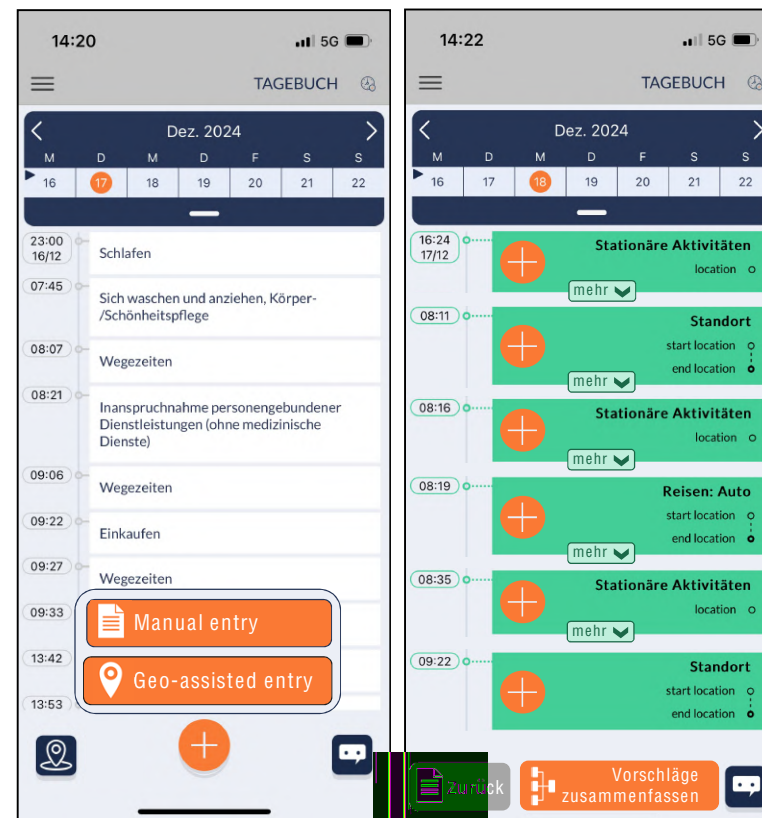


Management summary ^(3/3)

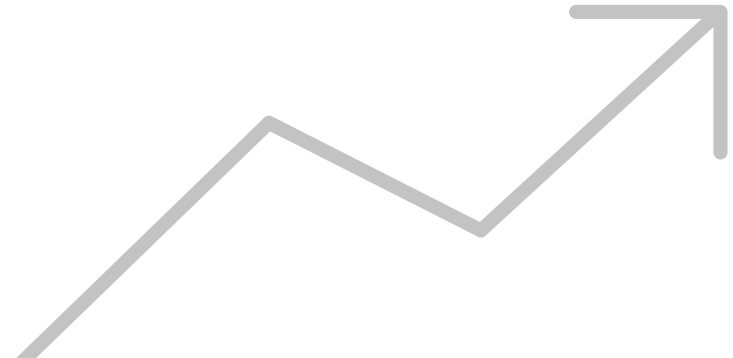
Results usability

- » The usability of the geo-assisted entry can be improved in different sections.
- » In general, neither the purpose of the suggestion list, nor the difference to the diary is clear to everyone
- » Users are confused where they are and what to do, they also don't know how to navigate between these two screens.
- » Entering a new activity in the geo-assisted function is too complicated. It takes too many taps and some users enter activities in the suggestion list with the manual entry (instead of the geo-assisted entry), as they use the wrong button (orange plus).
- » We recommend a redesign which makes navigation easier to understand and offers a more direct way to enter activities. This should also make the purpose of the suggestion list more obvious.

Suggested redesign



Most important findings



Benefit and purpose of the geo-assisted entry

Perceived purpose and benefit depend on the entry behaviour of users

- » Test persons (TP) preferred to enter activities immediately during the day (not all at once in the evening)
- » It often **took a while until geotracking results were available** in the app
 - » TP could not use the geo-assisted entry immediately
- » TP did not see an advantage of the geo-assisted entry and rated it as **unuseful**
- » To those TP who entered their activities at once in the evening the recalling benefit of the **geo-assisted entry was becoming clearer** and they consequently perceived the function as **helpful**

Benefit and purpose of the geo-assisted entry

Benefit depends on users' activities and movements

- » The geo-assisted entry is **helpful for activities during which locations change** and those which take place in a different location than the activity before
 - » The geolocation suggestions then provide a **thought support to recall these activities**
- » If activities take place in the same location as the activities before, the geo-assisted entry is not an advantage but a hurdle:
 - » If no location change happens, the **suggestion list contains only one long suggestion for that time period.**
 - » Users then have to **divide** the one long suggestion **into several** shorter **activities**
 - » In this case it is easier to use the manual entry

Benefit and purpose of the geo-assisted entry

The geo-assisted entry is a helpful third entry option

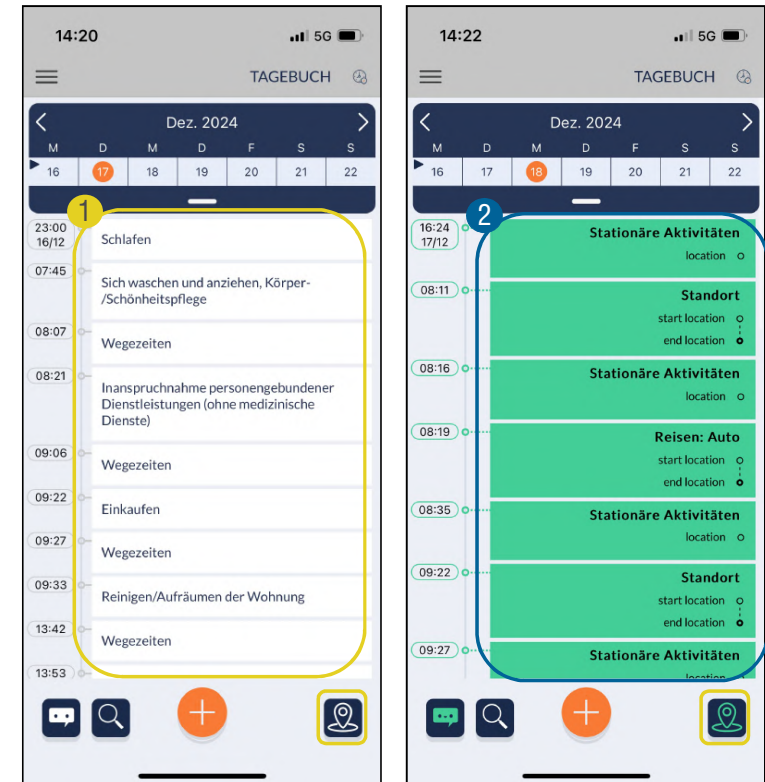
- » Besides the manual entry and the “ongoing activity” entry, **the geo-assisted entry is a useful third entry option**, making diary entries easier and reducing respondent burden
- » If more information would be provided automatically in the suggestions, it will be even more helpful, e.g.
 - » Mode of movement (car, bike, train,...)
 - » Start and end location (supermarket, train station, ...)



Usability – suggestion list

Suggestion list and navigation is not clear

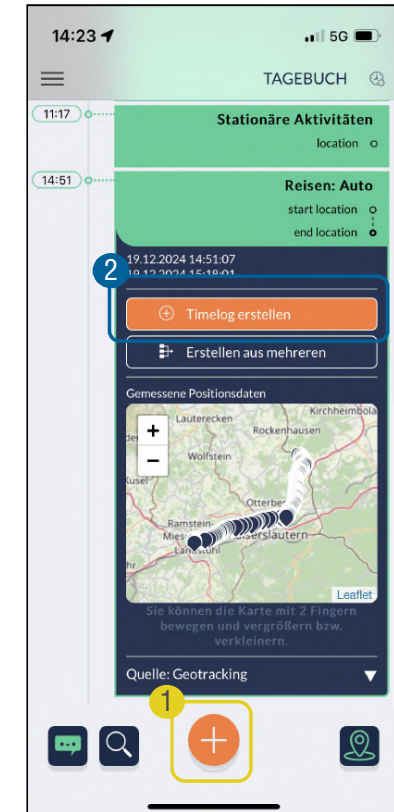
- » The difference between the diary itself (white (1)) and the suggestion list (green (2)) is not clear
 - » TP are confused because of different lists and expect activities to be added in the suggestion list
 - » They lose oversight of what their real diary is
- » The geolocation icon is recognized and understood correctly, but how to get back to the diary remains unclear
 - » We recommend a redesign for navigation between diary and suggestion list and also a redesign of the suggestion list



Usability – add activity

Wrong way of entering suggestions

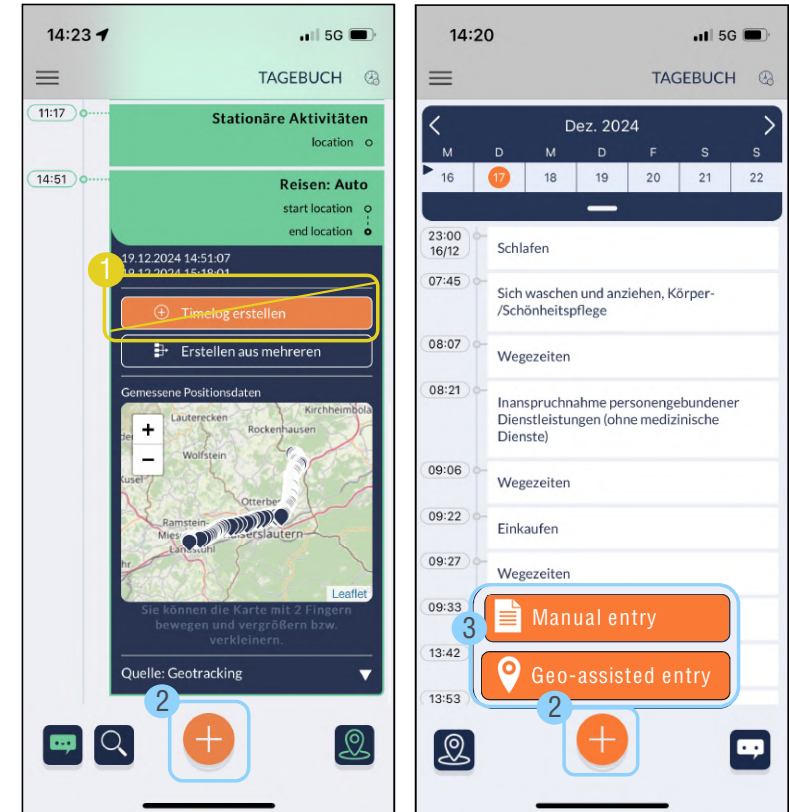
- » When asked to enter a new activity using the geo-assisted entry, TP use the big orange plus button (1) instead of “add timelog” (2)
- » This means they are doing a manual entry but believe they are using the geo-assisted entry, as they start in the suggestion list
- » “Add timelog” is not visible in the list, only in detailed view



Usability - redesign

Suggestions for a redesign (1/2)

- » We recommend to remove the “add timelog” function (1)
- » By tapping the orange plus button (2) users would be able to choose between manual entry and geo-assisted entry (3)



Usability - redesign

Suggestions for a redesign (2/2)

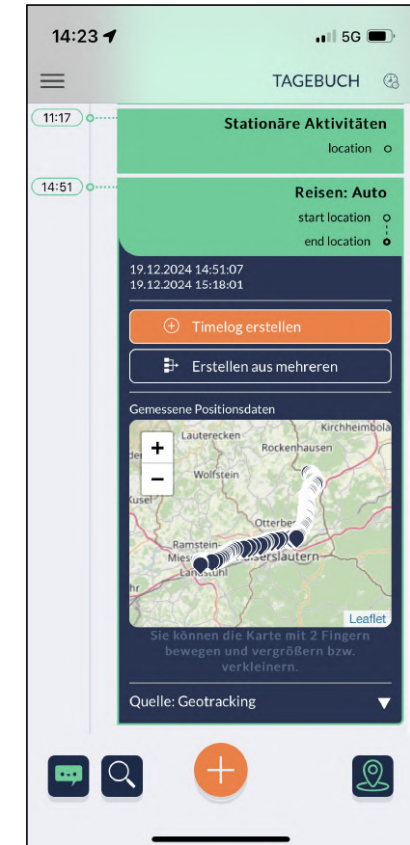
- » In the screen for geo-assisted suggestions the buttons to enter new activities are directly visible (not only in the detailed view) (1)
- » The orange plus button in the middle is removed. Entering an activity in the manual mode is not possible anymore from that screen
- » A <back> button is added to improve navigation and orientation (2)



Usability – detailed view

Same information given twice

- » Many TP **do not see an added value** in this detailed view, however for others the additional information makes the tracked geodata better understandable and helps finding the correct suggestions
- » The layout of the presented information is criticised (too small font size, size of map, ...)
- » Partly **irrelevant data** is shown (time shown twice, ...)



Usability – gaps and overlapping

No flexibility makes it hard to correct gaps

- » Each activity has to start and end exactly one minute after or before the next activity
- » Overlaps and gaps are not accepted
- » If there is a gap you cannot tap on it in the diary but have to fill it with a manual entry
- » We recommend
 - » to allow tolerances in the diary of up to 5 minutes, i.e. that no error message is shown if activities produce an overlap or a gap of 5 minutes or less in the diary.
 - » To make gaps tapable to be filled easier



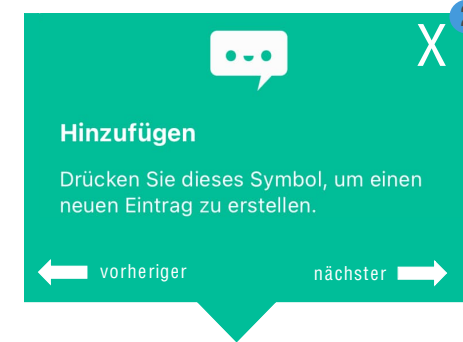
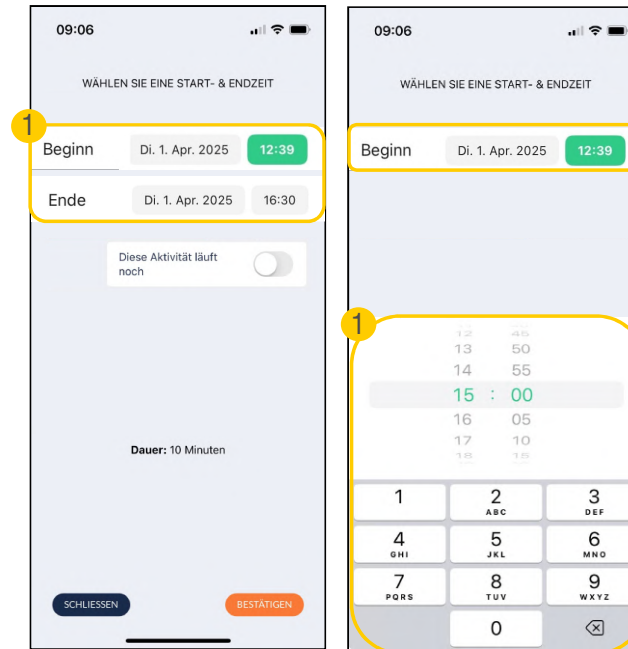
Trustworthiness

- » To raise trustworthiness it is important to name the sponsor, to provide the app with a name related to the sponsor and to inform about data usage and the purpose of tracking and the survey.
- » We recommend to point out that geotracking is only a supportive technology and that the tracked data itself is not of interest for the NSI, as many users suspected otherwise.
- » In general, users allow for geo tracking if they see a benefit in it. Therefore it would be good to provide some analysis and statistics to increase users' benefit.

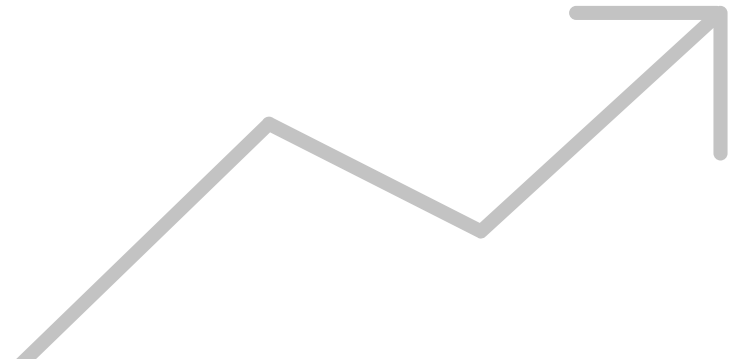
General usability

Some elements can be improved

- » The usability of selecting the screens to adjust start and end time (1) as well as choosing activities via the search bar or categories can be improved.
- » The assistant should have a visual option to be closed (2)
- » The function to deactivate geotracking should be found easier and be more prominent, e.g. in the burger menu

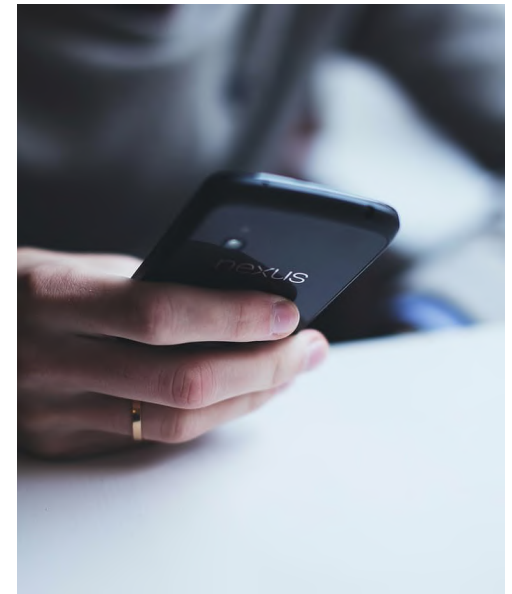


Background & objectives



Smart Survey Implementation (SSI) project

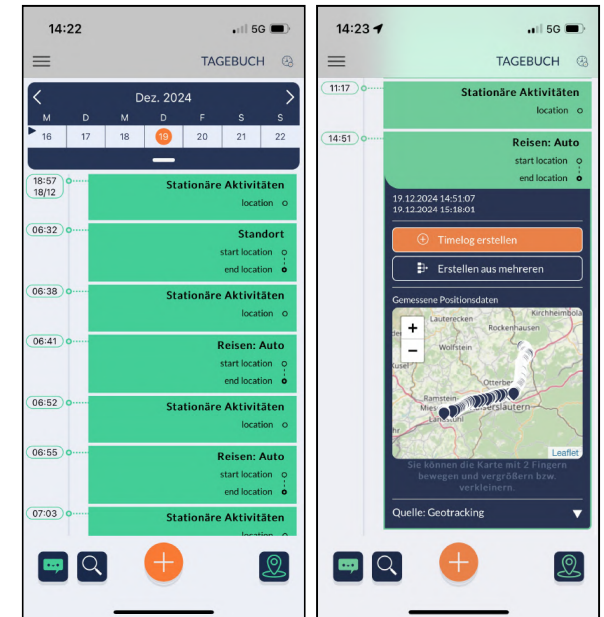
- » Was announced and is mainly financed by the European Union
- » Smart survey = surveys that combine smart sensor data with traditional survey questionnaires
- » Main goals (besides others) are to develop and test shared smart microservices for statistical production systems and optimize push-to-smart recruitment and motivation strategies
- » Its project consortium includes 11 project partners
- » Runs from May 2023 – April 2025
- » Is organised in 5 work packages



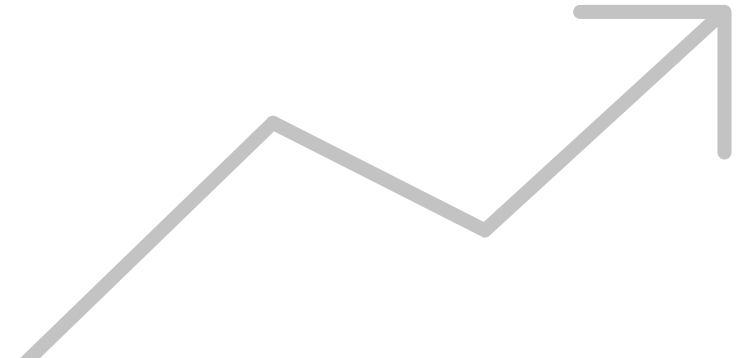
Symbolic image: <https://pixabay.com/de/photos/smartphone-technologie-handy-lg-1281632/>

Usability test within SSI

- » Goal of this SSI subproject is the development of a microservice which collects location data for the use in time-use diaries
- » This microservice should support people filling out the time-use diary by showing locations and times where they spent time and therefore generate more reliable data
- » It is planned to use this microservice for the time use survey
- » Main goal of this usability test was to test a first version of a new entry function, based on the geolocation microservice



Methodology & participants



Two approaches: focus group & interviews

Focus group with internal experts

- » Goal: discussing usability issues in detail
- » Participants: 5 experts for usability, internal from Destatis
- » Study structure:
 - 1) field phase (using the app in practice, at least two days),
 - 2) focus group (90 minutes, 3rd of December, carried out hybridly: inhouse in Wiesbaden and remote)
- » Time period: End of November – 3rd of December 2024

Interviews with potential users

- » Goal: collecting usability issues in practice
- » Participants: 14 potential users, external, 95 Euro incentive
- » Study structure:
 - 1) preliminary interview (20 minutes, carried out remotely),
 - 2) field phase (two days, directly after preliminary interview),
 - 3) interview about experience (45 minutes, carried out remotely, directly after field phase)
- » Time period: 9th – 20th of December 2024

Procedure of interviews with potential users



Overview of participants

Participant	Gender	Age	Education*
TP01	female	50	Low
TP02	female	35	Higher
TP03**	male	54	Low
TP04	male	19	Low
TP05	male	67	Higher
TP06	female	39	Low
TP07	female	58	Medium
TP08	female	47	Medium

TP01 – TP08 were not asked to put in diary entries at a specific point in time

Participant	Gender	Age	Education*
TP09	male	35	Higher
TP10	male	19	Medium
TP11	female	55	Higher
TP12	male	54	Medium
TP13	male	24	Medium
TP14	male	33	Higher
TP15	female	42	Low

TP09 – TP15 were asked to put in diary entries collectively in one badge for the first day of the test. On the second day they could choose freely.

*Classification of the educational level
Higher = University degree, higher vocational training
Medium = Graduation, vocational training
Lower = 9 or 10 years of school, no graduation

**App could not be tested in practice due to technical problems with the app; cancelled after initial interview

Distribution of acquisition criteria

Age distribution

as targeted

Age range	<26	26-49	50-65	>65
Number of participants	3	6	5 (4*)	1

Gender distribution

as targeted

Gender	Male	Female
Number of participants	8 (7*)	7

Education distribution

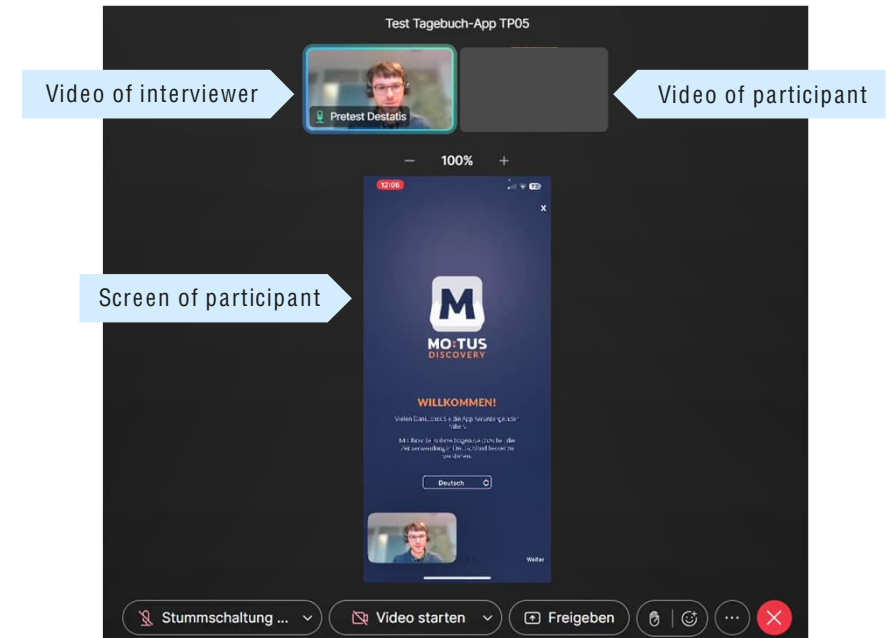
as targeted

Educational level	Low	Medium	Higher
Number of participants	5 (4*)	5	5

*Number of participants with the exclusion of participant TP03.

Test set up – first and second interview

- » Participant: takes part via smartphone from home, uses videocall software *WebEx* as app, shares screen while using and talking about the *Motus* app (see next slides)
- » Interviewer: participates via desktop from office
- » Technical Support: participates via desktop from office without video and sound, responsible to record interviews and to take notes

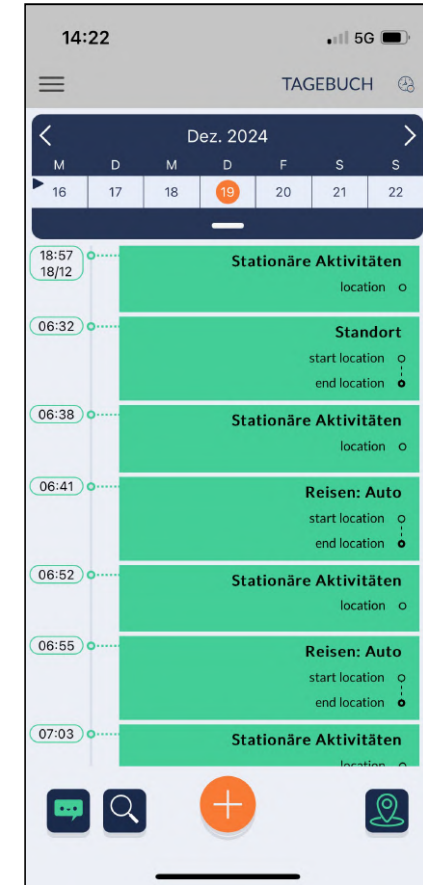


Screenshot of the Videocall, perspective of the technical support

Test set up – Motus app

Technical information regarding the app *Motus Discovery*

- » The newly developed microservice was integrated in the Motus Discovery app by hbits*
- » Still in development at the time of testing
- » Only tested on iOS (iPhone) due to technical issues
- » At the time of testing, location data was not always provided reliably (delayed in time and not perfectly accurate)
 - » For better results participants 9 to 15 were asked to fill out the diary at the end of the first day, when location data of the day was more likely to be completely available

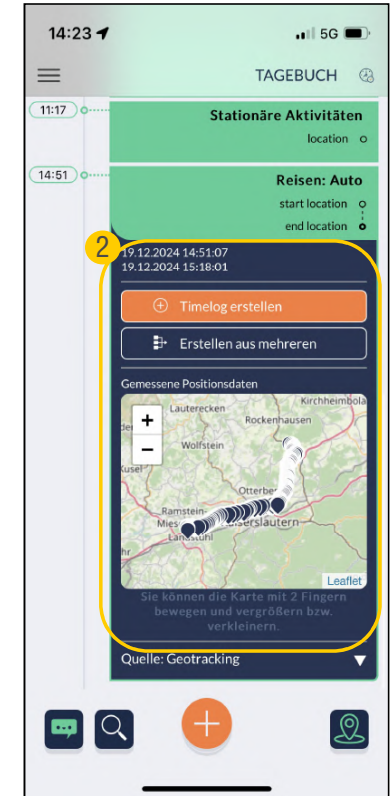


* hbits is one of 11 SSL project partners. They provided the app infrastructure (backoffice, frontend). They also brought the app with its new integrated microservice to the app store.

New microservice

The app can be used to enter activities of a day in a diary.
Two options are available to do so:

- » The manual entry (1)
 - » Manually enter activities
- » The geo-assisted entry (2)
 - » Based on tracked geolocations by the microservice, it shows the start and end time of whereabouts and movements as suggestions
 - » Location and movement data is available in addition as a thought support



Click path of the manual entry

1. Diary

2. Empty diary entry

3. Time selection

4. Empty diary entry

5. Main & secondary activity selection

6. Filled out diary entry

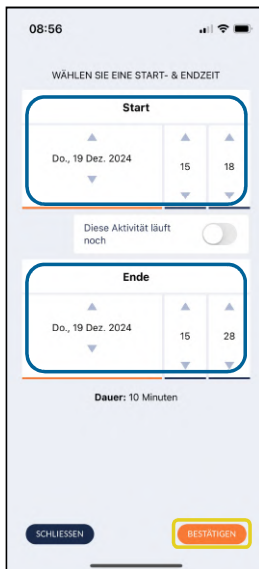
7. Diary



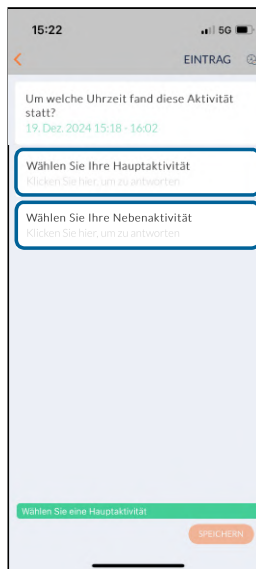
Tap on plus button



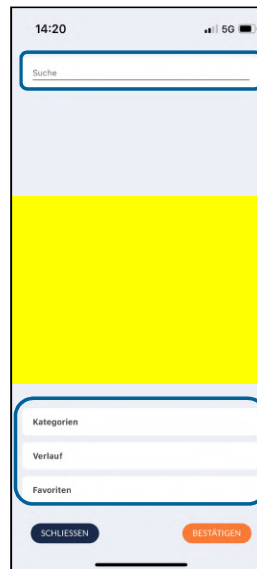
Tap on time



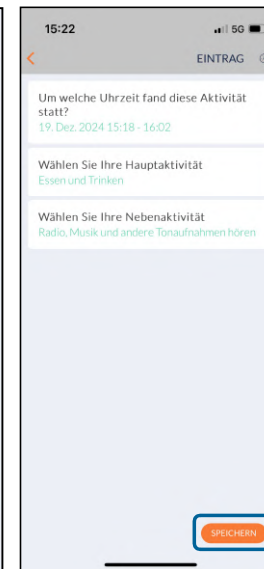
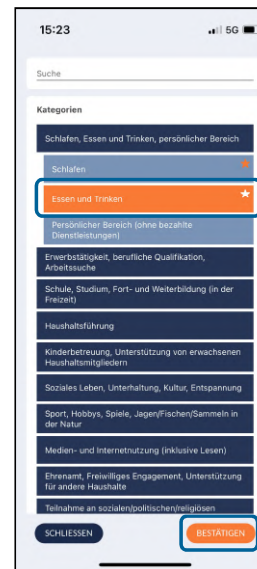
Adjust start and end time and confirm



Add main & secondary activity



Select main & secondary activity from category list & confirm



After saving the entry, it is shown in the diary



Click path of the geo-assisted entry

1. Diary

2. Geolocation suggestions

3. Detailed suggestion view

4. Diary entry showing times

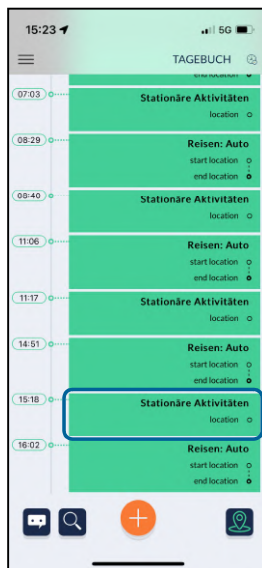
5. Main & secondary activity selection

6. Filled out diary entry

7. Geolocation suggestions



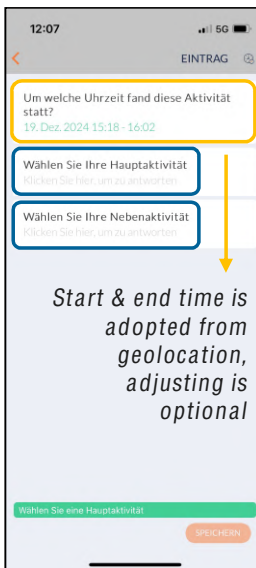
Tap on geolocation button



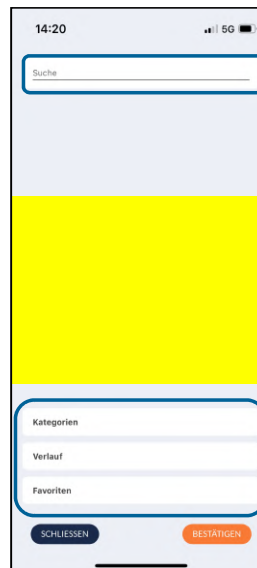
Open detailed suggestion view



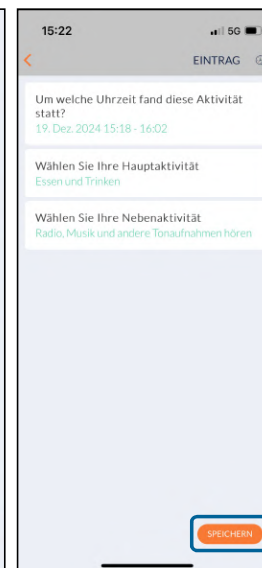
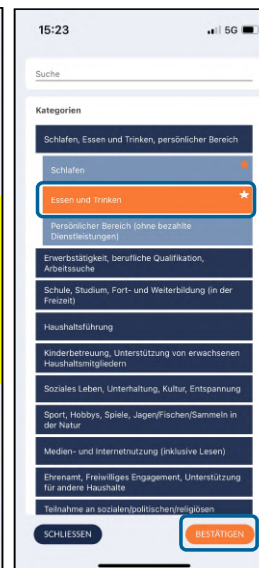
Tap on "create timelog"



Add main & secondary activity



Select main & secondary activity from category list

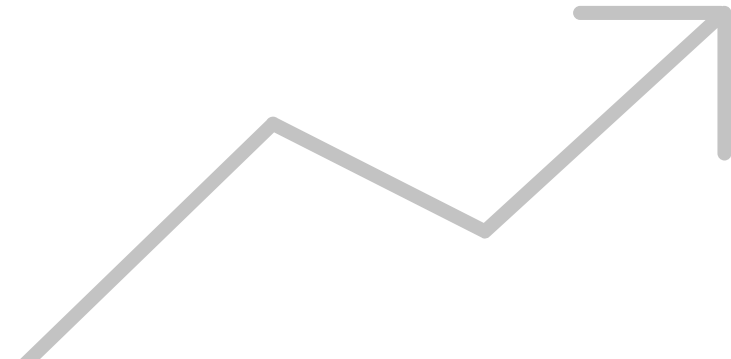


Safe filled out entry



Geolocation suggestions are displayed

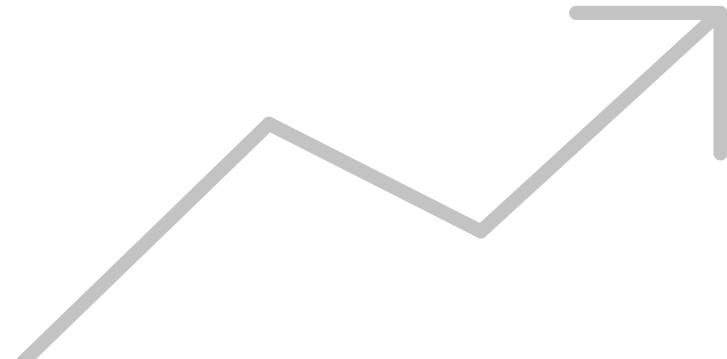
Results in detail



Reading and interpretation notes

- » Qualitative research considers small samples and therefore we do not provide results in the form of percentages. Instead we use the terms occasionally/few, sometimes/some and often/many to give an orientation about the frequency of observed behaviours and evaluations.
- » The findings presented come from the interviews with potential users. Learnings from the focus group are indicated with [FG].
- » Qualitative research does not only show results, it also asks for the “why” underlying a result and therefore gives a better understanding of the processes leading to a result.
- » Test persons (TP) may have been more motivated to use the app than the general population, as they were paid to participate. This might have had influence on the results.
- » Quotations are used for a more visual presentation of the results. The abbreviation of a test person (e.g. TP01) is used to indicate who made a statement.

First impressions



General impressions after use

(beginning of second interview, after two days of use)

- » Some TP mention that they overall had relatively **good experiences** with the app
- » Some TP need some time to **become familiar with the app** and how it works
- » Many TP raise **criticism on the geo-assisted entry**, as well as the procedure of **creating an activity**
 - » Both are described as complicated, counterintuitive and burdensome
 - » Many report problems while using the geo-assisted entry, as well as the manual entry
 - » At current state of development, no discussant would generally use the geo-assisted entry [FG]

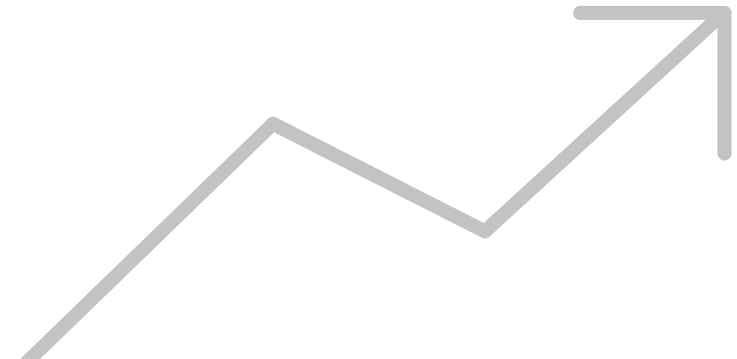
"My experience was generally quite good, it worked quite well [...], although things could be done differently at some points." (TP12, m54)

„I needed a short warm-up period.“ (TP10, m19)

"I had some prior experience from other apps, if I had done this for the first time, I would have had to deal with it for much longer." (TP06, f39)

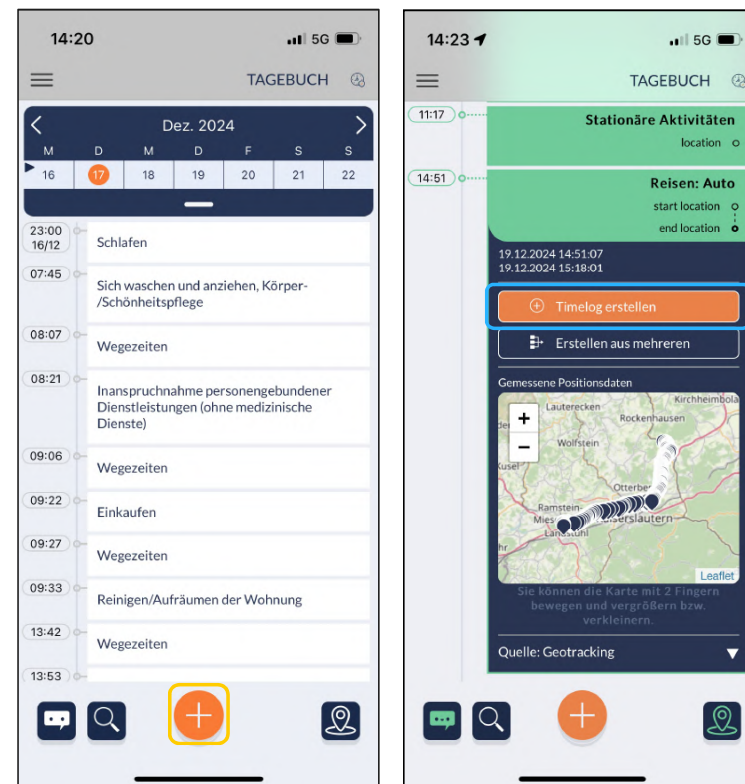
"Making entries from location data at the end of the day is quite a lot of fiddling" (TP09, m35)

User behaviour: entry creation



User behaviour: entry creation

- » When do participants actually create entries: continuously during the day or all at once in the evening?
- » Do participants prefer the manual input option or the geolocation suggestion to create entries?



User behaviour: when are entries created?

- » When given free choice when to create entries
 - » Most TP tend to create entries in the meantime during the day, mostly immediately after the respective activity happened
 - » Few TP create entries as both a bundle in the evening and during the day, how it suits them
 - » Only few TP create entries in the evening as a bundle
- » Reasons
 - » Few TP fear not being able to remember all activities correctly at the end of a day
 - » Some TP simply have time in between and want to get it done

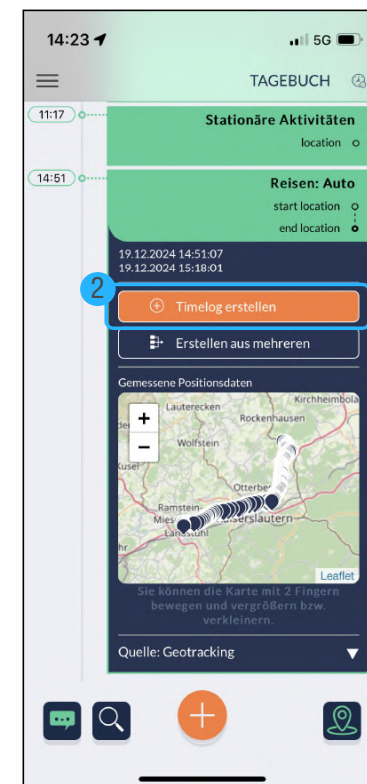
"I also added diary entries in the meantime, after work and then bundled again in the evening." (TP13, m24)

"Adding it in the evening... I would have had to lie, I wouldn't have been able to reconstruct events in such detail and it would have taken a long time." (TP08, f47)

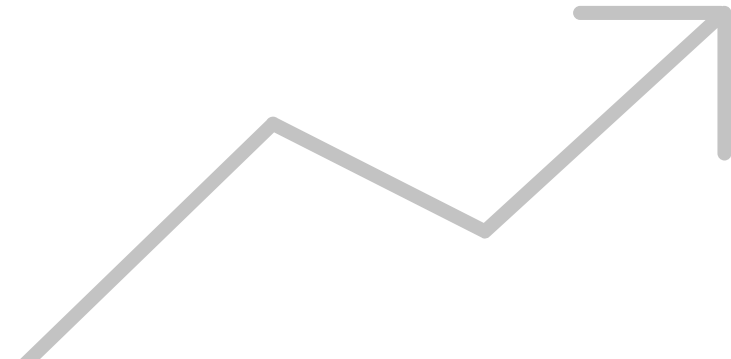
"Doing it during the day is like a challenge: I created the entry and then it was done." (TP04, m19)

User behaviour: preferred entry option

- » When given free choice how to add new activities to the diary ...
 - » Most TP mainly use the manual entry method (1)
 - » Few TP prefer the geo-assisted input (2)
 - » Few TP use both options and have no preference
 - » Discussants would prefer using the manual entry [FG]
- » [Reasons for the preference are presented in the following chapters]

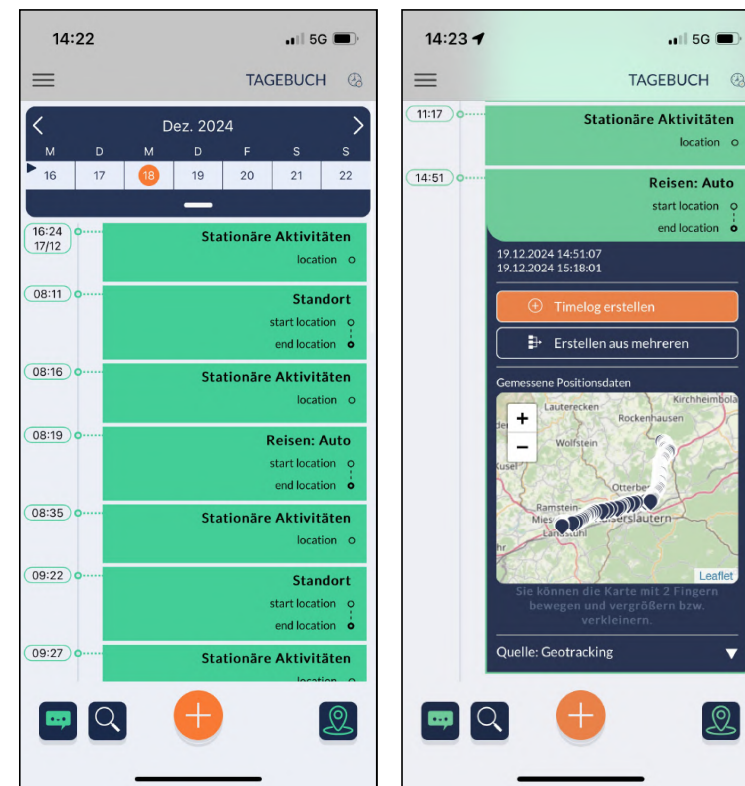


User assessment of the geo-assisted entry



User assessment of the geo-assisted entry

- » Is the geo-assisted entry considered useful?
- » Why and why not do TP consider the geo-assisted entry and suggestions useful?



