



User testing of the new Household Budget Survey 2022

Documentation report 2023

TALL

SOM FORTELLER

NOTATER / DOCUMENTS

2024/19

Nina Berg and Karianne Lund

In the series Documents, documentation, method descriptions, model descriptions and standards are published.

© Statistics Norway

Published: 8 March 2024

ISBN 978-82-587-1943-1 (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

In this publication, Statistics Norway presents documentation of user testing conducted during the development of a new electronic data collection solution for the Household Budget Survey 2022.

Karianne Lund and Nina Berg from Division for Methods have planned and carried out the tests. Gezim Seferi from Division for Social Surveys and Maciej Warchal from Division for User Insight and Web Development assisted with execution and expert evaluation. Esben Berg Aasgaard from Division for Corporate Communication contributed with design sketches and expertise. Kristin Egge-Hoveid from Division for Income and Living Conditions Statistics and Aina Holmøy from Division for Social Surveys participated in planning and analysis. Together, they have, at times, been part of the user testing team and have been integral to the larger development team for the new Household Budget Survey. The entire development team for the Household Budget 2022, along with interviewers and the Service Centre, has also contributed throughout the user testing process.

Statistics Norway, 5 March 2024

Arvid Olav Lysø

Abstract

Statistics Norway (SSB) has monitored Norwegian households' consumptions since 1958. The Household Budget Survey (HBS) is one of SSB's most complex social surveys and has traditionally been collected using paper diaries and a combination of in-person and telephone interviews. In 2022, SSB introduced a significant change by adopting a digital approach, encouraging participants to self-report via an app, which included a feature for receipt scanning. This change aimed primarily to reduce the respondent burden, counteract declining response rates, and meet a general need for modernization. An extensive series of tests was conducted to ensure the app's functionality met the diverse needs of respondents, focusing on user-friendliness and inclusivity, as this report describes.

The development of a new data collection tool was guided by the potential and challenges of integrating "smart technology" into survey processes. Feedback from our tests indicated that the majority of participants found the app convenient and easy to use, reflecting the prevalence of mobile technology use among working-age individuals in Norway. Many, but not all, are experienced mobile users and understand the interactive language of touch screens, scrolling, swiping, and the like. However, older participants showed a higher degree of hesitancy and encountered more technical issues, suggesting a digital divide that could impact participation rates among certain demographic groups.

With the transition from interviewer-administration to respondents' self-completion of the survey in the app, we saw a clear need to simplify the task of responding. The length of the questionnaire and the diary-keeping period were reduced, and English was offered as an alternative language. However, it is a challenge to communicate in-app the survey tasks to do the respondents. We know that the more information required, the harder it is to communicate this information. For respondents the tests suggested that interpretation of the response task is a greater challenge than handling the app itself. Technology offers us many new possibilities to guide and direct respondents through the survey, but an app has limitations in conveying nuanced instructions. Interviewers have made a significant difference in communicating the task. They are crucial in recruitment and have a more central role in motivating, guiding, and providing technical support to respondents than has been common in telephone and web surveys. They have also been important in assuring respondents that it is safe to open the survey link received via SMS from SSB.

Even though the majority carried their mobile phones with them most of the time, they did not record their purchases as promptly as we had hoped. Many forgot receipts from stores, and there was often a delay between shopping and recording. This affects memory and results in less detailed reporting. Less frequent record-keeping, along with superficial reporting for other household members, leads to less detailed and accurate reporting. Scanned receipts are checked less frequently than desired. Overall, respondents show less motivation to report as accurately as before. Whether the volume and details of scanned data can compensate for less accurate manual recording remains inconclusive for now. Further analysis of survey data from the new Household Budget Survey 2022 will provide a basis for an independent evaluation and report.

Through extensive user testing, SSB has gained valuable insight and in-depth knowledge of how the new Household Budget Survey app functions when interacting with respondents. Here, we share a summary of this work.