User assessment of the geo-assisted entry

- » Many participants consider the geo-assisted entry suggestions as **rather unuseful, as they add activities directly and continuously during the day**. Therefore, the purpose of the **functionality does not become clear and they did not want to use it**.
- » Some state that there is no added value as they still knew where they just were and what they did
- » Few further criticise that the geolocation suggestions take too much time until they are available. There was no (complete) list available at the time of entry. Therefore, they could not use the function directly when they wanted to enter an activity.
- » It only has limited added value [FG]

"For me that is unnecessary, I do not have to see where I am, I know where I am right now." (TP08, f47)

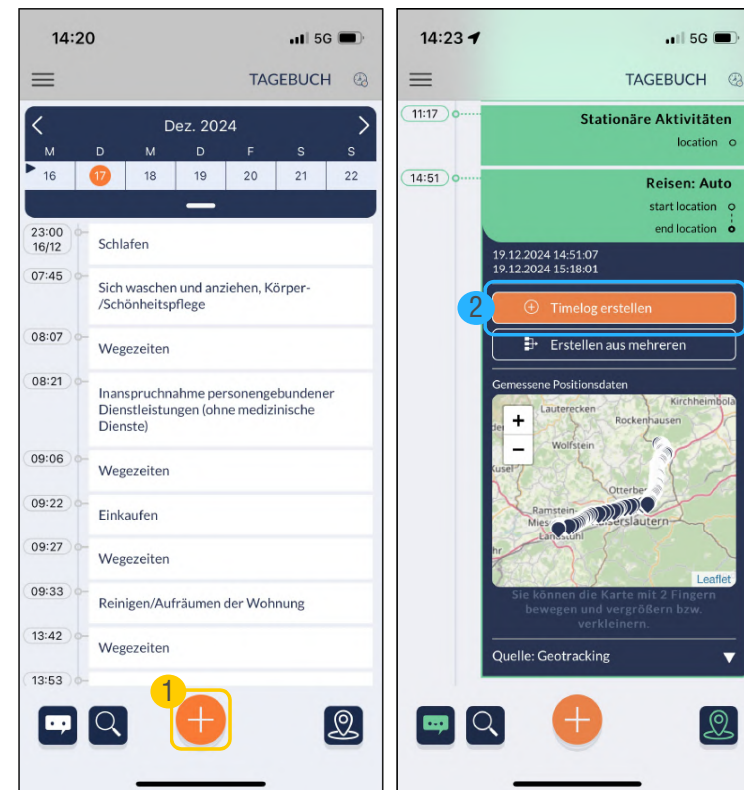
"Once the location has been recorded, I can enter the activity afterwards, I think it works better then. But I entered the activity beforehand, that's the problem, because I haven't even seen it [geolocation suggestion] yet." (TP05, m67)

User assessment of the geo-assisted entry

- » Many participants consider the geo-assisted entry as **rather useless**
 - » Many TP state the manual input (1) is easier and less complicated to use than the geo-assisted entry (2): tapping the orange plus needs less steps
 - » More support is expected for the actual creation of entries, instead of only showing information [FG]
e.g. automatically creating a transit time activity including a suggestion for the supposed mode of transport

"Thinking about what belongs together, where was I, is more complicated than manually entering from X to Y I did this or that." (TP15, f42)

"That is no added value for me. It should be simpler. Using the location was more complicated and confusing than creating a normal diary entry, [...]." (TP07, f58)



User assessment of the geo-assisted entry

- » When entering activities **collectively in one badge in the evening**, the **benefit of the geo-assisted entry becomes obvious**
 - » The full list of geolocations is available in the evening (delays do not play an important role) and gives an orientation and thought support while adding activities
 - » Accordingly, most TP who were supposed to fill in all activities in the evening find it to be helpful, **simplifying recalling information** regarding activities
 - The geo-assisted entry serves as a **thought support**, to remember activities [FG]
 - » If the suggested times are correct, they are rated as useful
 - » Few TP use the detailed view to identify travelling times and location changes

"I definitely prefer using the locations. Otherwise I wouldn't have been able to recap the day in the evening." (TP14, m33)

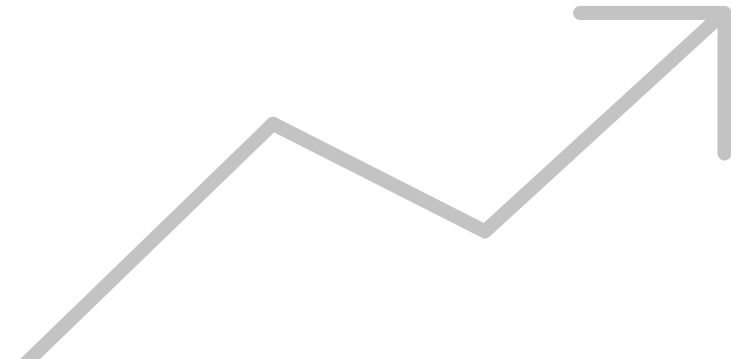
"I didn't fill in until the evening, which was a bit more difficult, but I also looked at the movement data, which was helpful." (TP12, m54)

"I used the movement geolocation data as the 'main component', then created the stationary activities around it if they matched. I sometimes also combined suggestions." (TP09, m35)

User behaviour and assessment

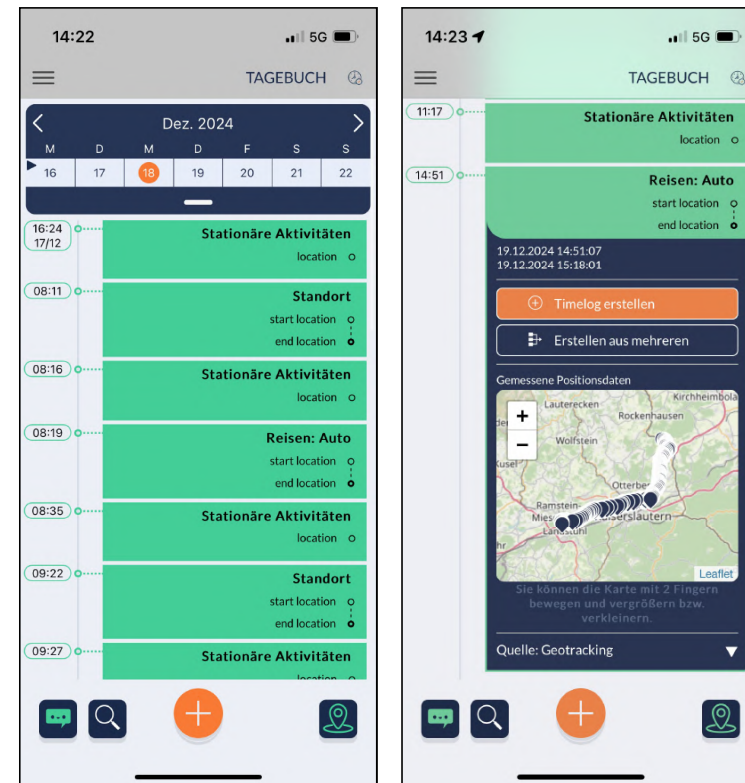
- » The geo-assisted entry is useful if...
 - » Suggestions, created on basis of geo-tracked data, are directly available
 - at the moment of testing delays occurred
 - » Users do **not add activities to their diary directly after the end of the activity**, as the geo-assisted entry then provides a reminder effect. This is especially true if locations have been changed.
- » As TP preferred adding entries directly during the day, the benefit of the geo-assisted entry is much likely to increase with ongoing development of the microservice, when geo-tracking data should be available directly without delays

Usability of the geo-assisted entry



Geo-assisted entry – usability

- » How do TP assess the user interface of the geo-assisted entry?
- » Which usability issues occur while using the geo-assisted entry?

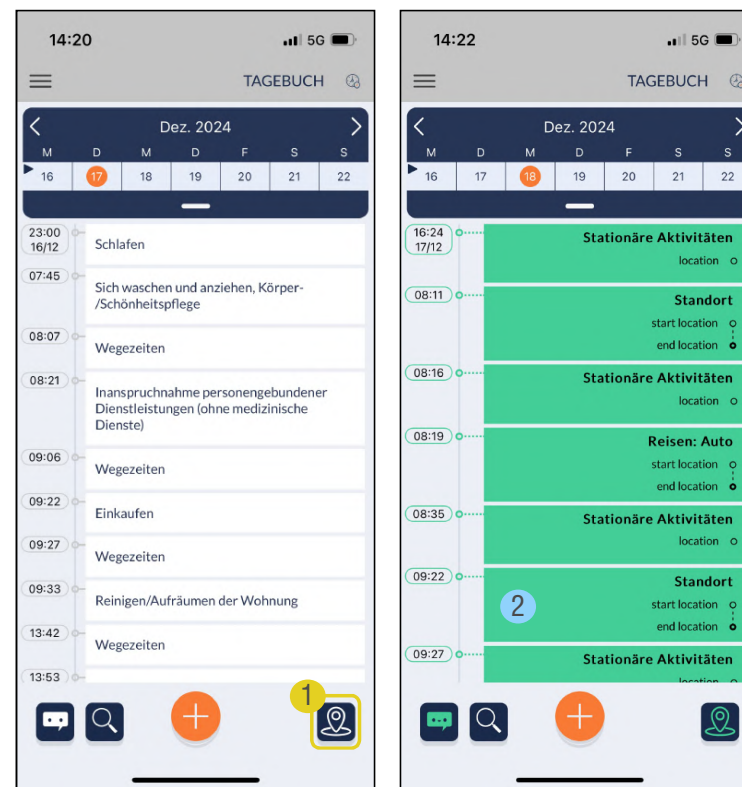


Geo-assisted entry – usability

Suggestion list - positive remarks

- » All TP find the icon "location" quickly when using the app for the first time (1)
- » Colours invite to tap the suggestion (2)

"The symbol on the bottom right - that reminds me of 'location'." (TP12, m54)

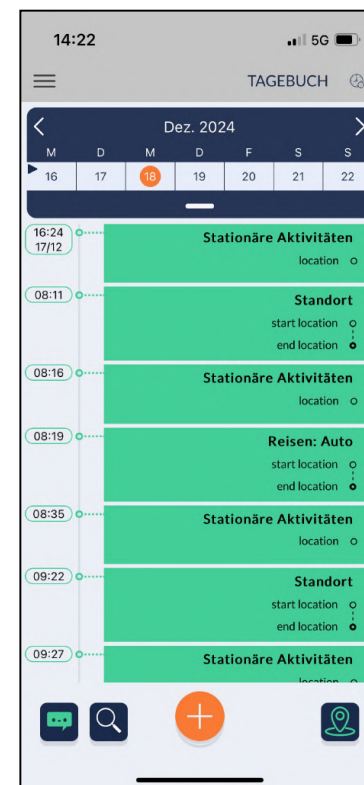


Geo-assisted entry – usability

Suggestion list - negative remarks (1/2)

- » The benefit of the displayed geolocations is not immediately clear to everyone
 - » For some TP it seems to have little to do with the diary itself
- » For some the list is confusing & cluttered
 - » The many suggestions are overwhelming for some TP (in case of many actual location changes or when several travel times were identified for only one route [FG])
 - » Unclear what is behind the entries: suggestions have to be opened to see what they refer to [FG]
- » Difference between 'location' and 'stationary activity' remains unclear

"With the green one [geolocation suggestions] I always have to jump in and see what is behind it." (TP05, m67)

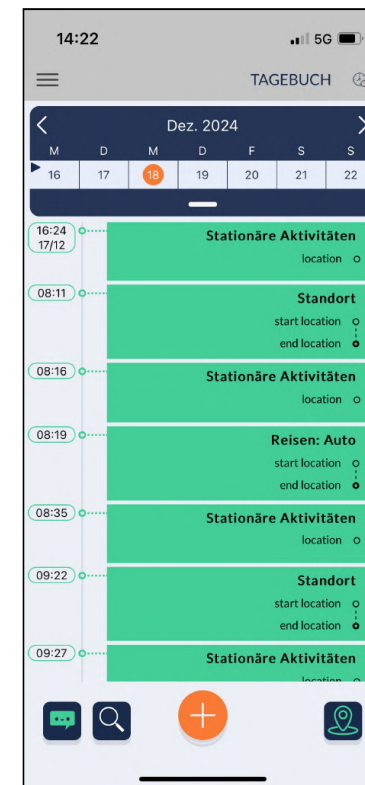


Geo-assisted entry – usability

Suggestion list - negative remarks (2/2)

- » Size of entries (tiles) does not correspond to the time length the entries actually represent
- » Missing visual distinction between the two types of geolocation data (stationary & route) [FG]
- » Unspecific information like “start/end location” has no added value, a meaningful term (e.g. “railway station”) was wished for [FG]
- » The times displayed are not always correct

“I was confused about what to do with this list because it didn't fit my entries at all.” (TP08, f47)



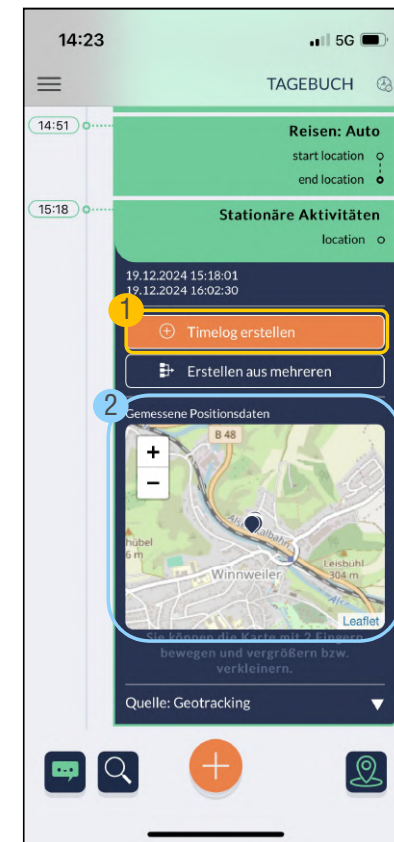
Usability – suggestion list

- » Providing more information derived from the tracking data and preallocating this information in the suggestions for diary entries (e.g. start location, end location, mode of travel) is likely to increase the benefit
- » A redesign with a more prominent button to add an activity from the suggestion list reduces the number of taps and makes the purpose of the list more obvious.
 - » See design suggestion on slides 56/57
- » Rename suggestions from „Standort“ [location] to „Reise“ [travel] for better differentiation to „Stationäre Aktivität“ [stationary activity]
- » Make all suggestions in the list have the same size (as small as possible). We don't recommend to adapt the size according to the duration of an activity (as suggested) as this would lead to a lot of unnecessary scrolling and would reduce the overview.

Geo-assisted entry – usability

Detailed view – positive remarks

- » The orange button “create timelog” catches the eye (1)
- » The map is useful for identifying what the entry refers to (2)
- » The detailed view makes the tracked geodata better understandable, as additional information is shown [FG]



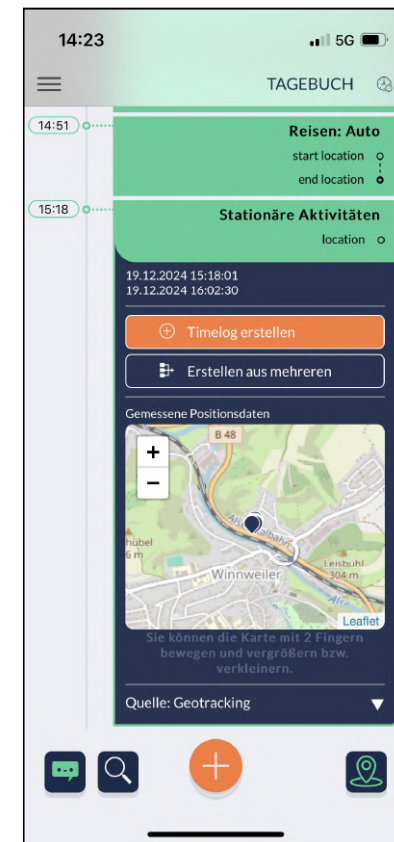
Geo-assisted entry – usability

Detailed view – negative remarks

- » Many TP do not see an added value in this view
- » Criticism about the layout:
 - » TP criticise the high number of pins on the map (for routes)
 - » Font size too small [FG]
 - » Map described as small, making its use fiddly [FG]
 - » TP criticise that the path is difficult to see
- » Partly irrelevant data is shown:
 - » e.g. date, time frame (already shown in the left column), data source, map origin [FG]
 - » The speed information is not required by TP

"Interesting, but the question is, what do I do with it?" (TP05, m67)

"You can click on the pins, but the information content is rather low" (TP14, m33)



Usability – detailed view

- » Leave out unnecessary information in detailed view and information which is already provided elsewhere
 - » Date, timeframe, speed
- » When detailed view is opened, the suggestion should automatically scroll to the top of the screen
 - » Start and end time of activity is visible
- » Tapping on a suggestion (green) should also open the detailed view, not only tapping on “mehr” [more]
 - » The “mehr” button is only to emphasise that there is more data available
- » Reduce number of pins in the map, use a line



Geo-assisted entry – usability

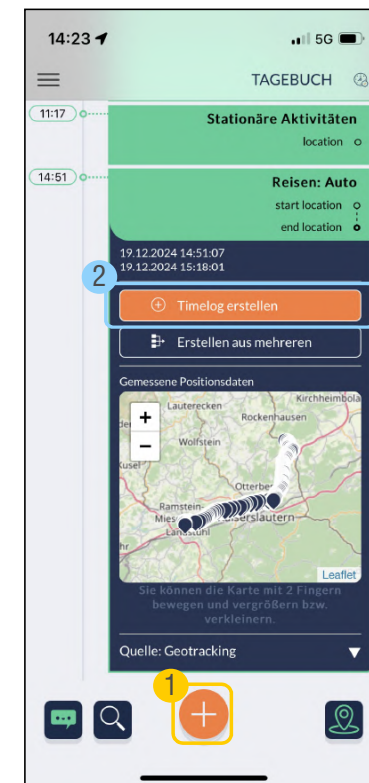
„Create timelog“ button

- » Some TP **do not understand the difference** between the two entry options:
 - » When they have opened the suggestion list, they use the **orange plus symbol** ⁽¹⁾ **instead of „create timelog“** ⁽²⁾ to create a new entry,
 - » They are then doing a manual entry but believe to do a geo-assisted entry or they think that there is a technical issue.

“This button [‘create timelog’] then creates a manual entry, although I have the location function opened, so it basically does the same thing. When you create an entry from the location data, it only provides the time.” (TP15, f42)

“I sometimes also used the location data to create entries, but I didn't realise what that was supposed to do, because if you go back to it afterwards [...], it still says ‘create activity’.” (TP06, f39)

“I always pressed the location button at the bottom right, but it seemed not to work properly, then I also pressed the plus button, and then it jumps back into the input mask.” (TP05, m67)



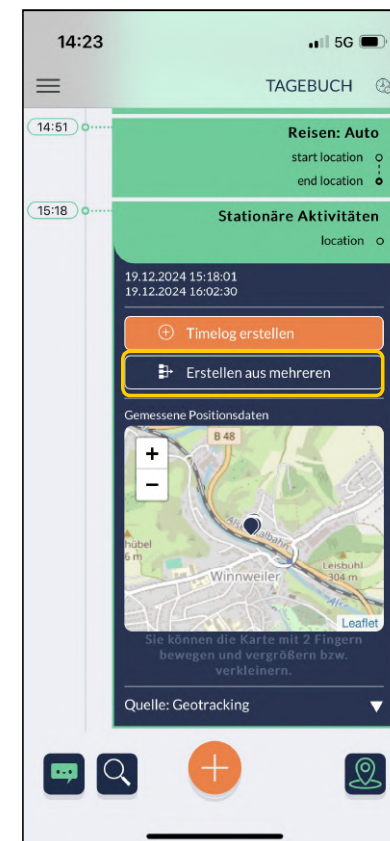
Geo-assisted entry – usability

Detailed view - „create from several entries“

Background: The function is used to combine different geolocation entries into one diary entry, e.g. in case a route has been split into several entries by the system

- » Only few TP understand the function by themselves, use it and describe it as useful
- » Some TP rate the function as useful after it was explained to them

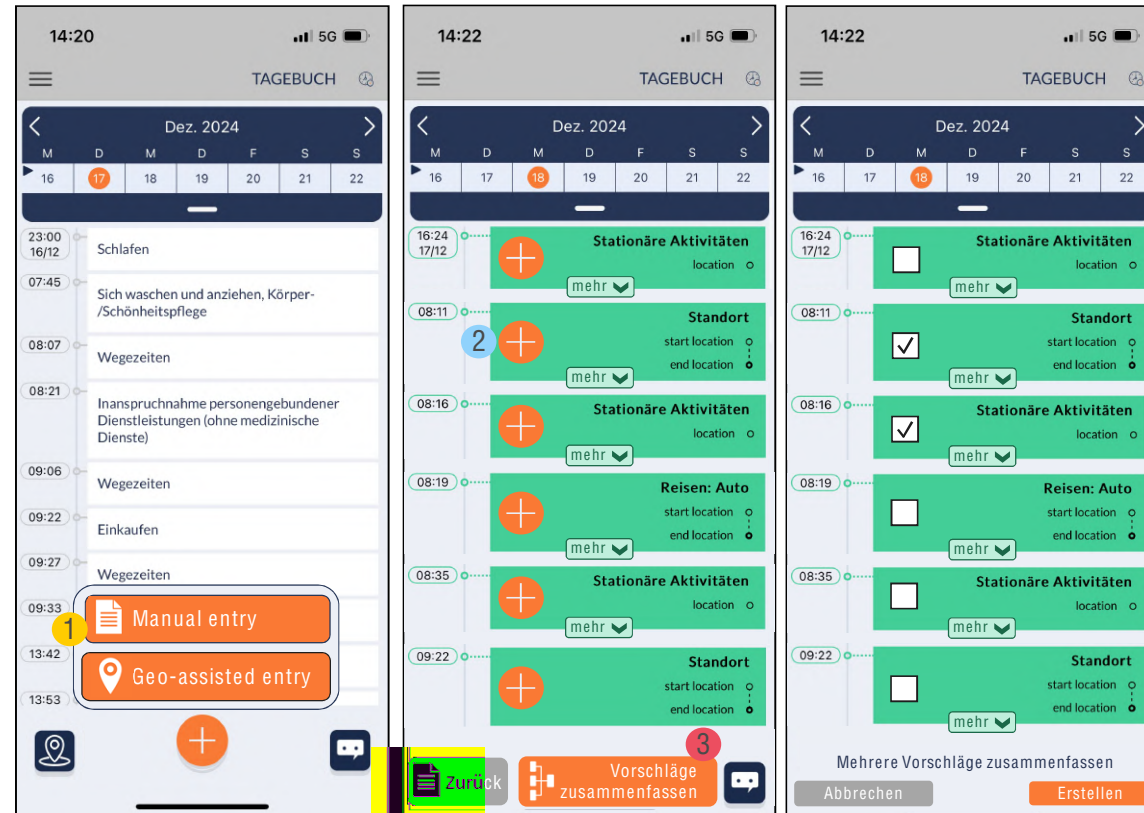
"It is unclear to me what is supposed to be summarised here." (TP05, m67)



Usability – entry of geo-assisted suggestions (1/2)

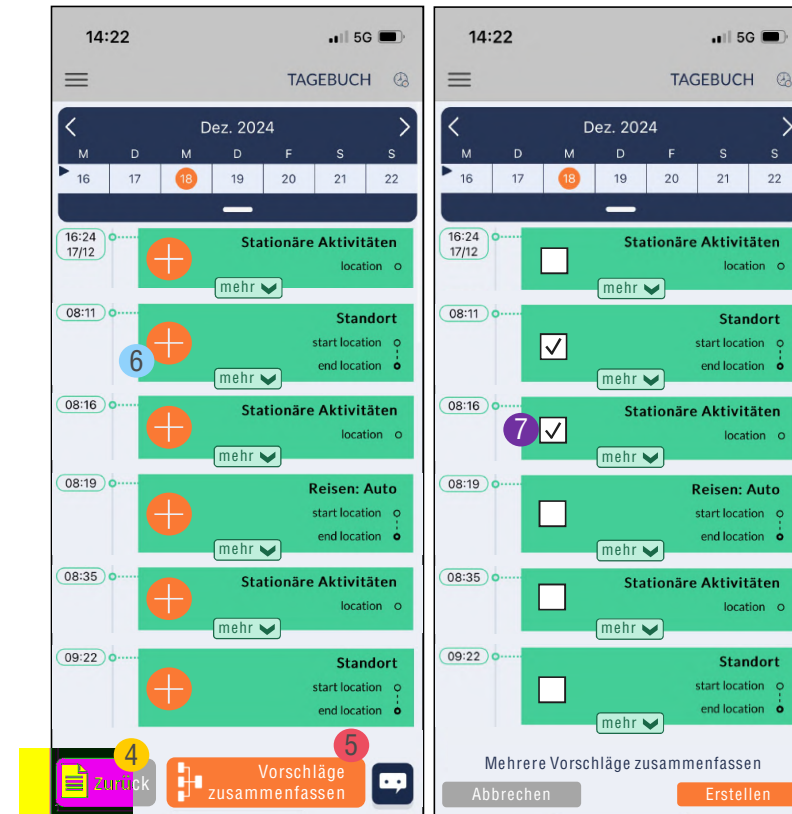
We recommend a redesign which reduces the number of taps and makes the purpose of the suggestion list more obvious

- » More direct way to enter a suggestion, more emphasis on the function of adding an activity to the diary
- » By tapping the plus, two options appear: manual and geo-assisted entry (1)
- » Navigation via plus button & following menu makes way for adding entries clear
- » Buttons to add a suggestion are placed more prominently on first level (2)
- » Buttons to create activity from several suggestions also more prominent (3)



Usability – entry of geo-assisted suggestions (2/2)

- » Use 'back' button to make navigation more clear (4)
 - » Instead of the former geolocation icon changing colours
- » Label the button 'adding from several suggestions' for better understanding (5)
 - » Delete orange plus (which was at that location previously), so it cannot be mistaken for the geo-assisted entry (5)
- » Always use the same orange plus symbol to add entries, including when adding via geo-assisted entry (6)
- » Checkboxes to create entries from several suggestions, moved to the same position as the plus symbol (7)
 - » Using the plus and checkboxes at the same time would be too confusing and crowded

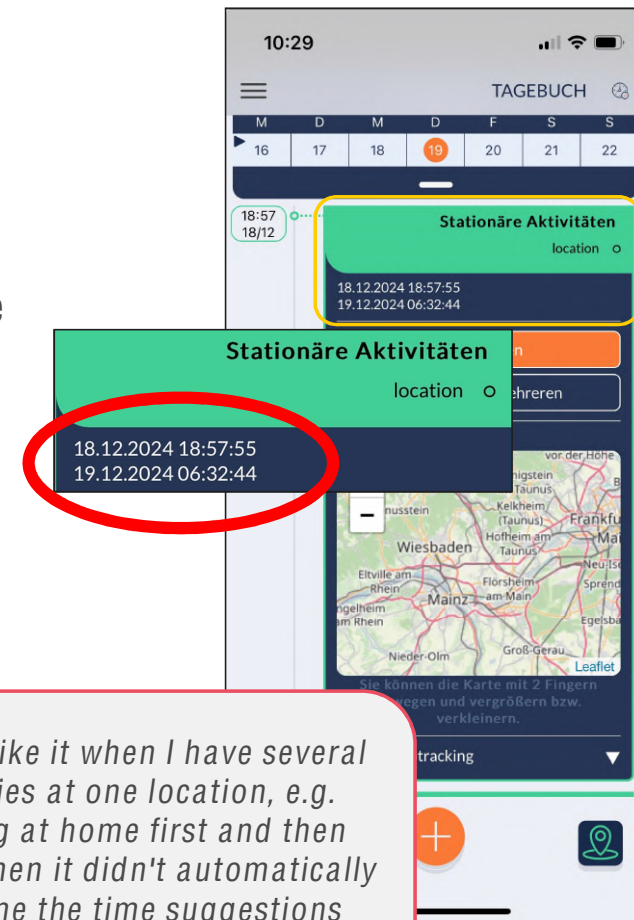


Geo-assisted entry – usability

Creating several activities from one geolocation

- » Creating several diary entries (activities) on the basis of one geolocation entry is confusing to some
 - » Several different activities may happen at the same location, while the corresponding geolocation suggestion does only show the start and end time of the entire time period
 - » User have to open the same geolocation suggestion several times and tap on „create entry“
 - » The times for the actual activities have to be recalled

→ In these cases it is easier to use the manual entry



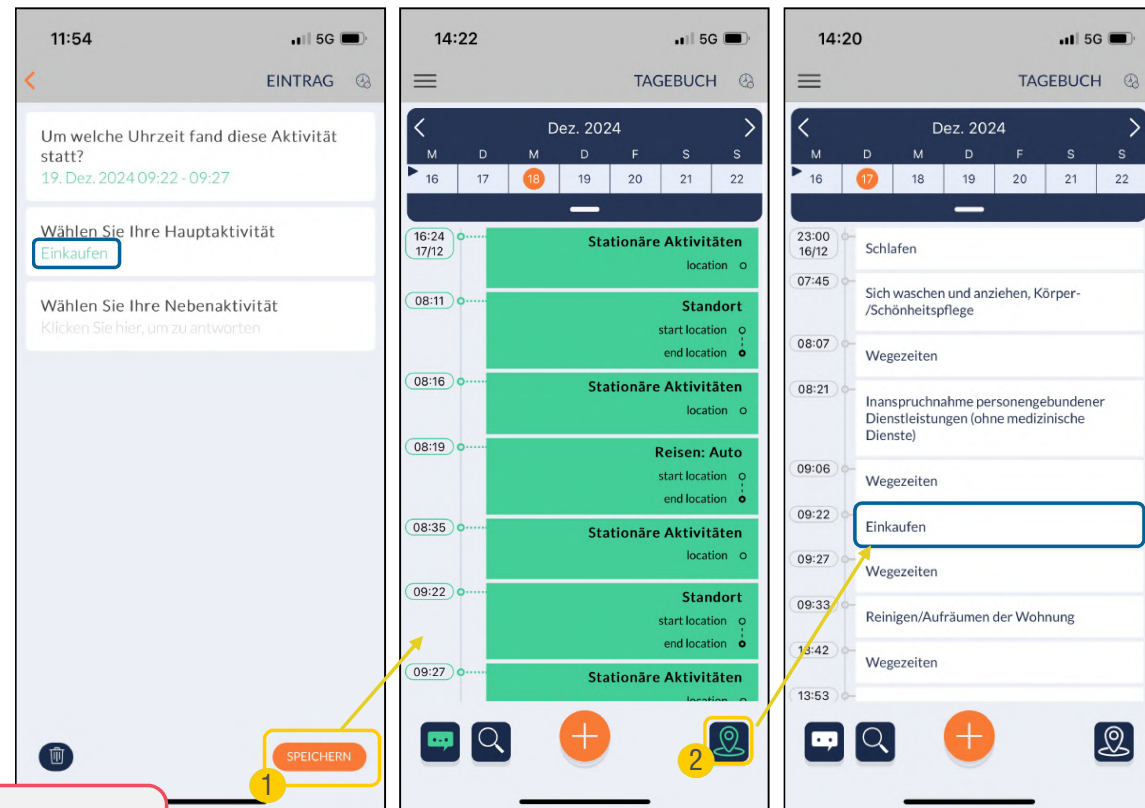
"Creating multiple entries for a location is cumbersome, as the time [pre-set time in input mask] always jumps back to the start time of the geolocation." (TP10, m19)

"I don't like it when I have several activities at one location, e.g. cooking at home first and then eating. Then it didn't automatically show me the time suggestions consecutively." (TP12, m54)

Geo-assisted entry – usability

Overview of location suggestions

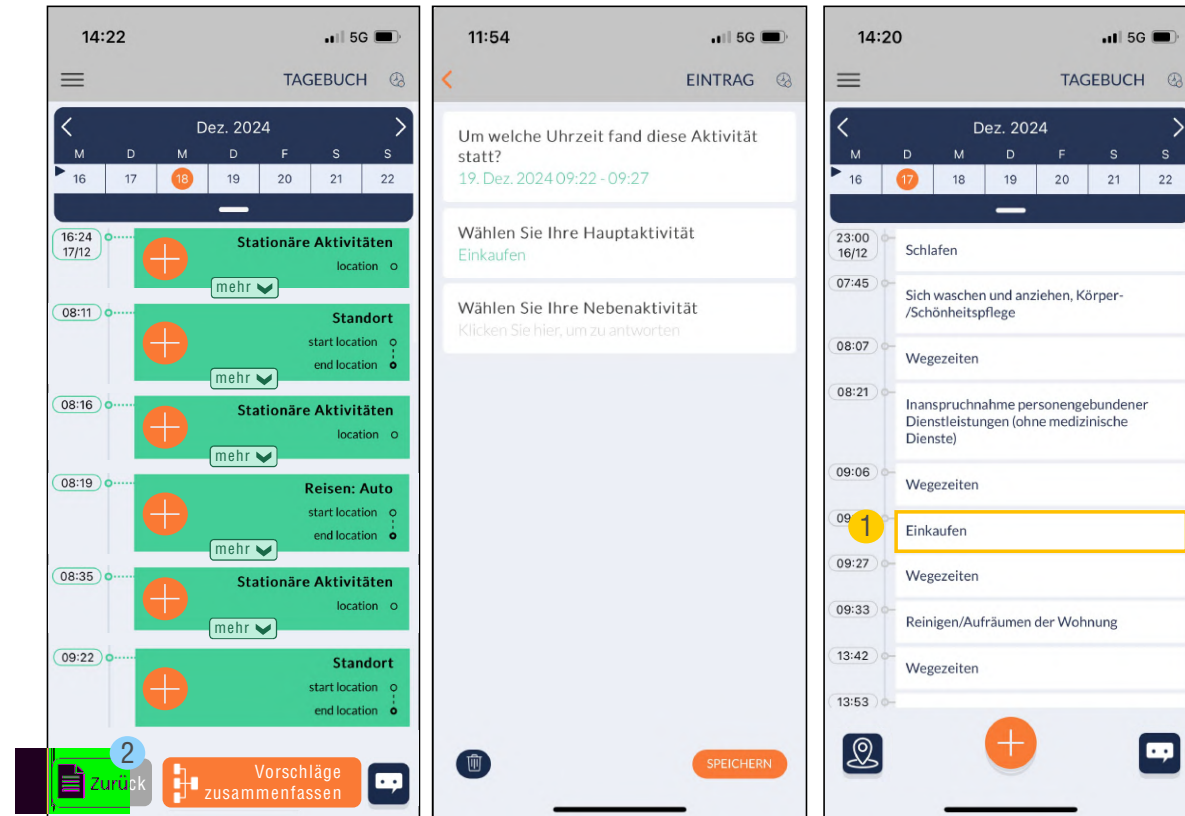
- » Many TP cannot find their way back to the diary after creating a diary entry via the geo-assisted entry
 - » After adding an entry via the geo-assisted entry (1), the geolocation suggestions are shown again
 - » Many are confused where to find the diary entry they just entered. It takes another tap on the geolocation symbol (2).
 - » TP are searching in the burger menu, tap on the diary icon at the top right, open the assistant or tap on the calendar



"Where can I see what I have entered?" (TP01, f50)

Usability entry flow

- » After adding an activity to the diary – no matter if a manual entry or geo-assisted entry is used – the diary should be presented
- » If possible, the new entry should be highlighted for a few seconds, to show users their new entry (1)
- » By adding a „back“ [zurück] button in the bottom left corner, navigation should be easier and better understandable (2)
- » All functionalities should be explained by the assistant



Geo-assisted entry – usability

Time indication and gaps

» Time indication:

- » TP criticise that reporting on minute level is too exact, too much of a burden (also in [FG])
- » TP generally criticise that a new activity must always start exactly one minute after the previous one. Time overlaps are not accepted, leading to error messages
- » Simultaneously, a time gap is not accepted, even if it is only about one minute: an indication that the gap should be filled appears in the diary

» Filling a gap:

- » Gap cannot be tapped, manually filling the gap is complicated
- » Must be specified to the minute (see above) by extra manual entry



Gaps and overlapping

Allow tolerances in diary of up to 5 minutes

- » If two activities have an overlap of less than 5 minutes, accept it without error message
 - » E.g. activity “working” ends at 16:50, activity “driving home” starts at 16:49
 - » If it is 5 minutes or more, show the warning message, whether the first entered activity should really be overwritten (see slide 82)
- » If there is a gap in the diary of less than 5 minutes, do not show it
 - » E.g. activity “working” ends at 16:45, activity “driving home” starts at 16:49
 - » If there is a gap of 5 minutes or more, show it in the diary with an error message (as it is) and make it tapable
 - Tapping on the gap should open the manual entry display with pre filled start and end time, according to the gap

Trust & privacy

- » Do TP consider the app trustworthy in general?
- » Which factors influence the trustworthiness of an app?
- » Are there privacy concerns regarding the tracking of the location?
- » What do TP think is the purpose of using the tracking?

Do TP trust the app? Why?

- » Most TP trust the app in general
- » Context is important for some TP to trust the app
 - » Sponsor of the app (Destatis)
 - » Purpose of the app and use of data
 - » Information about the survey
- » Some TP have no doubts at all
 - » Tech companies have data anyway
 - » Their data is not of interest for others
 - » There cannot be any negative consequences other than advertising
- » Few TP would not install such an app (they are doing it only for the test)



„Google already knows everything anyway... So I don't care.“ (TP14, m33)

Factors that increase trustworthiness

- » Sponsor: TP trust Destatis as sponsor
- » Name: the name of the app should have a connection to the sponsor, that would raise trust
 - » This was not the case for the Motus app
- » Information about data usage in the beginning
- » Design
 - » More fluent design
 - » Possibility to name activities on one's own and of one's own

Factors that increase trustworthiness

- » Not showing private locations like the home or work place in such high detail. An approximate location (circle with larger radius around the actual location) would be sufficient and ensure privacy [FG]
- » PIN: to access the diary, but not to log in every time
- » Performance:
 - » Less crashes and errors
 - » The better the performance of functionalities, the higher the trust

Situations in which TP do not want to be tracked

- » Some TP are ok with tracking all the time, otherwise they would not download the app
- » Some TP only do not want to be tracked in situations which could have huge negative consequences
 - » Doing something illegal
 - » Cheating on one's partner
- » TP know that they can deactivate geolocation tracking
 - » Way to deactivate is long and complicated

Assumed reason for location tracking

(as asked in the first interview)

- » Many assume that the NSI gets the data in one way or another
 - » Some presume that the app (NSI) connects locations with logged activities
 - » Many mention the movement behaviour being tracked
 - » Some believe the logged activities are being controlled
- » Few assume the app automatically creates entries based on the geolocation
- » Few have no exact presumption at all

"For me, the location doesn't matter, what I've done is more important. But you probably need the information, otherwise you wouldn't be recording it." (TP08, f47)

„Entries might also be checked. If I enter that I am doing household chores and the location data shows that I am in the car travelling to Hamburg, then something is wrong." (TP12, m54)

„If I enter a 3h travel time, the app wants to know whether I was driving through the city or to Hanover." (TP14, m33)

"Aha, I don't need to write anything myself. When I'm at the gym then it [the app] knows that I'm doing sports." (TP01, f50)

When do TP allow geo tracking in general?

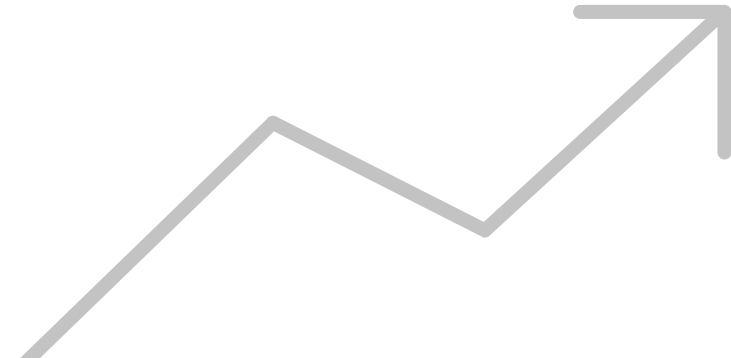
- » Wide range from „always“ to very restricted
- » Most need a sense behind it and an advantage, like...
 - » Navigation (maps)
 - » WhatsApp (to share location)
 - » To order a taxi/ uber
 - » Public transport
 - » Payback and supermarket app (to see and get the correct discounts)

"It gives you [Destatis] added value, because you know, that I always go shopping at 'Edeka' in the nearby village, for example." (TP13, m24)

Trust & privacy

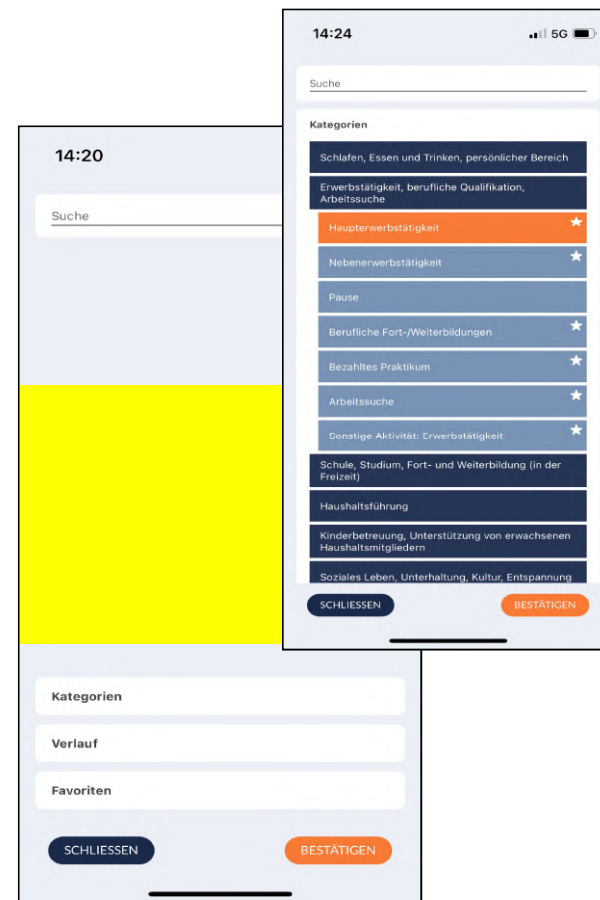
- » To raise trustworthiness ...
 - » Clearly name the sponsor of the survey
 - » Give the app a name which is related to the sponsor
 - » Provide information on data usage and the purpose of tracking as well as the purpose of the survey for those who are interested
 - Point out that geotracking is only a supportive technology and that the tracked data itself is not of interest for the NSI, only the diary entries
 - » Create an easy way to deactivate geo tracking, e.g. in the burger menu
- » Add additional analysis of data to create a benefit for users
 - » Hours and kilometres spent traveling with which mode of travel
 - E.g. by bike: 3 hours, 45 kilometres
 - » Time spent for different main categories, e.g. work, household, child care, ...

Selecting activities



Selecting activities

- » How do test persons select activities when entering a diary entry?
 - » Do they prefer the search function, category list, history or favourites?
- » What are issues when selecting activities?

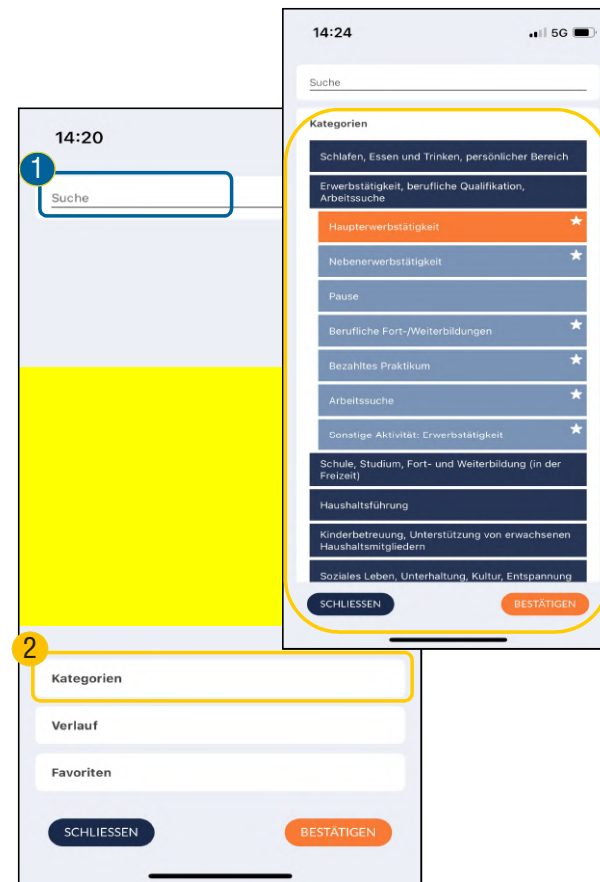


Selecting activities

- » Search function and selection via categories are used equally
 - » Some primarily choose activities via the category list
 - » Some primarily use the search function
- » Some TP overlook the search bar

"I had an appointment at the town hall, I did not know what to enter as a search term. But I then found 'visiting authorities' in the categories." (TP13, m24)

"I primarily used the search function because it was faster and always gave a matching result." (TP09, m35)



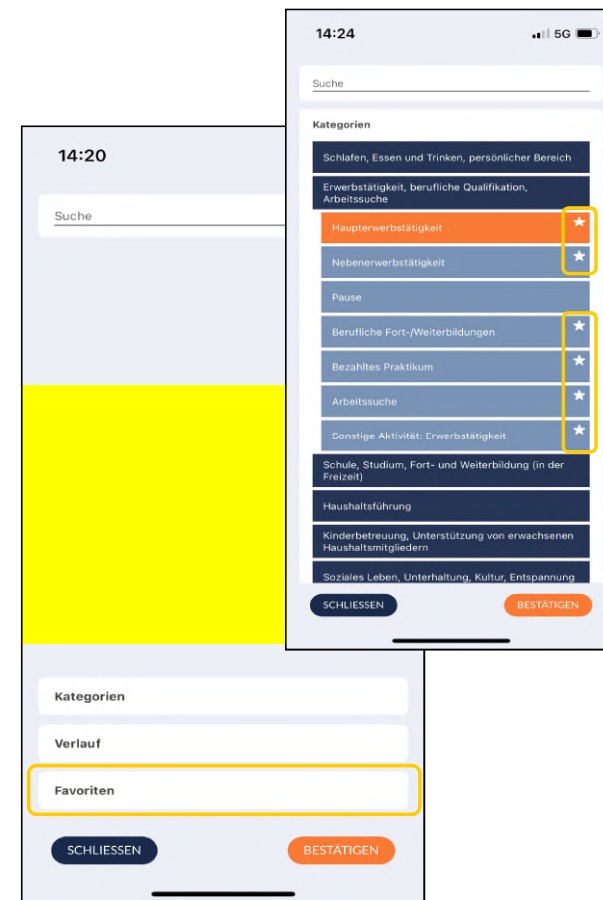
Selecting activities

„Favourites-function“

- » Most participants do not use the option to add categories as favourites
 - » Added value stays unclear for few
 - » Function (button) not perceived by few
 - » Few assume that often used categories are automatically added to favourites

"I also used favourites and added everything that occurs frequently: food, drink, setting the table, etc." (TP13, m24)

"I thought regularly used categories are marked with a star. I did not get that you have to tap on the white stars. Then they turn orange and it is a favorite." (TP09, m35)



Selecting activities

Further difficulties with finding correct categories

- » For some TP the categories are a bit confusing and they need time to orientate themselves
- » Fitting categories are hard to find in some cases
 - » Watching videos on YouTube (searched for category under „media usage“)
 - » Listening to podcasts
 - » Conversation with others (non-household members)
 - » Sometimes, similar sounding activities that fit less have to be chosen. That casts doubt on the added value of your own answer [FG]

“That list is a bit overwhelming, you have to get used to it.” (TP13, m24)



Selecting activities

Further difficulties – secondary activity and free text

- » Many TP ask what is meant by “secondary activity”
 - » Some do not understand why the same activities are available here
- » Free text entry
 - » TP expect to be able to enter a text that then appears in the diary, which is not the case.
 - » Some think it does not work.