NOTE: Please note that this report is a condensed translation of an internal publication at Statics Norway from 2023, see [Bruketesting av ny forbruksundersøkelse \(ssb.no\)](https://www.ssb.no/brukertesting-av-ny-forbruksundersokelse). The full Norwegian version includes descriptions of each test round. In the English version this is not included as understanding of the Norwegian language and culture is a prerequisite for comprehension. The translation process was supported by AI technology and verified by SSB for accuracy.

Contents

User testing of the new Household Budget Survey 2022	1
Preface	3
Abstract	4
Contents	5
1. Introduction	6
1.1. Background.....	6
1.2. The purpose of user testing	6
1.3. The measurement instrument	6
1.4. Organizational structure	7
1.5. Participants at SSB	8
2. Theoretical and methodological background	10
2.1. The cognitive perspective on the response process	10
2.2. Method of implementation	12
2.3. Analysis methods	17
2.4. Strengths and weaknesses of user testing	18
3. Tests	20
3.1. Total sample for all test rounds	22
3.2. Documentation.....	23
4. What we have learned	24
4.1. Summary of the tests	24
4.2. Methodological challenges	29
Glossary and abbreviations	35
References	37
Appendix A: The HBS team	39
Appendix B: Illustration of the HBS app	40
Appendix C: Reasons for non-response HBS 2022	47
List of figures	48
List of tables	49

1. Introduction

1.1. Background

Statistics Norway (SSB) has been mapping the annual consumption of Norwegian households since 1958, with the previous survey being conducted in 2012. In the Household Budget Survey (HBS), we collect information on the actual consumption of various goods and services from a representative sample of Norwegian households. These households should be representative of different household groups to be used as the basis for calculating the consumer price index and are also included in the National Accounts. The survey results were published towards the end of 2023. In the previous survey, data was collected through a combination of in-person and telephone interviews, along with a period of recording household expenses and receipts in an envelope, which was submitted to SSB for scanning. The Household Budget Survey is conducted in several other European countries as well. The development work in Norway has been guided by the need for comparability over time and across countries and to meet Eurostat's requirements for data collection and reporting.

For the 2022 Household Budget Survey, it was decided to develop an app that can be used to scan receipts. The main purpose of this transition was to make the task of responding easier for the respondents and thus counteract declining response rates. SSB aimed to meet user expectations and offer a more user-friendly way to participate in the survey. At the same time, SSB wanted to exploit new technology to reduce the extent of manual coding and scanning at SSB. Importantly, the measures were intended to counteract declining response rates without compromising the quality of data and statistics.

Below, we provide a description of the method, analyses, results, and recommendations delivered by the user testing team (see Table 1.1 page 8) in connection with the development of the HBS app. The work was carried out from January 2021 to June 2022.

To assist the reader, we have compiled key terminology and abbreviations in the "Glossary and Abbreviations" section at the end.

1.2. The purpose of user testing

When developing new questionnaires and data collection solutions, user testing aims to identify potential measurement errors that could compromise data quality. The primary objective of user testing is to gain insight into respondents' experiences with the survey, develop respondent-driven designs to reduce the response burden, and understand respondents and the potential sources of measurement errors. By understanding respondents and what may lead to measurement errors, we can create better solutions. We can design simpler and more unambiguous questions and data collection methods, resulting in more precise and comparable answers to ensure high participation rates and good data quality. More details about what user testing entails can be found in Chapter 2 on theoretical and methodological background.

1.3. The measurement instrument

The Household Budget Survey is a household survey with a measurement instrument consisting of several parts:

1. Recruitment interview conducted over the phone with a contact person (the "reference person"¹) for the household.
2. A diary for recording ongoing expenses for one week (the reference week).
3. A questionnaire about fixed expenses and significant purchases made in the past year.

A dedicated user testing team worked on the development and advisory aspects related to the content and design of all parts of the survey (a detailed description of the team and its work follows below).

The recruitment interview conducted over the phone was carried out using a dedicated system called Blaise, and this was not changed from the previous survey. The diary and the web questionnaire were integrated into a common electronic data collection solution, referred to as the "web app" or HBS app. We developed the HBS app as a Progressive Web App (PWA). This means it is an application accessible through a link in a web browser on either a mobile phone, PC, or tablet and can be saved to the device's home screen. This solution was chosen because users can open the link directly with a tap and log in with ID-porten/bank-ID² without having to download the app from AppleStore or GooglePlay, which some were reluctant to do at the time. A PWA solution allowed users to switch between using the solution on all their devices. A central new feature offered in the diary is the ability to automatically read receipts optically through the camera of a mobile phone/tablet using Optical Character Recognition (OCR). However, this scanning function can only be used on "smart" devices with a camera, i.e., devices with sensor or smart technology. This can be worked around by using a device with a camera and transferring the data to a device without a camera, such as a PC, or by manually recording receipts.

When we refer to the "measurement instrument" in this report, we focus on the digital tool that respondents "see," i.e., the HBS app. We do not delve into the underlying system that makes it all work, often referred to as the "back-end" of the entire data collection system.

A new interdisciplinary team was assembled to develop the data collection solution. For some team members, knowledge about the survey, data collection in general, and the use of apps in data collection were new concepts. Nevertheless, we clearly have been able to draw on generalizable experiences from other surveys regarding how we communicate with respondents and their experience of the "user journey," i.e., being part of surveys from start to finish. Experiences from the transition to web-based data collection for several of SSB's social surveys in recent years have also served as valuable insight into our work.

1.4. Organizational structure

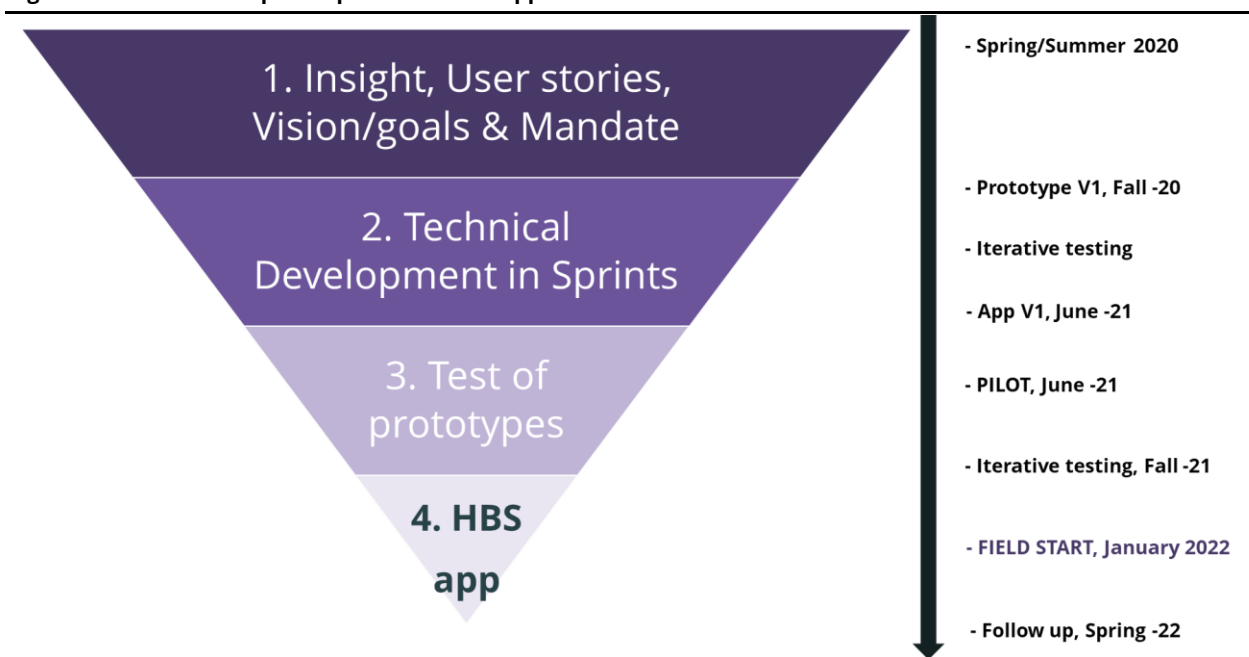
As mentioned, the development of the new Household Budget Survey was organized based on "agile" principles. The HBS Team comprised employees from various parts of the organization to achieve interdisciplinary synergy between development and user testing. The progress of the work was guided in an agile manner through short "sprint" periods involving planning, technical development, testing, evaluations, and revisions, which were repeated in an iterative process. This aligns with best practices for user testing, where we are also accustomed to developing questions/surveys, testing, evaluating, and revising in an iterative process³.