11:54 5G

EINTRAG

Um welche Uhrzeit fand diese Aktivität statt?
19. Dez. 2024 09:22 - 09:27

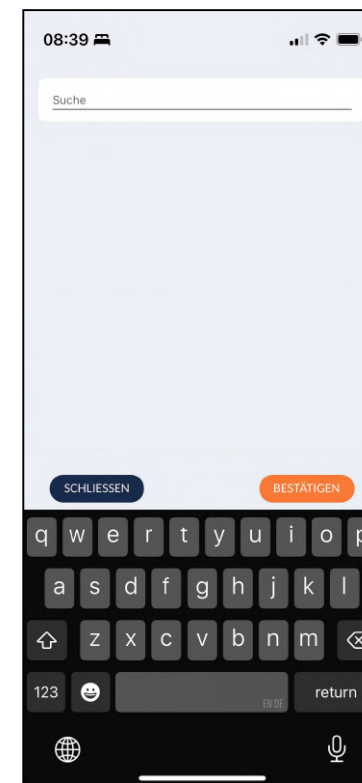
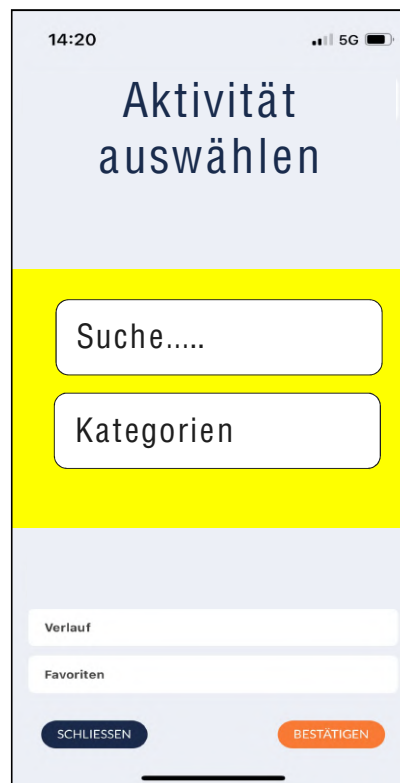
Wählen Sie Ihre Hauptaktivität
Einkaufen

Wählen Sie Ihre Nebenaktivität
Klicken Sie hier, um zu antworten

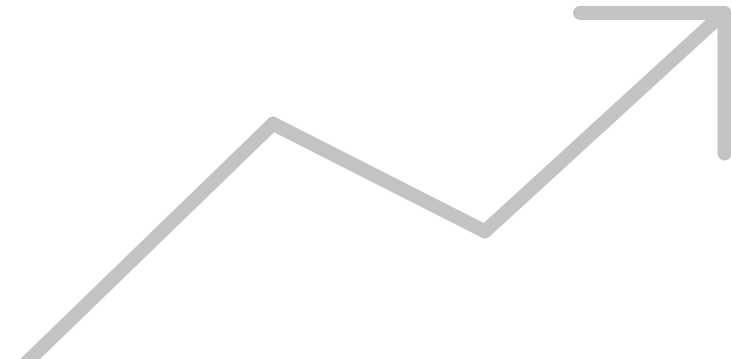
SPEICHERN

Selecting activities

- » Use space to put the search bar more prominently in the middle of the screen
 - » By tapping the search bar button
 - the search bar should move to the top of the screen
 - a keyboard should open
 - the cursor should be in the bar, ready to type in search terms
 - categories and favourites should not be shown
- » Also put categories more prominently into the middle of the screen
 - » By tapping the categories, the search bar is not shown



General usability issues



General issues

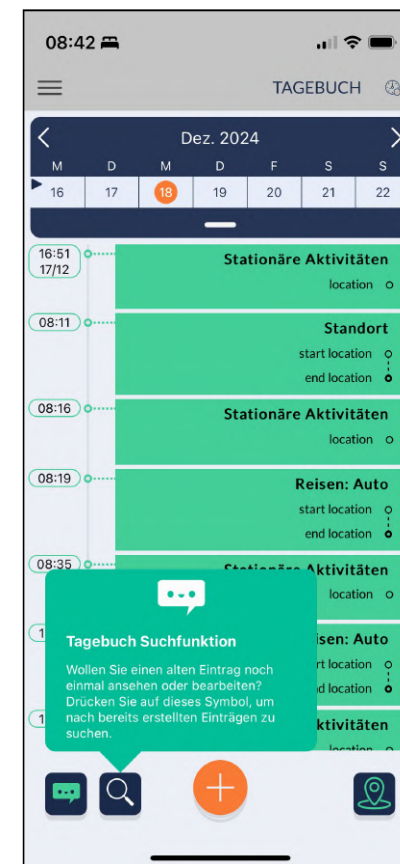
» Assistant

- » Not clear how to close it
- » Some TP tap through all the speech bubbles

» Search function (magnifier symbol)

- » A TP has problems closing it and therefore cannot reach the location suggestions
- » The “x” symbol to close is too small

"I clicked wildly on the screen to get rid of it. It reacts very badly." (TP01, f50)



General issues

Creation of an entry

Positive

- » Some TP used function 'Aktivität läuft noch' [ongoing activity]

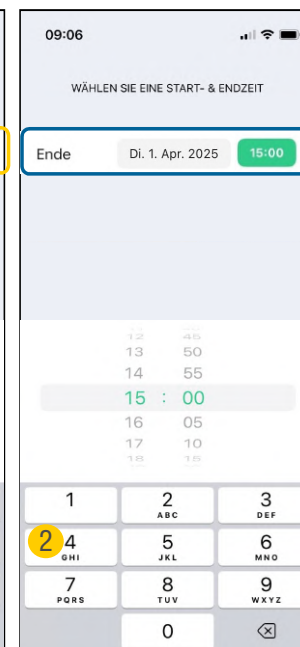
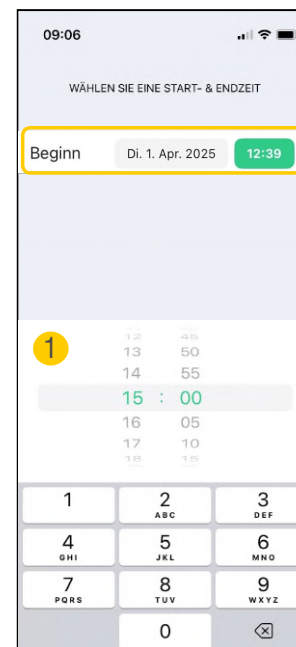
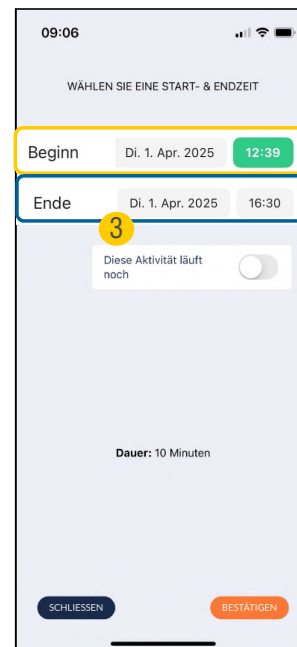
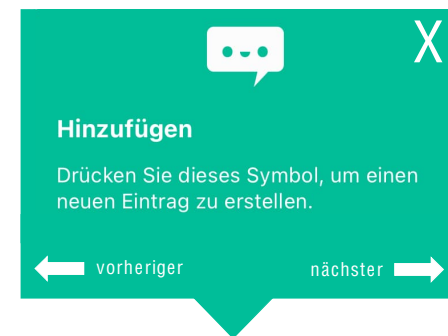
Negative

- » Many TP have problems with the fact that the start time of the very first entry cannot be changed,
 - » trying to change it results in an error message
- » Some TP do not realise that they can also enter times via the keyboard and do not only have to operate it with the arrow keys
- » Few TP are surprised about the pre-set end of the first activity (0.10 am)
- » Few try to enter a time in the future and receive an error message
- » One TP overlooks that the end of the activity can be specified

The screenshot shows a mobile app interface for creating an activity entry. At the top, the time is 09:06. Below the status bar, the text 'WÄHLEN SIE EINE START- & ENDZEIT' is displayed. The 'Start' section has a date picker set to 'Do., 19 Dez. 2024' and time pickers for '06' and '52'. Below this is a toggle switch for 'Diese Aktivität läuft noch' which is currently turned off. The 'Ende' section has a date picker set to 'Do., 19 Dez. 2024' and time pickers for '07' and '02'. Below the 'Ende' section, the duration is set to 'Dauer: 10 Minuten'. At the bottom, there are two buttons: 'SCHLIESSEN' (Close) and 'BESTÄTIGEN' (Confirm).

Assistant and time entry

- » Redesign assistant:
 - » Add arrow for <next> and <previous>
 - » Also add an <x> to close assistant
- » Redesign time entry: arrows are inconvenient to use (need to be tapped too often)
 - » Better use scrolling wheel as known from alarm clock (1)
 - » Use number pad (2) and calendar (3) for time selection



General issues

Creation of an entry - input confirmation

- » The repeated confirmation of the input ('Eingabe einfügen') is perceived as annoying
- » It is not comprehensible in which cases the confirmations are needed

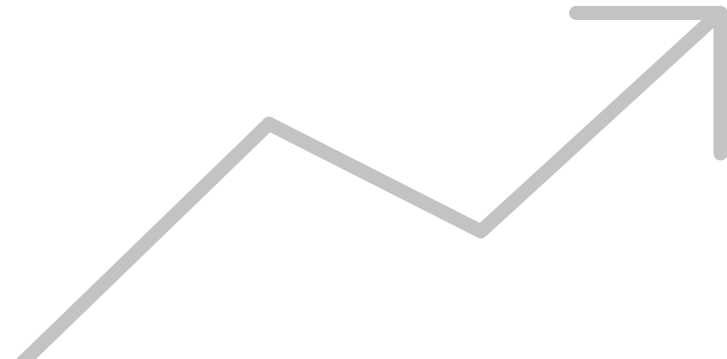
"It was laborious and not very intuitive." (TP06, f39)

Recommendation

- » Make sure that the screen is only displayed when an existing diary entry is (partly) overwritten by the new entry
- » Change headline to 'Bestehenden Tagebucheintrag überschreiben?' [overwrite existing diary entry]
- » Use space to explain what is going on

The screenshot shows a mobile app interface for creating a diary entry. At the top, the status bar shows the time 09:07 and signal strength. The app header has a close button (X) and the title 'Zeiteingabe einfügen' in orange. Below the title, there are two input fields. The first field is labeled 'Startzeitpunkt kürzen auf' and contains the date and time '19.12.2024 06:52 - 19.12.2024 08:05'. The second field is labeled 'Einkaufen' and contains the date and time '19.12.2024 07:02'. Below these fields, there is a section labeled 'Einfügen' with a text input field containing 'Essen/Getränke vor- und zubereiten, Tisch decken'. At the bottom, there are two buttons: 'ABBRECHEN' (dark blue) and 'EINFÜGEN BESTÄTIGEN' (orange).

Technical issues and bugs



Technical issues & bugs

- » Loading issues
 - » Saving activities takes long (2TP)
 - » Synchronisation is slow, sometimes leads to app crash (7TP)
- » Location tracking issue: wrong location (Brussels) shown
 - » TP closed app due to battery issues from the continuous tracking (TP8)
- » Geolocation suggestions: stationary activities displayed twice (same start time, different end times) (1TP)
- » 'Stop running activity' does not work (tapping on button not working) (2TP)
 - » Restarting the app does not help
 - » Deleting the entry of the running activity works

Technical issues & bugs

- » Geolocation function: when creating entries via geolocation, 'Invalid date' is shown instead of actual times

The screenshot shows a mobile app interface for creating an entry ('EINTRAG'). The first question is 'Um welche Uhrzeit fand diese Aktivität statt?' (At what time did this activity take place?). The answer field displays 'Invalid date - Invalid date' in green text, which is highlighted by a yellow rectangular border. Below this are two more questions: 'Wählen Sie Ihre Hauptaktivität' (Select your main activity) and 'Wählen Sie Ihre Nebenaktivität' (Select your secondary activity), both with a prompt to 'Klicken Sie hier, um zu antworten' (Click here to answer). At the bottom, there is a green button labeled 'Wählen Sie eine Hauptaktivität' (Select a main activity) and an orange button labeled 'SPEICHERN' (SAVE).

Technical issues & bugs

Quality of geolocation data

- » Geolocation data quality turns out to be faulty, not accurate
 - » Is observed in many cases and explicitly criticised by few*
- » Location tracking
 - » Sometimes inaccurate or wrong locations
- » Path tracking
 - » Paths divided into several entries, making overview confusing
 - » Paths not displayed at all
 - » Paths and corresponding time tracking often starts too late (parts cut off), making it necessary to correct times when creating an entry
 - » Transport mode prediction contains errors: TP travelled by foot, but the suggestion said bike
- » Geolocations & paths appear with significant delay

„Overall, it was very inaccurate. My journey to work was tracked well, while other things were not tracked correctly.“ (TP15, f42)

“Under movement data it said ‘cycling’, but I didn’t cycle. I don’t know why it said that.” (TP12, m54)

Contact

Your contact person

Johannes Volk

johannes.volk@destatis.de

+49 611 75 4785

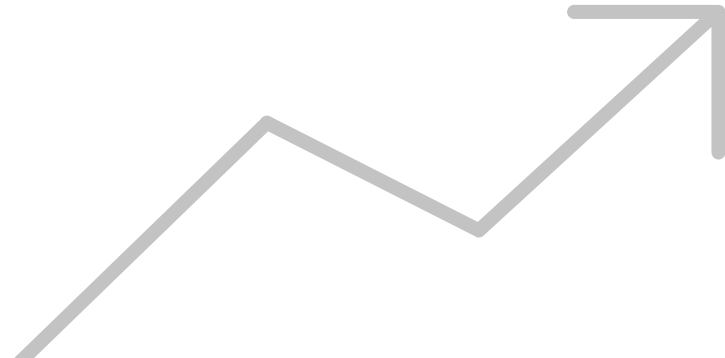
The team behind the survey

Johannes Volk

Lasse Häufglöckner

Anja Sommer

Sarah Schöffling





USABILITY TEST GEO-ASSISTED ENTRY FUNCTION

Country report: Italy - WP 2.3

V.1.0 17. 04.2025

Smart Survey Implementation-Grant Agreement Number: 101119594 (2023-NL-SSI)



Disclaimer: Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or Eurostat. Neither the European Union nor the granting authority can be held responsible for them.



Table of contents

3 Smart Survey Implementation (SSI)

4 Objectives and research questions

5 Geoservice small scale usability test in brief - ITALY

6 Methodology and participants

13 Findings

14 Preparatory activities findings

15 The intake meeting: main findings

16 Technical issues

17 Willingness, opinions and conditions for sharing geodata

19 Using geolocation service for time use survey

20 Testers experience of the app

21 Final feedback from testers

22 Overall attitude to the app

23 The evaluation of the geoservice for the diary completion

24 Preferences and insights for improvement

26 Technical issues

27 Trust and privacy

28 Final attitude to the geolocation tracking

29 Contacts

Smart Survey Implementation (SSI) project

- ✓ As part of the EU-funded **Smart Survey Implementation (SSI)** project, ISTAT contributed to testing a **geo-assisted entry function** designed to improve the quality and usability of time-use diary data.
- ✓ This smart feature, tested within Work Package 2.3, uses geolocation data collected via smartphone to suggest locations and times, supporting respondents in completing their diaries more accurately and with less effort.

Objectives and research questions

- ✓ The test focused on two main research questions:
 - whether users perceive a tangible benefit from using the geo-assisted entry function
 - how usable and intuitive they find this functionality within the context of completing a time-use diary

- ✓ The usability test aimed to assess both the **perceived benefit** and the **ease of use** of this feature. Participants used a prototype version of the app, which integrated the geolocation microservice, over two days—one with automatic geolocation-based suggestions, the other with manual input. Through interviews, observation, and feedback questionnaires, the test gathered insights into user experience, technical issues, and conditions for sharing location data.

Geoservice small scale usability test in brief - ITALY

- ✓ The Motus Discovery app was used
- ✓ Testers were internal Istat personnel, selected through a screening questionnaire to vary in age, sex, educational level and type of device
- ✓ Use of own device
- ✓ Test period: 17 february to 21 march
- ✓ Carried out inhouse in Istat Central Office (Rome)
- ✓ No incentives

METHODOLOGY AND PARTICIPANTS

Overview of participants

Participant	Gender	Age	Education*
TP01	female	33	Higher
TP02	female	35	Higher
TP03	female	39	Higher
TP04	male	49	Higher
TP05	female	49	Higher
TP06	male	51	Higher
TP07	female	51	Higher
TP08	female	55	Higher
TP09	male	56	Higher
TP10	male	60	Higher
TP11	male	60	Higher

Age distribution

Age range	26-49	50-65
Number of participants	5	6

Gender distribution

Gender	Male	Female
Number of participants	5	6

*Classification of the educational level:

Higher = University degree, higher vocational training

Medium = Graduation, vocational training

Lower = 9 or 10 years of school, no graduation

Activities carried out by using an app



Geolocation apps
are used on **daily basis**
by most of the
respondents

Other activities:

✓ **Most common activities:**

- Reading or writing emails
- Taking photos or recording videos
- Posting content
- Playing games
- watching videos, or listening to music

✓ **Least Common Activities:**

- Shopping,
- online banking,
- health tracking

Small Scale Test

In Italy, the small scale test activities follow a two-stages approach:

1. Preparatory Activities



- a. Brainstorming meeting
- b. Focus group

2. Testing Activities



- a. Intake meeting (first interview)
- b. Field phase
- c. Feedback questionnaire (second interview)

Preparatory activities

Brainstorming

17.02.2025

- **Target:** 8 participants belonging to the expert team of the Data collection Directorate in charge of the test
- **Task:** independent use of the Motus Discovery app for few days
- **Objective:** to identify potential painpoints to be discussed in the focus group

Focus group

06.03.2025

- **Target:** 5 participants belonging to the expert team of the Data collection Directorate NOT involved in the project (IT developers and questionnaire designers)
- **Task:** independent use of the Motus Discovery app for few days
- **Objective:** to explore pain points identified during app use, to be investigated during testing activities

Testing Activities

a. Intake Meeting (first interview)

3 hours

March 18, 2025

In person meeting, 11 participants involved:

- Plenary section: explanation of the project and the field phase
- One to one participant observation: app download and collection of the first impressions during predefined task performance (first interview)

b. Field Phase

2 days

March 19 and 20, 2025

Independent use of the time diary:

- Day 1: using location data for entries
- Day 2: using manual input for entries

c. Feedback questionnaire (second interview)

1 day

March 21, 2025

Opinions on the field phase:

- evaluation of the geoservice for the diary completion (positive and negative aspects)
- trust and privacy
- final attitude to geolocation tracking

Insights from Intake Meeting

Plenary Session



Organization of the one to one participants observation

Interviewer	Test Person	Time	Interviewer	Test Person	Time
 Barbara D'Amen	TP XXX	11:00-11:30	 Samanta Pietropaoli	TP XXX	11:00-11:30
	TP XXX	11:30-12:00		TP XXX	11:30-12:00
 Barbara Lore	TP XXX	11:00-11:30	 Serena Liani	TP XXX	11:00-11:30
	TP XXX	11:30-12:00		TP XXX	11:30-12:00
	TP XXX	12:00-12:30		TP XXX	11:30-12:00
 Katia Bontempi	TP XXX	11:00-11:30			
	TP XXX	11:30-12:00			

FINDINGS

Preparatory activities findings

Device use

- Various smartphone models (Samsung, Huawei, Realme) were used without major problems

Initial access and login

- The addition of a “show password” icon was suggested. Some users found it difficult to change the password.

Session persistence

- Some users found that the session remained open even after closing the app, while others were disconnected unexpectedly.

Battery consumption

- High battery consumption was reported, probably due to active geolocation.

Geolocation and synchronisation

- Some found that data synchronisation was slow after re-opening the app.

Notifications

- Not everyone received useful notifications, raising doubts about their functionality.

Logging of trips

- Some users did not see the routes recorded by the app, while others confirmed that the app suggested trips and transport correctly.

Usefulness of geolocation

- It was clarified that the tracking is used to help the user remember activities performed throughout the day.

THE INTAKE MEETING

MAIN FINDINGS

TECHNICAL ISSUES

Issues related to the App download

✓ Technical issues in app download:

few respondents reported the button «next» is not so readable and written in too small characters

Issues related to the Authorization to Geo-tracking

✓ Difficulties in understanding how to allow the geo tracking:

most of the respondents reported little or no difficulty in understanding how to authorize the app for tracking.

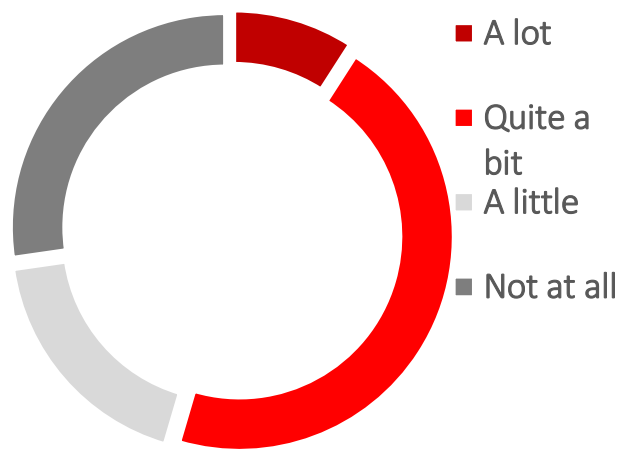
✓ Technical issues related to allow the geo tracking:

one respondent highlighted that the app's home screens were not displayed in their entirety immediately after downloading. The information at the bottom was partially overlapped.

Willingness, opinions and conditions for sharing geodata

✓ **Willingness*:**

(How open are you to sharing location information with apps?)



✓ **Opinions:**

(What do you usually think when you are asked to grant permission to track your location?)

- Some respondents are **generally cautious** about location tracking, granting permission only for **trusted purposes**, like official surveys, or for wellknown app, like google maps.
- Other respondents prefer **to limit access to location data** and are more selective with apps, concerned about privacy and data misuse.

"I have a suspicious reaction, but it is mitigated by knowledge. In this case the app is recommended by Istat, so I trust. Otherwise, I am suspicious and limit data sharing to a minimum. I often use the most restrictive security settings. I grant permission only while I am using the app."

Willingness, opinions and conditions for sharing geodata

✓ **Conditions:**

(Under what conditions do you usually share information about your location?)

Respondents share their location only when necessary for app functionality, such as navigation or activity tracking. They prioritize trust in the app and make sure that the purpose is clear, avoiding sharing on social apps or when privacy could be compromised.

“It depends on the reasons why this information is required, such as for use of navigation or searches such as restaurants in the area, not for use by social apps.”



Using geolocation service for time use survey



✓ **ICON TO VIEW GEODATA:**

Participants understood immediately how to identify the icon to view the data, while only few figured it out after several attempts.

✓ **GEOLOCATION DATA (GREEN BOXES):**

Most participants understood immediately how to click on the green boxes for viewing geodata, whilst few required help or understood after several attempts.

✓ **CREATE TIMELOG (ORANGE BOXES):**

Several participants figured out the orange boxes for creating a timeline after multiple tries, while others understood immediately.

✓ **CREATE FROM MULTIPLE (BLUE BOXES):**

Only few participants successfully created a path by connecting multiple movements, while the majority did not.

✓ **ASSISTANT BOTTON:**

Nobody use the assistant botton or found it helpful.

...**FINALLY**

Despite some difficulties, all participants were able to create diary entries by using geolocation service

Testers experience of the app

APP FEATURES AND GEOLOCATION DATA

- ✓ Most participants rated the app functionalities as **highly intuitive or moderately intuitive** and the **geolocation data easy to find**



SUGGESTED IMPROVEMENTS

- ✓ adding an **icon to return** from georeferencing to the diary mode
- ✓ using the **hamburger menu** (on the top of the screen) could be more intuitive for accessing data

USING GEOLOCATION DATA FOR DIARY ENTRIES

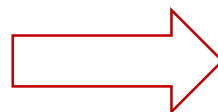
- ✓ Some participants found it **extremely easy**, whilst some others rated it as **somewhat easy** and **moderately easy**



- ✓ rename the option "**create timelog**" with "**add an activity**"
- ✓ the botton "+" is a bit confusive and could be removed from the green diary

FIRST IMPRESSIONS OF THE GEOLOCATION SERVICE

- ✓ Overall, participants reported a **positive first impression of using geolocation service** for creating activities in the diary
- ✓ Geolocation data has been considered by participants as **supportive data**



- ✓ add a summary of the registered activities
- ✓ in case of a lot of stationary activities, that happened in the same place for a long time, geolocation data could be not so helpful

FINAL FEEDBACK FROM TESTERS

Overall attitude to the app

- ✓ All testers recorded their activities at a later time
- ✓ The geolocation-based mode is by far the preferred option to create the diary
- ✓ The testers found the geolocation data helpful in reconstructing the sequence, places and timing of their daily movements

"I could enter the activities at the end of the day. Without the geodata I would surely have missed something"
- ✓ By relying on geodata, testers felt more confident about the accuracy of their completion

"I think it has made the process of entering activities faster, smoother and more accurate"
- ✓ For some, it has provided an insight into how they organise their day-to-day activities and how they can improve

The evaluation of the geoservice for the diary completion

On a scale from 0 to 5...

how easy was it to add moving activities to the diary?

how useful did you find the possibility to have your location automatically tracked by the app?

how intuitive did you find the geolocation feature?

0 = «not at all» 5 = «very»

- ✓ On average, testers found this task moderately **easy** (3,2), with a modal rating of 4. Two testers scored 0 and one of them failed to record activities in the diary
- ✓ The **intuitiveness** of this feature was rated 4,1 on average, with no score lower than 3.
- ✓ The geoservice was found to be quite **useful**, with an average score of 4.

Preferences...

- ✓ The possibility to draw on automatically recorded information was appreciated

“With the geolocation system, it is easier to remember what you have been doing during the day...”

“A summary of the activities carried out during the day is useful.”

- ✓ Three different ways to classify activities resulted to be a valid help for the diary completion

“Searching for the type of activity was easy, both by entering a keyword and by surfing the hierarchical classification. I used the latter most often.”

“The suggestions section made it possible to quickly classify displacements that had already been entered”



...and insights for improvement

✓ Colours seem to be the weak point of the app

“Texts are written in a font that is too light to be easily read on a blank screen”

“I would use a different background colour to highlight the activities already entered”

“When merging or separating activities, you cannot see any change”

“I would use two different colours for static and moving activities to visually distinguish them”

✓ Icons, commands and feedback should be improved

“The button for accessing geolocalisation is a commonly used symbol for this purpose. However, I initially thought that clicking on that button would activate geolocation instead of displaying the list of trips.”

“When using the position function, it is not easy to understand how to combine or divide periods.”

“I would replace the wording ‘Create timelog’ with ‘Add an activity’ in order to make the meaning of the action clearer. In general, I would avoid talking about time, and refer directly to activities.”

“When I fill in an activity that is reported in the green boxes, there should be an indication that it has been successfully completed”

“In my opinion, ‘switch between different timelines’ and ‘normal and tentative modes’ are not explanatory enough.”

Technical issues

- ✓ Burdensome notifications
- ✓ Battery consumption due to the geolocation use
- ✓ Slow and not always efficient when updating information
- ✓ Weak GPS signal may be a problem

“In some areas with a weak GPS signal, the accuracy may be reduced. This may force the user to check and correct the data, which can be annoying.”

“When I was on the metro, the app geolocated me in Brussels.”

“It did not always record the timestamps”

Trust and privacy

- ✓ Privacy concerns could be a drawback, as automatic location tracking means that users' movements are constantly monitored.

"I have a feeling of being under control"

- ✓ I would be useful to create an easy way to deactivate geo tracking

"I would like to be able to easily turn off geolocation at any time, without having to go into the application settings."

- ✓ The logout functionality should be made clearer

"When you close the app, it remains active in the background. From a security perspective, this should be avoided and the respondent should receive a notification advising to log out."

Final attitude to the geolocation tracking

After this experience, almost all testers would give their consent to the geotracking, but only if

- ✓ Clarity and accurate information about the use of geolocation data is provided
- ✓ The registration adds real value to the user experience
- ✓ It is a certified statistical survey and that privacy is respected
- ✓ The data will not be shared with third parties
- ✓ The location information will remain on their mobile phone
- ✓ The security measures in place, the retention period and the possibility to withdraw consent are guaranteed at all times

Project Coordinator:

Monica Perez - perez@istat.it

Test coordinators:

Barbara D'Amen- barbara.damen@istat.it

Barbara M. R. Lorè- lorè@istat.it

Supporting team:

Katia Bontempi – katia.bontempi@istat.it

Serena Liani – serena.liani@istat.it

Samanta Pietropaoli – pietropaoli@istat.it

Vincenzo Triolo – triolo@istat.it

Appendix D : Detailed results from French experiments on mode effects

To what extent is the mode selection driven by outcome versus layout? Analysis from the 2023 TUS test web app

October, 2024

Fanny Mikol, Insee – based on discussions with Danielle McCool, Utrecht University

Hypothesis

The UX/UI analysis of the 2023 TUS test web application in France highlighted potential improvements for the button interface compared to the existing app. In particular, it was found that some key buttons (such as "sleeping") were suboptimally placed and should be moved to the first or second level of the interface for better accessibility. To expand on this practical insight, we conducted a comprehensive analysis examining the relationship between the ergonomic characteristics of each button in the interface layout and the frequency of its corresponding activity in users' web declarations.

We hypothesize that **multiple factors influence the selection of a given activity k in the web diaries** (k ranging from 1 to 130, corresponding to the total number of "elementary" activities in the TUS nomenclature):

- i. The **frequency of the activity** in the general population increases the probability of an individual selecting this activity k in their web diary during their recorded day.
- ii. The **"ease of access" of the corresponding button** in the interface increases the individual's likelihood of clicking on it.

Model

To test these hypotheses, we modeled the probability of an individual i selecting an activity k during their day, Π_{ik} as a function of:

- i. The frequency of this activity f_k as observed in the paper diaries. We assume that paper diaries accurately represent the "true" frequency of each activity. While this is a significant assumption, it provides a necessary reference point for comparing activity frequencies in the web diaries.
- ii. Characteristics of the corresponding button layout:
 - The depth of the button in the interface hierarchy, i.e., the number of successive clicks required to reach the activity: $rank_k$
 - For panels with more than 6 buttons, a binary variable indicating whether the button is at the bottom of the list, potentially requiring scrolling and thus less visible: is_low_k .
 - A binary variable indicating whether the button label contains the word "Other". Indeed our goal was to check whether such catch-all categories encouraged or discouraged selection: $other_k$.
- iii. Characteristics of the individual X_i :
 - sociodemographic factors: sex, age, family status, income, residence type)...
 - a binary variable indicating the individual's general propensity to use buttons instead of other means (search bar or open field) for selecting activities: above or below the median propensity of the whole sample. We indeed believe this characteristic could capture the web navigation style of each individual.

To process the model, we use a **logistic specification**:

$$\Pi_{ik} = \alpha + \log[f_k / (1-f_k)] + \text{rank}_k + \text{is_low}_k + \text{other}_k + X_i + \varepsilon_{ik} \quad (1)$$

Here the $\log[x / (1-x)]$ specification for the activity frequency f_k variable is used to ensure homogeneous terms on both sides of the equation. And instead of the raw frequency f_k , we use an adjusted frequency f_adj_k that equals f_k for activities k detected in paper diaries, and f_min for activities never encountered in paper diaries. Such adjustment avoids zeros in the \log function that could otherwise cause the model to fail to converge. f_min is the minimum frequency observed for an activity in the paper diaries, corresponding to the situation where only one individual (out of the 793 individuals who filled a paper diary) recorded the activity. Hence, the adjusted frequency can be written as: $f_adj_k = \text{Max}(f_min, f_k)$. ε_{ik} are the residuals of the regression model.

Database

The database used to run the model contains a total of 10,518 rows. Each row represents a unique combination of an individual who completed a web diary (670 individuals) and an eligible activity (132 activities in the TUS nomenclature). The variable of interest I_{ik} is then constructed as follows: $I_{ik} = 1$ if individual i practiced activity k during their recording day, and $I_{ik} = 0$ otherwise. This structure allows for a comprehensive analysis of all possible individual-activity combinations, even though most individuals did not engage in most activities on their recording day (resulting in a large number of rows associated with $I_{ik} = 0$).

To address potential selection effects that may occur because not everyone in the population sample completed a web diary, we calculated a weight for each individual to correct for such bias. For each participant, the weight corresponds to the inverse probability of them belonging to the web diary respondent group, given certain observed characteristics (sex, age, family status, income, digital equipment, residence type, and geographic location). The same weighting procedure was also applied to the paper diary respondents to ensure comparability of activity frequencies between paper and web diaries.

The table 1 below displays some basic statistics from this database, illustrating the distribution of variables across (i) individuals and (ii) activities.

Table 1 - Key statistics of the web respondents database: 760 individuals * 132 activities

Variables related to individuals (weighted by individual weights correcting for web selection)		
Proportion of women		52,6 %
Proportion of people aged 35 or older		74,9 %
Proportion of those in single-parent families		12,6 %
Share of people with a monthly household income $\geq 3000\text{€}$		50,7 %
Share of individuals living in a house (vs apartment)		76,5 %
Frequency of the use of "selector" (buttons)	Mean	77,8 %
	Q1	72,7 %
	Median	93,3 %
	Q3	100,0 %
Variables related to activities		
Weighted frequency of activity practice among WEB diaries respondents	Mean	7,3 %
	Q1	0,3 %
	Median	1,6 %
	Q3	4,9 %
	Max	96,6 %
Weighted frequency of activity practice among PAPER diaries respondents	Mean	7,5 %
	Q1	0,1 %
	Median	1,0 %
	Q3	5,4 %
	Max	98,7 %
Rank of the activity in the buttons interface hierarchy	1st	3,8 %
	2d	19,7 %
	3d	68,9 %
	4th	7,6 %
is_low (i.e. whether the button is at the bottom of a list of 6 or more items)		19,70 %
other (i.e. whether the button label contains the word "Other")		5,30 %

Specifications

Several specifications of the logistic regression were tested to check the sensitivity of the results:

- Global specification: all 10,518 observations of the raw database are considered.

- (b) Exclusion of the activity “*Other personal care activities*”: previous analyses of the 2023 TUS test¹ showed that this category was often chosen by respondents on the digital app as a replacement when they were unsure where to click. In particular, this button seems to have served as an “escape” option for leisure-related activities that respondents struggled to categorize. As a result, this activity is significantly over-represented in the digital app’s outcomes compared to paper diaries outcomes. Including this category leads the model to deal with extreme values that may deteriorate its results. Consequently, we also tested the model after excluding the observations related to the “*Other personal care activities*”.
- (c) To account for the fact that we are dealing with rare events – given that in average, an elementary activity k (among the 132 elementary activities of the TUS nomenclature) is practiced by only 7 % of web respondents – we implemented specific models to handle this data bias:
 - A logistic regression model incorporating Firth’s correction to the likelihood. This general approach reduces small-sample bias in maximum likelihood by applying a likelihood penalty.
 - The “ROSE” method (Random Over-Sampling Examples), a bootstrap-based technique for binary classification in the presence of rare classes. This approach generates artificial observations based on the observed covariate distribution to achieve a balanced 50/50 split of the variable of interest.

Results and conclusion

Table 2 displays the results of these various logistic regressions.

Table 2 - Logistic regressions modeling the probability for individual i to practice activity k

Covariates	Standard logistic model				Logistic model adjusted for the existence of rare events			
	Coverage: all activities		Coverage: all activities, "Other personal care activities" excluded		Firth's method		ROSE method*	
	Estimate	sign.	Estimate	sign.	Estimate	sign.	Estimate	sign.
Intercept	0,38	*	0,34	.	0,35	.	2,63	***
Log-odds transformation of the weighted frequency of activity practice among paper diary respondents: <i>log[f_adj_i / (1-f_adj_i)]</i>	0,78	***	0,79	***	0,79	***	0,74	***
Sex: female	0,28	***	0,29	***	0,29	***	0,26	***
Aged 35 or older	0,18	***	0,17	***	0,17	***	0,11	***
Living in a single-parent family	-0,21	***	-0,22	***	-0,22	***	-0,21	***
Monthly household income >= 3000 €	0,08	**	0,09	**	0,09	**	0,06	**
Living in a house (vs an apartment)	-0,08	*	-0,08	*	-0,08	*	-0,06	*
Frequency of use of the "selector" in the web-app above the median of the web respondents population	-0,08	**	-0,11	***	-0,11	***	-0,07	***
Rank of the activity in the button interface hierarchy	<i>réf: 1st</i>							
2d	-0,42	*	-0,84	***	-0,85	***	-0,71	***
3d	-1,06	***	-0,94	***	-0,95	***	-0,75	***
4st	-1,13	***	-1,05	***	-1,06	***	-0,93	***
other (i.e. whether the button label contains the word "Other")	1,34	***	-0,26	**	-0,26	**	-0,23	***
is_low (i.e. whether the button is at the bottom of a list of 6 or more items)	-0,88	***	-0,36	***	-0,36	***	-0,24	***

Signif. Codes: '***' 0.001; '**' 0.01; '*' 0.05; '.' 0.1

* Unlike the three other regression models which are weighted, the regression using ROSE method is unweighted.

Note that weighting the observations in these three regression models has very little impact on the results.

We find that **all these methods yield coefficients that are consistent in their sign and significance, with relatively similar magnitudes**, except for the coefficient associated with the “*Other*” variable. Indeed, when we include the activity “*Other personal care activities*” in the sample (1st model), this leads to a highly significant positive coefficient for the variable identifying buttons whose label contains the word “*other*”. However, once we remove the observations corresponding to this activity from our sample, this same coefficient becomes significantly negative. This suggests that, except for the “*Other personal care*” activity and all else being equal, web respondents are less likely than paper respondents to select activities whose corresponding buttons contain the word “*other*”. In other words, buttons labeled with “*Other...*” do not seem to be used as “catch-alls” options by web respondents, contrary to what one might have expected, with the notable exception of the “*Other personal care activities*” button.

¹F. Mikol : « *Analysis of the 2023 Methodological Test – French TUS* » and S. Quantin and L. Court: « *Measurement effects in the Time Use test survey* », Contributions to the ESSNet Smart Surveys WP2 deliverable.

In all four models, the coefficient associated with the log-odds transformation of the frequency of activity practice among paper diary respondents (the “reference” frequency) is, unsurprisingly, highly significant and positive. This confirms, as expected, that the frequency of each elementary activity in web diaries closely matches those observed in paper diaries.

One interesting finding is that respondents who made extensive use of the buttons (“*selector*”) in the web app reported, on average, fewer distinct activities compared to other respondents. Indeed, while the quick-access buttons in the web app were designed to facilitate data entry, they might create a sort of “cognitive shortcut”: users who heavily rely on these buttons tend to stick to a limited set of familiar activities, compared to those who more frequently use the search bar or free-text field – tools that lead to a greater diversity in activity selection, as users initiate the description in their own words. This suggests that the convenience of buttons might paradoxically reduce the diversity of reported activities.

An obvious finding is that the position of buttons in the interface layout significantly influences the probability of recording the corresponding activities, all else being equal: the deeper a button is placed in the interface, the less likely it is to be selected. Similarly, for lists containing 6 or more buttons, those situated at the bottom are less frequently clicked, due to their reduced accessibility.

As a conclusion, this work demonstrates the significant impact of the web application’s ergonomics on activity selection patterns, particularly regarding the buttons interface layout. While buttons prove to be an efficient tool for recording simple or familiar activities, alternative input methods should also be enhanced for more specific activities. The search bar, in particular, could be promoted as an effective tool to counter these interface-induced biases and ensure more comprehensive activity recording.

QUESTIONING THE IMPACT OF ENDOGENOUS SELECTION IN AN AB/BA CROSSOVER TRIAL: THE EXAMPLE OF THE TIME USE TEST SURVEY

January 24, 2025

Abstract.

We consider the impact of possible endogenous participation (Missing Not At Random nonresponse scheme) on the results of the analysis of a crossover trial conducted by INSEE to investigate potential measurement effects related to an internet collection of the Time Use Survey compared to a paper collection. To do this, we conduct a sensitivity analysis [Rosenbaum, 2002a, 2010] on the conclusions obtained by comparing, after matching, the durations reported in the first inquiry from similar but possibly different respondents. The aim is to quantify the extent of an unobserved bias necessary to overturn the conclusions obtained, for example if participation is correlated with the unobserved variable of interest and the mode of collection. We find that internet data collection impacts the reported durations of sleep, meals, travel, and leisure for a limited number of respondents. These results are robust to possible endogenous selection and confirm those obtained from the analysis of the crossover trial assuming missing at random nonresponse.

Keywords. Mixed-mode surveys, crossover trial, endogenous selection, sensitivity analysis

1 Introduction

The time use survey is a household survey designed to collect people’s time use in order to carry out quantified analyses of the time spent at work, on domestic tasks, on leisure activities, and so on. A methodological test prior to the introduction of mixed-mode collection for this survey was set up in 2023 to question the existence of possible measurement effects linked to Internet collection. Usually, in this case, the protocol randomly assigns a collection mode to two representative samples of the population in order to compare, after matching, the responses given by the respondents. The protocol for the methodological test implemented by INSEE is different: it involves an AB/BA crossover trial on a sample of residents from 2,100 dwellings. The same people record their daily schedule in both the paper and digital diaries, one week apart. The dwellings are randomly assigned to a sub-sample defining the order of completion (digital diary then paper or the reverse). In this way, unlike the analysis of matched data, the results under different modes of collection are recorded for the same individual. In this respect, a crossover trial is more powerful because the intra-individual variance is often lower than the inter-individual variance. However, these advantages may be offset by disadvantages that are not present in the analysis of matched data, such as the need to pay attention to the length of the period between the two collection modes, a possible measurement effect of the previous collection mode on future responses (carryover effect), etc.

While these disadvantages can be reduced by adopting an appropriate design, non-response issues can not be excluded from a cross-over trial, as with any test involving a survey. A respondent may not take part in the whole test for many reasons, such as their direct experience of the collection mode on a given day, but also their comparison with the collection mode proposed earlier in the test, and their links with the variable of interest. In this type of protocol, most articles assume either that the probability of a planned observation being missing does not depend on the value that would have been observed, i.e. that the data are missing at random (see for example [Patel \[1985\]](#)), or rely on the absence of non-response in the first interrogation [[Ho et al., 2012](#), [Basu and Santra, 2010](#)].

In this study, we investigate the impact of a possible endogenous selection on the conclusions, as to the presence of a measurement effect, obtained by wrongly assuming that the missing data are MAR. To do this, we first compare the results obtained by the classic cross-test analysis with those obtained by analysing only the responses given in the first period between respondents using different modes. Cross-test analysis assumes that non-participation is Missing Completely At Random (MCAR) or Missing At Random (MAR), but is based on responses given by the same person. The analysis of matched pairs is based on the responses given by similar people and assumes the absence of different unobservable characteristics between these matched people, i.e. the absence of a selection bias (also known as a hidden composition bias linked to differences in an unobservable characteristic which is itself linked to the variable being evaluated). In both cases, the problem of endogenous selection (or MNAR non-response) remains. However, it is possible to take account of the impact of this possible hidden composition bias on the results of an analysis of matched data, by conducting a sensitivity analysis [[Rosenbaum, 2002a, 2010](#)]. The results obtained under the assumption of the absence of endogenous selection bias will be compared with those that relax this assumption. This consists of quantifying the extent of an unobserved bias necessary to invalidate the conclusions obtained by comparing the responses (after matching) of comparable but possibly different respondents.

2 Results assuming no endogenous selection

The protocol used to test the possible measurement effects of the collection mode on the results of the Time Use survey is a cross-over trial, or a test-retest survey, also known as a crossover design. A cross-over design differs from a conventional test in that each surveyed unit is randomly assigned not to a single collection mode, but to a sequence of collection modes. The justification for this type of survey design is as follows. Suppose we want to compare the impact of two collection modes A and B on the same surveyed unit. One possible approach is to ask the survey questionnaire to be completed first with collection mode A, then with collection mode B by the surveyed unit. However, such a protocol is potentially biased, because the answers given may reflect an impact linked to the time elapsed between the two interviews, unrelated to the collection mode considered.

To overcome this difficulty, the order in which the questionnaires are completed is randomly assigned to two sub-samples of surveyed units, each of which is representative of the population studied. In the case of the time-use test survey, the units in the first sub-sample (referred to in the following as AB) respond to the questionnaire first by internet, then 7 days later by paper; the units in the second sub-sample BA respond to the questionnaire first by paper before responding to the questionnaire by internet, 7 days later. It is therefore expected that the activities carried out during the first survey will be more or less the same as those carried out during the second survey¹. Because it includes two interview periods (and two collection modes) this protocol is called an AB/BA crossover trial².

2.1 Methodological approaches for analysing an AB/BA cross-over trial

In an AB/BA crossover test, unlike the usual test surveys, responses using different collection methods are recorded for the same person: the comparison of responses between different collection methods is therefore intra-individual (*within-subjects analysis*). From this point of view, such a protocol theoretically makes it possible to estimate a measurement effect with greater precision, for a given sample size, which makes it possible to implement a test at lower cost. Indeed, for a given variable, the intra-individual variance is often lower than the inter-individual variance that would result, for example, from a test comparing the responses of different respondents after random allocation to a collection method (*between-subjects analysis*). However, this advantage is offset by the fact that it is more difficult to use the test results. The first disadvantage is that the responses given to the second question may reflect a persistent impact of the first collection method: this is known as a carry-over effect. In our study, for example, the responses to the second survey may reflect a learning effect linked to the collection method used the first time. The Internet survey questionnaire is based on lists of activities and a prompt. Their use in the first interview can have an impact on the nature of the activities declared in the second interview when filling in the paper questionnaire, allowing for example a more precise coding than that resulting from a free “untrained” field, as is the case in the first interview via the paper questionnaire. Similarly, the responses may reflect an effect of the different mode of collection between the first and second interviews (treatment by period interaction). In our study, for example, such an effect could be caused by the respondent’s weariness during the second interview, leading him or her not to capture all the time slots and associated activities with the same attention as during the first interview³. Usually, such effects are assumed to be neutralised upstream by the design of the cross-over test with the compliance of a sufficiently long latency period between the two interrogations, known as the washout

¹The interviewers were asked to suggest that the collection day for both surveys be a working day.

²This name is used to differentiate it from another crossover trial protocol which would also involve the AA and BB processing sequences

³With two periods, it is not possible to identify these two effects separately.

period, which is, as we mentioned, 7 days in the test for the Time Use survey. *In this study, we will make this assumption.* Another difficulty lies in ensuring that respondents take part in the entire test. Non-response cannot be ruled out in a cross-over test, as in any test involving a survey. A respondent may not take part in the entire test for a number of reasons, such as their direct experience of the collection method on a given day, but also their comparison with the collection method proposed earlier in the test, and their links with the variable of interest. This greatly complicates the analysis, even in the presence of an appropriate design which makes it possible to exclude the hypotheses of a carry-over effect or a different collection mode effect between the first and second interrogation. This difficulty in interpreting the results is the main methodological focus of our study.

A common model for analysing an AB/BA cross-over trial (*within-subjects*)

We assume that the answer given by a respondent $i, i = 1, \dots, n$ at period $j, j = 1, 2$ can be represented by a random variable Y_{ij} . For the ease of notation, we note $t = t(i, j)$ the collection mode for respondent i at a period j . A model widely used in the analysis of an AB/BA crossover trial can be written as follows:

$$Y_{ij} = \mu + \pi_j + \tau_t + (\tau\pi)_{tj} + \xi_i + \epsilon_{ij} \quad (1)$$

where μ is a mean, π_j is the period effect, τ_t is the measurement effect of collection mode t and $(\tau\pi)_{tj}$ their interaction. Classically, to ensure parameter identification, it is assumed that $\pi = \pi_1 = -\pi_2$, $\tau = \tau_A = -\tau_B$, and $(\tau\pi)_{t1} = 0, (\tau\pi)_{B2} = (\tau\pi) = -(\tau\pi)_{A2}$. ξ_i is an independent effect (*random effect*) linked to the respondent.

Usually, $(\tau\pi)$ is often identified as the carryover effect. As we have said, we will assume for the rest that the design of the test allows it to be considered as null due to the time interval imposed between the two interrogations.

This individual random effects model can be estimated using panel econometrics. If, for an individual i , one of the two books (Y_{i1}, Y_{i2}) is not completed, it is still possible to include it in the analysis. This approach is only valid, from a theoretical point of view, in two cases (to our knowledge): (i) if the missing observations are missing completely at random (MCAR) or missing at random (MAR), according to the classification proposed by [Rubin \[1976\]](#), or (ii) if the non-response is endogenous (MNAR) but in this case, only if only the responses given in the second period are missing [on this approach, see [Ho et al., 2012](#)].