The work of the user testing team was adapted to the technical development process of the data collection solution in the HBS Team. Figure 1.1 illustrates this process.

¹ See definition/description in Glossary and Abbreviation, page 39.

² A national secure method for login to public services online.

³ See Development process/Process model in chapter 2, page 14.

Figure 1.1 The development process for the app

The user testing team began by assessing the information needs in the 2nd quarter of 2020. Following this, design sketches (prototypes) were developed and tested. After the first version of the app was developed and the piloting of the survey was conducted during the summer of 2021, we evaluated respondents' experiences with the app and participating in the survey. The app was then revised again before the data collection for the new Household Budget Survey commenced in January 2022⁴. Following the field start, the app was re-evaluated in the spring of 2022, and minor updates to the app and survey communication¹ were implemented from March 2023⁵.

The sprint periods were 2-3 weeks in length throughout the project period. This was primarily adapted for technical development and was suitable for testing individual elements as they were being developed. However, for qualitative methods and testing of the entire concept, this was a very short cycle for planning, recruitment, testing, analysis, and reporting. In practice, we found that the iterative testing process often required larger teams or longer periods, both for technical development and user testing, to be carried out effectively. The review of the tests revealed that we did not always manage to complete a full iterative cycle, including the revision of technical functions, even though we could separate ourselves from the sprint periods. We have tried to find satisfactory solutions within the resources available to us without deviating too much from the working methods and principles that we otherwise have good experience with.

1.5. Participants at SSB

The work on developing a new data collection solution for the Household Budget Survey was carried out by an interdisciplinary team. The following areas of expertise were represented:

- Statistics Division
- Technical Development
- Data Collection

⁴The data collection period for the Household Budget Survey runs for one year, from January through December 2022.

⁵ Adjustments or improvements to survey communication (from the interview corps or through emails and SMS) after data collection has commenced are common practices to increase the response rate. However, there have been no changes to the design of the front end of the app itself.

- Web Development and Interaction Design
- Methods Division

The team responsible for planning and conducting user tests has also been interdisciplinary, consisting of participants from both within and outside the project groups for the Household Budget Survey⁶. See Table 1.1:

Table 1.1 Participants in the user testing team for the Household Budget Survey

Name	Section	Role in the team
Kristin Egge-Hoveid	Division for Income and Living Conditions Statistics	Product Owner and Project Manager for the new Household Budget Survey
Esben Berg Aasgaard	Division for Corporate Communication	Designer (Sketching App & Testing and Advisory)
Gezim Seferi	Division for Social Surveys	Recruitment, Moderator Assistant, and Advisory
Frode Larsen	Division for User insight and web development	Interaction Design and Advisor
Maciej Warchal	Division for User insight and web development	Interaction Design and Advisor
Karianne Lund	Division for Methods	Planning and Conducting User Tests and Moderator
Nina Berg	Division for Methods	Responsible for Planning and Conducting User Tests and Moderator

Karianne Lund and Nina Berg planned and conducted the tests. Throughout the process, we have had assistance in conducting the tests from Gezim Seferi, Esben Berg Aasgaard, Kristin Egge-Hoveid, Maciej Warchal, and Frode Larsen. Others in the HBS team at large have also provided advice within their respective areas of expertise in some tests or parts of the work. The work was led by Nina Berg.

⁶For an overview of participants in the project groups for the development of the new Household Budget Survey and who are referred to as the "Household Budget Team" in this report, please refer to the following section: **Feil! Fant ikke referansebildet.**, side 43.

2. Theoretical and methodological background

In the development of a new data collection solution for the Household Budget Survey, the user testing team worked on both development, user testing⁷, and continuous evaluation of the measurement instrument. The goal of this work was to reduce the response burden for those who responded and reduce potential measurement errors that could undermine data quality. Our theoretical approach to the development work was based on cognitive psychology and well-established qualitative methods and implementation principles, which we will describe here.

2.1. The cognitive perspective on the response process

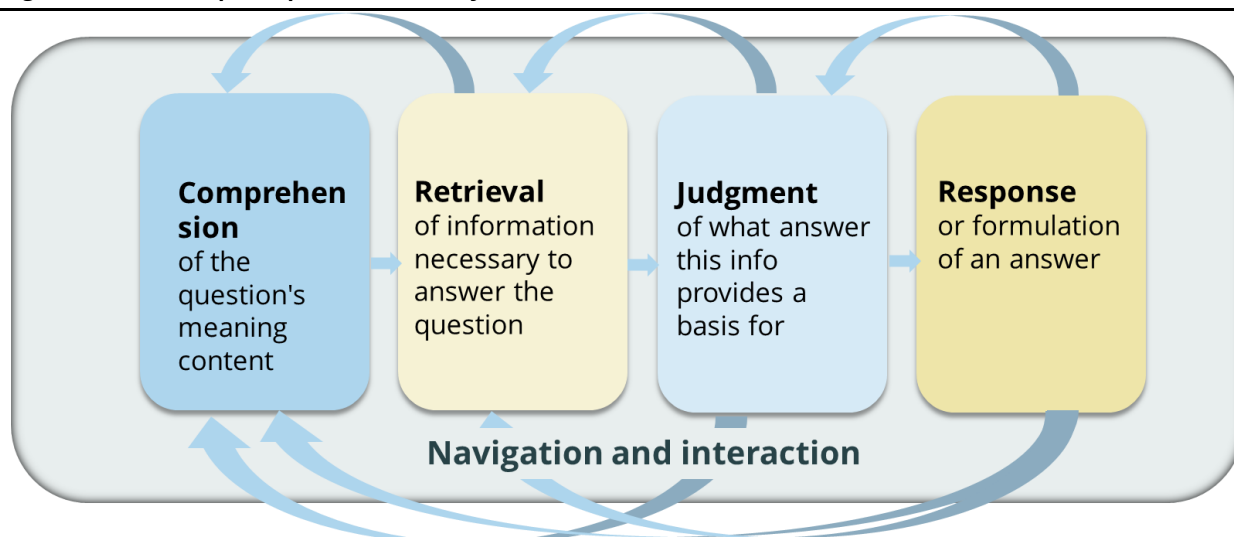
In the process of developing questionnaires and data collection solutions, we applied a cognitive perspective on the response process, often referred to as a respondent perspective (Tourangeau, 2000). Such a perspective aims to reduce the response burden for survey respondents and enhance the quality of their responses. The goal is to develop respondent-driven designs for data collection.

There are many examples where the needs of those who respond have not been well taken into account in the design of questionnaires. Those who design the questions and create the forms often do so based on their own understanding of the phenomenon they are investigating and their data needs. They tend to use technical language, which often differs from the language of the survey target group. Often, they are not aware of whether respondents have the same understanding of the language and whether they perceive the questions in the way it is intended.

Since the 1970s and 1980s, attention to the cognitive response process has increased. This approach to the response process acknowledges that respondents go through a complex cognitive process for each question they are asked in a survey. This respondent perspective replaced the earlier, more mechanical view of the response process, where data collection was seen more as the collection of data that were ready for "harvesting."

The American psychologist and survey methodologist Roger Tourangeau (2000) developed the response process model that we have based our work on, as shown in Figure 2.1. Since we were working with a digital tool where navigation and interaction with the app are essential, we added "navigation and interaction" as a backdrop to Tourangeau's response process model.

⁷ User testing, as described in this report, does not include technical testing. Technical testing was handled by those working on technical development in the HBS Team. Therefore, we do not describe this work here, even though the user testing team, along with the rest of the HBS Team, has been involved in it.

Figure 2.1 The response process model adjusted

Source: Adjusted after Tourangeau, R., Rips, L. J., Rasinski, K. (2000), The psychology of survey response.