In the Time Use test survey, 29 % (i.e. 27.5 % for the paper/internet sample and 29.8 % for the internet/paper sample) of the 2,100 households surveyed, for a total of 1070 people, agreed to take part in the test survey. Of these people, 1,045 were actually able to take part in the test because they had internet access (529 in the internet/paper batch and 516 in the paper/internet batch). We will consider that refusal to take part in the test does not depend on the sub-sample⁴, so that the assignment to a sequence of collection modes can always be assumed to be random between the two sub-samples⁵. However, contrary to the protocol, not all the people who agreed to take part completed both diaries, and some ended up completing neither or only one of them. As a result, at the end of the test, while 77 % of paper/internet respondents completed the questionnaire in the first period, only 54 % completed both methods, a drop of 23 percentage points. (cf. table 1). Conversely, while 64 % of internet/paper respondents answered the questionnaire in the first period, 57 % answered using both methods, i.e. 7 percentage points less.

⁴The sequence of collection methods is not known in advance by the people surveyed

⁵The two sub-samples are therefore no longer assumed to be representative of the original population, but of a sub-population which is nevertheless similar between the two sub-samples.

Table 1: Types of diaries completed by participating respondents, by batch (%)

	Paper-Web (516 ind.)	Web-Paper (529 ind.)	Total (1,045 ind.)
None	16	18	17
Web diary missing	23	18	20
Paper diary missing	7	7	7
Both diaries completed	54	57	56

Scope : participants with Internet access, France.

Note: Participation in the test is defined when the completed diary is usable (at least 4 time slots filled in). The batches considered are the actual batches observed and not the assigned batches: a very small number of people surveyed did not follow the assigned order of collection methods.

Non-response at the first interview is significant and rules out an approach allowing MNAR data. In addition, the higher non-response in the second interview when the respondent answered the paper diary in the first interview raises the question of a possible endogenous selection. For example, respondents with a large number of different activities on the same day (and therefore a large number of time slots to be recorded) may have decided out of weariness not to participate in the second survey, especially if the paper diaries were deemed time-consuming to record. Consequently, in the absence of a theoretical solution for estimating parameters in the presence of endogenous selection, the relevance of the results of the cross-test analysis should be questioned.

Matching analysis (*between-subjects*)

It is also possible to use a cross-over test, restricting the analysis to respondents in the first period. This analysis involves comparing the answers given by different respondents using different collection methods. However, these respondents will be considered similar because they are matched on observed characteristics. The analysis of matched pairs assumes the absence of different unobservable characteristics between these matched individuals, i.e. the absence of selection bias - since only respondents are considered - also known as hidden composition bias, linked to differences on an unobservable characteristic linked to the variable being evaluated. Thus, in both analyses, the hypothesis of a possible endogenous selection bias (MNAR non-participation) weakens the results.

Optimal pair matching quality

The relevance of the approach in the case of analysis by matching depends on the quality of the matching process, since we will be questioning the possible existence of a composition bias *within the pairs*. It is therefore necessary first to discuss the quality of the matching carried out in this analysis.

Traditional propensity score matching tends to create treatment and control groups with similar distributions for the observed variables. However, although the units surveyed within each pair thus formed have a similar (estimated) propensity score, they may nevertheless differ significantly on specific covariates. In order to create more similar pairs of surveyed units, we construct a distance that penalises significant differences in observable characteristics, and then create pairs of units that are as similar as possible, using an optimisation algorithm. A precise description of the matching process is provided in Appendix A. Without going into too much detail, however, two important points should be noted.

The duration of the various activities declared by a surveyed unit depends on the day of

the week on which the questionnaire is completed (weekend or not) and its main situation (employed, unemployed, retired or early retired, unable to work, studying, living at home or other situation). It is therefore essential to ensure that, within each pair, the units surveyed are identical with regard to these observed characteristics. As far as the day of the week is concerned, an exact match is required. To take account of the main situation, a penalty is taken into account in the distance referred to above. If two surveyed units k and l do not have the same main location, the penalty function used adds to the initial distance 10 times the maximum distance observed (on the previously calculated distance) if the two units differ on their main location. This penalty is taken into account by the optimisation algorithm as follows: if an exact match is possible, it will be considered; otherwise, a match as close as possible to an exact match will be made. Finally, we also include the usual socio-demographic variables available, such as gender, age, housing type, household type and income structure in 2021. In total, our matching is based on 63 covariates.

Table 2: Matching quality (% of pairs with exact match) on some of the characteristics observed

Household characteristics	
Dwelling	
House	67
Priority urban district ⁶	95
Social housing	83
Homeowner	67
Household income structure in 2021	
at least one unemployed benefit	63
at least one retirement pension	87
number of inhabitants with salaried income	54
Household type in 2021	
To be a single employee or not	84
To be a single retiree or not	96
To be a retired couple or not	95
To be a employee couple without children or not	85
To be a employee couple with children or not	69
Individual characteristics	
Sex	76
Age	8
French nationality	96
The main employment situation (employment, study, unemployment, retired, etc.)	98
Matrimonial situation (married, single, cohabitation, etc.)	71
Questionnaire completed on a working day	100

The purpose of table 2 is to illustrate the quality of the match obtained by specifying, for each characteristic observed, the percentage of pairs where the match is exact. As expected, after matching, 100 % of people who responded online on a weekday were matched with a person who had also responded to the paper questionnaire on a weekday. Similarly, it is interesting to note that for the main situation, which distinguishes between people who are employed, studying, unemployed, retired or disabled, 98 % of the pairs are made up of an Internet respondent whose situation is similar to the paper respondent to whom he or

she is matched. The quality of the matching on this characteristic was important because the main situation of the person surveyed is correlated with the nature and duration of the activities declared. Moreover, 100 % of the people in employment who responded via the internet were matched with a respondent to the paper questionnaire who was also in employment; it will therefore be possible to focus our analysis of working hours on people in employment only, in this case relying on an exact match⁷. Generally speaking, whatever the covariate, at least 63% of the pairs are made up of identical respondents in terms of the characteristic considered.

In summary, these results show that after matching, within each pair, the units responding by internet and those responding by paper do not differ more than would have been expected if the mode of collection had been randomly assigned (see Appendix B for a more formal presentation of this assertion).

2.2 Assuming the absence of endogenous selection, the results are consistent

In this section we present estimates, **under the hypothesis of the absence of endogenous selection bias**⁸, of the measurement effect of Internet data collection on the durations of different types of activities. These are obtained using the crossover test approach and the matching approach. Before discussing these results, however, it is useful to explain how all daily activities are grouped together in the various aggregates studied. As we have explained, some activities (sleeping, eating, leisure and travelling) will be analysed in detail later. Knowing in which aggregates they are found will therefore shed light on the detailed results obtained.

Among the aggregates presented, physiological and personal time includes time spent sleeping, eating (but not preparing meals), hygiene and personal and medical care. Work and study time includes hours spent at work for those surveyed in employment, hours spent in training (at school, at work or in activities other than sport and the arts⁹), breaks between these activities, and time spent on trade union activities or job-seeking. The **time spent on housework and leisure activities** totals the time spent on all household tasks within the home, as well as on shopping for consumer goods and personal artistic activities. While it includes the time devoted to one's pet, it does not include the time devoted to one's children, which is counted as time devoted to **looking after other people** in the household or in another household. Time spent on **sociability** corresponds mainly to time spent entertaining friends, chatting or telephoning. Leisure activities include those associated with culture (reading, television, shows, etc.) or sport, as well as walks and hikes; these last two activities are not counted as *travelling*.

All these durations constitute a partition of the total aggregate duration declared. The duration of the period covered corresponds to the time between the start time of the first time slot declared (normally 4 a.m. on the day of the survey) and the end time of the last time slot declared (normally 4 a.m. the following day). These two times may therefore differ if the activities carried out during certain time slots are not declared by the respondent.

The table 3 details, for each duration, the estimates of (i) an additive and constant measurement effect of Internet data collection, obtained by analysing the data from the cross-over test, and (ii) an average measurement effect, obtained by analysing only the timetables declared in the first period after matching respondents who responded using different data collection methods. Thus, assuming the absence of endogenous selection

⁷Our matching also ensures that 100 % of internet respondents who are studying or unemployed are also matched with a paper respondent in a similar situation.

⁸or similarly the absence of hidden composition bias

⁹such as driving lessons, cookery lessons, etc.

in participation in the crossover trial ¹⁰, detailing one’s timetable using the internet questionnaire would, for example, reduce the total declared journey time by 20 minutes, with a 95 % confidence interval (CI) equal to $[-27; -14]$. Likewise, if we assume the absence of compositional bias after matching, declaring one’s timetable via the internet would reduce the total journey time recorded over the day by 32 minutes (95 % CI $[-43; -20]$).

Table 3: Estimates of the measurement effect of internet data collection over different activity durations

	Crossover Constant effect	Matching Mean effect
Personal and physiological time (PPT)	-1 [-20; 18]	-5 [-36; 25]
Working and studying	-8 [-24; 8]	-22 [-50; 7]
Housework and recreational activities	-7 [-16; 3]	6 [-11; 23]
Care	4 [0; 8]	4 [-4; 11]
Sociability	25 [17; 33]	27 [14; 40]
Leisure	-51 [-64; -38]	-45 [-67; -24]
Travelling	-20 [-27; -14]	-32 [-43; -20]
Total aggregate duration	-53 [-73; -34]	-67 [-102; -32]
Time period covered	-27 [-44; -11]	-44 [-73; -14]

Note: The crossover test estimates are based on the durations declared by 866 people who completed at least one diary, standard errors robust to serial correlation and heteroscedasticity. [General Feasible Generalized Least Squares Analysis Wooldridge, 2002, chapitre 10]. The matching analysis is based on 311 pairs. As expected, the variance of the estimates is lower in the crossover analysis and leads to narrower confidence intervals. Statistically significant effects at 95 % in grey.

Generally speaking, whatever the duration considered, the two methods produce similar results¹¹. Assuming the absence of endogenous selection, our results show that responding via the Internet would have a significant impact on the reported duration of socialising, leisure activities and travelling (and the total duration of activities). Furthermore, with the internet questionnaire, the time schedules would not be completed entirely, so that the period covered would be less than that which would have been observed if the respondent had used a paper questionnaire.

However, are these consistent results relevant? If non-response is endogenous or if, after matching, individuals within the same pair differ on an unobserved characteristic, the estimates obtained with these two approaches are both biased. As a first step, it is therefore necessary to discuss the possible existence of an endogenous selection bias for the crossover analysis or a composition bias (resulting from a possible endogenous selection) for the matching approach. If necessary, this will justify carrying out a sensitivity analysis on our conclusions.

¹⁰as a reminder, we also assume the absence of a *carryover* effect or the interaction of a period effect with a collection mode effect.

¹¹With, as expected, a smaller variance of estimators in the crossover analysis, leading to smaller confidence intervals.

2.3 Is an endogenous selection bias possible?

Before trying to highlight a possible endogenous or compositional selection bias, it may be useful to illustrate the problem(s) we are facing with an example. To do this, let's consider two students of the same age, the same nationality, living in the same type of household, etc. The paper respondent reports a longer sleep duration than the internet respondent. Several explanations are possible.

- First, the internet respondent may simply have slept badly: the observed difference in duration is then due to chance, i.e. to an unobserved characteristic not correlated with the collection method and with participation.
- It may also be a measurement effect of the internet collection: the two respondents are strictly identical, but the internet respondent did not detail his entire schedule, for example, the time slot from 1 a.m. to 4 a.m. because he did not connect.
- Finally, one of the two may have a “student job” in the evening which in fact reduces their sleep duration: the difference is this time due to an unobserved characteristic correlated with the measured quantity. This unobserved characteristic may also be correlated with the collection mode and participation, for example if students who agreed to participate with the paper diary less frequently have a job on the side than respondents using the digital diary.

In this last case, comparing the responses of the two students after matching does not allow us to highlight a causal effect of the collection mode on the declared sleep duration. But it should be emphasized that this is not possible in the crossover-trial approach, even though in this type of test, the respondents are “their own control”. Indeed, by integrating the responses given by the respondents who only completed their schedule with a single collection mode, non-participation is no longer independent of the quantity studied.

Therefore, whatever approach is taken, the difficulty for the analysis is that we do not know what situation is reflected by the data. It may be considered to take advantage of a question supposedly not impacted by a measurement effect to highlight the existence of a possible composition bias (after matching) or selection bias (for the crossover trial).

A response assumed to be unaffected by the collection mode

If the schedule detailed by the respondent with the internet questionnaire does not cover the expected 24-hour period, it is consistent to anticipate that the durations of some activities will be lower than those that would have been observed with the paper questionnaire. This consistency refers to the possibility that the collection mode affects different responses in a known direction.¹² Similarly, it is possible that we expect that the mode of collection will affect the response to one question but not another, and we would like to exploit the anticipated lack of effect to provide information about the possible existence of endogenous (or hidden compositional) selection bias.

To be useful, a response not affected by the collection method must be correlated with an unobserved characteristic. In the Time Use test survey, the respondent is asked, at the end of the paper and digital diaries, for their estimate of the total duration of journeys made during the detailed day. This duration, which does not result from an aggregation of the declared travel times, excludes an impact of the *design* of the digital questionnaire, of the post-collection coding of the activities of the paper questionnaire, of the non-completion of

¹²This consistency can be exploited to increase the insensitivity of conclusions to the presence of endogenous selection bias in a sensitivity analysis (which formally requires taking into account the multiplicity of tests from the same database, see [Rosenbaum \[2002a\]](#), §9, and [Rosenbaum \[2020\]](#), §18).

the entire diary, etc. It may, in this sense, be possible to assume that the response to this question is not impacted by the collection mode used.

However, the total duration of travels made during the day is probably strongly correlated with a variable of interest in our analysis: the duration of travels calculated by aggregating the corresponding times declared in the completed diaries. It is also plausibly associated with unobserved characteristics such as the existence of a job (for students), possibly also very far from home. Therefore, the total duration of travels made during the day declared by the respondent can be used to distinguish a measurement effect of the collection method used (i) from an endogenous selection bias in the crossover-trial approach and (ii) from a hidden composition bias in the matching approach.

It can be demonstrated that an unaffected response can provide a consistent and unbiased test¹³ of the presence of an unobserved characteristic with which it is correlated (see Rosenbaum [1989a,b] and Rosenbaum [2002a], §6). It should be kept in mind, however, that a bias on an unaffected response can only reflect part of a larger difference on an unobserved characteristic. More precisely, if such a bias results from differences on an unobserved characteristic, then these differences are always at least as large as the bias itself. [see Rosenbaum, 1989a, for a more formal discussion]¹⁴.

Participation correlated with the importance of travels made during the day

Table 4 presents the results obtained by the crossover-test analysis and by the matching analysis if we consider the total travel time during the day. Regardless of the methodological approach chosen, and in a consistent magnitude, the total travel time during the day reported at the end of the questionnaire is significantly lower for digital diaries compared to paper diaries. In other words, online respondents and paper respondents differ in terms of the time spent on travel on the day of the survey, while it is unlikely that this difference is caused by a measurement effect of the collection mode.

¹³It may be useful to recall here what is meant by consistent and unbiased testing, notions which are perhaps less well known than their counterparts for an estimator. A test of a null hypothesis H_0 against an alternative hypothesis H_1 is *consistent* if the test “works” when the sample sizes are sufficiently large. A test is *unbiased* if it is “oriented in the right direction” regardless of the sample size. For a test to be a 5 % significance level test of H_0 , the chance that the associated P-value is less than 0.05 must be at most 5 % when H_0 is true. The *power* of a test of the null hypothesis, H_0 , against an alternative hypothesis, H_1 , is the probability that H_0 is rejected, when H_1 is true. If we conduct a test at the 5 % threshold, then the power of the test is the probability that the P-value is less than or equal to 0.05 when H_0 is false and H_1 is true. It is therefore appropriate to have a test with the highest possible power. From this point of view - and therefore more formally - a test is *consistent* against the alternative hypothesis H_1 if its power tends to 1 when the sample size increases, that is to say that the rejection of H_0 in favour of H_1 is almost certain if H_1 is true and the sample size is large. A test is an *unbiased* test of H_0 against H_1 if the power is at least equal to the significance level when H_1 is true. At the 5 % significance level, a test is unbiased against H_1 if the power of the test is at least 5 % when H_1 is true.

¹⁴This explains that “controlling” the response not affected by the collection method cannot exclude the existence of a composition bias.

Table 4: Is an endogenous or compositional selection bias possible?

	Crossover Constant effect	Matching Mean effect
Total travel time declared	−9 [−3;−15]	−15 [−25;−5]

Note: The crossover test estimates are based on the durations declared by 866 people who completed at least one diary, standard errors robust to serial correlation and heteroscedasticity. [General Feasible Generalized Least Squares Analysis Wooldridge, 2002, chapitre 10]. The matching analysis is based on 311 pairs. As expected, the variance of the estimates is lower in the crossover analysis and leads to narrower confidence intervals. Statistically significant effects at 95 % in grey.

This result suggests the existence of an endogenous selection bias. It is therefore necessary to question the sensitivity of the conclusions obtained. For this, we implement a sensitivity analysis to the results of the matching approach.

Some hypotheses/possible explanations for the hidden selection/composition bias:

- Paper respondents travel more during the day than internet respondents, for example for social activities outside their home (cf. significant link between “sociability” and collection mode, table 3),
- the notion of “journeys” in this question is understood differently by paper respondents and internet respondents. Does the internet application explicitly ask to distinguish between travelling and walking? Our measure of aggregated travel time does not include walks outside the home, travels during work, outings with a pet.
- Paper respondents do not have the same definition of “travel”, including for example times for walks, strolls, etc. unlike internet respondents. → what about the fact that the declared journey time remains lower than the aggregated journey time (coding problem linked to errors?)

3 Sensitivity analysis on matched pairs

In a random experiment, the probability of a unit in the sample being surveyed by Internet is the same as that of being surveyed *via* the paper questionnaire. Consequently, apart from by chance, no type of person is over-represented in the group surveyed by internet or the group surveyed *via* the paper questionnaire. If the observed difference in responses is too large to be due to chance, i.e. too large to be plausibly attributed to an unlucky sequence of coin flips, then the existence of a measurement effect is demonstrated. In our study, respondents are randomly assigned to a first collection mode, so that the protocol resembles a random experiment. However, as not all people responded, those who did respond in one or other group may be different. Respondents may differ on observable characteristics, but it is difficult to rule out the possibility that they do not also differ on a characteristic that cannot be measured, such as having a student job. In this case, comparing their declared length of working time would lead us to conclude that there is a measurement effect, even in the absence of an effect of the mode of collection on the response given.

In what follows, we will refer to the possible existence of such an unobserved characteristic as a hidden composition bias¹⁵.

¹⁵The proposed example is similar to a selection effect on an unobservable in survey theory. We do not

The sensitivity analysis model proposed by Rosenbaum and implemented in this study provides a partial response to this possibility. The aim is to estimate how “large” the unmeasured characteristic should be in order to modify the conclusions drawn regarding the existence of a measurement effect, assuming the absence of hidden bias. In this section, we first recall in a unified framework - for continuous and binary variables - how random inference is performed in a pairwise randomised experiment (section 3.1). In this respect, this section completes the presentation for the single case of binary variables presented by Court and Quantin [2024]. We then describe, within this unified framework, how to carry out a sensitivity analysis (section 3.2). In the third sub-section we return to the interpretation of the deviation from the hypothesis of the absence of hidden composition bias in the case of the analysis of pairs of respondents. In particular, we detail the link between the sensitivity analysis parameter, introduced in the previous sub-section, and the correlation between the unobserved characteristic and the collection mode on the one hand, and the correlation between the unobserved characteristic and the response given on the other (section 3.3). The aim of this sub-section is to make our results easier to understand for readers more familiar with the econometric approach to evaluation, which usually questions the presence of an unobserved characteristic in terms of its correlation with the treatment and the variable under study. Finally, we present the notion of effects attributable to the treatment in the case of a binary variable of interest, which will enable us to propose another quantification of the measurement effect of the collection mode (section 3.4).

As we will see in the presentation of the results, this quantification sheds light on the magnitude of the measurement effect in addition to, and sometimes more explicitly than, an average effect. Nevertheless, the interested reader will find in Appendix D a theoretical presentation of the Hodges-Lehmann estimators (within the framework of the Wilcoxon’s signed rank statistic) necessary for estimating an additive effect, including within the framework of sensitivity analysis. These various parts are highly theoretical; a non-formalised and more intuitive summary of the approach is therefore proposed in section 3.5.

All these sections draw heavily on the comprehensive presentations of books and articles by Rosenbaum [2002a], Rosenbaum and Silber [2009], Rosenbaum [2010].

Notations, treatment effect and treatment assignment

In this study, we consider I pairs, $i = 1, \dots, I$, each with two respondents $j = 1, 2$, one responding via the internet $Z_{ij} = 1$ (called “treated”), the other one responding online $Z_{ij} = 0$ (called “control”), matched on observed characteristics \mathbf{x}_{ij} , so that $\mathbf{x}_{i1} = \mathbf{x}_{i2}$ et $Z_{i1} + Z_{i2} = 1$, for each pair i . However, within each pair, individuals may differ on an unobserved characteristic u_{ij} , so that $u_{i1} \neq u_{i2}$. Using Neyman [1923] and Rubin [1974] notations, each surveyed person has two potential responses, the one r_{Tij} if he or she responds online and r_{Cij} if he or she responds *via* the paper diary. So, the observed response for individual j in pair i is $R_{ij} = Z_{ij}r_{Tij} + (1 - Z_{ij})r_{Cij}$. Treatment effect $r_{Tij} - r_{Cij}$ corresponds to the measurement effect of the alternative collection mode (internet) compared to the reference collection mode (paper). It is not observed for any of the ij units surveyed, mainly because our analysis in this section is restricted to responses

use this terminology, however, because our problem does not concern the representativeness of respondents in relation to the survey population, but rather the relevance of the comparison between similar respondents from the point of view of observed characteristics alone. Nevertheless, it is clear that the existence of a selection effect on unobservables may imply a hidden composition bias between the two groups of respondents. Furthermore, as we shall see, the inference will not be based on assumptions of distributional independence. This is why we have not retained the term “violation of the independence assumption conditional on observables”, as classically evoked in evaluation econometrics. Schematically, however, these two notions are of course very similar.

given during the first period only.

We note $\mathbf{R} = (R_{11}, \dots, R_{2I})^T$, $\mathbf{Z} = (Z_{11}, \dots, Z_{2I})^T$, $\mathbf{r}_T = (r_{T11}, \dots, r_{T2I})^T$, $\mathbf{r}_C = (r_{C11}, \dots, r_{C2I})^T$ and $\mathbf{u} = (u_{11}, \dots, u_{2I})$ vectors of dimension $2I$ associated with the above quantities. The sharp null hypothesis of no treatment effect from Fisher [1935] states that the responses of each respondent ij is unchanged if they respond by internet, i.e. $H_0 : r_{Tij} = r_{Cij}, \forall i, j$ or $H_0 : \mathbf{r}_C = \mathbf{r}_T$. If H_0 is true, then $\mathbf{R} = \mathbf{r}_C$.

We note $\mathcal{F} = \{(r_{Tij}, r_{Cij}, \mathbf{x}_{ij}, u_{ij}), i = 1, \dots, I, j = 1, 2\}$ and \mathcal{Z} the set of 2^I possible assignments to the treatment \mathbf{Z} so that $\mathbf{z} \in \mathcal{Z}$ if $z_{ij} = 0$ or 1 and $z_{i1} + z_{i2} = 1$ for each i . Finally, the number of elements in a set S is denoted by $|S|$; so $|\mathcal{Z}| = 2^I$.

3.1 Random inference in randomised experiments

In a randomised pair-wise experiment, random assignment to the treatment ensures that $\forall i, \Pr(Z_{ij} = 1 \mid \mathcal{F}, \mathcal{Z}) = 1/2$ and $\Pr(\mathbf{Z} = \mathbf{z} \mid \mathcal{F}, \mathcal{Z}) = 1/|\mathcal{Z}| = 2^{-I}$ for all $\mathbf{z} \in \mathcal{Z}$. As the Fisher [1935] formula states, random assignment is the basic element of inference in such a randomised experiment, in the sense that the distribution of any test statistic, $t(\mathbf{Z}, \mathbf{r})$, under the null hypothesis H_0 , is its permutation distribution:

$$\Pr(t(\mathbf{Z}, \mathbf{R}) \geq k \mid \mathcal{F}, \mathcal{Z}) = \Pr(t(\mathbf{Z}, \mathbf{r}_C) \geq k \mid \mathcal{F}, \mathcal{Z}) = \frac{|\{\mathbf{z} \in \mathcal{Z} : t(\mathbf{Z}, \mathbf{R}) \geq k\}|}{2^I} \quad (2)$$

Because, under H_0 , $\mathbf{R} = \mathbf{r}_C$ is fixed conditional on \mathcal{F} and on \mathbf{Z} being evenly distributed on \mathcal{Z} . Thus, in randomised inference, in the case of a randomised experiment, it is knowledge of the process of random assignment to the treatment that ensures that the distribution of the test statistic is known, which makes it possible to test, in this context, the null hypothesis of no effect of the treatment.

In practice, however, as the number of pairs increases, the size of \mathcal{Z} increases and makes it difficult to calculate (2) directly. In this case, it is usual to use an asymptotic approximation of (2). It is also possible to calculate exactly (i) by taking advantage of the simplification specific to the statistic used (see Rosenbaum [2002a] and Court and Quantin [2024] for example for the McNemar's test) or (ii) without using the whole \mathcal{Z} set from the characteristic function of the test statistic [Pagano and Trichler, 1983].

In the following, we illustrate how to implement an asymptotic approach for several test statistics commonly used in matched pair analysis¹⁶. Despite their variety, it is indeed possible to present this approach within a unified framework. More specifically, we discuss the McNemar's test, the Wilcoxon's signed rank statistic, Stephenson's statistics, Rosenbaum's U-statistics and Hubert-Maritz's M-statistics.¹⁷ While these statistics differ according to the nature of the variable of interest (binary, discrete or continuous), the associated tests have different levels of power.¹⁸ particularly when only some of the treated units respond to the treatment. Consequently, the choice of one test statistic rather than another can have an impact on the robustness of the conclusions. As we will see in the results, we will in fact use several test statistics to discuss the robustness of the conclusions in the context of sensitivity analysis for continuous variables.

We note $V_i = Z_{i1} - Z_{i2}$, $Y_i = R_{i1} - R_{i2}$, $Y_{Ci} = R_{Ci1} - R_{Ci2}$, and we consider $q_i \geq 0$ a function of the absolute value of the difference in responses within a pair, $|Y_i|$, such as

¹⁶The interested reader can find out how to calculate exact distributions for a small number of pairs in the reference articles.

¹⁷Find in Appendix C a short description of U-statistics of Rosenbaum (and their link with Stephenson statistics, Stephenson [1981]) and M-statistics [Huber, 1964, Maritz, 1979, 1981]

¹⁸and in particular from the point of view of sensitivity analysis [Rosenbaum, 2010].

$q_i = 0$ if $|Y_i| = 0$. With this notation, $V_i Y_i = (Z_{i1} - Z_{i2})(R_{i1} - R_{i2})$ corresponds to the difference in response between the processed unit and the control unit within each pair, and under H_0 , $Y_i = Y_{Ci}$.

Most statistics in matched pair analysis are of the form $t(\mathbf{Z}, \mathbf{R}) = \sum_{i=1}^I \text{sign}(V_i Y_i) q_i$ or a linear combination of such a statistic, with $\text{sign}(w)$ being 1, 0, or -1 respectively if $y > 0$, $y = 0$ and $y < 0$. Then, the Wilcoxon's signed rank statistic can be deduced by assuming that q_i is equal to the rank of $|Y_i|$ if $|Y_i| > 0$ [see for instance [Rosenbaum, 2010](#), §2.3.3]¹⁹. In the special case where R_{ij} is a binary variable, Y_1 is 1, 0 or -1 ; considerate $q_i = |Y_i| = 0$ for matching pairs, $q_i = |Y_i| = 1$ for discordant pairs, allows us to use the McNemar's test [see for instance [Rosenbaum, 2002a](#), §2.4.3].

For all these statistics, under the null hypothesis H_0 , $Y_i = Y_{Ci} = r_{Ci1} - r_{Ci2}$ and q_i are fixed in (2), conditionally to \mathcal{F} , and in a randomised experiment, V_i is 1 or -1 with a probability of $1/2$ for each value. And so, the distribution of $T = t(\mathbf{Z}, \mathbf{R}) = \sum_{i=1}^I q_i \text{sign}(V_i Y_i)$ in (2) is that of the sum of I discrete independent random variables, $\text{sign}(V_i Z_i) q_i$ which take the values q_i or 0 with equal probabilities, and $E(T | \mathcal{F}, \mathcal{Z}) = \sum q_i / 2$ and $\text{var}(T | \mathcal{F}, \mathcal{Z}) = \sum q_i^2 / 4$.

3.2 Sensitivity analysis model

The previous distribution was deduced from the random assignment to treatment that takes place in a randomised experiment. However, in our study, it is the surveyed units that are randomly assigned to a collection mode, but not the respondent units. In our study, after matching on observed characteristics \mathbf{x}_{ij} , the probabilities of being treated $\pi_{ij} = \Pr(Z_{ij} = 1 | \mathcal{F})$ may differ within a pair $\pi_{i1} \neq \pi_{i2}$, precisely because two similar people may differ on an unobserved characteristic $u_{i1} \neq u_{i2}$ so that after matching, $\Pr(Z_{ij} = 1 | \mathcal{F}, \mathcal{Z}) \neq 1/2$.

Sensitivity analysis model from [[Rosenbaum, 2002a, 2010](#)] considers “deviations” of various magnitudes from a random allocation of the treatment - the exact meaning of which is explained below - and their impact on the inference of a possible treatment effect. Specifically, the sensitivity analysis model assumes that the surveyed units $i1$ and $i2$ have unknown probabilities $\pi_{ij} = \Pr(Z_{ij} | \mathcal{F})$ but such that for two units with the same observed characteristics $\mathbf{x}_{i1} = \mathbf{x}_{i2}$ they may differ in their odds ratios of responding by Internet by a factor of at most $\Gamma \geq 1$, i.e. formally:

$$\frac{1}{\Gamma} \leq \frac{\pi_{i1}/(1 - \pi_{i1})}{\pi_{i2}/(1 - \pi_{i2})} \leq \Gamma \quad \text{when} \quad \mathbf{x}_{i1} = \mathbf{x}_{i2} \quad (3)$$

for $i = 1, \dots, I$.

Furthermore, the model restricts the distribution of \mathbf{Z} to \mathcal{Z} by conditioning on $Z_{i1} + Z_{i2} = 1, \forall i$; i.e. by integrating the pairwise structure [see [Rosenbaum, 2002a](#), §4].

As [Rosenbaum \[2002a, 2020\]](#) shows, this modelisation is similar if we state

$$\Pr(\mathbf{Z} = \mathbf{z} | \mathcal{F}, \mathcal{Z}) = \frac{\exp(\gamma \mathbf{z}^T \mathbf{u})}{\sum_{\mathbf{b} \in \Omega} \exp(\gamma \mathbf{b}^T \mathbf{u})} = \prod_{i=1}^I \frac{\exp(\gamma z_{i1} u_{i1} + \gamma z_{i2} u_{i2})}{\exp(\gamma u_{i1}) + \exp(\gamma u_{i2})}, \mathbf{u} \in [0, 1]^{2I} \quad (4)$$

¹⁹Il est possible par des considérations similaires d'étendre cette analogie aux statistiques de Stephenson [[Rosenbaum, 2007a, 2010](#), §16] et aux U-statistiques de Rosenbaum [[Rosenbaum, 2011](#)]. De même, si $\psi(\cdot)$ est une fonction impaire, $\psi(-y) = -\psi(y)$, alors $\sum_{i=1}^I \psi(V_i Y_i) = \sum_{i=1}^I \text{sign}(V_i Y_i) q_i$ avec $q_i = \psi(|Y_i|)$ permet de considérer les tests basés sur les M-statistiques [[Rosenbaum, 2007b, 2014](#)]. On trouvera en annexe C une description sommaire des U-statistiques de Rosenbaum (et leur lien avec les statistiques de Stephenson, [Stephenson \[1981\]](#))

where $\mathbf{z} \in \Omega$, and $\gamma = \log(\Gamma) \geq 0$. Despite its complexity, we note that I terms in (4), i.e. $\Pr(Z_{ij} \mid \mathcal{F}, \mathcal{Z}) = \exp(\gamma u_{ij}) / (\exp(\gamma u_{i1}) + \exp(\gamma u_{i2}))$ are bounded by $1/(1 + \Gamma)$ et $\Gamma/(1 + \Gamma)$ [Rosenbaum, 2002a, §4.2].

Meanwhile, when $\Gamma = 1$, or $\gamma = 0$, then $\pi_{i1} = \pi_{i2}$ in (3) and $\Pr(\mathbf{Z} = \mathbf{z} \mid \mathcal{F}, \mathcal{Z}) = 2^{-I}$ in (4). It means that when $\Gamma = 1$ or $\gamma = 0$, we find ourselves in the situation of a randomised experiment, i.e. with no hidden compositional bias. Consequently, Γ is a measure of the degree of deviation from a study with no hidden bias.

To test for the existence of a treatment effect, as we discussed in the previous section, we need to know the distribution of the test statistic T under the hypothesis H_0 . However, in the presence of a hidden composition bias, this distribution is not known, since the distribution of \mathbf{Z} is not known (in (4), for instance, values u_{i1} and u_{i2} are unknown). However, as Rosenbaum [2002a] shows, previous sensitivity analysis model allows us to deduce, at a fixed Γ , a bounding box for this unknown distribution of the test statistic T , by the distributions of two statistics, noted below T^+ and T^- , whose parameters are known and depend on Γ .

$$\Pr(T^+ \geq k) \geq \Pr(T \geq k \mid \mathcal{F}, \mathcal{Z}) \geq \Pr(T^- \geq k), \forall k \quad (5)$$

Usually, T^+ is the sum of I random independent variables, where the i^{th} variable equals a q_i value with the probability $\Gamma/(1 + \Gamma)$ and with the probability $1/(1 + \Gamma)$ and T^- the sum of I random independent variables defined in a similar way by reversing the roles of $\Gamma/(1 + \Gamma)$ and $1/(1 + \Gamma)$; the q_i value depends on the test statistic used.

When $I \rightarrow \infty$, the probability $\Pr(T^+ \geq k)$ can be approximated using a Normal approximation of the distribution of T^+ with $E(T^+) = \frac{\Gamma}{1+\Gamma} \sum_{i=1}^I q_i$ and $\text{var}(T^+) = \frac{\Gamma}{(1+\Gamma)^2} \sum_{i=1}^I q_i^2$, and a similar approximation for T^- ²⁰.

This bounding of the distribution, under the null hypothesis, of the test statistic T makes it possible to obtain bounds on the p -value associated with a value k of the test statistic, at a given Γ . These therefore define an interval of possible values for the p -value²¹ which reflects the uncertainty about our conclusions linked to a possible hidden composition bias, the magnitude of which is characterised by Γ . By varying Γ , the range of possible values of the p -values increases. There is therefore a Γ_{max} value for the magnitude of the hidden composition bias, beyond which the hypothesis of no measurement effect cannot be rejected. Discussing the possibility, after matching, of such a magnitude Γ_{max} of the bias makes it possible to argue about the relevance or otherwise of the existence of a treatment effect.

The sensitivity analysis proper will therefore consist of considering several values of Γ and studying how the inferences change. The conclusions will be considered sensitive to the existence of a hidden bias if values of Γ close to 1 lead to inferences that are very different from those obtained by assuming that the study is free of hidden bias. Otherwise, our conclusions will be considered insensitive if high values of Γ are required to alter them (the higher the Γ required, the lower the sensitivity).

²⁰However, as Rosenbaum [2002a] reminds us, such an approximation requires that the number of discordant pairs increases as I increases, which is the case for the statistics presented in the next section.

²¹When $\Gamma = 1$ (which characterises the absence of hidden compositional bias), the two statistics have an identical distribution and the interval is reduced to one value.

3.3 Interpreting the deviation from the assumption of no hidden compositional bias

The general expression (3) of the sensitivity analysis model quantifies the amplitude of the hidden composition bias considered, Γ , as an odds ratio of the probabilities of being treated by the two matched units. To simplify interpretation, the amplitude of the hidden composition bias can be expressed as an interval of possible values of the probability $\theta_i = P(Z_{i1} = 1, Z_{i2} = 0 \mid \mathcal{F}, \mathcal{Z}, Z_{i1} + Z_{i2} = 1)$ of observing, within a pair, the surveyed unit 1 (for example) responding by internet. In the absence of hidden composition bias, $\theta_i = 1/2$. In the presence of hidden composition bias, θ_i can be bounded by :

$$\frac{1}{1 + \Gamma} \leq \theta_i \leq \frac{\Gamma}{1 + \Gamma} \quad (6)$$

$\Gamma = 1$ means $\theta_i = 1/2$. When $\Gamma = 1.5$, θ_i is bounded by 0.40 and 0.60, i.e. a small deviation from a situation that would be similar to a random assignment *after matching*, as opposed to $\Gamma = 3$, since θ_i would then be bounded by 0.25 and 0.75.

Amplification of a sensitivity analysis

In the sensitivity analysis, the parameter Γ describes the magnitude of the association between u_{ij} and Z_{ij} and the associated sensitivity analysis is used to determine the range of possible inferences. The limits of (5) are reached for an unobserved characteristic u_{ij} that is highly correlated, or even perfectly correlated, with potential responses r_{Cij} [see Rosenbaum, 2002a, and below].

However, the interpretation based on such a covariate is not necessarily the most appropriate if a quasi-perfect correlation between u_{ij} and r_{Cij} is unlikely.

An amplification of the [Rosenbaum and Silber, 2009] sensitivity analysis offers a different but equivalent interpretation of the Γ parameter, based on two parameters:

- Λ which controls the magnitude of the association between the unobserved characteristic u_{ij} and Z_{ij} ,
- Δ which controls the magnitude of the association between u_{ij} and r_{Cij} , in the same way as traditional econometric evaluation methods which question the impact on results of the existence of an unobserved characteristic correlated with the treatment and the outcome variable.

Indeed, because u_{ij} is correlated with Z_{ij} and r_{Cij} , it implies association between Z_{ij} and r_{ij} in the absence of adjustment on u_{ij} . More precisely, $\Lambda \geq 1$ defines the extent of the association between $Z_{i1} - Z_{i2}$ and (u_{i1}, u_{i2}) : it states $\Pr(Z_{i1} = 1 \mid \mathbf{x}_{ij}, u_{ij}, \mathcal{Z})$ is at least $1/(1 + \Lambda)$ and at most $\Lambda/(1 + \Lambda)$ ²². Likewise, $\Delta \geq 1$ represents the extent of the association between $Y_{Ci} = R_{i1} - R_{i2} = r_{Ci1} - r_{Ci2}$ under H_0 and $(u_{i1} - u_{i2})$: it states $\Pr(Y_{Ci} > 0 \mid \mathbf{x}_{ij}, u_{ij}, \mathcal{Z})$ is at least $1/(1 + \Delta)$ and at most $\Delta/(1 + \Delta)$.

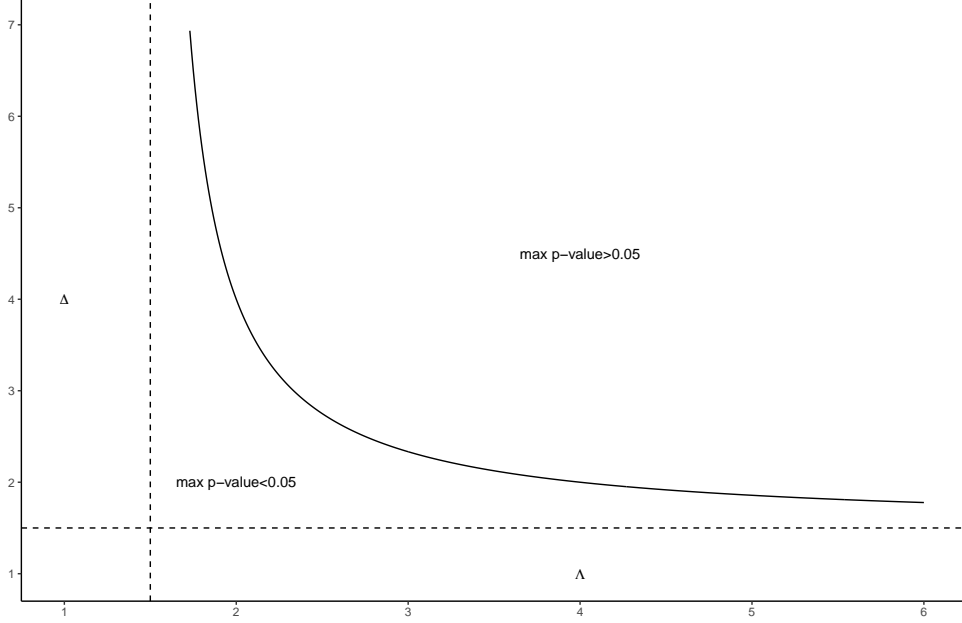
Formally, in the interpretation of a sensitivity analysis, a bias of magnitude Γ is equivalent to a sensitivity analysis model defined from (Λ, Δ) as soon as :

$$\Gamma = \frac{(\Lambda\Delta + 1)}{(\Lambda + \Delta)} \quad (7)$$

²² Γ and Λ are close, but different : Γ refers to $\pi_{ij} = \Pr(Z_{ij} \mid r_{Tij}, r_{Cij}, \mathbf{x}_{ij}, u_{ij})$, whereas Λ refers to $\Pr(Z_{ij} \mid \mathbf{x}_{ij}, u_{ij})$. So, Γ is based on u_{ij} and r_{Cij} , and implicitly on Y_{Ci} , whereas Λ is only based on u_{ij} and not on r_{Cij} i.e. without including Y_{Ci} .

So, an unobserved characteristic that would double ($\Lambda = 2$) the chances of being treated in a pair and quadruple ($\Delta = 4$) the chances of a positive difference between responses is identical to $\Gamma = 1.5$. But the same applies to any (Λ, Δ) values such as $1.5 = (\Lambda\Delta + 1)/(\Lambda + \Delta)$, for instance $(\Lambda, \Delta) = (4, 2)$ or $(\Lambda, \Delta) = (2.5, 2.75)$. Figure 1 represents (a part of) all possible (Λ, Δ) values for $\Gamma = 1.5$, and, by that, all (Λ, Δ) values for which the upper bound of the p-value of the equation (5) is less than 0.05 if $\Gamma_{\max} = 1.5$.

Figure 1: Amplification (Λ, Δ) for $\Gamma = 1.5$



Equation (7) defines the correspondence between a sensitivity analysis based on one parameter and one based on two parameters. As we mentioned earlier, it is interesting to note that a one-parameter analysis is also the limit of a two-parameter analysis, in the sense that $\Delta \rightarrow \infty$ dans (7) leads to $\Lambda \rightarrow \Gamma$. So, bounds of (5) can be read as the special case of a two-parameter sensitivity analysis model where the unobserved characteristic u_{ij} is quasi-perfectly correlated to Y_{Ci} .

Another important point of equation (7) is that it allows us to define which unobservable characteristic u_{ij} is relevant from the point of view of bias in random inference. Indeed, if $\Gamma = 1$, then the possible values of (Λ, Δ) are the curves $(1, \Delta)$ where $\Delta \in (0, \infty[$ and $(\Lambda, 1)$ where $\Lambda \in (0, \infty[$. It means a bias can exist only if u_{ij} is correlated simultaneously with Z_{ij} and r_{Cij} .

Finally, table 5 gives for different Γ , the range of possible values of θ_i and a couple of possible values for (Λ, Δ) .

Table 5: How to interpret Γ ?

Γ	Range of possible values of θ_i		Λ	Δ
1	0.50	0.50	1	1
1.1	0.48	0.52	1.40	1.80
1.25	0.44	0.56	2	2
1.5	0.40	0.60	2	4
2	0.33	0.67	3	5

Note : from [Rosenbaum, 2017a, chapter 9].

To sum up, Γ is a measure of the deviation from the situation where there is no hidden composition bias after matching. In the absence of hidden composition bias, the probability that the first person in a pair is treated is $1/2$. If there is a hidden compositional bias, $\Gamma = 2$ for instance, it means, this probability may be as low as $1/3$ or as high as $2/3$ (i.e., it is bounded by $1/(1 + \Gamma)$ and $\Gamma/(1 + \Gamma)$). It represents quite a gap from $1/2$. It is sometimes more useful to express the deviation from the assumption of no hidden bias as a function of the magnitude of the association between an unobserved characteristic and the treatment and variable of interest considered. So, for example, $\Gamma = 2$ is also identical to the existence of an unobserved characteristic which would increase the chances of being treated, within a pair, by a factor of 3 and those of observing a higher value of the variable of interest by a factor of 5.

3.4 Treatment attributable effects - a simplified presentation of the case of a binary variable of interest

The sensitivity analysis model described above makes it possible to question the impact of the existence of a hidden composition bias on the conclusion of a significance test of the p-value effect. In the case of continuous variables, as shown by Rosenbaum [2002a, 2010], it is also possible, using the same sensitivity analysis model, to estimate an interval of possible values for a treatment effect when the hypothesis of the absence of hidden composition bias is relaxed by a magnitude of Γ . This interval of possible values of the estimated effect reflects the uncertainty linked to the existence of an unobserved characteristic, and its amplitude depends on the value of the Γ parameter considered in the sensitivity analysis. To do this, the approach is based on Hodges-Lehmann estimators [Hodges and Lehmann, 1963, Lehmann, 1975], which invert the adjusted test statistic, i.e., for example in the case of a constant additive effect τ_0 , by considering $t(\mathbf{Z}, \mathbf{R} - \tau_0 \mathbf{Z})$. As an example, we formally describe the approach, with the Wilcoxon's signed rank statistic as the test statistic, in Appendix D.

This method could also test other hypotheses, such as a zero effect for certain treated units and a non-zero effect that differs two by two for the others. However, in addition to the multiplicity of hypotheses to be tested, as Rosenbaum [2010] points out, it is also difficult to understand the associated confidence intervals in a space that is no longer one-dimensional, and therefore to interpret the results obtained. To get around this difficulty, it is usual to estimate the average effect of the treatment on those treated. However, although the average effect of the treatment takes account of the heterogeneity of the effect, it provides only an imperfect description of it and is not very robust, particularly as the approach is based on the mean of the differences statistic, which is one of the least effective statistics. Rosenbaum [2002b] proposes instead to focus on so-called treatment-attributable effects which, as we shall see, offer a synthetic analysis of the existence of effects that could be attributed to the treatment, without making any assumptions about their magnitude at

the individual level. Before going into more detail about the methodology used, we will first explain the notion of an effect attributable to the treatment introduced by Rosenbaum [2001].

Let's consider as "success" the occurrence of a particular event, for example the fact that a surveyed unit did not complete its entire 24-hour schedule, i.e. a binary variable. The number of effects attributable to the treatment is the number of events among the treated units that would not have occurred if they had not responded via the internet. This estimate, as we shall see, can be made on the assumption that there is no hidden composition bias after matching ($\Gamma = 1$), but also by allowing this assumption to be relaxed ($\Gamma > 1$).