The model illustrates that during the interpretation phase, inadequate text or guidance can create problems for the respondent. For example, providing insufficient information or using unclear or unfamiliar words and concepts can lead to respondents not understanding the task and/or interpreting it differently when providing their answers.

Once the question is interpreted, the retrieval of information begins. Respondents must think about what information they need based on their interpretation. Factors that make this process challenging include questions about events that occurred a long time ago, as memory fades and becomes less reliable in time.

After retrieving the information, it's time for evaluation. The more abstract the question, the more challenging it is to arrive at an answer. If the questions are also complex, it becomes difficult to process or estimate and arrive at a response. For example, when multiple factors are involved, a kind of average must be calculated. Some respondents may master this, while others may struggle. This doesn't necessarily mean that the answers of those who master it are more accurate than others. Additionally, people often tend to provide socially acceptable answers.

Reporting is the final step in the response process. Respondents may feel that their answers do not fit the provided response options. Common issues include missing response options or difficult terminology in the response options. It's also not uncommon for measurement units to not align with question asked, such as being asked about frequency (e.g., "how often") while response alternatives are given in proportions (e.g., "a large part of the time").

We've illustrated that navigation and interaction underlie Tourangeau's model because a majority of Norwegian respondents today answer surveys on mobile devices and must be familiar with the digital language of mobile devices. This digital language includes knowing where to find elements on a mobile screen, how to navigate from one screen to another, swipe horizontally, scroll up and down on a page, interpret symbols and interactions, identify clickable elements, and understand how to change settings. These functionalities are increasingly becoming part of an interactive digital language for PCs and tablets as well. Therefore, respondents' navigation and interaction within the app's "user interface" significantly influence the entire response process.

In general, it is believed that these are the steps in a process that most respondents go through when answering questions. In Figure 2.1 the process appears linear, but it often requires revisions along the way which the arrows attempt to indicate. Not all questions are equally demanding to answer. Many items can be answered spontaneously, and the closer the respondent is to the topic or when the behaviour being assessed occurred, the easier it is to respond. For example, in Norway, questions about age typically require

less thought than questions about the household's total income. Evaluating abstract phenomena is difficult for most because it requires retrieving information and selecting what is relevant. Often, it can involve aspects the respondent hasn't considered before or requires estimation or approximation, further complicating the task. Respondents may choose to refrain from answering or respond with "don't know" if possible. Responses can also be influenced by previous answers in the questionnaire because respondents often seek consistency throughout the survey.

To achieve good data quality, it must be straightforward for respondents to go through all the steps in the response process for each question in the survey. Questions and navigation must be simple, clear, and unambiguous. If questions demand a lot from respondents in any of the phases (interpretation, retrieval, judgement, or reporting), the response burden can become high for those who respond. This may lead to respondents using mental shortcuts to formulate their answers quickly and complete the survey. This is often referred to as "satisficing." It can result in unit non-response, where respondents abstain from participating, partial non-response where respondents skip individual questions or sections of the survey, or measurement errors where respondents do not invest the required effort to provide precise answers to survey questions. Among these, unsystematic measurement errors can be the most problematic, as they can be difficult to identify and calculate the effect of.