By definition, for each unit surveyed, the potential responses r_{Tij} and r_{Cij} therefore take the value 1 if the event occurs (i.e. in the event of "success") and 0 otherwise. In what follows, we assume that responding via the Internet can cause an event but cannot prevent it if the event would have occurred if the respondent had used the paper questionnaire, i.e. $r_{Tij} \geq r_{Cij}$ ²³. For example, responding via the internet may lead a respondent not to complete their entire diary, but it is assumed that this event would necessarily have occurred if the same had been the case when responding with the paper questionnaire. This assumption is necessary for the inference based on the distribution of \mathbf{Z} . As $r_{Tij} \geq r_{Cij}$, a person responding by internet who does not complete all of his or her diary ($r_{Tij} = 0$) would not have completed it in full either by responding to the paper questionnaire ($r_{Cij} = 0$). Likewise, if a person responding with paper complete his/her whole questionnaire ($r_{Cij} = 1$), he/she would have done the same online ($r_{Tij} = 1$). So, assumption such as $\delta = \delta_0$ is *compatible* with data if $\delta_{ij} = 0$ when ($Z_{ij} = 1$ and $R_{ij} = 0$) or ($Z_{ij} = 0$ and $R_{ij} = 1$) [see Rosenbaum, 2001, for the use of compatible assumptions], and *incompatible* otherwise. An incompatible hypothesis can be rejected with certainty, i.e. with a type 1 error of zero. If a hypothesis is true, then it is compatible for all \mathbf{Z} , whereas a false hypothesis may be compatible for some \mathbf{Z} and not for others. Since we are trying to test a null hypothesis based on the distribution of \mathbf{Z} , we need to be able to ensure that it corresponds to a compatible hypothesis.

We note $\delta_{ij} = r_{Tij} - r_{Cij}$ the mode effect and δ the associated vector of $2I$ dimension. The attributable effect of the mode is $A = \sum_{i,j} Z_{ij}(r_{Tij} - r_{Cij}) = \sum_{i,j} Z_{ij}\delta_{ij}$ ²⁴. Furthermore, when we note $T = \sum_{i,j} Z_{ij}R_{ij}$ the number of "successful" internet respondents, then $T - A = \sum_{i,j} Z_{ij}r_{Cij}$ corresponds to the number of people responding by internet who would have been successful even without responding by internet. In the following, r_{Ci+} denotes the number of people in pair i who would have known about the event even if they had not responded via the internet.

Consider the test of the null hypothesis $H_0 : \delta = \delta_0$ against the alternative $H_1 : \delta \geq \delta_0$.

If the null hypothesis is incompatible, then it is rejected with a type 1 error of zero. If the null hypothesis is compatible, then under $H_0 : \delta = \delta_0$, r_{Cij} is known for all i, j , because $r_{Cij} = R_{ij} - \delta_{0ij}Z_{ij}$. By noting $A_0 = \sum_{i,j} Z_{ij}\delta_{0ij}$, the number of attributable effect to the mode among units responding by internet associated with δ_0 , $T - A_0 = \sum_{i,j} Z_{ij}r_{Cij}$, i.e. $T - A_0 = \sum_i B_i$ is the sum of I random independent binary variables $B_i = \sum_j Z_{ij}r_{Cij}$.

²³Note that such a model cannot be verified or disproved because r_{Tij} and r_{Cij} are never observed simultaneously.

²⁴ A is a quantity that is not observed since r_{Tij} and r_{Cij} are not, and its value depends on the assignments to the Z_{ij} collection mode: it is therefore a random variable.

In absence of a hidden composition bias,

$$\Pr(B_i = 1) = \pi_i = \frac{r_{Ci+}}{2} \quad \text{avec } \Gamma = 1$$

In the presence of a hidden compositional bias, $\Pr(B_i = 1)$ is unknown but can be bounded by [Rosenbaum, 2002b] :

$$\bar{\pi}_i = \frac{\Gamma r_{Ci+}}{\Gamma r_{Ci+} + 2 - r_{Ci+}} \geq \Pr(B_i = 1) \geq \frac{r_{Ci+}}{r_{Ci+} + \Gamma(2 - r_{Ci+})} \quad (8)$$

Noting $\beta(k, \boldsymbol{\pi})$, the probability to have at least k success among I tries with $\boldsymbol{\pi} = (\pi_1, \dots, \pi_I)^T$, then $\Pr(\sum_i B_i \geq k) = \beta(k, \boldsymbol{\pi})$ and it corresponds to the significance level of the null hypothesis test for the statistic $k = T - A_0$.

And so, in absence of a hidden composition bias, the method to test $H_0 : \boldsymbol{\delta} = \boldsymbol{\delta}_0$, is to calculate $r_{Cij} = R_{ij} - Z_{ij}\delta_{0ij}$ under H_0 , then r_{Ci+} , to obtain $\pi_i = \bar{\pi}_i = \frac{r_{Ci+}}{2}$, then $\Pr(\sum_i B_i \geq k) = \beta(k, \boldsymbol{\pi})$.

In the presence of a hidden composition bias, i.e. if $\Gamma > 1$, as π_i is unknown due to an unobservable characteristic, $\Pr(\sum_i B_i \geq k) = \beta(k, \boldsymbol{\pi})$ cannot be calculated but it can be topped by $\beta(k, \bar{\boldsymbol{\pi}}) \geq \beta(k, \boldsymbol{\pi})$ according to (8) after determined r_{Ci+} .

In both cases, when $I \rightarrow \infty$, $\beta(k, \bar{\boldsymbol{\pi}})$ may be approximated²⁵ by

$$\beta(k, \bar{\boldsymbol{\pi}}) \rightarrow 1 - \Phi\left(\frac{k - \sum_i \bar{\pi}_i}{\sqrt{\sum_i \bar{\pi}_i(1 - \bar{\pi}_i)}}\right) = 1 - \Phi\left(\frac{T - A_0 - \sum_i \bar{\pi}_i}{\sqrt{\sum_i \bar{\pi}_i(1 - \bar{\pi}_i)}}\right) \quad (9)$$

General expression of $\bar{\pi}_i$ in (8) may be explained by considering under H_0 all possible values of r_{Ci+} . There are four possible types of pair, the characteristics of which are shown in the table 6 below. D and C indicates a discordant and concordant pair respectively. $D(+, -)$ represents a discordant pair where the treated unit experiences the event, but the control unit does not.

Table 6: Structure of concordant and discordant pairs under H_0

		Number				under H_0			
		Z_{i1}	Z_{i2}	R_{i1}	R_{i2}	r_{Ci1}	r_{Ci2}	r_{Ci+}	$\bar{\pi}_i$
$D(+, -)$	n_{D+}	1	0	1	0	1	0	1	$\Gamma/(\Gamma + 1)$
$D(-, +)$	n_{D-}	1	0	0	1	0	1	1	$\Gamma/(\Gamma + 1)$
$C(-, -)$	n_{C-}	1	0	0	0	0	0	0	0
$C(+, +)$	n_{C+}	1	0	1	1	1	1	2	1

Note : D and C represent discordant and concordant pairs, and $+$ and $-$ represents if the unit actually experience the event.

The number T of people responding via the internet who are “successful” corresponds to $n_{D+} + n_{C+}$. Suppose we want to test the hypothesis that there are A_0 effects attributable to the treatment among the n_{D+} discordant pairs $D(+, -)$, where the treated unit is the one experiencing the event (success). If we estimate $r_{Cij} = R_{ij} - Z_{ij}\delta_{0ij}$ under H_0 , there would be $r_{Ci+} = n_{D+} - A_0$ people in pair i who would have experience the event regardless of the collection mode. So, according to table 6, under H_0 , (i) $\bar{\pi}_i = 1$ for n_{C+} concordant

²⁵An exact expression of the test can be implemented from the binomial distribution.

pairs, (ii) $\bar{\pi}_i = 0$ for n_{C-} concordant pairs, and (iii) $\bar{\pi}_i = \Gamma/(1+\Gamma)$ for the $n_{D+} + n_{D-} - A_0$ still discordant pairs under H_0 . So, $\sum_i \bar{\pi}_i = n_{C+} + (n_{D+} + n_{D-} - A_0)\Gamma/(1+\Gamma)$ is the expectation of $T - A_0 = (n_{D+} + n_{C+} - A_0)$. Finally,

$$\frac{T - A_0 - \sum_i \bar{\pi}_i}{\sqrt{\sum_i \bar{\pi}_i(1 - \bar{\pi}_i)}} = \frac{n_{D+} - A_0 - (n_{D+} + n_{D-} - A_0)\Gamma/(1+\Gamma)}{\sqrt{(n_{D+} + n_{D-} - A_0)\Gamma/(1+\Gamma)^2}} \quad (10)$$

which allows us to calculate $\beta(k, \bar{\pi})$.

It is also possible to test that there are A_0 effects attributable to the collection method among the I , and not simply among the n_{D+} discordant pairs; we will not test these hypotheses in our study. Similarly, a confidence interval for the number of effects attributable to the collection method can be calculated: for all these questions (which are more complex to implement), we refer the interested reader to [Rosenbaum \[2002b\]](#). Nevertheless, we would like to stress that there is univocity in the interpretation of the test in the sense that two compatible hypotheses, δ_0 and $\bar{\delta}_0$, leading to the same number A_0 of treatment effects, have an analogous conclusion. More generally and intuitively, [Rosenbaum](#) demonstrates a “relation of order” between two hypotheses compatible with $\delta_0 - \bar{\delta}_0 \geq 0$, meaning δ_0 is greater than $\bar{\delta}_0$ because it attributes more effects to the collection mode. If $H_0 : \delta = \delta_0$ is rejected against $H_1 : \delta \geq \delta_0, \delta \neq \delta_0$ and if δ_0 is greater than $\bar{\delta}_0$ then $\bar{\delta}_0$ is also rejected.

3.5 Summary of sensitivity analysis

The sensitivity analysis model assumes that two people matched on certain observed characteristics may differ in their chances (*odds-ratio*) of being treated, within the pair they form, by a factor of no more than Γ .

Thus, $\Gamma = 1$ corresponds exactly to the assumption of no hidden compositional bias, i.e. to the random inferential approach [[Fisher, 1935](#)]. $\Gamma > 1$ means, conversely, that we do not really know the distribution of the probability of being treated, but that the deviation from random inference is limited in its extent by Γ . A high value of Γ thus allows a large deviation from the hypothesis of no hidden compositional bias.

As the distribution of the probability of being treated is not known when $\Gamma > 1$, the inference does not provide a p -value, but an interval of p -value. Similarly, it does not provide an estimate of the measurement effect of the collection mode, but an interval of possible values for the measurement effect of the collection mode, the length of this interval increasing as Γ increases. An inference is said to be sensitive to a bias of magnitude Γ when the interval qualitatively includes different inferences, for example both rejection and acceptance of the null hypothesis at the 5 % threshold. By varying Γ , the sensitivity analysis therefore makes it possible to question the extent of the hidden composition bias required to modify the conclusions obtained under this hypothesis, based on the data. Admittedly, this does not completely resolve the disagreement between the assertion that there is a measurement effect and that there is a bias due to an unobserved covariate. Nevertheless, it is one thing to say that a tiny, barely perceptible bias in the assignment of modes could explain an association, than to say that only a large bias could do so. That said, where possible, we will complete this analysis by attempting to provide a reasonable explanation for the measurement effect.

4 Results

We now detail the results of our analysis of the measurement effect on four activity durations: sleeping time, travel time, leisure time and meal duration. These durations were chosen

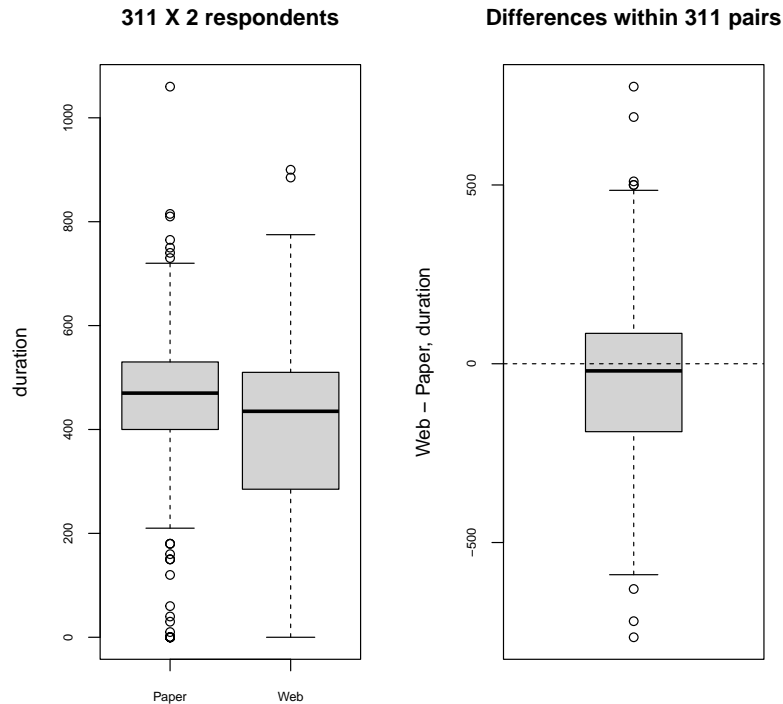
by the Household Living Conditions Division because they are the historical variables of interest in the Time Use Survey.

4.1 A detailed example of sensitivity analysis: sleeping time

Figure 2 represents the differences in sleeping time declared by the internet respondents and by the respondents with the paper diary in the 311 matched pairs.

In 32 % of our 311 pairs, the difference (internet minus paper) in sleeping time declared by the two respondents do not exceed 1 hour. However, the median difference is -20 minutes. Moreover, while in 25 % of pairs, Internet respondents declared 85 minutes *more* sleeping time than paper respondents, 25 % of Internet respondents declare 190 minutes *less* sleeping time than the people surveyed on paper to whom they were matched. Thus, the distribution is asymmetrical to the left and despite a peak fairly close to zero, large negative differences are more frequent than large positive differences. This indicates that when very low sleep durations are observed, they are more frequently reported by people responding by internet. As suggested by the distribution of differences in duration within the pairs in Figure 2, assuming no hidden composition bias, the mode of collection could therefore have a significant effect on the duration of sleep declared by some respondents, with perhaps a smaller effect or none at all for most other respondents.

Figure 2: Difference in sleeping times (internet - paper)



If $\Gamma = 1$, as explained in section 3.2, we assume the absence of hidden composition bias. Under this hypothesis, responding via the internet would lead respondents to declare nearly an hour less sleep on average (-53 min., CI 95 % [-79 ; -26], table 7). A similar result (-56 min. CI 95 % [-72 ; -40]) is obtained with the crossover analysis, i.e. assuming MCAR or MAR non-response.

Table 7: Sensitivity analysis on the mean effect of collection mode on sleeping time

Γ	p -value	Mean effect		I.C. 95 %	
	Upper lim.	Lower lim.	Upper lim.	Lower lim.	Upper lim.
1.0	<0.001	-53	-53	-79	-26
1.1	0.001	-61	-44	-88	-18
1.2	0.008	-69	-36	-96	-10
1.3	0.031	-76	-30	-104	-3
1.4	0.091	-83	-23	-112	4
1.5	0.210	-90	-17	-119	10

When $\Gamma = 1.2$, one assumes the existence of a hidden composition bias, of reduced magnitude: the probability of assignment to the internet collection mode within a pair of one of the two units is no longer 0.5, but lies between $(1 + \Gamma)^{-1}$ and $\Gamma/(1 + \Gamma)$, i.e. 0.45 and 0.55 (see Section 3.3). Under this hypothesis, it is still possible to conclude that there is a measurement effect of the Internet collection method on the reported sleep time, of between -69 min. and -36 min. (CI 95 % [-96;-10]). However, if $\Gamma \geq \Gamma_{max} = 1.4$, it is no longer possible to reject the hypothesis that there is no effect of measurement. Indeed, for example, an unobserved covariate which, after matching, would double the chances of responding online and triple the chances of a negative difference (cf. section 3.3) could explain the observed association.

Table 8: Sensitivity analysis (Γ_{max}) on total duration and number of sleeping slots

Statistics	total duration	number of slots
Mean	1.4	2.0
Wilcoxon	1.3	2.0
Stephenson (5,5,5)	1.5	1.9

Considering other test statistics (Wilcoxon’s signed rank statistic, Stephenson’s statistic, see Appendix C) in order to increase the power of the sensitivity analysis [Rosenbaum, 2002a, §15] does not alter the conclusions obtained (see table 8): the hypothesis of a measurement effect of the internet collection mode on sleeping time is rather sensitive to the existence of a hidden composition bias. If we consider the number of sleeping slots declared²⁶, the hypothesis of a measurement effect of the collection mode is less sensitive to the existence of a composition bias: whatever the test statistic used, Γ_{max} is close to 2. This means that the probability of assignment to the internet collection method within a pair of one of the two units would have to be less than 0.33 or greater than 0.67 (instead of 0.5) to explain the observed association. This result should come as no surprise; in fact, the hypothesis of a measurement effect of the collection mode on sleeping time is more sensitive to the greater heterogeneity of the declared durations, which could be explained in part by an unobserved covariate.

This observation leads us to consider quantifying the measurement effect as a number of “events” (effects) attributable to the data collection method, i.e. more precisely by considering as “an event due to the Internet data collection method” the fact of declaring less than 2 sleeping time slots (see section 3.4). In fact, the diaries must detail the respondent’s schedule over 24 hours from 4 am to 4 am the next day. It is therefore likely (and expected) that each person will declare at least 2 sleeping time slots, for example between 4 am and

²⁶as a continuous variable of interest

6 am the first day and then from 11 pm to 4 am the following day. However, in our 311 pairs, 28 % of internet respondents declared less than 2, compared with 14 % of paper respondents. Only a hidden composition bias of a similar magnitude to that for the number of sleeping slots ($\Gamma \geq \Gamma_{max} = 1.8$ by sensitivity analysis with the McNemar's statistic) could explain such an association.

More precisely, after matching, 96 of the 311 pairs (i.e. 31 %) are discordant: one of the two respondents declare fewer than two slots. In 70 pairs (i.e. 73 % of the discordant pairs), it is the Internet respondent who declare fewer than 2 slots. Of these 70 pairs, how many can be attributed to the collection method? In other words, how many of them would not have been discordant if the Internet respondent had used the paper diary? Before determining this number of effects attributable to the collection method, we illustrate the approach implemented as presented in section 3.4.

Suppose we test $H_0 : \delta = \delta_0$ corresponding to a number of effects attributable to the collection method $A_0 = 10$ among the discordant pairs. First, we suppose there is no hidden composition bias, i.e. $\Gamma = 1$. Then under H_0 , there are $70 - 10$ discordant pairs who would still have been discordant if the Internet respondent had used the digital diary ($n_{D+} - A_0$ to use the notation from section 3.4). As there are 26 discordant pairs ($n_{D-} = 96 - 70$) where the paper respondent is the one declaring less than 2 slots, we can write (see equation 10) :

$$\frac{T - A_0 - \sum \bar{\pi}_i}{\sqrt{\sum \bar{\pi}_i(1 - \bar{\pi}_i)}} = \frac{70 - 10 - (70 + 26 - 10) \times 1/2}{\sqrt{(70 + 26 - 10) \times (1/2) \times (1/2)}} = 3.66314$$

So, the null hypothesis of A_0 attributable effects to the mode is largely rejected at the 5 % threshold rejected ($3.66 > 1.65 = \Phi(0.95)$). By varying A_0 , we can state with 95 % confidence that, in the absence of hidden composition bias, 43 % of the 70 discordant pairs where the Internet respondent declared less than 2 sleeping slots are attributable to the internet collection mode. In other words, responding by Internet would have had an impact on the responses of 10 % of the 311 pairs.

Assuming the existence of a hidden composition bias of magnitude $\Gamma = 1.4$, with a similar approach, we could still be sure (at 95 %) that responding by Internet has an impact on 5 % of the 311 pairs. Finally, Γ would have to reach 1.9²⁷ so that we could not reject the hypothesis that none of the discordant pairs is attributable to a measurement effect.

The collection method therefore seems to have a significant effect on the sleeping time declared by some respondents, and a smaller effect, or even none at all, for most participants. One possible explanation could be that some respondents did not complete the entire logbook by internet, leading them to declare less than 2 sleeping slots over the 24 hours. Indeed, whatever the method of collection, 90 % of respondents began to detail their sleep schedule at 4 am. However, only 74 % of Internet respondents completed it by the requested end time (4 am the following day), compared with 87 % of respondents using the paper diary.

A sensitivity analysis on the probability of not completing the diary by the requested end time produces results similar to those presented above. After matching, of the 311 pairs, 105 (37 %) are discordant on their end time, and of these, 76 (72 %) are due to the Internet respondent. In the absence of hidden composition bias, we would be 95 % sure that the mode of collection caused at least 43 % of the discordant pairs (where the Internet respondent does not have the expected end time). This represents 14 % of the 311 pairs, a

²⁷i.e., after matching, an unobserved covariate which would triple the chances of responding by internet and increase by a factor of 5 those of declaring less than 2 sleeping slots could explain the results obtained.

result in line with the previous estimate. Finally, $\Gamma = 1.9$ would be required for no effect to be attributed to the mode of collection. Such similarities are not surprising. Indeed, in 70 % of the pairs where the number of sleeping slots declared in the digital diary is less than 2, the diary is not completed by the requested end time.

4.2 Impact on travel time

In the absence of hidden composition bias, respondents report on average 32 min (CI 95 % $[-43; -20]$) less travel time when responding on the internet (see Table 3). This measurement effect is not very sensitive to the hypothesis of a hidden composition bias since $\Gamma_{max} = 1.7$ ²⁸(see table 9). Is such a deviation from the random assignment hypothesis (needed to not reject the hypothesis of no measurement effect) plausible after our matching?

Table 9: Sensitivity analysis Γ_{max} on total travel time and number of travel slots

Statistics	total duration	number of travel slots
Mean	1.7	1.3
Wilcoxon	1.6	1.2
Stephenson (5,5,5)	2.0	1.2

In section 2.3 we showed that, assuming there was no measurement effect on the total time spent travelling during the day, as requested at the end of the questionnaire, it was not possible to reject the hypothesis of the existence of a hidden composition bias²⁹. Indeed, on average, Internet respondents report 15 minutes less (CI 95 % $[-25; -5]$) of total travel time over the course of a day than paper respondents (see Table 4). Thus part of the measurement effect³⁰ of Internet data collection on total travel time calculated by aggregating time slots would in fact correspond to a bias due to an unobserved characteristic³¹. Can we be more conclusive with an intra-individual comparison of these two outcomes of travel times?

To do this, let's first compare, for respondents with a paper questionnaire, the total travel time obtained by aggregating the times of the corresponding time slots in the questionnaire with the one declared at the end of the questionnaire, as shown in figure 3a. With the paper questionnaire, on average, the total travel time obtained by aggregation exceeded the one declared at the end of the questionnaire by 13 minutes (95% CI $[7; 19]$). This discrepancy could be due to a memory problem: questioned at the end of the survey, respondents could underestimate their total travel time for the day³². A second possible explanation is a problem in coding the activities declared on the paper questionnaire. Each respondent specifies the corresponding activity for each time period. An algorithm developed by INSEE then codes the activity according to its wording. This means, for example, that the activity

²⁸With other statistics, $\Gamma_{max} = 2.0$. Conversely, the hypothesis of a measurement effect on the number of travels is sensitive to a hidden composition bias.

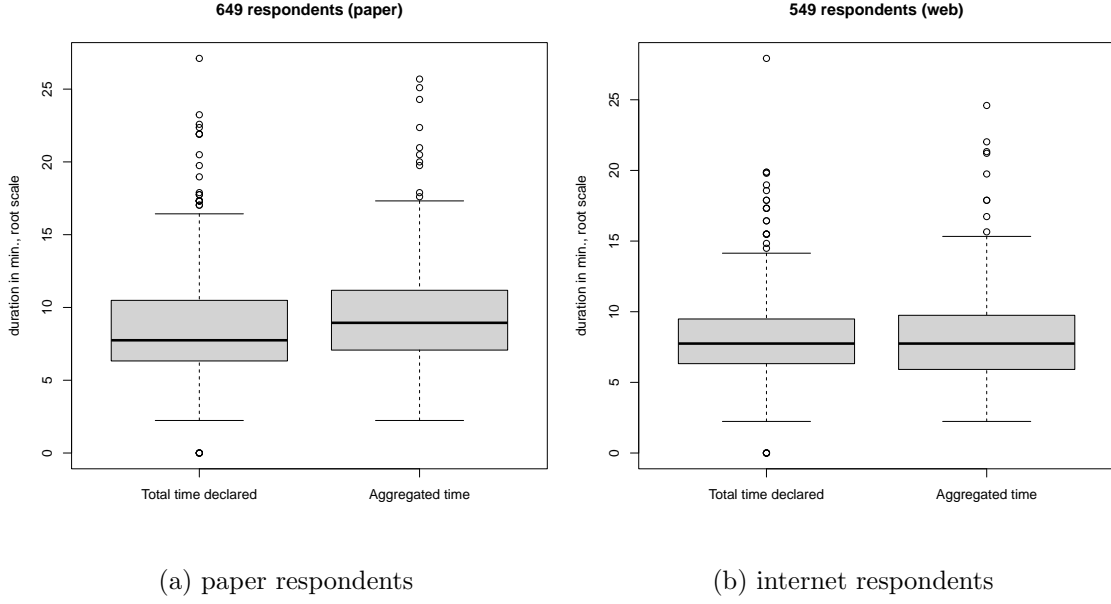
²⁹As a reminder, as explained, the approach implemented provides a consistent and unbiased test for the presence, after matching, of an unobserved characteristic correlated with total travel time.

³⁰estimated assuming the absence of compositional bias

³¹The minimum deviation, Γ_{max} with respect to random assignment, i.e. the magnitude of the composition bias to reject the hypothesis of an association between the collection mode and the total travel time over a day declared at the end of the questionnaire is $\Gamma_{max} = 1.1$. However, this magnitude is too small to reject the hypothesis of a measurement effect on total travel time calculated by aggregating the times for the corresponding time slots: this means that a measurement effect must always be considered plausible [see Rosenbaum, 2023, for a more precise statement and the theoretical considerations supporting this assertion].

³²The qualitative tests carried out prior to the test survey revealed that, whatever the data collection mode, respondents did not sum up the travel times once the notebooks had been completed in order to answer this question.

Figure 3: Aggregated and declared travel times for paper and internet respondents



“going out shopping” may not be split into two activities (the outward and return journey to get there and the time spent shopping) but counted as a single commuting activity.

To decide between these two explanations, we propose two analyses. Firstly, for respondents to the digital diary, we will compare the total travel time obtained by aggregating the durations of the corresponding time slots in the diary with the one declared at the end of the questionnaire. With the digital diary, the respondent first chooses whether the time period he enters corresponds to a journey. If not, a selector allows them to specify the type of activity (see Figure 11, Appendix F). To return to the previous example, this type of design of the digital notebook is more likely to lead the respondent to separate the time spent going to the supermarket from the time spent shopping. Consequently, the absence of any significant difference between the two travel times could indicate the absence of a memory problem. Secondly, we will consider the “total duration of time slots to be coded” - we will specify in the corresponding section what we mean by this - in the paper diaries as a measure of the intensity of the “treatment”. Intuitively, if the differences between the travel times within the pairs result from a coding problem, it is expected that they will be all the more marked as the total duration of the time slots to be coded in the paper diary is greater.

A memory problem?

Figure 3b compares the travel times declared at the end of the questionnaire for respondents using the digital diary with those obtained by aggregating the corresponding time slots in the diaries. While the difference is significant for paper respondents, as we have pointed out, it is not significant for internet respondents (Figure 3b). Thus, under the hypothesis that there is no difference in the composition of memory capacities between the two types of respondent, such a finding would tend to validate the hypothesis of a coding problem rather than that of cognitive differences.

A codification issue in the paper diary?

The second approach more explicitly questions the hypothesis of a coding problem. It is

based on a measure of the intensity of “treatment” by considering the duration of time slots corresponding to “other travels”, i.e. time slots in the paper diary which do not correspond to home-work travels but where the respondent mentioned, for example, having used a means of transport.

More specifically, with the paper diary, the respondents must specify for each time period the activity carried out, with a freely worded description (see Figure 12a, Appendix F). They must also specify the location of the activity or the means of transport used to carry it out, if applicable. They are also asked whether the time period corresponds to a home-work commute (see Figure 12b, Appendix F). As part of this test, once the paper diary had been collected, the interviewers were asked to specify for each time period and on the basis of all this information whether it corresponded to (i) an activity carried out at home, (ii) a home-work journey, (iii) an activity carried out in another place or (iv) whether it corresponded to another journey. As mentioned above, these elements are then used by a learning algorithm to code the activity. On the basis of this information, the algorithm must be able to distinguish between journeys other than the home-work journey to another location (such as dropping children off at school), which are included in the nomenclature under the category “journeys”, and walks outside the home (whether or not they are motorised), journeys during work or visits to friends’ homes³³, which, for example, are not to be listed as a route but associated with other activity categories in the nomenclature. Therefore, the hypothesis considered here is that the longer the time slots listed as “other journeys” (and therefore to be coded by the algorithm), the greater the impact of collecting data using the paper questionnaire on total travel times³⁴.

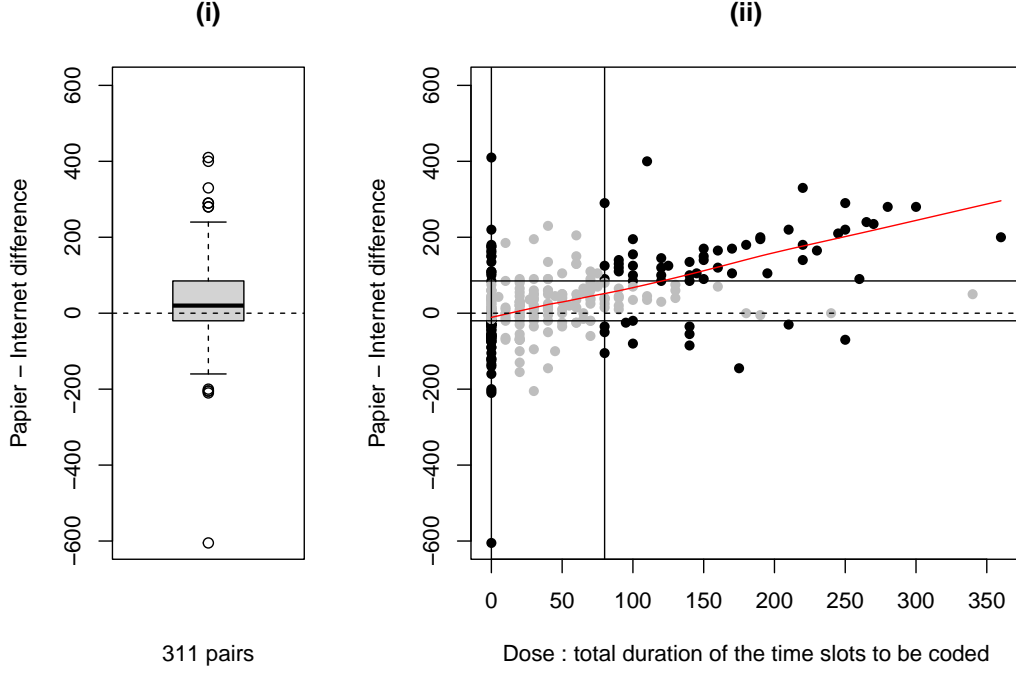
Two analyses of the measurement effect of paper collection

Figure 4 (i) shows the difference in travel times between the paper and digital diaries, within each pair, calculated by aggregating the duration of the corresponding time slots after coding the paper diary. As we mentioned at the beginning of this section, using the Wilcoxon’s signed rank statistic, random inference, i.e. in the absence of hidden composition bias, rejects the null hypothesis of no effect of paper data collection on travel times. However, we restate the approach and the results so that the reader can later make the connection with the approach that takes account of the intensity of the “treatment”.

³³which would only be entered for a single time period, with no distinction being made between the associated journey times.

³⁴calculated by aggregating the time slots identified as such

Figure 4: Differences in travel times (paper - internet) within 311 pairs



Note: Figure (i) shows the difference in travel times (paper - internet) within the 311 pairs. Figure (ii) represents these differences within the pairs, Y_i as a function of the total duration d_i of the time slots “other travel” in the paper respondent’s diary, contrasting in black those above or below the 1st and 3rd quartiles of each distribution. The red line corresponds to the associated *lowess* regression.

In the absence of hidden composition bias, as explained in section 3.1, the analysis is based on random inference, i.e. by considering a random assignment of respondents to a collection mode in each of the 311 pairs, for example by flipping a coin 311 times. There are 2^{311} possible assignments to a collection mode, $\mathbf{k}_1, \dots, \mathbf{k}_{2^{311}}$, whose distribution probability is $\mathbf{p} = (2^{-311}, \dots, 2^{-311})$ since each assignment has the same probability of occurrence. Comparing respondents with the digital diary and the paper diary, ignoring the total duration of the time slots to be coded, i.e. the “treatment doses”, with the Wilcoxon’s signed rank statistic leads to a p -value for the null hypothesis of no effect of the mode of collection of $1.4 \times 10^{-8} < 0.0001$. From the point of view of random inference, this means that less than one assignment in 10,000 could lead to the observation of such a high statistic, under the hypothesis of no effect of collection mode and in the case of random assignment of collection modes.

However, the protocol used did not correspond to a random assignment of the collection mode to the respondents but to the surveyed ones, which requires us to consider the impact of a hidden composition bias after matching on the inferences obtained, with a sensitivity analysis. When $\Gamma > 1$, the sensitivity analysis no longer considers a single probability of distribution of the 2^{311} possible treatments $\mathbf{k}_1, \dots, \mathbf{k}_{2^{311}}$, but a set \mathcal{P}_Γ of distribution probabilities \mathbf{p} . Each distribution probability $\mathbf{p} \in \mathcal{P}_\Gamma$ gives a p -value for the Wilcoxon’s test. We are looking for the highest p -value, \bar{P}_Γ , since we do not know the true distribution probability associated with the assignment to the collection mode. When $\Gamma = 1.7$, the maximum p -value is $\bar{P}_\Gamma = 0.049$. It means that a hidden characteristic which would double the odds of responding on paper and multiply by 8 the odds of observing a positive travel time difference could explain the differences observed.

We have another comparison in this study, this time looking at the difference in travel times between the paper respondent and the internet respondent within a pair, as a function of the total duration of the time slots to be coded in the paper diary. Do we observe greater differences in travel times in pairs where the respondent with the paper questionnaire declared a high volume of time slots to be coded?

Figure 4 (ii) shows the difference in journey times within a pair as a function of the total length of the other journeys declared by the respondent in the paper diary. The red line corresponds to the associated regression *lowess*: this moves away from 0 as the total duration of the time slots to be coded increases. This result suggests a tendency for pairs to show a high difference in travel time when the respondent with the paper questionnaire declares a significant amount of time in the time slots to be coded.

In the same figure, the vertical and horizontal lines correspond to the 1st and 3rd quartiles of the two distributions under consideration. Thus, the points (in black) at the top right of the graph represent pairs within which the difference in duration is important³⁵ and for which the total duration of the time slots to be coded is high³⁶.

Crosscut test focuses on the sub-populations defined by the quartiles, for which the corresponding number of pairs is given in table 10. The odds ratio estimated by this test is $12.8 = (48 \times 48) / (12 \times 15)$, revealing a strong association between the total duration of the time slots to be coded declared by the respondent to the paper questionnaire and the differences in journey times observed within the pairs. The p -value associated with this test obtained by random inference is 2.5×10^{-10} .

Table 10: Number of pairs associated with the *crosscut* test in Figure 4 (ii)

Duration differences	Total duration d_i of activities to be coded		
Y_i	$d_i \leq 0$	$d_i \geq 80$	Total
$Y_i \geq 85$	12	48	60
$Y_i \leq -20$	48	15	63
Total	60	63	123

Note : The table reports the number of pairs of the four subsets defined according to the difference (Y_i) in travel times between the paper respondent and the internet respondent on the one hand and the total duration (d_i) of the activities with the status to be coded of the paper respondent on the other hand. So, there are 48 pairs where $Y_i \geq 85$ and $d_i \geq 80$. They correspond to the 48 black dots in the upper right corner of the figure (ii) from Figure 4. The pairs associated with the gray points are not included in the figures reported in this table.

Such a calculation corresponds to a fictitious protocol of random assignment of the total durations of the time slots to be coded between the different respondents to the paper diary. There are $311!$ different assignments, $\mathbf{h}_1, \dots, \mathbf{h}_{311!}$ of total durations of time slots to be coded between pairs, whose distribution probability is $\mathbf{p}' = (1/311!, \dots, 1/311!)$ since each assignment has the same probability of realization in a random experiment. In other words, in a random experiment, odds of declaring a total time d of time slots to be coded in a pair rather than d' are the same for all pairs. Test p -value 2.5×10^{-10} estimates that less than one random assignment over 100,000 could have produced a *crosscut*'s statistic as high as the one we got (if the hypothesis H_0 of no effect of the collection mode is true in our random experiment).

It is however necessary to question this result since the total duration of time slots to be

³⁵i.e. greater than the 3rd quartile of the distribution of differences in journey times observed within pairs.

³⁶i.e. above the 3rd quartile of the corresponding distribution.

coded were of course not randomly assigned between the pairs. The sensitivity analysis model presented in section 3.2 can be adapted by assuming, this time, that the chances of a total duration d instead of d' to be coded differ from one pair to another by a factor at most $\Gamma' \geq 1$ [for an explicit formalization, see Rosenbaum, 2016] (where $\Gamma' = 1$ corresponds to a random experiment). Within this hypothesis, there is not just one distribution probability anymore but a set $\mathcal{P}'_{\Gamma'}$ of distribution probabilities \mathbf{p}' .

Each $\mathbf{p}' \in \mathcal{P}'_{\Gamma'}$ gives a p -value for the *crosscut*'s test and we are looking for the highest p -value, $\bar{P}'_{\Gamma'}$. When $\Gamma' = 5.7$, $\bar{P}'_{\Gamma'} = 0.049$, i.e. a hidden characteristic which would multiply by 10 the odds of declaring a high total duration to be coded and multiply by 10 the odds of observing a positive difference of travel time, could explain the differences we got.

Evidence factors

In the above, the first analysis focuses on the differences between a paper respondent and an internet respondent. The second analysis asks whether there are significant differences among pairs where the paper respondent reports a significant amount of time in time slots to be coded, i.e. whether the differences in travel times are correlated with the total duration of the time slots to be coded. Each seems to provide a convincing result on the existence of an effect of data collection with the paper questionnaire on the duration of reported journeys, but each remains subject to a possible composition bias. However, it should be noted that on this point they are not identical. Indeed, the first analysis is sensitive to an unobserved bias *between respondents with two collection modes* which determines who responds with the digital diary and who responds with the paper diary: for example, the fact of making longer journeys daily, as suggested by the result on the total journey time reported at the end of the questionnaire. However, it is not sensitive to a bias that determines among respondents to the digital diary who declares more or less duration among the “other journeys”. It is exactly the opposite for the second analysis. In this, unlike for example the comparison of the results of a sensitivity analysis with different test statistics for which the hidden composition bias is of the same nature, the two analyses differ. Can we therefore combine them in order to strengthen our conviction, for example by obtaining results less sensitive to the presence of a hidden composition bias, than those obtained by considering them separately?

Answering this question involves questioning the independence of the two analyses. Graphs from Figure 4 suggest conflicting intuitions on this question. Visually, differences from Figure 4 (i) do not “anticipate” graph from Figure 4 (ii). However, if we project points from Figure 4 (ii), we can find Figure 4 (i). In a sense, the first analysis could appear to be “independent” of the second. From another point of view, the two analyses appear to be strongly correlated. The aim is therefore to be able to combine the results of the sensitivity analyses of these two approaches *without any additional assumption* on their possible relationships, in particular independence, in the hope that the conclusions of these two comparisons reinforce each other. In particular, the additional assumption that we most want to *avoid* is that the person most likely to respond with the paper diary - for example because they make long journeys every day - *is not* also the most likely to declare a high total duration of activities to be coded: in fact, it is entirely possible that this is the case.

Rosenbaum [2017b] shows³⁷ that the upper bounds of the two p -values obtained from the two previous analyses (evidence factors) can be combined as if they came from independent analyses. This is not to consider that the two analyses are independent, but to emphasize that from the point of view of the p -value, the most unfavourable case corresponds to the

³⁷Without going into details in this working document - see Rosenbaum [2020].

situation of independence, even if the joint distributions are strongly correlated. More formally, he establishes the following proposition:

Proposition 1 *If H_0 is true, if bias from the first analysis is at most Γ and bias from second analysis is at most Γ' , then the pair of upper bounds of p -values $(\bar{P}_\Gamma, \bar{P}'_{\Gamma'})$ is stochastically larger than the uniform distribution on the square $[0, 1] \times [0, 1]$, which implies that $\Pr(\bar{P}_\Gamma \leq \alpha, \bar{P}'_{\Gamma'} \leq \alpha') \leq \alpha\alpha'$*

Intuitively, this proposition states that the conclusions of the two previous sensitivity analyses can reinforce each other as we illustrate with the following results.³⁸.

Table 11: Combined P -value of evidence factors of an effect of paper collection on travel times

Wilcoxon Γ	Crosscut \rightarrow Wilcoxon \downarrow	Crosscut test Γ'					
		1	4	5.7	6	7	∞
		0.000	0.005	0.049	0.064	0.127	0.906
1	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1.3	0.000	0.000	0.000	0.000	0.000	0.000	0.001
1.5	0.005	0.000	0.000	0.002	0.002	0.004	0.025
1.7	0.049	0.000	0.002	0.012	0.014	0.023	0.118
1.9	0.192	0.000	0.008	0.035	0.041	0.065	0.351
∞	1.000	0.000	0.028	0.109	0.127	0.202	1.000

Note : p -values for two evidence factors combined with a meta-analysis technique, using the truncated product of Zaykin et al. [2002] at 0.2.

Table 11 reports the results obtained by combining the p -values associated with the Wilcoxon's statistic for different values of Γ for the first analysis, with those associated with the crosscut's statistic for different values of Γ' for the second analysis. Combination of the two p -values for each pair (Γ, Γ') is done with meta-analysis techniques as if both analyses were independent and based on two different studies, on two different samples, etc. More precisely, method to combine the two p -values is their truncated product which is a generalization of Fisher's method of combining p -values. Fisher's method consists of taking the product of the two p -values and is interested in its distribution to obtain a new p -value. The truncated product considers the product of the p -values that are not greater than a truncation value here 0.2.

The first result is that, for biases of magnitude greater than those considered for each of the analyses taken separately, respectively $\Gamma = 1.9 > 1.7$ and $\Gamma' = 6 > 5.7$, the combined p -value 0.041 allows us to reject the hypothesis of an absence of effect of paper collection on the declared travel times. This means that the conclusions of each analysis reinforce each other and that by considering them jointly we have a stronger (and from this point of view quantifiable) conviction of the existence of a collection effect in the sense that the biases to reject this hypothesis must be greater. But even more, we can question the relevance of our results, if for example we consider that the second analysis is completely uninteresting because it is completely biased, that is to say with a value of $\Gamma' \rightarrow \infty$. It turns out that the hypothesis of an effect of the collection mode would still be rejected if $\Gamma = 1.5$ (p -value combined = 0.025). This means that, in this case, we still have a convincing result even if

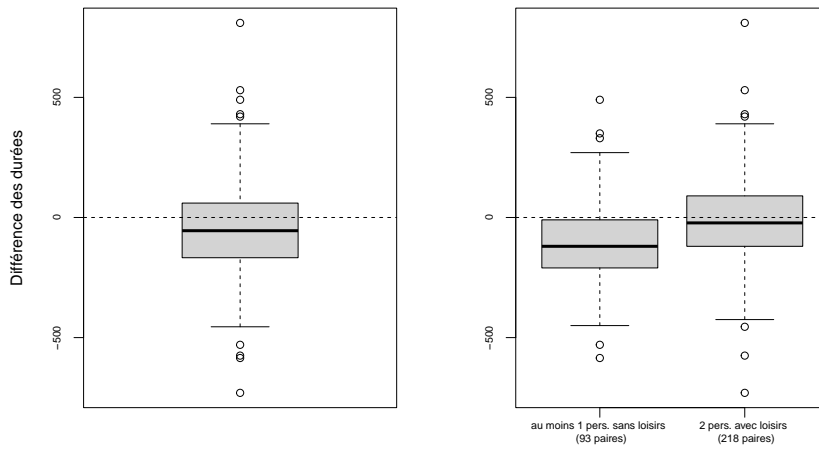
³⁸It should be noted that this proposition does not apply to the combination of all types of analyses. The validity of the approach implemented in our case is detailed in appendix E. For analysis designs that allow the constitution of evidence factors, see Rosenbaum [2021].

we do not take into account our second analysis. Similarly, if the first analysis is completely biased, the second still allows us to reject the hypothesis of no effect for $\Gamma' = 4$.

4.3 Impact on leisure time

Regarding leisure time, in the absence of hidden composition bias, respondents report on average 45 min ($[-67; -24]$ at 95 %) less leisure time when they respond on the internet as illustrated in Figure 5 (left). This gap would reflect differences in the probability of declaring at least one leisure activity (see Figure 5 right): indeed, 25 % of online respondents do not declare any leisure activity whereas they are only 7 % of respondents using the paper notebook in that case.

Figure 5: Difference in leisure time (internet - paper)

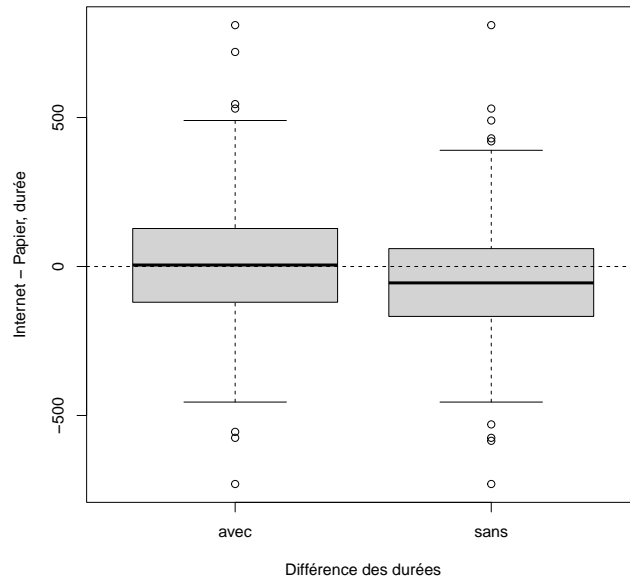


By applying an approach similar to that implemented for sleeping time, it is possible to test the hypothesis of an effect of internet data collection on the probability of not declaring leisure time and to estimate the proportion of effects attributable to this data collection mode. The sensitivity analysis reveals that the hypothesis of an effect of the internet mode is very little sensitive to the hypothesis of a hidden composition bias since $\Gamma_{max} = 2.8$. In other words, even considering that the probability of assignment to the internet data collection mode within a pair is only 0.30 instead of 0.5, we could still conclude that there is a significant measurement effect.

After matching, 87 pairs out of 311 (i.e. 28 % of them) are discordant, including 71 due to the absence of leisure time slots declared by the respondent via the Internet (i.e. 82% of the discordant pairs). In the absence of hidden composition bias ($\Gamma = 1$), we are 95 % certain that the Internet data collection mode caused 62 % of the 71 discordant pairs, i.e. 14 % of the 311 pairs in total. Even considering the existence of a significant hidden composition bias, for example with $\Gamma = 2$, we can still state with 95 % certainty that Internet data collection leads to not declaring leisure time in 6 % of the 311 pairs.

How can we explain such an effect of the Internet data collection mode on the absence of declaration of leisure activities? In the Time Use Survey, each activity is integrated into more general categories and subcategories. The “Leisure” category includes sports activities, outdoor activities (such as walks), and cultural activities outside, but also at home activities such as reading, watching television, playing video games. These latter activities, although personal and solitary, are therefore not included in the nomenclature in the category

Figure 6: Differences in leisure duration (internet - paper) within the 311 pairs - with or without correction of the activity code

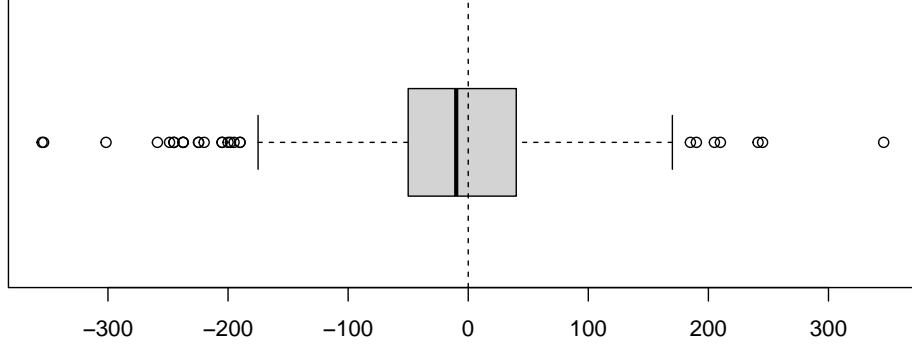


“Personal and physiological time” which rather groups together the activities of sleep, meals or time devoted to personal hygiene. The problem is that “Personal and physiological time” category includes a subcategory whose name “other personal activities” can be confusing. Indeed, with the online notebook (see figure 11 Appendix F), respondents mainly respond with a *selecter* (80 % of activities) which first displays the general categories (Personal time, Work and studies, Domestic tasks, etc.) and then a detail of the subcategories. Online, the “Leisure” category appears at the bottom of the page presenting the general categories, i.e. *after* the “Personal (and physiological) time” category. Such ergonomics of the activity selector might discourage respondents to scroll through the categories to reach the “Leisure” category located lower down. Therefore, it would be likely that respondents would choose the subcategory of “other personal activities” within the category “Personal and physiological time” displayed first, instead of browsing those proposed in the “Leisure” category, to declare the time slots corresponding to time spent watching television, reading, playing video games, etc.

In fact, respondents via the Internet report fewer leisure activities such as “watching television” and “doing nothing” compared to respondents with the paper diary. Conversely, respondents with the digital diary more frequently report time slots in the “other personal activities” subcategory of the “Personal and physiological time” category (4.5 % of activities reported) than respondents with the paper diary (less than 0.1 % of activities reported).

Following this hypothesis, it is possible to correct this potential classification error by imputing all the “other personal activities” as leisure activities in the digital diary. As illustrated in figure 6, with this new coding, the effect of the Internet data collection mode previously estimated is no longer significant, tending to confirm the hypothesis of an effect linked to the *design* of the questionnaire.

Figure 7: Differences in meal durations among 311 pairs (internet - paper)



Note : this graph represents meal durations distribution (internet - paper) among 311 pairs, using the reciprocal transformation proposed by [Rosenbaum \[2022\]](#). Schematically, 90 % of the values are not modified by this transformation. This allows a representation of distinct extreme observations without them leading to a visual compression of the central part of the distribution. This transformation does not affect the symmetry of the distribution: formally in the case of a random assignment, in the absence of treatment effect the distribution remains symmetrical. These properties ensure that the ranks of the absolute values of the differences are unchanged, such that the associated signed rank statistics, such as the Wilcoxon's statistic, are unchanged by the transformation.

4.4 Impact on meal duration

As we can see on Figure 7, in absence of hidden bias, respondents report on average 16 minutes (95% CI [-28;-4]) less meal duration when responding online. However, this effect is sensitive to the hypothesis of hidden composition bias ($\Gamma_{max} = 1.2$).

If we consider the number of time slots devoted to meals, the hypothesis of a measurement effect of Internet data collection is not very sensitive to a hidden composition bias, since, whatever the test statistic used, Γ_{max} is close to 2 (See Table 12).

Table 12: Sensitivity analysis Γ_{max} on total duration and number of meals

Statistic	total duration	number of slots
Mean	1.2	2.0
Wilcoxon	1.1	1.8
Stephenson (5,5,5)	1.2	2.2

In fact, 11 % of online respondents report more than 3 meals per day (which may be considered as an expected number of meals) compared to 25 % for paper respondents. Conversely, a similar proportion report exactly 3 meals regardless of the data collection mode (47 % online *versus* 49 % by paper questionnaire). Completing the diary on paper has a definite effect on the number of meals reported, but this is restricted to a small proportion of respondents. Overall, these results suggest a limited or no impact on the total time spent on meals, but a more frequent reporting, with the paper diary, of short time slots for “meals” which could correspond to a coffee break, *snacking*, etc.

5 Conclusion

Our study aims to address the difficulties posed by the analysis of a crossover trial conducted to investigate the existence of measurement effects in a survey with different data collection modes. As we have pointed out, the presence of potential endogenous selection weakens the conclusions drawn from the analysis of a crossover trial, as such an analysis needs to assume that non-response is MAR or MCAR. Nevertheless, it is possible to discuss the existence of measurement effects by comparing the responses, given in the first questioning, by similar individuals surveyed with different data collection modes. To this end, it is appropriate to conduct a sensitivity analysis [Rosenbaum, 2002a, 2010] on the conclusions obtained under the assumption of no bias related to potential endogenous selection.

Such an analysis was applied here to the Time Use test Survey, conducted in 2023, to investigate the existence of a measurement effect on reported durations for different activities with a digital diary instead of a paper diary. Our results highlight that it is difficult to reject the hypothesis of no measurement effect of the data collection mode on the reported duration of certain activities, even in the presence of endogenous selection. However, this measurement effect would often only impact the durations of a limited number of respondents. More precisely, the under-reporting of sleep hours would be correlated with the non-completion of the entire diary. Indeed, online some respondents do not report their last activities of the day, which reduces the covered period reported and likely the declared sleeping time. Regarding travel time, our results suggest that, with a paper diary, difficulty in coding activities related to journeys other than those from home to work would lead to overestimating the total reported travel time. For instance, a time slot dedicated to “going outside to do some shopping” may be coded as a trip and as such may include the time spent on purchases. The under-reporting of leisure activities online might be explained by the interface’s selector ergonomics, which could lead respondents to classify their leisure activities under another category. Finally, the under-reporting of the number of meals online could reflect a difference in reporting behaviour for short breaks, such as the time spent having coffee during the day outside of meal times. However, the magnitude, even if limited, of these effects can still be questioned. Indeed, the approaches implemented in this study do not account for possible interference between surveyed units within a household: for instance, our estimates do not include the potential completion of a diary by or with the help of a third party.

References

- S. Basu and S. Santra. A joint model for incomplete data in crossover trials. *Journal of Statistical Planning and Inference*, 140:2839–2845, 2010.
- L. Court and S. Quantin. Discuter l’existence d’un effet de sélection dans un cadre multimode grâce à une analyse de sensibilité : application aux enquêtes annuelles de recensement. *Documents de travail Insee*, M2024/03, 2024.
- R. A. Fisher. *The Design of Experiments*. Oliver and Boyd, 1935.
- Weang Kee Ho, John N.S. Matthews, Robin Henderson, Daniel Farewell, and Lauren R. Rodgers. Dropouts in the AB/BA crossover design. *Statistics in Medicine*, 31:1675–1687, 2012.
- J. Hodges and E. Lehmann. Estimates of location based on rank tests. *Annals of Mathematical Statistics*, 34:598–611, 1963.
- P.J. Huber. Robust estimation of a location parameter. *Annals of Mathematical Statistics*, 35:73–101, 1964.
- E.L. Lehmann. *Nonparametrics : Statistical methods based on ranks*. San Francisco : Holden-Day, 1975.
- J. Maritz. *Distribution-Free Statistical Methods*. London: Chapman & Hall, 1981.
- J. S. Maritz. Exact robust confidence intervals for location. *Biometrika*, 66:163–166, 1979.
- Jerzy Neyman. On the application of probability theory to agricultural experiments. Essay on principles. Section 9. *Roczniki Nauk Rolniczych*, Tom X:1–51, 1923. Réimprimé en anglais dans *Statistical Science*, 1990, 5, 463–480.
- M. Pagano and D. Tritchler. Obtaining Permutation Distributions in Polynomial Time. *Journal of the American Association*, 78:435–440, 1983.
- H.I. Patel. Analysis of incomplete data in a 2-period crossover design with reference to clinical trials. *Biometrika*, 72:411–418, 1985.
- Paul R. Rosenbaum. The role of known effects in observational studies. *Biometrics*, 45: 557–569, 1989a.
- Paul R. Rosenbaum. On permutation tests for hidden biases in observational studies. *Annals of Statistics*, 17:643–653, 1989b.
- Paul R. Rosenbaum. Effects Attributable to Treatment: Inference in Experiments and Observational Studies with a Discrete Pivot. *Biometrika*, 88(1):219–231, March 2001.
- Paul R. Rosenbaum. *Observational Studies*. Springer Series in Statistics, 2 edition, 2002a.
- Paul R. Rosenbaum. Attributing Effects to Treatment in Matched Observational Studies. *Journal of the American Statistical Association*, 95(457):183–192, 2002b.
- Paul R. Rosenbaum. Confidence Intervals for Uncommon but Dramatic Responses to Treatment. *Biometrics*, 63:1164–1171, June 2007a.
- Paul R. Rosenbaum. Sensitivity Analysis for m-Estimates, Tests, and Confidence Intervals in Matched Observational Studies. *Biometrics*, 63:456–464, June 2007b.
- Paul R. Rosenbaum. *Design of observational studies*. Springer Series in Statistics, 2010. doi: 10.1007/978-1-4419-1213-8.

- Paul R. Rosenbaum. A New u-Statistic with Superior Design Sensitivity in Matched Observational Studies. *Biometrics*, 67:1017–1027, 2011.
- Paul R. Rosenbaum. Weighted M-statistics With Superior Design Sensitivity in Matched Observational Studies With Multiple Controls. *Journal of the American Statistical Association*, 109(507):1145–1158, september 2014.
- Paul R. Rosenbaum. The crosscut statistic and its sensitivity to bias in observational studies with ordered doses of treatment. *Biometrika*, 72:175–183, 2016.
- Paul R. Rosenbaum. *Observation and Experiment : An introduction to causal inference*. Harvard University Press, 2017a.
- Paul R. Rosenbaum. The General Structure of Evidence Factors in Observational Studies. *Statistical Science*, 32(4):514–530, November 2017b.
- Paul R. Rosenbaum. *Design of Observational Studies*. Springer Series in Statistics, 2 edition, 2020.
- Paul R. Rosenbaum. *Replication and Evidence Factors in Observational Studies*. Chapman and Hall/CRC, 2021.
- Paul R. Rosenbaum. A New Transformation of Treated-Control Matched-Pairs Differences for Graphical Display. *The American Statistician*, 76(4):346–352, 2022.
- Paul R. Rosenbaum. Sensitivity analyses informed by tests for bias in observational studies. *Biometrics*, 79:475–487, 2023.
- Paul R. Rosenbaum and Jeffrey H. Silber. Amplification of Sensitivity Analysis in Matched Observational Studies. *Journal of the American Statistical Association*, 104(488):1398–1405, December 2009.
- D.B. Rubin. Estimating causal effects of treatments in randomized and non randomized studies. *Journal of Educational Psychology*, 66(5):688–701, 1974. URL <https://doi.org/10.1037/h0037350>.
- D.B. Rubin. Inference and Missing Data. *Biometrika*, 63:581–592, 1976.
- W.R. Stephenson. A general class of one-sample parametrics test statistics based on subsamples. *Journal of the American Association*, 76:960–966, 1981.
- Jeffrey M. Wooldridge. *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge, 2 edition, 2002.
- D.V. Zaykin, L.A. Zhivotovsky, P.H. Westfall, and B.S. Wier. Truncated product method combining P-values. *Genetic Epidemiology*, 22:170–185, 2002.

Annexes

A Optimal Pairwise Matching

Propensity score matching tends to create treated and control groups that have similar distributions for the observed variables. However, while the surveyed units within each pair thus formed have similar (estimated) propensity scores, they may nevertheless differ greatly on specific covariates within each pair. To create more similar pairs of surveyed units, we construct a distance that penalizes large differences in propensity scores, and then we create pairs of units that are as similar as possible, using an optimization algorithm.

In a multivariate framework, it is important to take into account the differences in units of measurement of the covariates in the choice of the distance used. The Mahalanobis distance generalizes to several variables the notion of measuring the distance in number of standard deviations while also taking into account the correlations between these variables. Formally, if $\hat{\Sigma}$ is the empirical covariance matrix of the variables \mathbf{x} then the Mahalanobis distance between two surveyed units, k and l associated with these covariates is $(\mathbf{x}_k - \mathbf{x}_l)^T \hat{\Sigma}^{-1} (\mathbf{x}_k - \mathbf{x}_l)$. However, if a covariate has outliers or a very stretched distribution, its standard deviation can be large and the Mahalanobis distance will tend to ignore this covariate in the matching. Moreover, with binary variables, the variance is maximum when the probability of the event is $\frac{1}{2}$, and minimum when it is close to 0 or 1. Therefore, the Mahalanobis distance gives greater weight to binary variables measuring rare events, i.e. with probabilities close to 0 or 1. In order to circumvent these difficulties, we favor the Mahalanobis distance of the rank. Concretely, this distance is calculated (i) by replacing the values of the covariates (each covariate being considered separately) by their ranks, by applying an average rank in the case of identical values, and (ii) by adjusting the covariance matrix of the ranks of the covariates with a diagonal matrix whose elements are the ratios of the standard deviations of the ranks of the unique values by the standard deviations of the ranks of the values including identical ones. The first step (i) makes it possible to limit the influence of outliers and long tails of the distribution. The presence of identical values reduces the variance of the ranks: the second step (ii) readjusts the variance-covariance matrix so that each covariate has its variance excluding identical values. Such a correction reduces the usual influence with the Mahalanobis distance of covariates presenting many identical values, such as binary variables measuring a rare event.

This distance is usually calculated between all the units surveyed whose difference in estimated propensity scores $|\hat{e}(\mathbf{x}_k) - \hat{e}(\mathbf{x}_l)|$ does not exceed, in absolute value, a caliper w , whose width w is here taken equal to half the standard deviation of the estimated propensity score, $\hat{e}(\mathbf{x})$. However, due to the small sample size, there is not necessarily a pairwise matching for which the condition $|\hat{e}(\mathbf{x}_k) - \hat{e}(\mathbf{x}_l)| \leq w$ between two matched units k and l is respected for all treated units. Therefore, to ensure a matching of each treated surveyed unit, we calculate the Mahalanobis distance of the rank between all units, to which we add a penalty function when the constraint imposed by the caliper is not respected. Formally, the penalty used in this study is $1000 \times \max(0, |\hat{e}(\mathbf{x}_k) - \hat{e}(\mathbf{x}_l)| - w)$. This penalty is null if $|\hat{e}(\mathbf{x}_k) - \hat{e}(\mathbf{x}_l)| \leq w$, but equals $1000 \times (|\hat{e}(\mathbf{x}_k) - \hat{e}(\mathbf{x}_l)| - w)$ if $|\hat{e}(\mathbf{x}_k) - \hat{e}(\mathbf{x}_l)| > w$.

The durations of the different activities reported by a surveyed unit depend on the day of the week on which the questionnaire is completed (weekend or not) and its main situation (employed, unemployed³⁹, retired or early retired, unable to work, studying, at home or other situation). It is therefore important to be able to ensure that within each pair, the surveyed units are identical on these observed characteristics. As regards the day of the

³⁹registered or not with Pôle Emploi.

week, an exact match is imposed. To take into account the main situation, a second penalty is added to the distance mentioned above. If two surveyed units k and l do not have the same main situation, the penalty function used adds to the distance 10 times the maximum distance observed (on the distance previously calculated) if the two units differ on their main situation.

Once the distance matrix is formed, an optimal pairwise matching is performed. Such a matching constitutes pairs composed of a treated survey unit and a control unit such that the sum of the distances within the pairs is minimal⁴⁰. Therefore, an optimal matching will seek to avoid penalized distances by respecting the caliper constraint as much as possible; when this is not possible, it will favor a matching where the violation of the caliper constraint is infrequent and as small as possible. Regarding the main situation, if an exact matching is possible, it will be considered; otherwise, a matching as close as possible to an exact matching will be performed.

B What does it mean to judge the quality of the match?

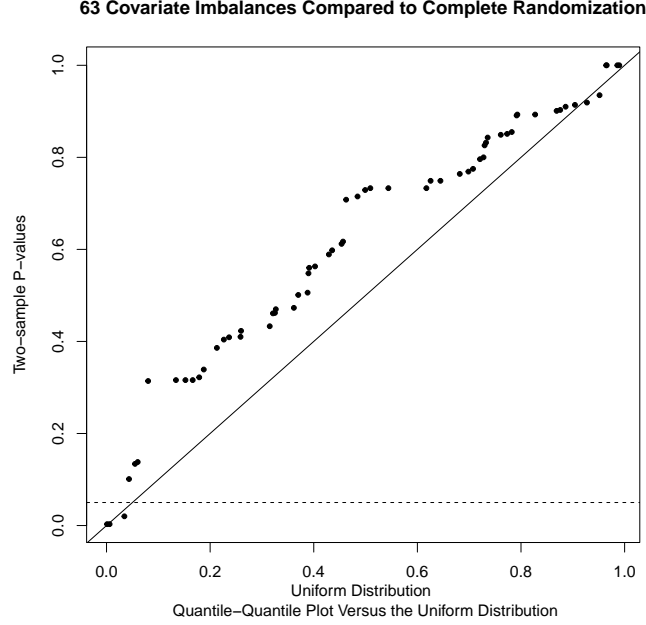
As explained in the methodological section devoted to sensitivity analysis, random inference to highlight a causal effect does not rely on the comparison of identical people⁴¹, but on the distribution of the probability of assignment to the collection mode. From this point of view, if the matching aims to reduce the differences between the two samples, the question of its quality arises in other words: have these differences been reduced sufficiently, that is to say compared to those that would have been observed if we had done a completely random experiment. A random experiment would have balanced the unobserved covariates as observed. In the case of matching, obviously, we can only be interested in the observed covariates.

In our matching, we consider 63 covariates. If, in a randomized experiment, we perform a randomization test on these 63 covariates, then the only reason to reject the null hypothesis of no effect is chance. Thus, we expect to obtain $63 \times 0.05 = 3.15$ P-values ≤ 0.05 . How does the quality of our matching compare to what we would have achieved with complete randomization?

⁴⁰This minimization problem is not immediate because the closest control unit to a processed unit may also be the closest control unit to another processed unit. A best-first choice or greedy algorithm will not generally result in finding the optimal pairwise matching.

⁴¹As Paul R. Rosenbaum points out in his presentation at the 2019 Committee of Presidents of Statistical Societies Awards, if we consider 63 binary covariates, we have $2^{63} = 9.2 \times 10^{18}$ possible categories, or as many types of people. But there are only 8.0 billion people on Earth. This means that for every person we find in a category, there are on average 1.2×10^9 categories with no people in them. This means that we know for sure that we will not compare identical people on 63 covariates.

Figure 8: Differences between treated and controlled subjects on the 63 covariates after matching compared to the situation that would have been observed after a random experiment



As illustrated in the graph 8, the balance obtained by matching for this study on the observed covariates is better than that which would have been obtained by random assignment. But, again, however, this is only for the observable covariates.

C Rosenbaum's U-statistics

We consider three integers $m, \underline{m}, \overline{m}$ with $1 \leq \underline{m} \leq \overline{m} \leq m < I$; the triplet $m, \underline{m}, \overline{m}$ defines the U-statistic considered. Its calculation is as follows :

- We consider subsets of m pairs. For each subset, the differences $V_i Y_i$ are ordered according to their absolute values $|Y_i|$,
- Once the ordering is done, we count the number of positive $V_i Y_i$, that is to say the number of differences (pairs) where the response provided by the treated unit is greater than that of the control unit, among those numbered $\underline{m}, \underline{m} + 1, \dots, \overline{m}$. Then we aggregate this result over all $\binom{I}{m}$ subsets.

To clarify things, we now detail some examples of U-statistics. The statistic $(1, 1, 1)$ is the sign statistic: it totals the number of pairs where the response provided by the treated unit is greater than that of the control unit. The statistic (m, m, m) , which corresponds to the Stephenson statistic [Stephenson, 1981], is interested in the $\binom{I}{m}$ subsets (Y_{i1}, \dots, Y_{im}) of m pairs and counts the number of highest positive differences (since $\overline{m} = m$). The statistic $(8, 7, 8)$ consider the $\binom{I}{8}$ subsets (Y_{i1}, \dots, Y_{i8}) of 8 pairs and counts the number of highest positive differences among the two highest $|Y_{ik}|$. Finally, the statistic $(20, 16, 19)$ it examines all subsets of pairs of size 20, ignores the sign of the highest $|Y_{ik}|$ value, and counts the number of positive differences Y_{ik} among the next 4 highest $|Y_{ik}|$ values. Note, as Rosenbaum [2011] said, the statistic $(2, 2, 2)$ is the Wilcoxon's signed-rank statistic.

Formally, U-statistics are signed rank statistics. Noting a_i the rank of $|Y_i|$ for all $i = 1, \dots, I$, the difference $V_i Y_i$ with rank a_i has the l^{th} highest value $|Y_i|$ in $\binom{a_i-1}{l-1} \binom{I-a_i}{m-l}$ subsets of size m [Rosenbaum, 2011]. And so, the statistic $(m, \underline{m}, \overline{m})$ is :

$$T = \sum_{i=1}^I \text{sign}(V_i Y_i) q_i$$

where

$$q_i = \binom{I}{m}^{-1} \sum_{l=\underline{m}}^{\overline{m}} \binom{a_i-1}{l-1} \binom{I-a_i}{m-l}$$

Note that, if $m = \underline{m} = \overline{m} = 2$, $q_i = \binom{a_i-1}{m-1} = \binom{a_i-1}{1} = a_i - 1$, so \tilde{T} is the Wilcoxon signed-rank statistic, except for a constant.

By varying m, \underline{m} and \overline{m} , we can control the degree of influence of observations of different magnitudes. To illustrate this, we represent in Figure 9, the values of $q_i / \max_{1 \leq j \leq I} (q_j)$ by a_i / I for different values of m, \underline{m} and \overline{m} and with $I = 300$. We therefore represent for different values m, \underline{m} and \overline{m} , the contribution of each pair to the statistic considered as a function of its rank. Stephenson statistic (5, 5, 5), and U-statistics (8, 7, 8) and (20, 16, 19) give more weight to high values of $|Y_i|$ and less weight to low values of $|Y_i|$ than the Wilcoxon's signed-rank statistic. The (20, 16, 19) statistic gives more weight to high values of $|Y_i|$, while reducing, unlike the (8, 7, 8) statistic, the influence of extreme values which may be outliers.

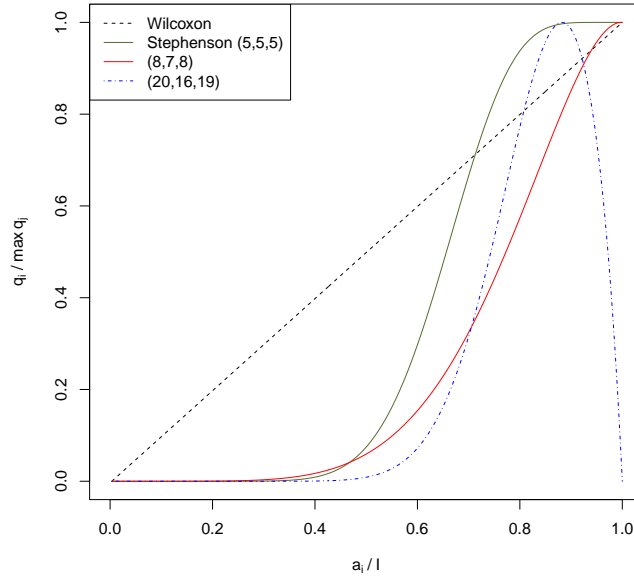


Figure 9: Representation of q_i , normalized to have a maximum of 1, as a function of a_i / I where a_i is the rank of $|Y_i|$

D Estimation of an additive and constant treatment effect

In this appendix, we first present the Hodges-Lehmann estimator of a constant additive effect associated with the Wilcoxon's signed rank statistic. Because this estimator relies on the distribution of the test statistic, it is not possible to calculate it in the presence of a

hidden composition bias. Therefore, we detail how the (5) framing of the distribution of the test statistic can be used to determine a range of possible values of the Hodges-Lehmann estimator. This range of possible values of the estimated effect reflects the uncertainty related to the existence of an unobserved characteristic, and its amplitude depends on the value of the parameter Γ considered in the sensitivity analysis.

D.1 The Hodges-Lehmann estimator of a constant additive treatment effect

There are many estimators of an effect τ if we assume the treatment effect to be additive and constant, that is, if we assume that $\mathbf{R} = \mathbf{r}_C + \tau\mathbf{Z}$. The estimator proposed by Hodges and Lehmann [1963], Lehmann [1975] is deduced from the test statistic used to test the null hypothesis H_0 of an absence of treatment effect. As we will see in the following section D.2, it is therefore natural that we can use the framework (5) of the distribution of the test statistic to deduce an interval of possible values of the effect, constant and additive, of the treatment.

To test the null hypothesis of an additive and constant effect $H_0 : \tau = \tau_0$, it is necessary to calculate the adjusted test statistic, $t(\mathbf{Z}, \mathbf{R} - \tau_0\mathbf{Z})$, then to test from this the null hypothesis of no treatment effect. Hodges and Lehmann They propose to determine to estimate an additive and constant effect of the treatment by the value τ_0 for which the adjusted test statistic T has the expected distribution. Schematically⁴², the Hodges-Lehmann estimator is the solution $\hat{\tau}$ to the equation $\bar{t} = t(\mathbf{Z}, \mathbf{R} - \tau_0\mathbf{Z})$ where \bar{t} is the the expected expectation of the test statistic in the absence of a treatment effect, i.e. after adjustment. This calculation is possible because in the absence of hidden composition bias and treatment effect, the expectation of the Wilcoxon's signed rank statistic is known.

To realize this, let us consider - in this section and in the following one - and only to simplify the expression of the results that there are no identical values of $V_i Y_i$. In the absence of a treatment effect, we therefore expect to sum half of the ranks of the I differences, since there is a one in two chance that the treated unit has a higher value than the control unit, in the absence of hidden composition bias. As total sum of the I is $\frac{I(I+1)}{2}$, so $\bar{t} = \frac{I(I+1)}{4}$.

Formally, Hodges and Lehmann [1963] define their estimator as:

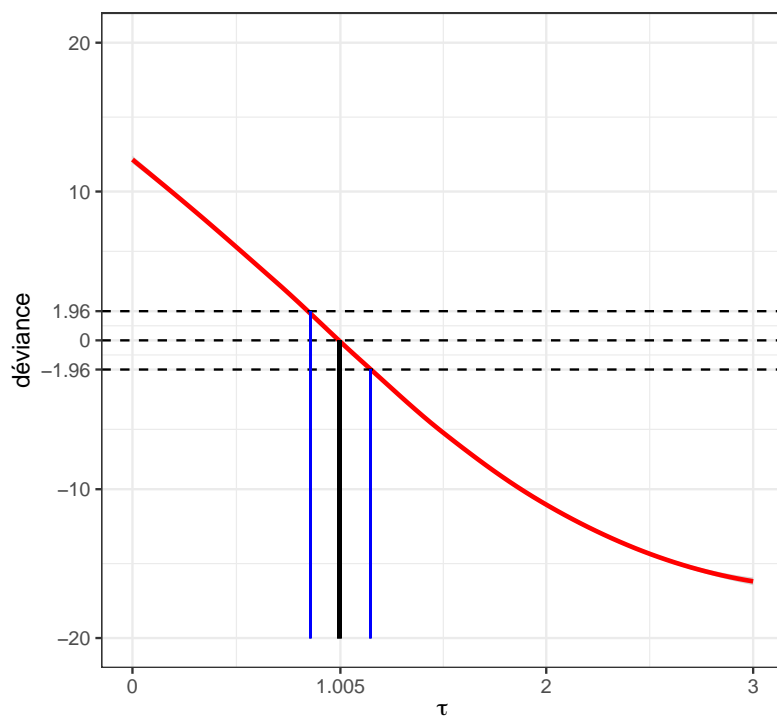
$$\begin{aligned} \hat{\tau} &= \text{SOLUTION}\{\bar{t} = t(\mathbf{Z}, \mathbf{R} - \hat{\tau}\mathbf{Z})\} \\ &= \frac{\inf\{\tau : \bar{t} > t(\mathbf{Z}, \mathbf{R} - \tau\mathbf{Z})\} + \sup\{\tau : \bar{t} < t(\mathbf{Z}, \mathbf{R} - \tau\mathbf{Z})\}}{2} \end{aligned}$$

to take account of the possibility that the test statistic only takes discrete values⁴³. Thus, in the absence of hidden compositional bias, it is possible to invert the test statistic in order to estimate an additive and constant treatment effect. As Graph 10 also shows, the approach can be extended to determine a 95% confidence interval for the estimated effect.

⁴²see [Rosenbaum, 2002a, section 2.7.2]

⁴³The Hodges-Lehmann estimator inherits the properties associated with the $t(.,.)$ test statistic used, in the sense that, for example, as soon as the test is convergent, the Hodges-Lehmann estimator is convergent [see Maritz, 1981, for more details]. A test is convergent, as in the case of the Wilcoxon's signed rank test, if the probability of rejecting a false hypothesis, whatever it may be, tends towards 1 as the sample size increases. Intuitively, this concept is therefore linked to the convergence of an estimator whose probability of being close to the true value tends towards 1 as the sample size increases.

Figure 10: An illustration of the Hodges-Lehmann estimator of the treatment effect and the corresponding confidence interval



Note : For each τ , constant additive effect of the treatment with H_0 , the curve (red) represents the gap between the wilcoxon's signed-rank statistic (centred and reduced) and the null value (corresponding to the expectation of the distribution $\mathcal{N}(0,1)$). This is how we determine the τ value which cancels out this deviation (black vertical segment) and the limits of the corresponding 95 % confidence interval (blue vertical segments).

D.2 Sensitivity analysis of the Hodges-Lehmann estimator

In the presence of a hidden composition bias, the distribution of the test statistic is not known and is no longer equal to $I(I+1)/4$, so that it is no longer possible to determine the Hodges-Lehmann estimator of an additive and constant effect. However, considering the sensitivity analysis model (3), the expectation of the adjusted statistic \bar{t} may be bounded by the expectations of T^- and of T^+ (defined in Section 3.2), i.e. by two known quantities: \bar{t}_{\min} and \bar{t}_{\max} . Formally, if we apply the formulae in section 3.2 (and still considering the absence of identical values for d_s), we get that:

$$\bar{t}_{\min} = \frac{p^- I(I+1)}{2} \quad \text{and} \quad \bar{t}_{\max} = \frac{p^+ I(I+1)}{2}$$

with $p^- = 1/(1 + \Gamma)$ and $p^+ = \Gamma/(1 + \Gamma)$.

By calculating $\hat{\tau} = \text{SOLUTION}\{\bar{t} = t(\mathbf{Z}, \mathbf{R} - \hat{\tau}\mathbf{Z})\}$ for all \bar{t} in $[\bar{t}_{\min}; \bar{t}_{\max}]$, we determine the interval of possible values of the Hodges-Lehmann estimators associated with an additive and constant effect. In practice, the minimum and maximum of $\hat{\tau}$ correspond to the Hodges-Lehmann estimators associated with \bar{t}_{\min} and \bar{t}_{\max} . In the end, the sensitivity analysis of the Hodges-Lehmann estimator consists of determining the range of possible values of this estimator for different Γ quantities.

D.3 Sensitivity analysis of the confidence interval of a Hodges-Lehmann estimator

As with any estimator, in the absence of hidden compositional bias, it is possible to determine a confidence interval for the Hodges-Lehmann estimator. Similarly, in the presence of a hidden compositional bias, it is possible to determine a *confidence interval for the interval of possible values* of the Hodges-Lehmann estimator.

Formally, this is obtained from the normal approximation of the distribution of T_τ^+ and T_τ^- . Its bounds correspond to:

$$\inf \left\{ \tau : \frac{T_\tau - E(T_\tau^+)}{\sqrt{\text{var}(T_\tau^+)}} \leq 1.96 \right\} \quad \text{and} \quad \sup \left\{ \tau : \frac{T_\tau - E(T_\tau^-)}{\sqrt{\text{var}(T_\tau^-)}} \geq -1.96 \right\}$$

E Evidence factors

In this section, we briefly justify the evidence factor approach implemented in section 4.2. To illustrate our point, we will use a fictitious example consisting of two pairs whose surveyed units, numbered 1 to 4, are presented in Table 13, which specifies for each the mode of collection and the total duration to be coded declared. In what follows, we will use the term “treatment” to refer to collection by paper diary and the term “dose” to refer to the total duration of the time slots to be coded.

Table 13: Four positions on the collection mode in two pairs

Structure		Treatment position		Travel duration
ID	Pair	Paper collection	Duration to be coded	
1	1	1	150	240
2	1	0	0	130
3	2	1	30	70
4	2	0	0	40

There are two different treatments and two different treatment doses or, to use the terms of Rosenbaum [2017b], 4 “treatment positions”: being the respondent with the paper diary

that declares 150 minutes “of other journeys”, being the respondent with the digital diary associated with this respondent, being the respondent with the paper diary that declares 30 minutes “of other journeys” and being the respondent with the digital diary associated with this one.

General structure

Let’s consider the experiment that would assign each of the 4 respondents to one of these treatment positions while keeping the composition of the pairs intact, since they have been constituted in such a way that the respondent with the paper diary and the Internet respondent are similar in terms of the socio-demographic characteristics observed.

Formally, it is useful to represent an assignment to a treatment position by a permutation matrix, which assigns a treatment position to 4 respondents $((1, 2, 3, 4)^T)$ matched in $(n = 2)$ pairs.

$$\mathbf{gn} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \\ 2 \\ 1 \end{bmatrix}$$

In the example above, the assignment to treatment \mathbf{g} assigns the first respondent to the paper collection mode with 30 min. of time to code, while keeping within the same pair the other respondent to whom he is matched and allocating him the internet collection mode. The set of possible assignments in the experiment we are considering is a finite \mathcal{G} subgroup of $(2n)!$ matrices of possible permutations. In fact, there are 2 ways of assigning the pairs to a duration to be coded, 150 min. to one pair and 30 min. to the other. Furthermore, within the 2 pairs, there are 2 ways of allocating the paper collection mode to one of the respondents, i.e. 2^2 . So there are no $4! = 24$ possible assignments in the treatment, because we want to keep the pair constitution, but $2 \times 2^2 = 8$ possible assignments. Noting $|\mathcal{S}|$ the number of elements in a finite set \mathcal{S} , there are $|\mathcal{G}| = (n)! \times 2^n$ possible assignments within n pairs.

A probability distribution on \mathcal{G} gives, to every permutation matrix $\mathbf{g} \in \mathcal{G}$, a probability $p(\mathbf{g})$. More precisely, a probability distribution p on \mathcal{G} verifies (i) $p(\mathbf{g}) \geq 0$ for all $\mathbf{g} \in \mathcal{G}$ and (ii) $1 = \sum_{\mathbf{g} \in \mathcal{G}} p(\mathbf{g})$ and an assignment \mathbf{G} to position treatment from this distribution has $\Pr(\mathbf{G} = \mathbf{g}) = p(\mathbf{g})$. In a random experiment, \mathbf{g} is picked randomly among \mathcal{G} ; so $p(\mathbf{g}) = |\mathcal{G}|^{-1}$ for all $\mathbf{g} \in \mathcal{G}$ and $p = (p_{\mathbf{g}_1}, p_{\mathbf{g}_2}, \dots, p_{\mathbf{g}_{|\mathcal{G}|}}) = (\frac{1}{|\mathcal{G}|}, \dots, \frac{1}{|\mathcal{G}|})^T$ ⁴⁴. If the different possible assignments modify the treatment position allocated to each unit, it is important to remember that under the Fisher’s assumption H_0 of no effect of the collection mode, the journey times are fixed; thus the different permutations do not change the journey times declared. The hypothesis H_0 assumes that person 1 would always declare 240 minutes of travel time under the 8 possible treatment positions; the same is true for the 3 other persons in our example.

This assignment of units to a treatment position can be related to the two analyses with random inference carried out in section 4.2 by considering two subgroups of permutations.

The first subgroup of permutations, \mathcal{K} of \mathcal{G} characterises the analysis in which treatment doses are ignored and we are only interested in the comparison of the respondent’s journey times with the digital diary and the paper diary. In this case, the associated random experiment assigns treatment positions while preserving the treatment doses allocated to

⁴⁴As we shall see later, in the case of a non-random experiment, we will consider a set \mathcal{P} of probability distributions p , this set reflecting our uncertainty, calibrated as part of a sensitivity analysis, about the true probability distribution.

each pair and modifying only the choice of the respondent who answers with the paper diary⁴⁵. In our case, there are 4 possible permutations, and if we consider n pairs, \mathcal{K} has $|\mathcal{K}| = 2^n$ elements. In fact, the Wilcoxon signed rank statistic used in this first analysis is invariant to $n!$ permutations of all pairs and only considers the 2^n permutations of collection mode within pairs.

There is a second subgroup of permutations, \mathcal{H} , of \mathcal{G} which corresponds to the random inference made in the second analysis. This only permutes the treatment doses between pairs, while retaining within each pair the respondent who answers with the paper diary. In our example, \mathcal{H} has only 2 elements and with n pairs, $|\mathcal{H}| = n!$ elements. The crosscut's statistic used in the second analysis is in fact conditioned by the collection modes assigned within each pair in the sense that one of the 2^n possible permutations of the collection mode within each pair is fixed and uses the $n!$ permutations of treatment doses between the pairs.

As this presentation of the two subgroups suggests, it is possible to show that for any $\mathbf{g} \in \mathcal{G}$, there is a unique representation $\mathbf{g} = \mathbf{h}\mathbf{k}$ where $\mathbf{h} \in \mathcal{H}$ and $\mathbf{k} \in \mathcal{K}$.

Sensitivity analysis for the no measurement effect test

In our study, the mode of collection or the “treatment doses” are not randomly assigned to the respondents. The sensitivity analysis model no longer considers *one* probability distribution but a set of possible probability distributions. Thus, for the first analysis, with a sensitivity parameter value Γ , we consider a set \mathcal{P}_Γ whose elements are probability distributions $\mathbf{p} \in \mathcal{K}$. Likewise, we consider $\mathcal{P}'_{\Gamma'}$ the set of probability distributions $\mathbf{p}' \in \mathcal{H}$ for each value Γ' of the sensitivity parameter.

Combining two sensitivity analyses

The starting point is therefore to consider two evidence factors such that (i) each assignment to a treatment position $\mathbf{g} \in \mathcal{G} = \mathcal{H}\mathcal{K}$ has a unique representation $\mathbf{g} = \mathbf{h}\mathbf{k}$ with $\mathbf{h} \in \mathcal{H}$ and $\mathbf{k} \in \mathcal{K}$, where \mathcal{H} is a subgroup of \mathcal{G} and (ii) the test statistic for the first factor $t(\mathbf{g}) = t(\mathbf{h}\mathbf{k})$ is \mathcal{H} -invariant, and so $t(\mathbf{h}\mathbf{k}) = t(\mathbf{k})$.

In our case, the two factors of evidence are the two analyses carried out separately. Furthermore, the Wilcoxon's signed rank statistic takes the same value whatever the treatment doses, since this statistic is not a function of the treatment doses. No other hypothesis is necessary, notably on the independence between \mathbf{H} and \mathbf{K} .

The null hypothesis H_0 of no effect is then tested twice: the first time with a marginal test using $t(\mathbf{k})$ (Wilcoxon's signed rank statistic, in this case) and the second time with a conditional test knowing $\mathbf{K} = \mathbf{k}$ using $t'(\mathbf{h}\mathbf{k})$ (crosscut's statistic in this case). This produces a pair of p -values upper bounds $(\bar{P}_\Gamma, \bar{P}'_{\Gamma'})$ of the real p -values, (P, P') . Formally, we assume $\mathbf{p} \in \mathcal{P}_\Gamma$ and we test H_0 with the statistic $t(\mathbf{g}) = t(\mathbf{h}\mathbf{k})$ which is invariant on \mathcal{H} (i.e. $t(\mathbf{h}\mathbf{k}) = t(\mathbf{k})$ for all \mathbf{h} and \mathbf{k}) to obtain the highest p -value: \bar{P}_Γ . We test H_0 again using the conditional distributions $\mathbf{p}' \in \mathcal{P}'_{\Gamma'}$ of $t'(\mathbf{g}) = t'(\mathbf{h}\mathbf{k})$ of \mathbf{H} knowing $\mathbf{K} = \mathbf{k}$ to obtain the highest p -value $\bar{P}'_{\Gamma'}$ of the conditional p -values.

In order to combine the two factors of evidence, i.e. the two sensitivity analyses carried out in this way, it is necessary to be able to limit the one resulting from the joint distribution. For this purpose [Rosenbaum \[2017b\]](#) establishes the following proposition on the p -values from the joint distribution.

Proposition 2 *If the hypothesis H_0 is true, if $\Pr(\mathbf{K} = \mathbf{k})$ is one of the probability distributions $\mathbf{p} \in \mathcal{P}_\Gamma$, and if $\Pr(\mathbf{H} = \mathbf{h} | \mathbf{K} = \mathbf{k})$ is one of the probability distributions $\mathbf{p}' \in \mathcal{P}'_{\Gamma'}$,*

⁴⁵Remember that under H_0 the declared journey times are fixed in Fisher's sense.

then for any joint distribution of the product knit, the pair of upper bounds on p -values, $(\bar{P}_\Gamma, \bar{P}'_{\Gamma'})$, is stochastically greater than the uniform distribution over the square $[0, 1] \times [0, 1]$, which implies that $\Pr(\bar{P}_\Gamma \leq \alpha, \bar{P}'_{\Gamma'} \leq \alpha') \leq \alpha\alpha'$ for all $0 \leq \alpha \leq 1$ and $0 \leq \alpha' \leq 1$.

This proposition implies that from a practical standpoint, it is possible to combine these two analyses $(\bar{P}_\Gamma, \bar{P}'_{\Gamma'})$ as if they were two independent p -values obtained from two uncorrelated analyses.

F An illustration of paper and digital diaries

Figure 11: An illustration of digital diary' selecter

Que faisiez-vous ?

Saisissez une activité 

 Temps personnel Dormir, se laver, manger...	 Travail et études Recherche d'emploi, formations	 Tâches domestiques Ménage, cuisine, courses...
 Soins aux enfants et aux adultes du ménage	 Aide non rémunérée à d'autres ménages	 Interaction sociale Lire, écouter de la musique, jeux...

Figure 12: An illustration of paper diary

Exemple de remplissage de ce carnet

► Notez vos occupations de manière détaillée :


Tâches domestiques. Subdivisez en lessive, vaisselle, raccommodage, etc.

Lecture (sauf études). Précisez ce que vous lisez (journal, roman, etc.).

Trajets. Distinguez les trajets des autres activités.

Travail. Inutile de détailler vos activités durant le travail, inscrivez simplement « je travaille ».

N'oubliez pas de noter les pauses entre les périodes de travail.



Décrivez vos différentes occupations de la journée : Indiquez les heures de début et de fin (plage horaire de l'activité) grâce à une accolade. Décrivez votre occupation.		Faites-vous autre chose en même temps ? (lecture, conversation, radio, TV...)	Avez-vous utilisé un ordinateur ou un smartphone pendant votre activité ? (ou tout autre objet connecté à internet)
18 h	10 } Je rentre du travail avec un collègue	Conversation	<input type="checkbox"/>
	20 }		<input type="checkbox"/>
	30 } Je me repose sur le canapé	J'écoute de la musique	<input checked="" type="checkbox"/>
	40 } Je poste des messages sur mon Facebook		<input checked="" type="checkbox"/>
19 h	50 } Je pars pour le supermarché	Radio	<input type="checkbox"/>
	10 } Je fais des courses pour dîner		<input type="checkbox"/>
	20 }		<input type="checkbox"/>
	30 } Trajet de retour du supermarché	Radio	<input type="checkbox"/>
20 h	40 } Je surveille ma nièce qui fait ses devoirs		<input type="checkbox"/>
	50 } Je prépare le dîner	Je garde ma nièce	<input type="checkbox"/>
	10 }		<input type="checkbox"/>
	20 }		<input type="checkbox"/>
	30 } Je mange avec ma femme, mon frère et ma nièce	Conversation	<input type="checkbox"/>
	40 }		<input type="checkbox"/>
	50 }		<input type="checkbox"/>
		Je range la cuisine	J'écoute la radio
	Je regarde les actualités sur mon téléphone	Je grignote des biscuits	<input checked="" type="checkbox"/>

6

(a) Entering activities

Lieu ou Moyen de transport	Cochez cette case s'il s'agit d'un trajet entre votre domicile et votre travail	En présence de qui (plusieurs réponses possibles)						Votre activité est dans un but (une seule réponse possible)			
		Personnes du ménage					Autres personnes que vous connaissiez	Personnel ou pour mon ménage	Profes- sionnel	Aide à un autre ménage	Bénévole, pour une association
		Seul	Votre conjoint	Votre père, votre mère	Enfant(s) du ménage	Autres personnes du ménage					
En voiture	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Chez moi	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
"	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
À pied	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Au supermarché	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
À pied	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Chez moi	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
"	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
"	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
"	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
"	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
"	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

7

(b) Entering travelling means



Appendix E: SSI Energy Use: Dongle Pilot

Jeldrik Bakker

Maaïke Kompier

Barry Schouten

Remco Paulussen

Daphne Debije

Noël Mingels

Gaston Hayen

CBS Heerlen
CBS-weg 11
6412 EX Heerlen
P.O. Box 4481
6401 CZ Heerlen
+31 45 570 60 00
www.cbs.nl

project number

17th of April, 2025

Index

1.	Introduction	5
1.1	Context and Background	5
1.2	Objectives and Scope	5
1.3	Relevance to Official Energy Statistics	5
1.4	Pilot Setup	6
1.5	Position within the SSI Programme	6
2.	Methodology	7
2.1	Study Design	7
2.2	Data Collection Tools and Procedures	7
2.3	Ethical Considerations and Privacy	8
3.	Data Handling and Preparation	9
3.1	Data Delivery and Format	9
3.2	Data Transformation and Aggregation	9
3.3	Merging and Quality Checks	9
3.4	Normalisation for Visualisation	10
3.5	Diary & Questionnaire Data Processing	10
4.	Results	11
4.1	Participation and Respondent Characteristics	11
4.2	Installation and Ease of Use	11
4.3	Perceptions of Energy Efficiency and Measures Taken	12
4.4	Insights from Diary-Linked Data	13
4.5	Perceptions of Privacy and Occupancy Effects	18
5.	Recommendations for Future Research	19
5.1	Summary of Key Findings	19
5.2	Recommendations	19
6.	Conclusion and Discussion	21
6.1	Feasibility and Added Value	21
6.2	Key Limitations	21
6.3	Looking Forward	21
6.4	Conclusion	22
7.	References	23
8.	Appendix: Testing Materials	Fout! Bladwijzer niet gedefinieerd.

1. Introduction

1.1 Context and Background

The Energy Data Donation (EDD) case study is part of the *Smart Survey Implementation (SSI)* project coordinated by Eurostat. SSI explores how “smart features” — such as sensor data, app-based participation, and data donations — can be implemented in official statistics. Within SSI, three use cases were selected: receipt scanning for the Household Budget Survey (HBS), geotracking for the Time Use Survey (TUS), and EDD for energy statistics. EDD is the final and most exploratory case study, aiming to reach maturity level 2 as defined in SSI deliverable D4.1. This maturity level requires comprehensive evaluation across methodological, IT, logistical, and legal-ethical dimensions.

The EDD case study investigates the feasibility of obtaining detailed energy data directly from households through data donation mechanisms. While the HBS and TUS pilots had clearly defined output requirements based on established ESS regulations, energy statistics remain more fragmented and less standardised. Nonetheless, both national and international stakeholders have a strong and growing interest in improving the frequency, granularity, and scope of household energy data.

The broader objective of EDD is not only to assess the feasibility of smart energy data collection but also to contribute to the generalisation of smart survey methods across domains. Smart surveys require substantial investments in infrastructure, methodology, and organisational workflows. EDD offers a valuable opportunity to assess which design elements can be reused or adapted efficiently in future smart surveys.

1.2 Objectives and Scope

The primary goal of EDD is to explore how household energy data can be collected via smart technologies in a way that is scalable, minimally burdensome, and suitable for statistical production. Specifically, EDD investigates:

- The technical feasibility of collecting energy data through a smart meter dongle.
- The integration of passively collected energy data with self-reported information such as energy diaries and questionnaires.
- The conceptual accuracy of energy donations in relation to official statistical needs.
- The potential for disaggregated statistics, such as separating household energy use from transport-related energy (e.g. electric vehicle charging).
- The practical and ethical implications of implementing smart features at scale.

1.3 Relevance to Official Energy Statistics

Energy statistics are part of the European Statistical System (ESS), with some outputs mandatory and others highly relevant to policy and research. Currently, Statistics Netherlands (CBS) delivers energy statistics on a yearly basis. However, the potential exists to accelerate this to monthly or even near-real-time statistics — especially now that CBS receives monthly meter readings from energy providers for most residential users.

In terms of granularity, ESS requirements demand disaggregation of household energy use into final uses such as heating, cooling, cooking, lighting, and water heating. Additionally, it is mandatory to distinguish energy used for transport (e.g. electric vehicles, battery systems) from general household consumption.

In terms of broadening, very little statistical information currently exists on household energy production via solar panels or use of heat pumps, even though these are rapidly becoming more common. Another emerging topic is indoor temperature monitoring, which is not required under current ESS obligations but is considered highly relevant by national users for understanding comfort and efficiency.

1.4 Pilot Setup

To explore the feasibility of EDD, a small-scale pilot was conducted in the Netherlands. Thirteen CBS employees were recruited to participate, of whom eleven completed the full study. The pilot combined three elements:

- Passive energy data collection via the HomeWizard P1 dongle and app.
- A short questionnaire capturing background information and perceptions.
- A self-reported diary logging presence and appliance usage over eight days.

Participants also exported a second set of meter data after one month to allow for comparison between short- and longer-term data collection.

This pilot focused on electricity and gas, with particular attention to households that produce energy (e.g. via solar panels) or have dynamic consumption patterns. While the sample size was small, the pilot allowed CBS to explore installation barriers, privacy perceptions, data quality, and the potential of combining passive measurements with diary data to distinguish energy usage by device and context.

1.5 Position within the SSI Programme

EDD contributes to the SSI programme as a test case for smart data donation, complementing the receipt scanning and geotracking use cases. As such, the EDD pilot provides critical insights into:

- Push-to-smart recruitment strategies and motivation.
- Respondent interface design and user experience (UI/UX).
- AI/ML potential for disaggregating energy use post-survey.
- Legal and ethical considerations, including privacy-by-design and proportionality.
- Technical requirements for secure and scalable data donation services.

The results of this pilot will contribute directly to the deliverables due in April 2025 and help inform the broader discussion on the future of smart surveys in Europe. By documenting not only what worked but also what needs refinement, EDD plays a key role in establishing best practices and transferable knowledge across national statistical institutes (NSIs).

2. Methodology

2.1 Study Design

Pilot Objectives and Scope

The primary goal of this pilot was to explore the feasibility of collecting detailed energy consumption data using a HomeWizard P1 meter, alongside a self-reported diary. The study aimed to assess the technical ease of collecting passive data, the usefulness of complementary diary information for identifying the use of specific household appliances and explore the feasibility for measuring energy consumption for official statistics.

Participant Recruitment and Sampling

A total of 13 colleagues from Statistics Netherlands (CBS) were initially recruited directly to take part in the pilot. One colleague did not participate after all and another was not able to install the dongle. Because of the limited sample size, the sample distribution was chosen to facilitate diversity in gender, age, and department in order to increase the likelihood of different kinds of energy-using households. While this approach limited the representativeness of the sample, it allowed the research team to gain rapid insights into installation challenges, data quality, and participant experiences.

Data Collection Period

The data collection period was split into two parts. The first part lasted for eight consecutive days. Respondents were asked to keep a diary throughout this period, recording:

- Their presence at home (e.g., which parts of the day they were away or at home).
- Which major electrical appliances were used (e.g., dishwasher, washing machine, tumble dryer).

Upon completing the eight-day data collection, participants were instructed to export the data from the HomeWizard P1 meter and email it to the researchers. The diary was either handed in personally or emailed. The second part ended one month after the first data collection started. Participants were instructed to export the data the HomeWizard P1 meter collected over this period and email it to the researchers as well.

2.2 Data Collection Tools and Procedures

Respondents were provided with the HomeWizard P1 Meter and a booklet containing the pre-questionnaire, installation instructions, diary and post-questionnaire (see Appendix X).

HomeWizard P1 Meter

The P1 meter is a device that connects directly to a digital smart meter via the P1 port. It records electricity and gas usage in 15-minute intervals, including:

- Electricity import (consumption from the grid).
- Electricity export (for households with solar panels).
- Peak usage
- Gas consumption, depending on the availability of gas readings in the smart meter.

The P1 meter was chosen for its relatively simple installation process, which, for most newer smart meters, involved plugging the dongle into the P1 port. However, one respondent with an older smart meter required an additional USB-C cable to complete the setup.

Energy App

Once connected, participants downloaded and linked the HomeWizard Energy app to their P1 meter. The app offered real-time insights into household energy usage, which many found helpful for monitoring daily consumption. At the end of the pilot, participants exported their data via the app. In order to be able to export data, respondents had to sign up for a one month subscription. These costs were reimbursed. The exported data was then sent to the research team for analysis.

Diary Completion

To complement the P1 meter data, participants maintained an energy diary. Each day, they noted:

1. Presence: Whether they were at home or away, allowing for an approximate link between energy usage and occupancy.
2. Appliance Usage: Which major household devices were run, along with estimated times.
3. Special Activities: Any noteworthy events that could explain lower- or higher-than-usual consumption periods.

2.3 Ethical Considerations and Privacy

Data Sensitivity & Informed Consent

Energy usage data can reveal detailed patterns about household occupancy and routines. Therefore all participants were fully informed about the purpose of the study and the types of data collected. They provided consent to share their energy usage data and diary entries. The potential privacy implications were explained, and participants were given the option to withdraw from the study at any time.

Data Security

For this pilot, data transfers relied on participants exporting their usage files via the Energy app and emailing them to the research team. While this approach was adequate for a small-scale pilot, it is not suitable for large-scale population monitoring. A more secure and automated data collection infrastructure would be essential for future expansions.

3. Data Handling and Preparation

3.1 Data Delivery and Format

All respondents submitted their data in comma-separated values (CSV) files. Each respondent provided two pairs of files:

1. After eight days: A separate CSV for electricity and a separate CSV for gas.
2. After one month: The same two-file format (electricity and gas).

Despite using the same HomeWizard P1 meter, individual data files sometimes varied in the number of columns. This was partly due to differences in respondents' home installations and meter configurations (for example, some meters provided additional tariffs or peak wattage information).

- Tariff Classes: Where respondents had multiple tariff classes active (e.g. day and night tariffs), the meter recorded separate readings. The CSV included meter values per tariff class. In most cases, two tariffs were present, but there was no follow-up question regarding potential price differences between those tariffs.
- Export to the Grid: For households with solar panels, the P1 meter recorded exported energy (i.e. electricity generated and delivered back to the grid). The data file thus contained separate columns for import (consumption) and export (production).
- Peak Wattage: A subset of respondents also supplied data on the peak wattage for three "groups." According to HomeWizard, these groups represent the three phases in an electrical installation, though the exact household wiring details were not investigated in this study.

3.2 Data Transformation and Aggregation

To facilitate analysis, several transformations were applied:

1. Electricity Import and Export: The original data provided meter values for each 15-minute interval. To calculate consumption, consecutive meter readings were subtracted, yielding the actual energy used (kWh) in that quarter-hour block. Where day/night tariffs existed, these readings were combined into a single "import" total for each time interval. The same approach was used for export data.
2. Peak Wattage: For respondents reporting peak usage across multiple phases (i.e. three columns of data), the values were summed into a single "peak wattage" column per time interval.
3. Gas Data: Gas readings only contained a one tariff. Following the same logic as electricity import, consecutive readings were subtracted to derive quarter-hourly gas consumption.

3.3 Merging and Quality Checks

Following the transformations, the electricity and gas files were merged into a single dataset for each respondent, matched on the timestamp. To check for completeness and validity, the number of records per respondent for the 8-day period was compared with the expected minimum number of records ($4 \times 24 \times 8 = 768$ records). One respondent had less records than we expected and had mistakenly exported daily totals instead of 15-minute intervals. Some individuals surpassed it (up to 5,280 records), because they had been using the P1 meter for a longer period and chose to export the entire data history.

3.4 Normalisation for Visualisation

To facilitate combined plotting of import, export, gas usage, and peak wattage on a similar scale, the data were normalised per respondent. Electricity Import and Export was divided by the combined maximum value per respondent. This resulted in the maximum kWh per respondent equal 1 and the minimum can't be less than zero. Similarly, gas consumption has been scaled. For Peak Wattage the minimum can be below zero when more energy is created with the solar panels than is being used at that moment. Therefore, it has been divided by the maximum absolute Peak Wattage to account for the negative values. This resulted in a range between -1 and 1. This normalisation enabled straightforward visual comparisons across different types of consumption/production data, without losing relative peaks or patterns within each participant's dataset.

3.5 Diary & Questionnaire Data Processing

All diary entries were transcribed into an Excel file and subsequently imported into R. Participants recorded which appliances they used, the approximate times of use, and the duration. In total 320 time measurements were recorded and 39 contained missing values for the time and/or duration (34 of those from just three respondents). Where the exact usage time was missing, these entries could not be plotted in time-series charts. However, they still contributed to overall counts of appliances usage.

The text descriptions of the appliances were standardised and grouped into broad categories for analysis and plotting:

- Heating/ Air conditioning (AC)
- Electric car charging
- Kitchen (general cooking appliances)
- Washing machine
- Tumble dryer
- Dishwasher
- Other (miscellaneous equipment)

Presence Data

Information on the number of people present in the household was extracted into a separate Excel file, split into four six-hour blocks (night, morning, afternoon, evening). This dataset was also imported into R, allowing for analysis of occupancy patterns in relation to energy consumption peaks.

Questionnaire Data

The questionnaire data were likewise transcribed into Excel and loaded into R. Variable formats (e.g. numeric, categorical) were verified to ensure accurate analyses.

4. Results

4.1 Participation and Respondent Characteristics

Of the 13 CBS colleagues initially recruited, 11 completed the full pilot. One participant inadvertently exported daily instead of quarter-hourly data, which limited their dataset for detailed analysis. The remaining ten provided sufficiently granular data across both the eight-day and one-month periods.

As shown in Table 1, the respondent group was balanced in terms of gender and included variation in age, household size, housing type, and energy label—providing a broad view of energy usage contexts within a small sample.

Table 1. Demographic and household characteristics

Measure	Count
<u>Gender</u>	
Female	5
Male	6
<u>Age</u>	
25-35	2
35-45	4
45-55	3
55-65	2
<u>Household size</u>	
2 persons	4
3 persons	2
4 persons	4
5 persons	1
<u>Housing type</u>	
Detached	5
Semi-detached	3
Terraced	3
<u>Energy Label</u>	
A	1
B/C	4
D/E	0
F/G	1
Don't know	5

4.2 Installation and Ease of Use

Most participants found the HomeWizard P1 meter straightforward to install: ten of the eleven reported it was "easy" or "very easy." One respondent encountered technical issues and required an additional USB-C cable due to an older smart meter. The installation instructions were considered clear by nearly all participants.

Feedback on the energy diary was mixed. Some respondents appreciated the increased awareness it provided, particularly in identifying high-consumption appliances. However, one participant noted that logging appliance usage consistently was burdensome. No technical faults or data quality issues were reported.

4.3 Perceptions of Energy Efficiency and Measures Taken

Before the pilot, participants were asked about their satisfaction with their home’s energy efficiency. Of the 11 who completed the study, five were (very) satisfied, four were neutral and two were dissatisfied.

Those who were satisfied generally had better energy labels or at least were aware of their label, suggesting a link between knowledge, label quality, and satisfaction. Respondents who were less satisfied often had lower energy labels or unknown labels. Since 2008, it is obligatory to have an energy label when buying a house in the Netherlands, meaning that older homes are more likely not to have a label. Furthermore, older houses are less energy efficient. Respondents who reported fewer energy-saving measures frequently felt their homes were already adequately efficient, whereas those less satisfied were often unsure which measures remained viable or beneficial.

Energy-saving measures most frequently taken included purchasing energy-efficient appliances, installing solar panels, and improving insulation (Figure 1). Those who were satisfied with their energy efficiency often reported that no further measures were necessary.

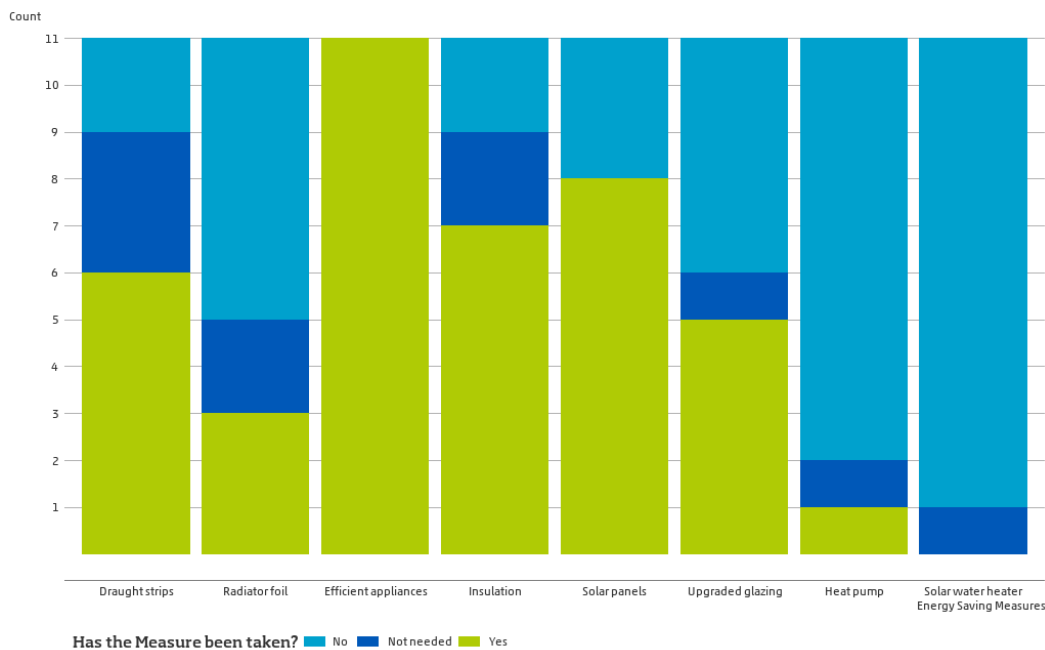


Figure 1. Measures taken to save energy.

The most frequently cited reasons for undertaking these measures were reducing costs, an investment, improving living comfort and it being better for the climate (see Figure 2). Being self-sufficient, a frontrunner or because of the experiences of others seem to play a smaller role.

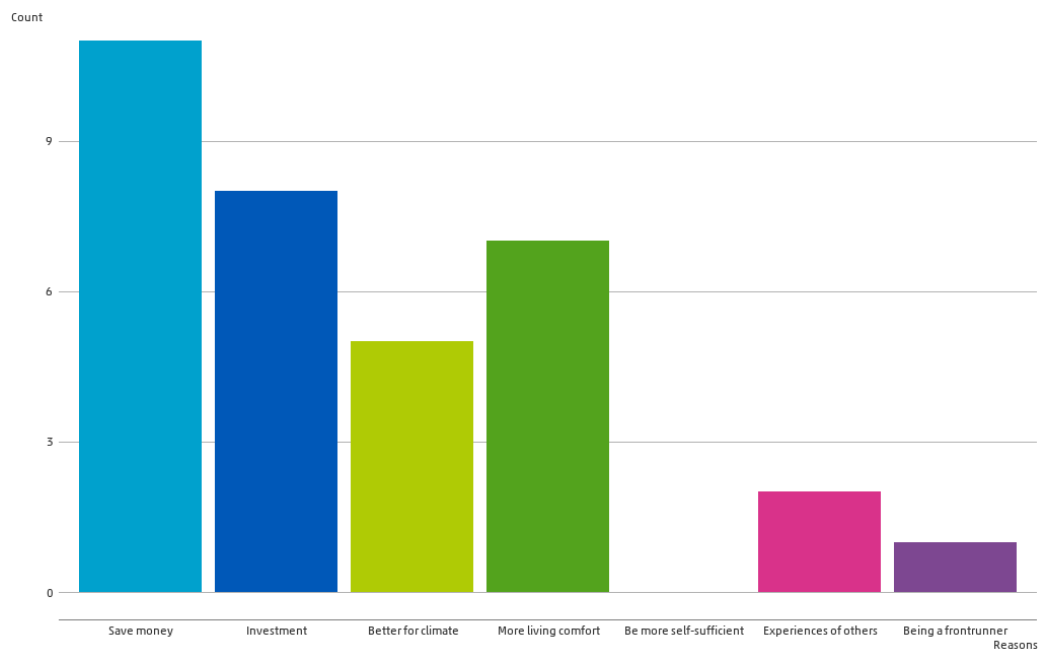


Figure 2. Reasons to save energy

4.4 Insights from Diary-Linked Data

During the first eight days of the study, participants logged the use of 320 energy-consuming appliances. Cooking and washing machines were most frequently reported, followed by dryers and dishwashers (Figure 3). Note that entries marked as “cooling” actually referred to heating via air-conditioning systems during the winter period.

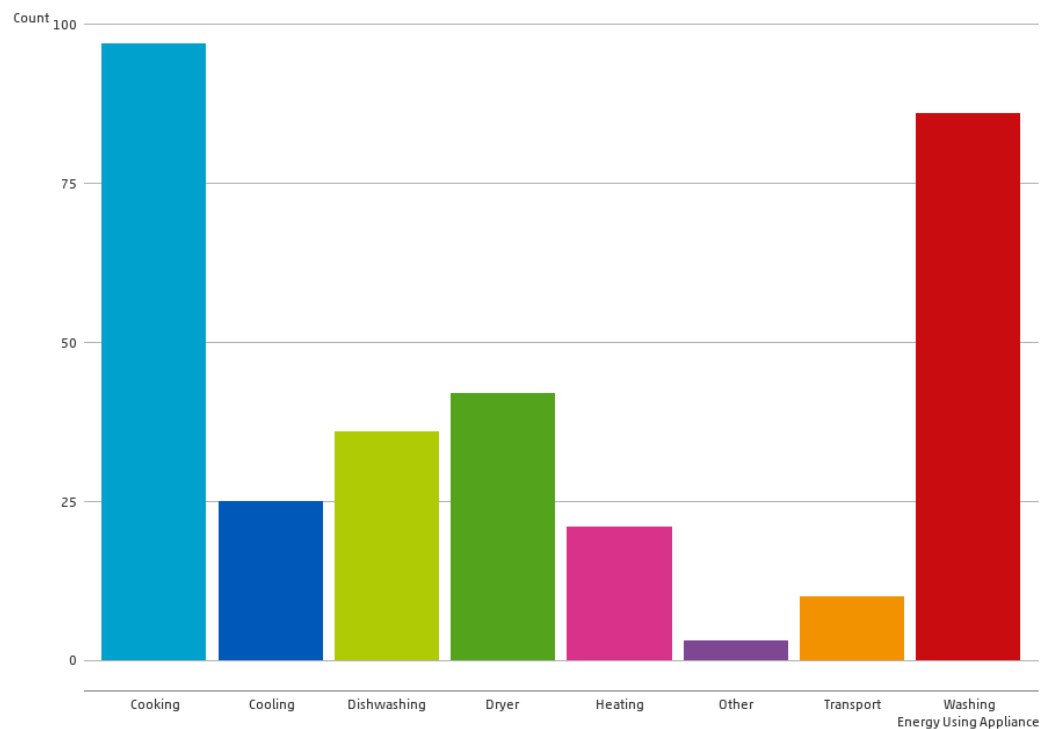


Figure 3. Total reported frequency of energy using appliances according to the diaries

4.4.1 Example Patterns and Insights

To illustrate the potential and limitations of quarter-hourly energy data, several examples were selected from across the pilot participants. These examples reflect common patterns as well as exceptional or complex cases, and help clarify what kinds of insights are feasible with the current setup.

Basic energy use with clear appliance activity

The first example (Figure 4a) shows a relatively straightforward case where energy use can be matched to specific appliances using the diary. Peaks in electricity usage align well with the start of washing machine, dryer, and dishwasher cycles. This is in line with expectations, as the initial phase of these appliances—particularly water heating—often requires the most energy. Cooking activity during dinner also corresponds to a visible peak. Gas consumption shows a strong peak in the morning, likely linked to heating, and smaller peaks at other moments, including during cooking. This example demonstrates how basic disaggregation is possible when a limited number of devices are used, their operation does not overlap and when diary information is available.

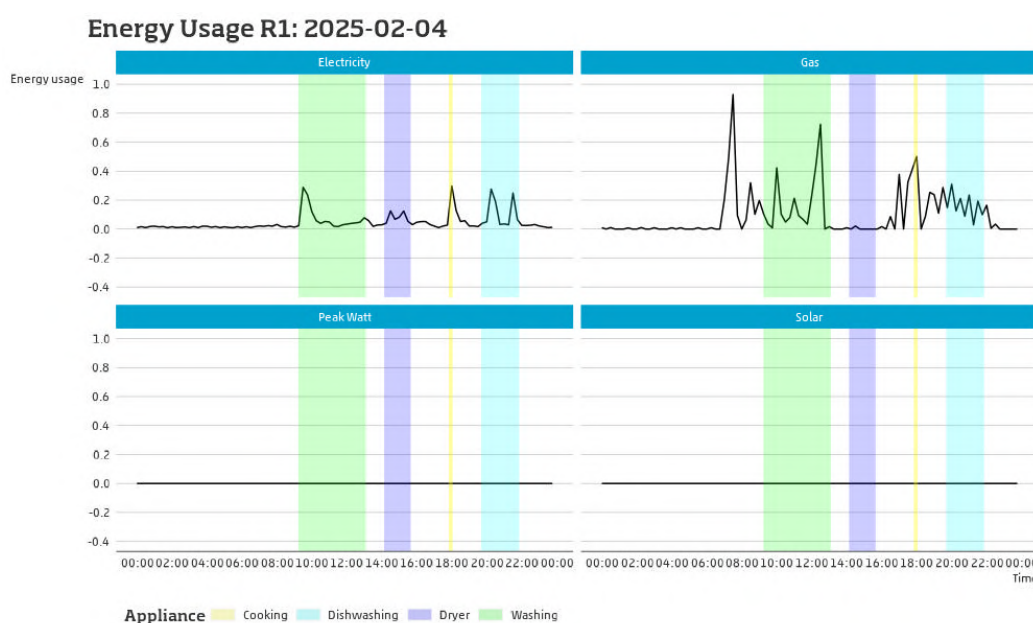


Figure 4a. General energy usage example for gas and electricity

Energy usage with solar production and peak power visibility

In a second example (Figure 4b), solar panel output and peak power data were also available. Solar energy was produced during the day, while most energy consumption occurred in the morning and evening, confirming a temporal mismatch that could overload the energy grid in decentralised systems. Notably, some periodic consumption—likely from a refrigerator or freezer—was only visible in the peak power data, not in the net energy import. During daylight, such appliances were likely powered directly by solar generation. This highlights the added value of peak power metrics for identifying hidden or background loads, especially in solar-enabled households.

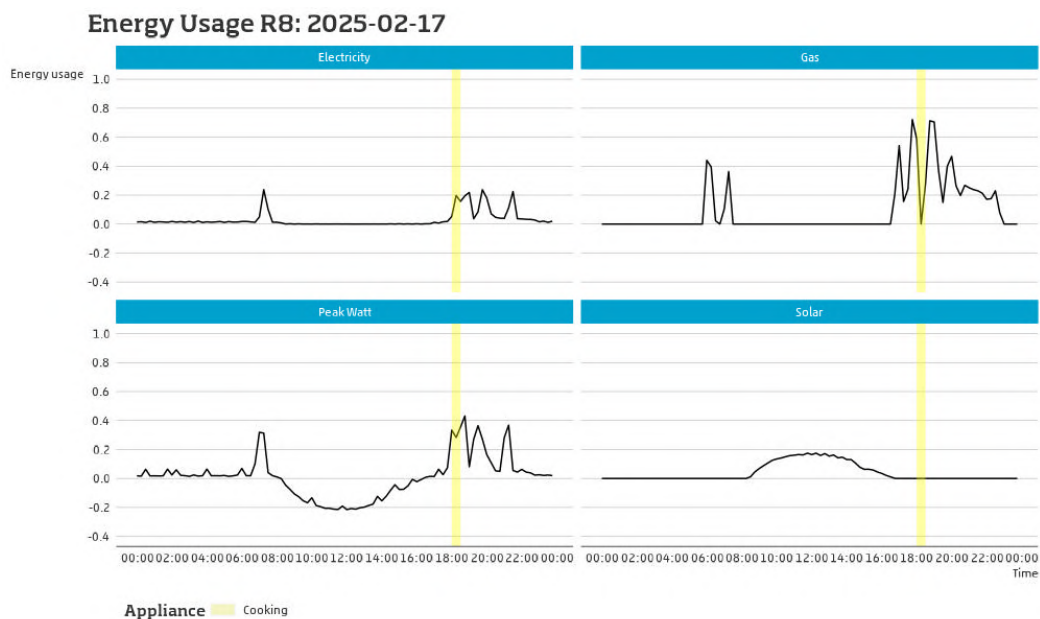


Figure 4b. General energy usage example for gas, electricity, peak power and solar energy.

Complex overlapping energy use patterns

Another example (Figure 4c) presents a much more complex picture, where many appliances were used throughout the day. Even with detailed diary data, it becomes almost impossible to attribute energy use to individual devices due to overlapping activities. This example shows the practical limits of both self-reporting and low-frequency data. It also illustrates the cognitive burden placed on participants when asked to track their energy use manually, especially if multiple appliances are used concurrently or repeatedly. These types of patterns suggest that fully manual methods are not sustainable for long-term or large-scale measurement efforts.

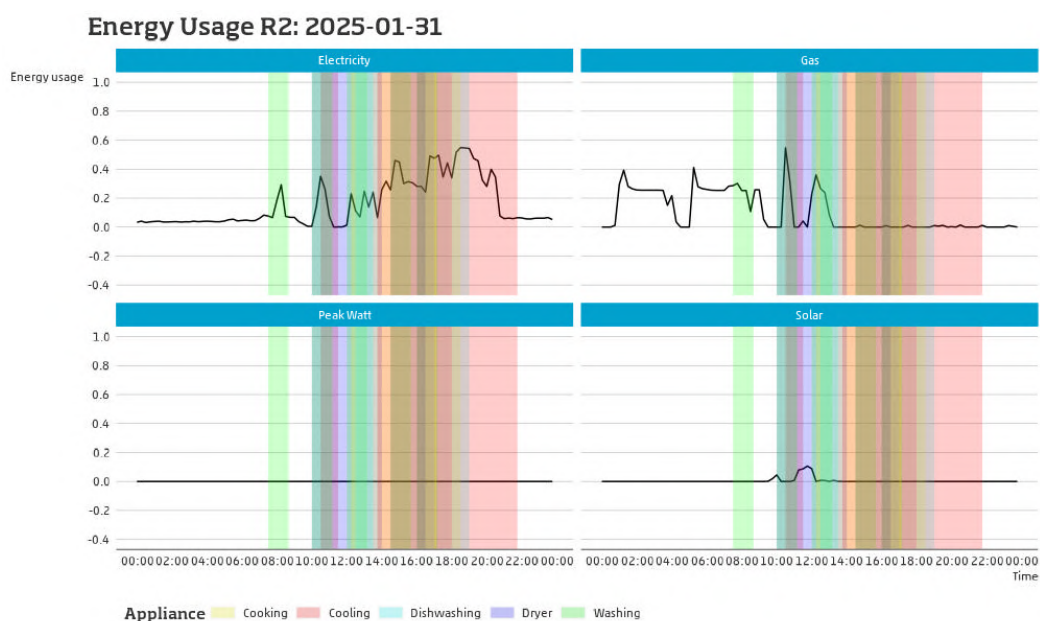


Figure 4c. Example of complexity of energy usage

Heating with air-conditioning systems

A more specific use case is shown in Figure 4d, where an air-conditioning unit was used for heating during a cold January day. This activity is clearly visible in the electricity data and aligns well with diary reports. However, without contextual information—such as temperature data—it would be difficult to determine whether the system was used for heating or cooling. This example underlines the need for integrating external contextual data, such as weather conditions, to improve the interpretation of energy usage.

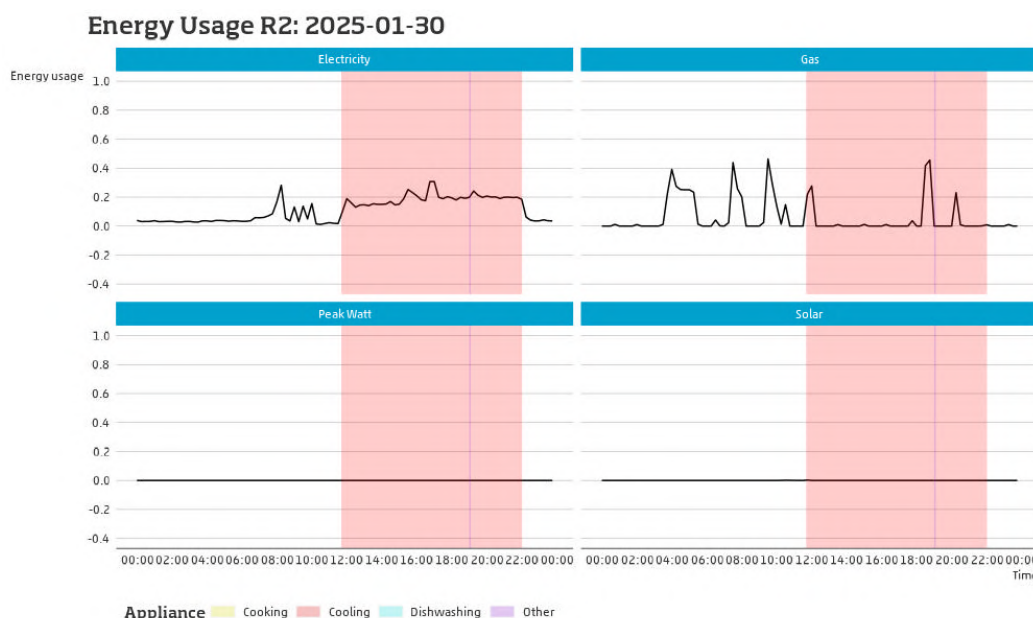


Figure 4d. Example of energy usage of the air-conditioning system for heating with electricity

Combined gas and electric heating

Figure 4e provides an example where gas and electricity usage peaks coincide, suggesting a combined or hybrid heating system. This pattern adds another layer of complexity to disaggregation, as energy usage for a single activity (e.g. heating) may draw from multiple sources. It also shows that energy patterns do not always follow clear one-to-one relationships between devices and energy types.

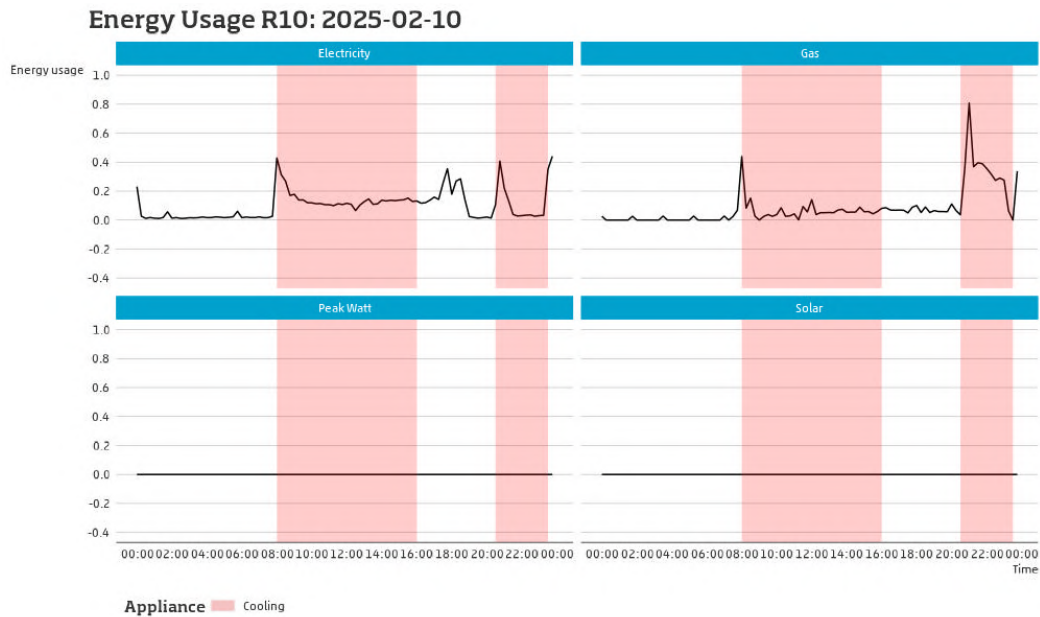


Figure 4e. Example of energy usage of the air-conditioning system for heating with gas and electricity

Irregular and highly volatile usage

Finally, Figure 4f illustrates a case of highly irregular and unpredictable energy use. The shape of the energy profile varies significantly from one day to another, likely driven by a mix of habits, occupancy, and device schedules. This example reinforces the idea that household energy behaviour is highly individual, and that generic disaggregation models are unlikely to perform well across all households without some form of local adaptation or calibration.

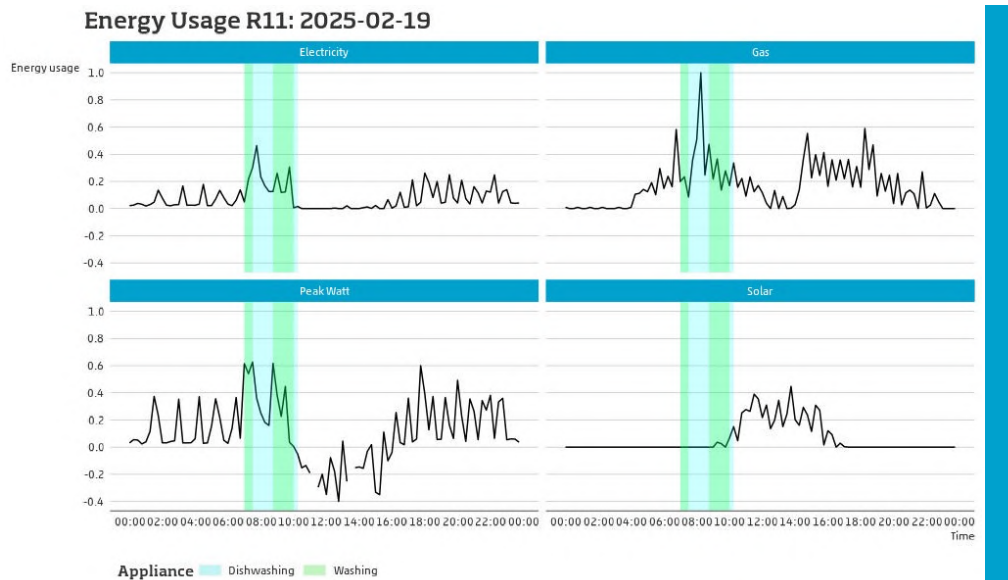


Figure 4f. Example the irregularity of energy data.

Appliance Peaks

These examples collectively show that 15-minute interval data allows for general consumption pattern analysis but lacks the granularity needed to reliably identify individual appliances. The addition of solar output and peak power metrics enhances interpretability in certain contexts, especially when devices draw power directly from solar generation. However, to move towards automated disaggregation, higher-frequency data and machine learning models—trained and possibly validated with user input—are essential.

Behavioural Change

Furthermore, three respondents reported adapting their habits during the study. For example, they scheduled energy-intensive appliances for daytime hours to optimise solar panel output or to avoid overlapping multiple appliances at once.

4.5 Perceptions of Privacy and Occupancy Effects

Half of the participants raised concerns about privacy, pointing out that real-time consumption data could reveal whether a household member was present or absent—especially in single-person dwellings.

Occupancy Effects

Most households saw greater electricity consumption in the mornings and evenings, aligning with diary reports of higher occupancy during those periods (see Figure 3). In one notable instance, a respondent recorded an unusually high number of occupants for one evening although this effect was not clearly visible in the energy use. It seems like energy consumption is mainly tied to that the number of occupants is less relevant to the energy consumption.

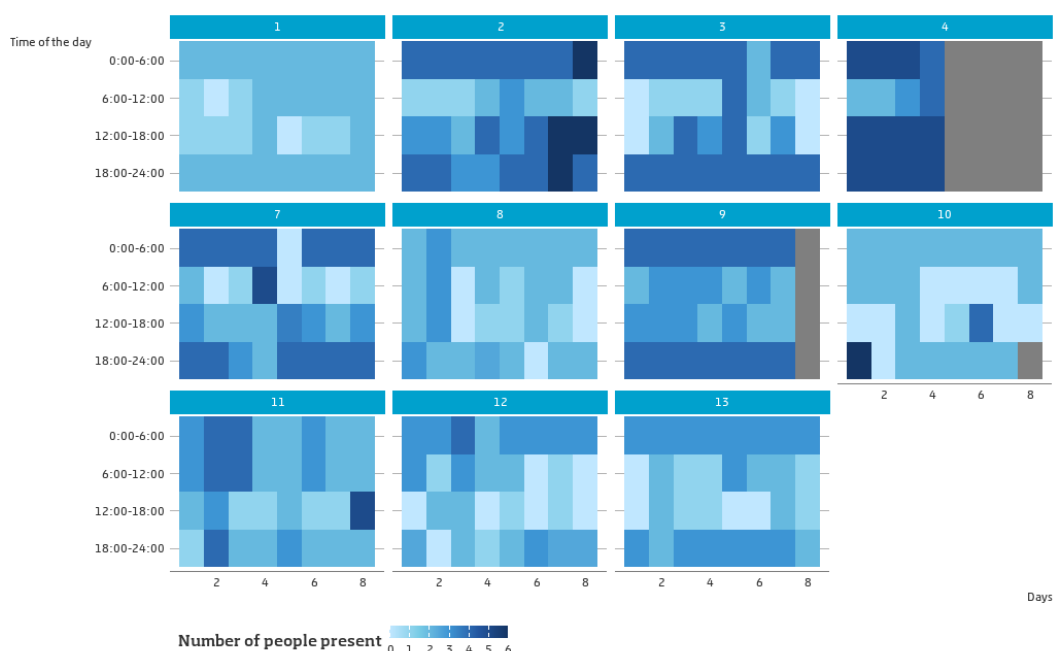


Figure 5. Number of people present during the pilot per respondent

5. Recommendations for Future Research

5.1 Summary of Key Findings

This pilot demonstrates that passive data collection through a smart meter dongle is technically feasible and well-received by participants. Once installed, the HomeWizard P1 meter required little effort from participants, and the accompanying app provided real-time insights that many found both engaging and informative. However, the study also highlighted critical limitations in using this method for producing detailed energy statistics that meet current and future statistical demands.

The most significant limitation is the lack of granularity: the data collected in 15-minute intervals does not provide enough resolution to automatically detect which appliances are being used or to disaggregate energy use by type (e.g. heating, lighting, cooking). This limitation prevents the transition to fully automated, diary-free statistics, which is a key long-term goal.

Moreover, while participants generally managed to install the dongle, some required assistance, and the current approach of manual data exports via email is not scalable. The diary element should be reduced to a minimum due to the response burden and privacy was also a concern, with several participants expressing discomfort about the personal nature of energy use patterns.

5.2 Recommendations

Based on the pilot results, the following recommendations are made for the continuation and scaling of the Energy Data Donation (EDD) initiative:

1. Do Not Scale Up the Current Setup

While the HomeWizard system served as a useful prototype, it is not suitable for population-wide monitoring. The reliance on a commercial third party introduces risks related to data control, transparency of preprocessing steps, and potential changes in the service offering. For official statistics, it is crucial that the entire data pipeline is transparent and under the control of the NSI.

2. Develop an In-House Solution

To meet privacy, quality, and operational standards, it is recommended to develop a proprietary dongle and accompanying mobile app. These should:

- Be user-friendly, modelled after the successful UX of the HomeWizard app.
- Allow for secure, automated data transfer (e.g. via a microservice).
- Include local processing where feasible to enhance privacy and minimise data transmission.
- Support software updates and model learning (e.g. for improved disaggregation over time).

3. Aim for a Rotating Energy Panel

A rotating Energy Panel seems to be the best way to measure this in the future. This panel should entail energy, gas, and possibly water data for a fixed period (e.g. one year). A dedicated onboarding phase—possibly supported by field interviewers—will be crucial for ensuring high participation and low selection bias. Interviewers can assist with the installation where needed and provide clear trustworthy communication about privacy, purpose, and societal value. Based on behavioural adaptation observed in this and similar studies, energy use patterns are likely to be temporarily affected by awareness during the first weeks. Therefore, the measurement period

should be long enough for participants to return to typical behaviour, allowing for more reliable long-term data. A one-year cycle per household—with potential seasonal effects captured—is recommended, followed by rotation. This will likely reduce the effect of increased energy awareness, but not completely mitigate it.

4. Include High-Frequency Measurement in Next Pilot

To determine the optimal data collection frequency, a follow-up study with higher-frequency sensors is essential. The work of Froehlich et al. (2009) on HydroSense shows that individual water appliances can be detected with 97.9% accuracy using pressure-based measurements sampled at up to 1,000 times per second. While this frequency may be more than necessary for energy use, it provides a useful upper bound. The goal is to identify a minimal viable frequency that balances accuracy with data minimisation.

5. Expand Scope to Include Water

Although smart water meters are not yet standard in the Netherlands, they are under development. Collaboration with water utilities can help prioritise installation of smart meters in panel households. This makes it possible to integrate water, gas, and electricity into one comprehensive resource use panel. This aligns with broader sustainability monitoring goals and will enrich modelling capabilities.

6. Address Shared Infrastructure and Multi-Household Contexts

In apartment buildings or homes with shared meters, individual household data cannot be isolated through the meter alone. However, research by Froehlich et al. (2009) suggests that aggregate signals can still be decomposed with sufficient accuracy, especially with training data and machine learning models. A subset of the panel should be selected to explicitly study this challenge.

7. Adopt Privacy-by-Design Principles

Because energy data can reveal detailed patterns about household routines, privacy risks must be taken seriously. The following measures are recommended:

- Local data processing, where possible, to ensure only required statistics on an aggregated level leave the household.
- Clear and comprehensive participant education, going beyond standard consent forms.
- A DPIA should be performed as soon as the technical and methodological design is drafted.

8. Combine ML with Participant Input

To minimize reliance on diary data while still maintaining accuracy, a hybrid system could be developed where a machine learning model makes initial predictions about appliance usage and the user is able to validate and adjust. Occasionally, the app can prompt users to validate or correct these predictions. This approach mirrors the mobility survey methods already in use and may serve as an efficient calibration mechanism.

6. Conclusion and Discussion

This pilot study marks an important first step in exploring the feasibility of using passively collected energy data for official statistics through the concept of Energy Data Donation (EDD). As the final and most exploratory case within the Smart Survey Implementation (SSI) project, EDD aimed to test whether real-time household energy data, collected via consumer-grade smart meter dongles, could reduce the reliance on traditional surveys while improving both the granularity and frequency of energy statistics.

6.1 Feasibility and Added Value

The study demonstrates that passive energy data collection is both technically feasible and acceptable to participants. The HomeWizard P1 dongle was relatively easy to install, and the accompanying app provided a level of real-time feedback that participants appreciated. Importantly, participants reported increased awareness of their energy use, and some even adapted their behaviour accordingly.

From a data integration perspective, it was possible to combine quarter-hourly gas and electricity usage with self-reported diary data on presence and device usage. This allowed for illustrative analyses of peak usage patterns and presence-activity correlations. However, despite this progress, the pilot also made clear that the data collected at 15-minute intervals is not sufficiently detailed to automatically distinguish between types of energy usage or devices. Diaries remain essential for interpretation, limiting the scalability and automation potential of the current method.

6.2 Key Limitations

Several limitations temper the results. First and foremost, this was a small-scale, convenience-based pilot with 11 complete cases. While sufficient for prototyping and identifying issues, the sample is not representative. Moreover, data was collected manually, with respondents emailing their exported files, which is impractical for any kind of large-scale rollout.

Second, the involvement of a commercial third party introduces data governance concerns. For statistical production, it is crucial that all data pipelines are under the control of the statistical office to ensure transparency, reproducibility, and long-term stability.

Third, while privacy concerns did not deter participation in this study, several respondents expressed unease about how detailed energy data could be used to infer personal routines. This reinforces the importance of adopting strong privacy-by-design measures, particularly as data frequency increases.

6.3 Looking Forward

To advance EDD from a conceptual pilot to a robust, production-ready approach, several strategic investments and design choices are needed:

- A high-frequency follow-up pilot must be conducted to determine the optimal sensor resolution required for device-level disaggregation. Insights from earlier research (e.g. Froehlich et al., 2009) suggest that much finer sampling may be needed than currently used. Furthermore, this pilot should be used to develop self-learning algorithm that reduce respondent burden over time. Finally the algorithm needs to be validated on the

app needs to be tested. This can be done in the same pilot or spread across multiple pilots. Careful attention needs to be given on the detection of behavioural change due to monitoring and how to adjust for this.

- A rotating national energy panel is the most likely end goal, possibly including water usage, with careful attention paid to participant recruitment, onboarding, and long-term engagement.
- A proprietary dongle and data infrastructure must be developed to ensure end-to-end control of the data collection process, ethical compliance, and integration into existing statistical workflows. An important element here is localized data aggregation and data minimisation strategies.
- Privacy must remain central to the design. Participants need to understand why data is collected, how it will be used, and what protections are in place. Transparent communication and participatory design will be essential in maintaining trust.

6.4 Conclusion

The EDD pilot has shown what is possible—and what is still out of reach. Passive measurement technologies hold great promise for reducing respondent burden and enabling more timely and detailed statistics. However, to realise this potential, more research is needed. Investments are needed for purpose-built infrastructure, privacy-conscious algorithms, and robust methodological frameworks.

7. References

Eurostat (2023). Smart Survey Implementation: Grant Agreement Number: 101119594.

Froehlich, J., Larson, E., Saba, E., Campbell, T., Atlas, L., Fogarty, J., & Patel, S. (2009).
HydroSense: Infrastructure-Mediated Single-Point Sensing of Whole-Home Water Activity.

Energiemeter onderzoek

Naam:



Fijn dat je ons wilt helpen bij het uittesten van dit onderzoek.

In dit boekje zitten alle materialen die je nodig hebt voor het onderzoek, behalve de P1 meter. Die heb je los ontvangen. We willen graag onderzoeken of we met deze P1 meter data kunnen verzamelen om nieuwe energiestatistieken te maken.

We willen je vragen om dit onderzoek thuis serieus uit te proberen – dus gewoon meedoen zoals je zou doen als je echt zou meedoen aan dit onderzoek. En als jou daarbij zaken opvallen horen we dat graag. Alle opmerkingen zijn welkom. Noteer ze s.v.p. in de evaluatie vragenlijst.

Als bedankje mag je de P1 meter houden aan het einde van het onderzoek.

Graag dit boekje na afloop meenemen naar kantoor in Heerlen om in te leveren.

Mocht je vragen hebben die niet tot het einde van het onderzoek kunnen wachten, stuur dan een Slack bericht naar Maaïke Kompier of mail naar userlab@cbs.nl.

Hartelijk bedankt!

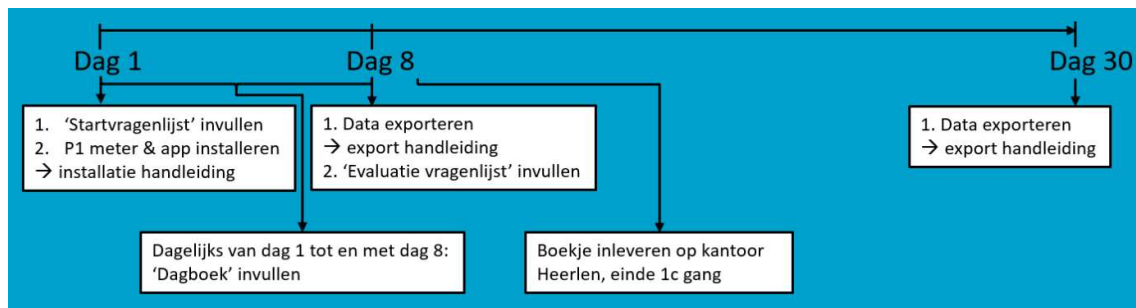
Inhoud van dit boekje:

1. Instructies	2
2. Startvragenlijst	3
3. Installatie handleiding	5
4. Dagboek	17
5. Export handleiding	25
6. Evaluatie vragenlijst	29

1. Instructies

In dit onderzoek verzamelen we data met een P1 meter voor de periode van 1 maand.

1. Vul eerst de **Startvragenlijst** in.
2. Volg daarna de **Installatie handleiding** om de P1 meter te installeren en de bijbehorende app te downloaden.
3. Vul vervolgens iedere dag het **Dagboek** in gedurende 8 dagen.
4. Volg na 8 dagen je de **Export handleiding** om de data te exporteren en met ons te delen (mail naar userlab@cbs.nl)
5. Vul ten slotte de **Evaluatie vragenlijst** in.
6. Neem het boekje mee naar kantoor in Heerlen en lever het in. We bespreken graag kort na hoe je het hebt ervaren.
7. Exporteer na 30 dagen de data nog eens en deel deze weer met ons (mail naar userlab@cbs.nl).



2. Startvragenlijst

We hebben een paar korte vragen voor je, om de resultaten later beter te kunnen duiden.

1. In wat voor soort woning woon je?

Als je in meer dan één woning woont, omschrijf dan degene waar je het meest tijd doorbrengt.

- ☐ Vrijstaande woning
- ☐ Half-vrijstaande woning
- ☐ Rijtjeshuis, tussen- of hoekwoning
- ☐ Boven- of benedenwoning
- ☐ Appartement of flat
- ☐ Ander soort woning

2. Wat is het energielabel van je woning?

- ☐ A, A+, of A++
- ☐ B of C
- ☐ D of E
- ☐ F of G
- ☐ Ik weet het energielabel van mijn woning niet.

3. Hoe tevreden of ontevreden ben je over het algemeen met je woning als het gaat om energiezuinigheid?

- ☐ Heel tevreden
- ☐ Tevreden
- ☐ Niet tevreden en niet ontevreden
- ☐ Ontevreden
- ☐ Heel ontevreden

4. Welke maatregelen zijn er genomen om te besparen op het energieverbruik?

	Ja	Nee	Niet nodig	NVT
Tochtstrips geplaatst of kieren gedicht				
Radiatorfolie aangebracht				
Zuinigere apparatuur of lampen gekocht				
Isolatie aangebracht aan dak, muren of vloer				
Zonnepanelen laten installeren				
Glas vervangen				
Warmtepomp geïnstalleerd				
Zonneboiler geïnstalleerd				
Andere maatregelen, namelijk: ...				

5. Wat waren de belangrijkste redenen om maatregelen te nemen om te besparen op jouw energieverbruik?

U kunt meerdere antwoorden kiezen.

- ☐ Geld besparen (op energierekening)
- ☐ Goede investering voor de woning
- ☐ Beter voor het milieu of klimaat
- ☐ Meer wooncomfort
- ☐ Minder afhankelijk zijn van energiebedrijven
- ☐ Ervaringen van burens, familie of vrienden.
- ☐ Ik wil hier graag in voorop lopen.
- ☐ Anders, namelijk: ...
- ☐ Niet van toepassing, ik heb geen maatregelen genomen

6. Jezelf meegerekend, uit hoeveel personen bestaat dan het huishouden waartoe je behoort?

Kinderen die niet op dit adres staan ingeschreven, NIET tot het huishouden rekenen.

7. Wat is jouw geboortedatum?

8. Wat is je geslacht?

- ☐ Man
- ☐ Vrouw
- ☐ Overig
- ☐ Wil ik liever niet zeggen

3. Installatie handleiding

Het installeren bestaat uit 3 stappen:

1. De P1 meter aansluiten
2. De HomeWizard Energy app installeren
3. Een Energy+ abonnement afsluiten in de HomeWizard Energy app

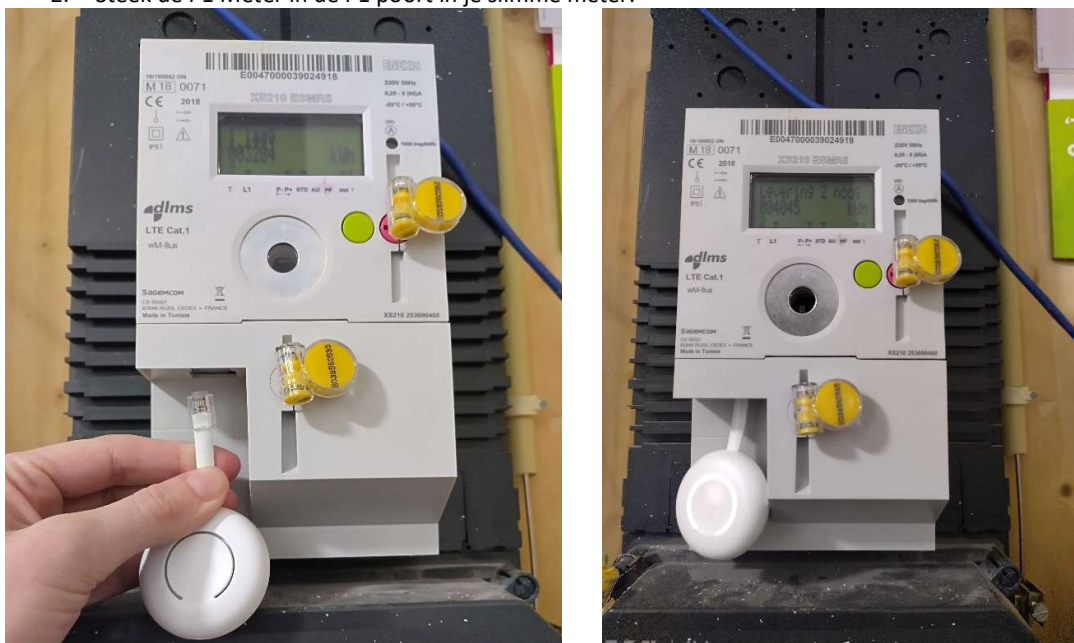
Mocht je tijdens de installatie vast lopen, stuur dan een bericht naar Maaïke Kompier via Slack.

1. Stap 1. De P1 meter aansluiten

1. Haal de P1 meter uit het doosje.



2. Steek de P1 Meter in de P1 poort in je slimme meter.

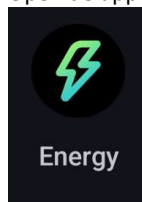


2. De HomeWizard Energy app installeren

1. Download de HomeWizard Energy app met onderstaande QR code:

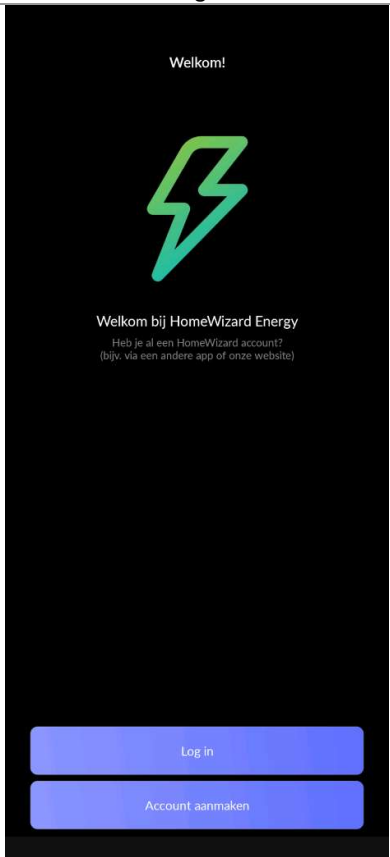


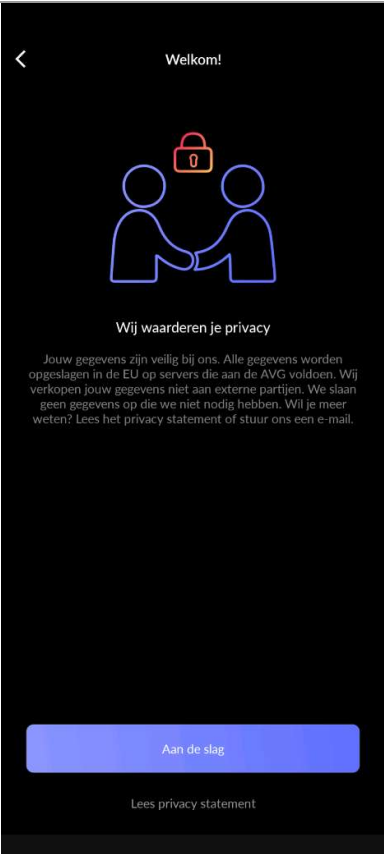
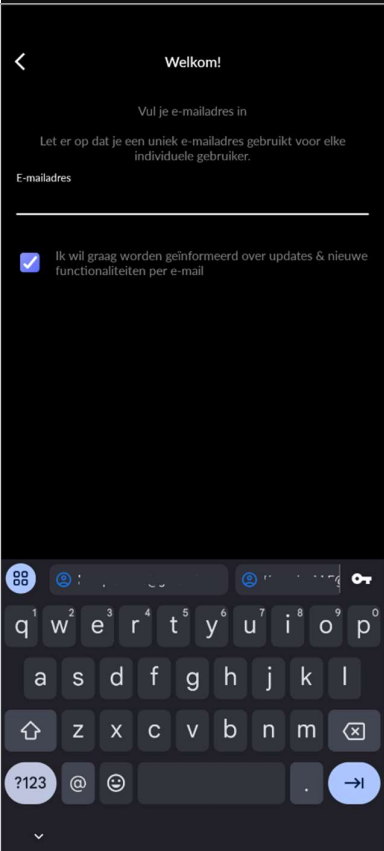
2. Open de app

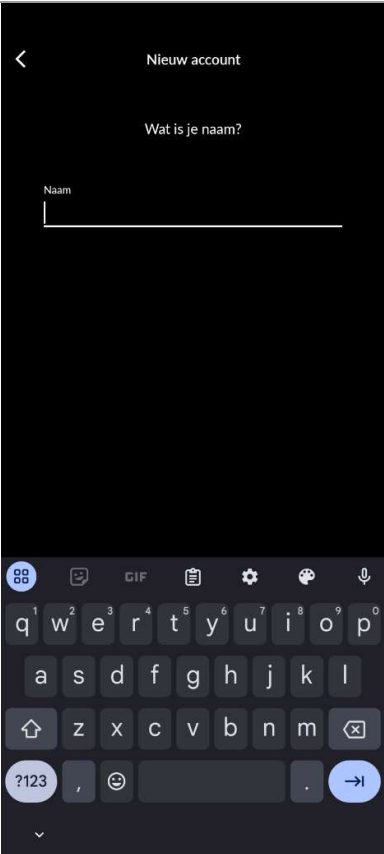





3. Volg de instructies in de app

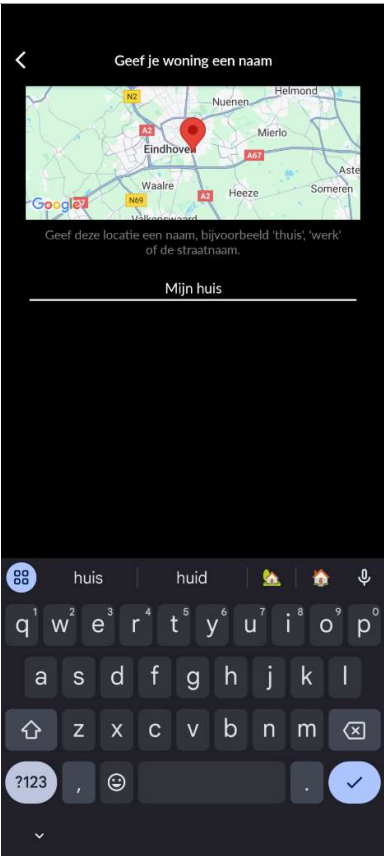
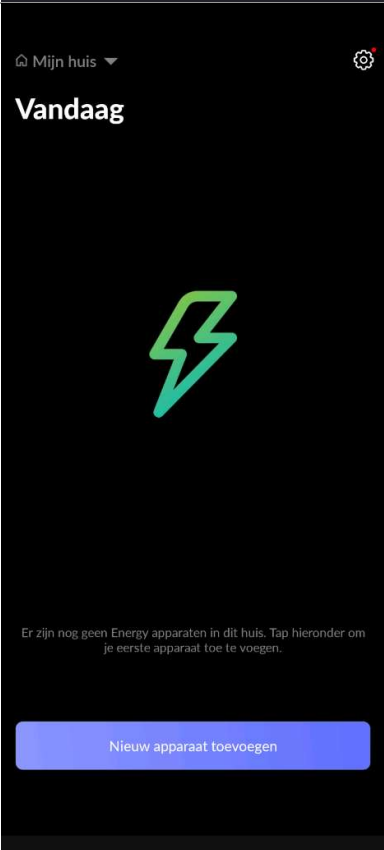
Hieronder volgen de schermabeeldingen van de instructies in de app.

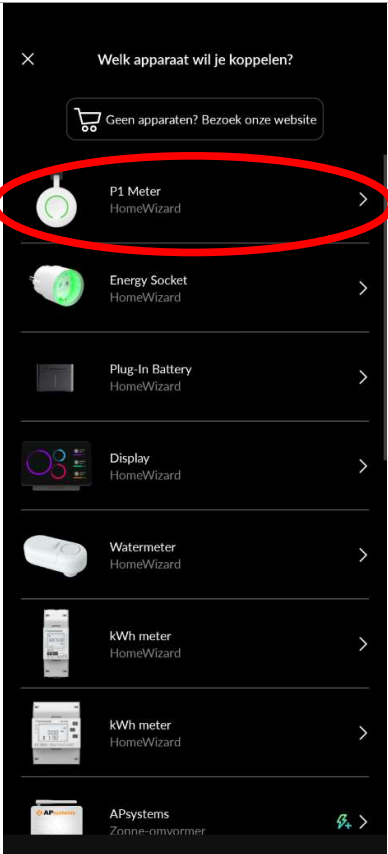

 The screenshot shows the welcome screen of the HomeWizard Energy app. At the top, it says 'Welkom!' in white. Below that is a large green lightning bolt icon. Under the icon, it says 'Welkom bij HomeWizard Energy' and 'Heb je al een HomeWizard account? (bijv. via een andere app of onze website)'. At the bottom, there are two blue buttons: 'Log in' and 'Account aanmaken'.	<p>Druk op Account aanmaken</p>
--	---------------------------------




	<p>Druk op Aan de slag</p>
	<p>Geef een e-mailadres op waarop je je inloggegevens wil ontvangen en druk op enter</p>

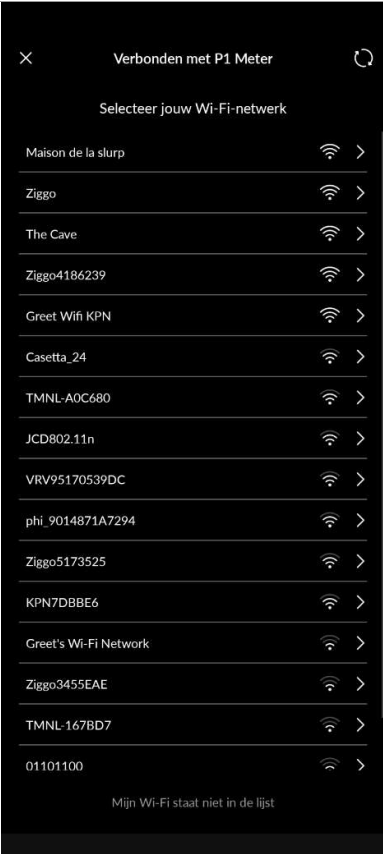

	<p>Geef een naam op en druk op enter</p>
	<p>Druk op Volgende</p>

 <p>Jouw locatie</p> <p>Waar woon je?</p> <p>We hebben je locatie nodig om de app in te stellen op de juiste tijdzone en zomer/wintertijd zodat we jouw grafieken correct kunnen weergeven.</p> <p>Gebruik huidige locatie</p> <p>Kies locatie handmatig</p>	<p>Kies de optie die je wil</p>
 <p>Nieuw huis aanmaken?</p> <p>Je eerste Energy apparaat aan het installeren? We maken een huis voor je aan en helpen je op weg.</p> <p>Heeft een ander lid van je huishouden al een of meerdere Energy apparaten ingesteld? Ze kunnen je uitnodigen zodat jullie beiden toegang hebben.</p> <p>Huis toevoegen</p> <p>Ontvang een uitnodiging</p>	<p>Druk op Huis toevoegen</p>

	<p>Geef je huis een naam en druk op enter</p>
	<p>Druk op Nieuw apparaat toevoegen</p>

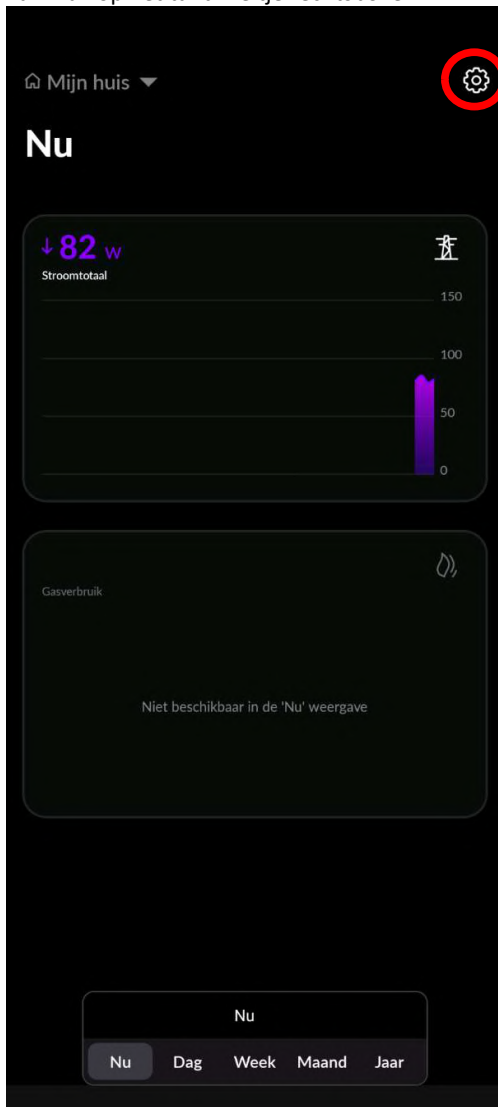
 <p>Welk apparaat wil je koppelen?</p> <p>Geen apparaten? Bezoek onze website</p> <p>P1 Meter HomeWizard</p> <p>Energy Socket HomeWizard</p> <p>Plug-In Battery HomeWizard</p> <p>Display HomeWizard</p> <p>Watermeter HomeWizard</p> <p>kWh meter HomeWizard</p> <p>kWh meter HomeWizard</p> <p>APsystems Zonne-omvormer</p>	<p>Druk op P1 Meter</p>
 <p>P1 Meter toevoegen</p> <p>Zorg ervoor dat je P1 Meter aan staat.</p> <p>Klik de P1 Meter met de P1 poort van je slimme meter. Als het status lichtje van de P1 Meter niet gaat branden, verbindt dan een externe Micro USB-adapter aan de USB poort van je P1 Meter.</p> <p>Volgende</p>	<p>Druk op Volgende</p>

 <p>P1 Meter toevoegen</p> <p>Koppel apparaat</p> <p>Houd de knop op de P1 Meter ingedrukt totdat deze blauw oplicht.</p> <p>Volgende</p>	 <p>Druk de knop op de P1 meter in tot hij blauw is en druk dan op Volgende</p>
 <p>Verbinding maken met apparaat</p> <p>De app Energy wil een tijdelijk wifi-netwerk gebruiken om verbinding te maken met je apparaat.</p> <p>P1meterSetup</p> <p>Annuleren Verbinden</p>	<p>Druk op Verbinden</p>

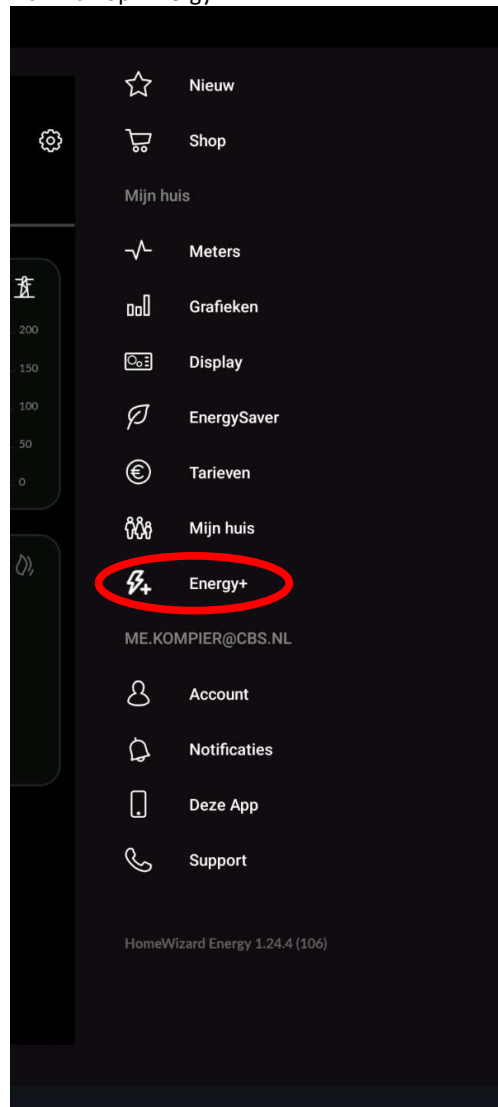
	<p>Druk op de naam van het gewenste Wi-Fi netwerk</p>
	<p>Je bent nu klaar met het instellen van de app.</p>

3. Een Energy+ abonnement afsluiten in de HomeWizard Energy app

1a. Druk op het tandwielletje rechtsboven



1b. Druk op 'Energy+'



2. Sluit een Energy+ abonnement af voor 1 maand. Deze kosten kun je declareren op het project PR002608.



4. Dagboek

Dag 1 – Datum:

1. Aanwezigheid in huis

Vul in hoeveel personen er op welk moment thuis waren. Dit geeft ons een beter beeld van de momenten waarop het energieverbruik waarschijnlijk hoger ligt.

Tijd	Personen thuis
Nacht (0:00 - 6:00)	
Ochtend (6:00 - 12:00)	
Middag (12:00 - 18:00)	
Avond (18:00 - 24:00)	

2. Gebruik van apparaten

Noteer elk groot apparaat dat je hebt aangezet, samen met het tijdstip en eventuele opmerkingen.

Je hoeft niet alle apparaten die je aan of uitzet te noteren, alleen de apparaten die veel energie verbruiken zoals een vaatwasser, kookplaat, oven, kachel, wasmachine of droger.

Tijdstip	Apparaat	Duur gebruik (indien bekend)	Opmerkingen

3. Speciale activiteiten

Noteer hier extra informatie over deze dag, die mogelijk invloed heeft gehad op jouw energieverbruik.

Bijvoorbeeld een bijzonder warme of koude dag, onderhoudswerkzaamheden of koken voor gasten.

4. Algemene opmerkingen

Noteer hier eventuele andere opmerkingen die je hebt.

Bijvoorbeeld over het gebruik van de P1 meter, problemen die je bent tegengekomen of het onderzoeksproces.

Dagboek

Dag 2 – Datum:

1. Aanwezigheid in huis

Vul in hoeveel personen er op welk moment thuis waren. Dit geeft ons een beter beeld van de momenten waarop het energieverbruik waarschijnlijk hoger ligt.

Tijd	Personen thuis
Nacht (0:00 - 6:00)	
Ochtend (6:00 - 12:00)	
Middag (12:00 - 18:00)	
Avond (18:00 - 24:00)	

2. Gebruik van apparaten

Noteer elk groot apparaat dat je hebt aangezet, samen met het tijdstip en eventuele opmerkingen.

Je hoeft niet alle apparaten die je aan of uitzet te noteren, alleen de apparaten die veel energie verbruiken zoals een vaatwasser, kookplaat, oven, kachel, wasmachine of droger.

Tijdstip	Apparaat	Duur gebruik (indien bekend)	Opmerkingen

3. Speciale activiteiten

Noteer hier extra informatie over deze dag, die mogelijk invloed heeft gehad op jouw energieverbruik.

Bijvoorbeeld een bijzonder warme of koude dag, onderhoudswerkzaamheden of koken voor gasten.

4. Algemene opmerkingen

Noteer hier eventuele andere opmerkingen die je hebt.

Bijvoorbeeld over het gebruik van de P1 meter, problemen die je bent tegengekomen of het onderzoeksproces.

Dagboek

Dag 3 – Datum:

1. Aanwezigheid in huis

Vul in hoeveel personen er op welk moment thuis waren. Dit geeft ons een beter beeld van de momenten waarop het energieverbruik waarschijnlijk hoger ligt.

Tijd	Personen thuis
Nacht (0:00 - 6:00)	
Ochtend (6:00 - 12:00)	
Middag (12:00 - 18:00)	
Avond (18:00 - 24:00)	

2. Gebruik van apparaten

Noteer elk groot apparaat dat je hebt aangezet, samen met het tijdstip en eventuele opmerkingen.

Je hoeft niet alle apparaten die je aan of uitzet te noteren, alleen de apparaten die veel energie verbruiken zoals een vaatwasser, kookplaat, oven, kachel, wasmachine of droger.

Tijdstip	Apparaat	Duur gebruik (indien bekend)	Opmerkingen

3. Speciale activiteiten

Noteer hier extra informatie over deze dag, die mogelijk invloed heeft gehad op jouw energieverbruik.

Bijvoorbeeld een bijzonder warme of koude dag, onderhoudswerkzaamheden of koken voor gasten.

4. Algemene opmerkingen

Noteer hier eventuele andere opmerkingen die je hebt.

Bijvoorbeeld over het gebruik van de P1 meter, problemen die je bent tegengekomen of het onderzoeksproces.

Dagboek

Dag 4 – Datum:

1. Aanwezigheid in huis

Vul in hoeveel personen er op welk moment thuis waren. Dit geeft ons een beter beeld van de momenten waarop het energieverbruik waarschijnlijk hoger ligt.

Tijd	Personen thuis
Nacht (0:00 - 6:00)	
Ochtend (6:00 - 12:00)	
Middag (12:00 - 18:00)	
Avond (18:00 - 24:00)	

2. Gebruik van apparaten

Noteer elk groot apparaat dat je hebt aangezet, samen met het tijdstip en eventuele opmerkingen.

Je hoeft niet alle apparaten die je aan of uitzet te noteren, alleen de apparaten die veel energie verbruiken zoals een vaatwasser, kookplaat, oven, kachel, wasmachine of droger.

Tijdstip	Apparaat	Duur gebruik (indien bekend)	Opmerkingen

3. Speciale activiteiten

Noteer hier extra informatie over deze dag, die mogelijk invloed heeft gehad op jouw energieverbruik.

Bijvoorbeeld een bijzonder warme of koude dag, onderhoudswerkzaamheden of koken voor gasten.

4. Algemene opmerkingen

Noteer hier eventuele andere opmerkingen die je hebt.

Bijvoorbeeld over het gebruik van de P1 meter, problemen die je bent tegengekomen of het onderzoeksproces.

Dagboek

Dag 5 – Datum:

1. Aanwezigheid in huis

Vul in hoeveel personen er op welk moment thuis waren. Dit geeft ons een beter beeld van de momenten waarop het energieverbruik waarschijnlijk hoger ligt.

Tijd	Personen thuis
Nacht (0:00 - 6:00)	
Ochtend (6:00 - 12:00)	
Middag (12:00 - 18:00)	
Avond (18:00 - 24:00)	

2. Gebruik van apparaten

Noteer elk groot apparaat dat je hebt aangezet, samen met het tijdstip en eventuele opmerkingen.

Je hoeft niet alle apparaten die je aan of uitzet te noteren, alleen de apparaten die veel energie verbruiken zoals een vaatwasser, kookplaat, oven, kachel, wasmachine of droger.

Tijdstip	Apparaat	Duur gebruik (indien bekend)	Opmerkingen

3. Speciale activiteiten

Noteer hier extra informatie over deze dag, die mogelijk invloed heeft gehad op jouw energieverbruik.

Bijvoorbeeld een bijzonder warme of koude dag, onderhoudswerkzaamheden of koken voor gasten.

4. Algemene opmerkingen

Noteer hier eventuele andere opmerkingen die je hebt.

Bijvoorbeeld over het gebruik van de P1 meter, problemen die je bent tegengekomen of het onderzoeksproces.

Dagboek

Dag 6 – Datum:

1. Aanwezigheid in huis

Vul in hoeveel personen er op welk moment thuis waren. Dit geeft ons een beter beeld van de momenten waarop het energieverbruik waarschijnlijk hoger ligt.

Tijd	Personen thuis
Nacht (0:00 - 6:00)	
Ochtend (6:00 - 12:00)	
Middag (12:00 - 18:00)	
Avond (18:00 - 24:00)	

2. Gebruik van apparaten

Noteer elk groot apparaat dat je hebt aangezet, samen met het tijdstip en eventuele opmerkingen.

Je hoeft niet alle apparaten die je aan of uitzet te noteren, alleen de apparaten die veel energie verbruiken zoals een vaatwasser, kookplaat, oven, kachel, wasmachine of droger.

Tijdstip	Apparaat	Duur gebruik (indien bekend)	Opmerkingen

3. Speciale activiteiten

Noteer hier extra informatie over deze dag, die mogelijk invloed heeft gehad op jouw energieverbruik.

Bijvoorbeeld een bijzonder warme of koude dag, onderhoudswerkzaamheden of koken voor gasten.

4. Algemene opmerkingen

Noteer hier eventuele andere opmerkingen die je hebt.

Bijvoorbeeld over het gebruik van de P1 meter, problemen die je bent tegengekomen of het onderzoeksproces.

Dagboek

Dag 7 – Datum:

1. Aanwezigheid in huis

Vul in hoeveel personen er op welk moment thuis waren. Dit geeft ons een beter beeld van de momenten waarop het energieverbruik waarschijnlijk hoger ligt.

Tijd	Personen thuis
Nacht (0:00 - 6:00)	
Ochtend (6:00 - 12:00)	
Middag (12:00 - 18:00)	
Avond (18:00 - 24:00)	

2. Gebruik van apparaten

Noteer elk groot apparaat dat je hebt aangezet, samen met het tijdstip en eventuele opmerkingen.

Je hoeft niet alle apparaten die je aan of uitzet te noteren, alleen de apparaten die veel energie verbruiken zoals een vaatwasser, kookplaat, oven, kachel, wasmachine of droger.

Tijdstip	Apparaat	Duur gebruik (indien bekend)	Opmerkingen

3. Speciale activiteiten

Noteer hier extra informatie over deze dag, die mogelijk invloed heeft gehad op jouw energieverbruik.

Bijvoorbeeld een bijzonder warme of koude dag, onderhoudswerkzaamheden of koken voor gasten.

4. Algemene opmerkingen

Noteer hier eventuele andere opmerkingen die je hebt.

Bijvoorbeeld over het gebruik van de P1 meter, problemen die je bent tegengekomen of het onderzoeksproces.

Dagboek

Dag 8 – Datum:

1. Aanwezigheid in huis

Vul in hoeveel personen er op welk moment thuis waren. Dit geeft ons een beter beeld van de momenten waarop het energieverbruik waarschijnlijk hoger ligt.

Tijd	Personen thuis
Nacht (0:00 - 6:00)	
Ochtend (6:00 - 12:00)	
Middag (12:00 - 18:00)	
Avond (18:00 - 24:00)	

2. Gebruik van apparaten

Noteer elk groot apparaat dat je hebt aangezet, samen met het tijdstip en eventuele opmerkingen.

Je hoeft niet alle apparaten die je aan of uitzet te noteren, alleen de apparaten die veel energie verbruiken zoals een vaatwasser, kookplaat, oven, kachel, wasmachine of droger.

Tijdstip	Apparaat	Duur gebruik (indien bekend)	Opmerkingen

3. Speciale activiteiten

Noteer hier extra informatie over deze dag, die mogelijk invloed heeft gehad op jouw energieverbruik.

Bijvoorbeeld een bijzonder warme of koude dag, onderhoudswerkzaamheden of koken voor gasten.

4. Algemene opmerkingen

Noteer hier eventuele andere opmerkingen die je hebt.

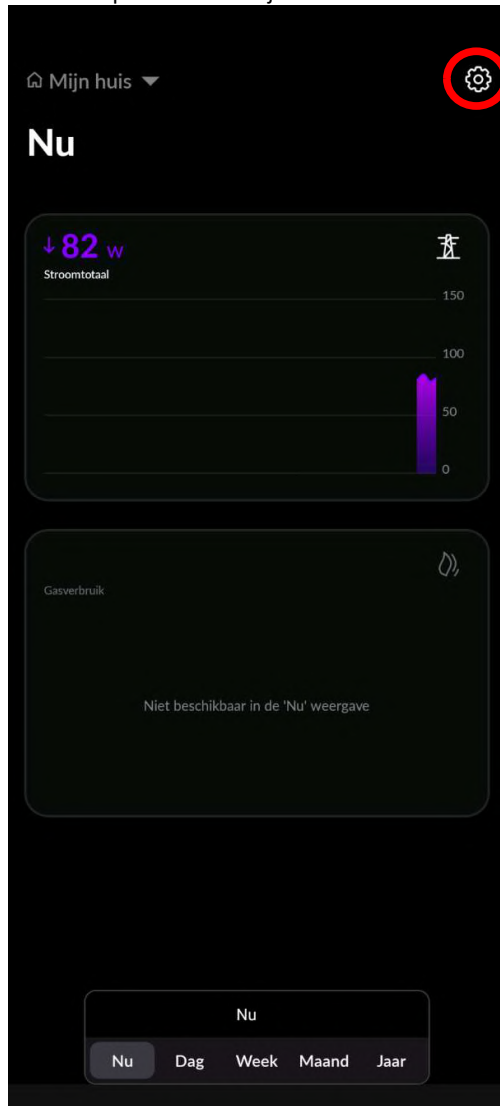
Bijvoorbeeld over het gebruik van de P1 meter, problemen die je bent tegengekomen of het onderzoeksproces.

5. Export handleiding

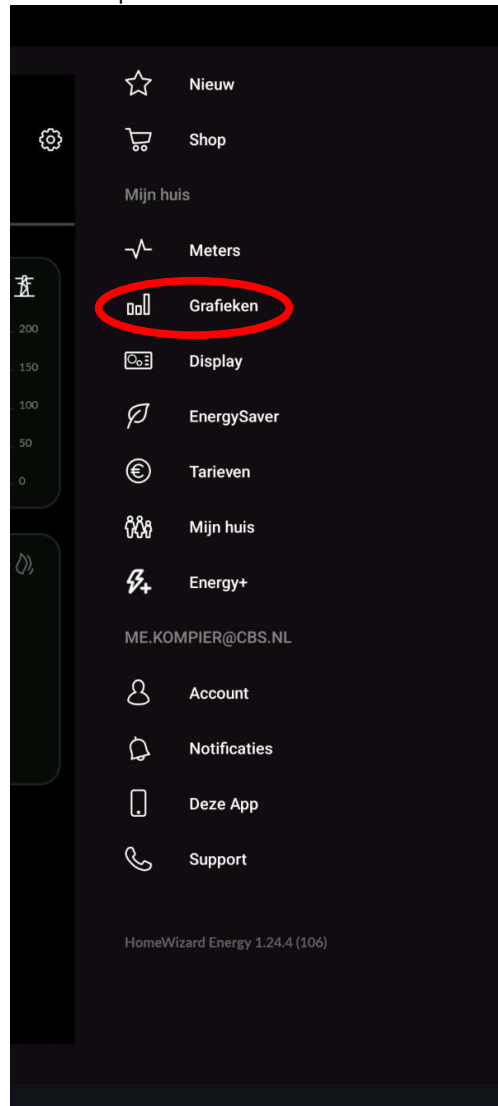
Hieronder volgt een instructie om de data te exporteren.

Doe dit een keer na 8 dagen en een keer na 30 dagen.

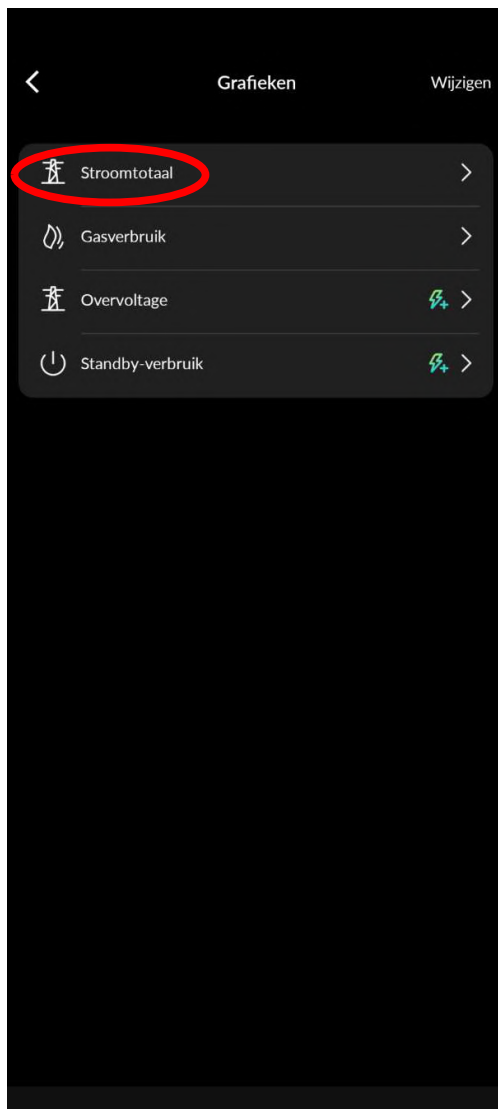
1a. Druk op het tandwielletje rechtsboven



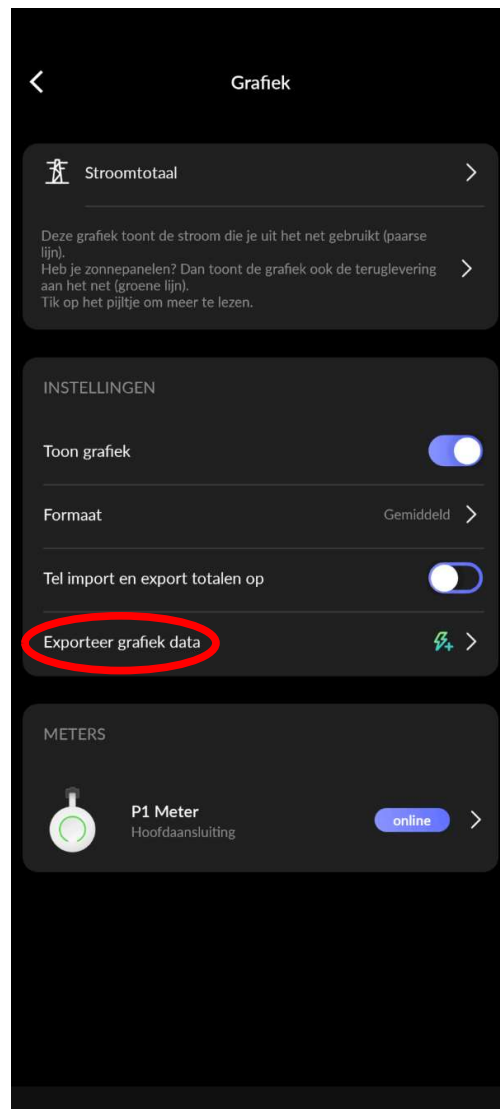
1b. Druk op 'Grafieken'



2a. Druk op 'Stroomtotaal'



2b. Druk op 'Exporteer grafiek data'



< Exporteer grafiek data

Exporteer je data naar Excel of soortgelijke programma's om de data voor jezelf op te slaan of verder te analyseren. Tik op het pijltje om meer te lezen. >

KIES GRAFIEK

Gasverbruik ☐

3a Stroomtotaal ☒

EXPORT BEREIK

3b Van 01 januari 2025 >

3c Tot 23 januari 2025 >

3d Resolutie 15 minuten >

3e Export data bestand (.csv)

3a. Selecteer 'Stroomtotaal'

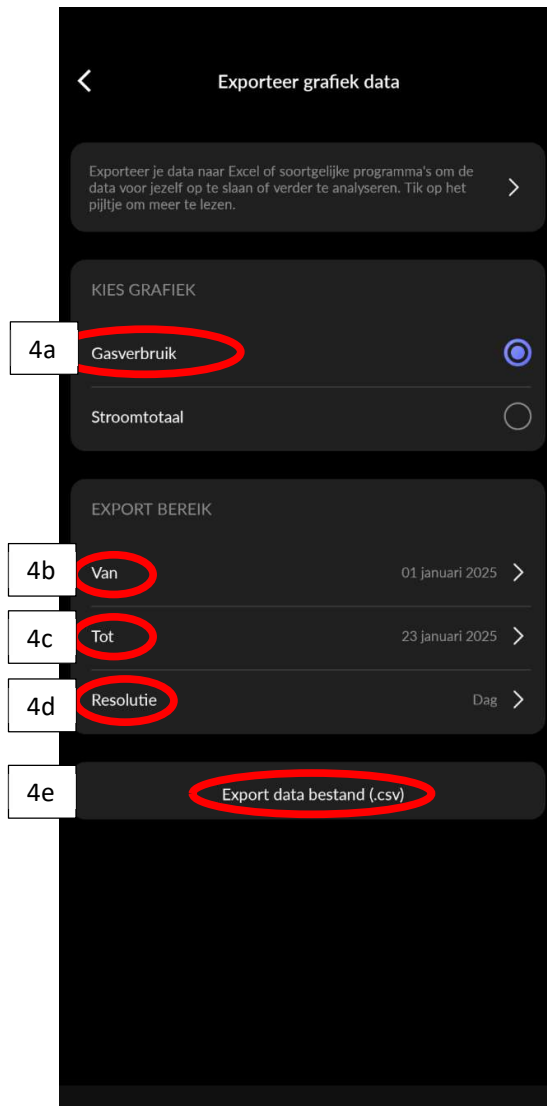
3b. Kies bij 'Van' de datum dat je de P1 meter geïnstalleerd hebt

3c. Kies bij 'Tot' de datum van vandaag

3d. Kies bij 'Resolutie' 15 minuten

3e. Druk op 'Export data bestand (.csv)'

3f. Stuur het bestand per mail naar userlab@cbs.nl



- 4a. Selecteer 'Gasverbruik'
- 4b. Kies bij 'Van' de datum dat je de P1 meter geïnstalleerd hebt
- 4c. Kies bij 'Tot' de datum van vandaag
- 4d. Kies bij 'Resolutie' 15 minuten
- 4e. Druk op 'Export data bestand (.csv)'
- 4f. Stuur het bestand per mail naar userlab@cbs.nl

6. Evaluatie vragenlijst

Tot slot hebben we nog een paar vragen voor je, over jouw ervaringen met de P1 meter.

1. Hoe makkelijk of moeilijk was het om de P1 meter aan te sluiten?

- ☐ Heel makkelijk
- ☐ Makkelijk
- ☐ Niet makkelijk, niet moeilijk
- ☐ Moeilijk
- ☐ Heel moeilijk

2. Waren de instructies duidelijk en voldoende om de P1 meter goed in gebruik te nemen?

- ☐ Ja
- ☐ Nee

2a. Als nee: wat voor instructies miste je?

3. Kun je beschrijven wat jouw ervaringen waren in de week dat de P1 meter data verzamelde?

4. Heb je jouw gedrag aangepast naar aanleiding van de metingen?

- ☐ Ja
- ☐ Nee

4a. Als ja: wat heb je aangepast?

5. Waren er dagen/dagdelen waar je in de app een onverwacht groot verbruik zag? Zo ja, kun je die (nu) verklaren?

6. Waren er dagen/dagdelen waarvoor data incompleet waren? Zo ja, welke data ontbraken?

7. Waren er dagen/dagdelen waarvoor de data naar jouw mening niet-plausibel waren? Zo ja, wat was volgens jou de afwijking?

8. Heb je één of meerdere keren een storing gehad? Zo ja, wat voor storing?

9. Vind je de metingen die de P1 meter doet privacy gevoelig?

- ☐ Ja
- ☐ Nee

10. Heb je tips hoe we het onderzoek doen met een P1 meter nog makkelijker of zinvoller kunnen maken?

11. Heb je nog andere tips of opmerkingen?

Ontzettend bedankt voor je deelname!