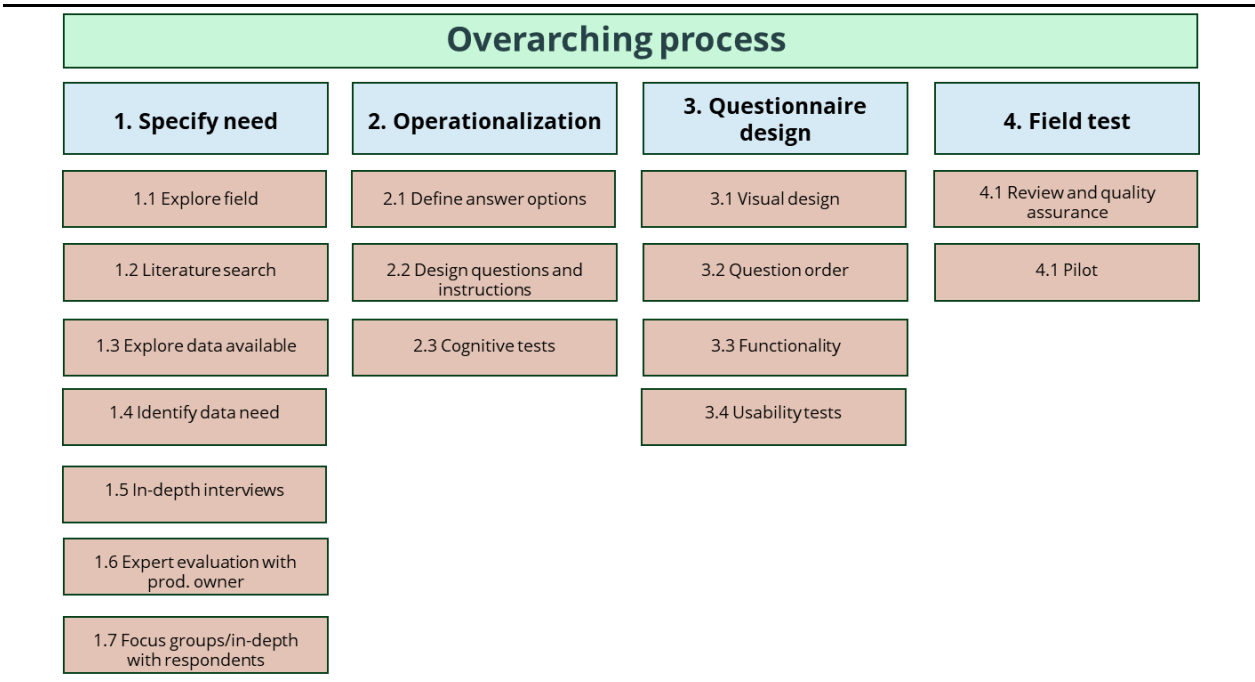
2.2. Method of implementation

The process model for developing measurement instruments

When developing new survey forms or measurement instruments, such as for the Household Budget Survey, the work is organized according to a systematic process, as seen in our process model in Figure 2.2. It is essential to emphasize that the development process is not linear but iterative. We begin with specifying information needs before moving on to developing questions and response options. After that, we work on questionnaire design and functionality, ending the process with testing the entire solution (see Figure 2.2).

In this case, we started by testing sketches of solutions early in the process. We made changes based on observations from the tests and retested accordingly. Subsequent changes were made as the work progressed throughout the development cycle. The model illustrates how the work began with specifying information needs (1), with additional elements and details added during operationalization and design, such as instructions (2.2), visual aids (3.1), and functionality (3.3). In the pilot phase (4.2), the entire measurement instrument, i.e., the HBS app, along with the data collection procedure, was tested, focusing on how all parts of the solution worked together in practice.

Figure 2.2 The process model for development of measurement instruments



Note: The process model for the development of the survey questionnaire is part of SSB's business model (GSPBM), which is based on what we refer to as the Total Error model or Total Error (TE).⁸

Throughout the development and testing process of the measurement instrument, we employed various forms of qualitative methods, which we will describe in more detail below. The approach or method we chose depended on the data needs and the stage we were in the development process. The development of the new Household Budget Survey essentially involved a transition from an existing paper-based questionnaire to self-completion on the web/in the app. We began by evaluating where there was the greatest need for improvement (1.6). When the survey questions for the different sections were ready, we conducted cognitive tests (2.3). When sketches of the technical solution for a function or feature were ready, we conducted various types of user tests⁹. Finally, we tested the entire solution through cognitive tests and usability tests, as well as a pilot survey before the actual survey went into the field (4.2).

An important premise for user testing is that the researchers' or surveyors' data needs must always be balanced against the respondents' access to data. In other words, user testing ensures that required data is collected in a way that takes into account respondents' motivation, abilities, and access to information. By considering the respondents' perspective, we ensure a solid foundation for high-quality data. This means that through user testing, we contribute to high-quality statistics.

Qualitative methods for user testing

We have various qualitative methods or types of user testing to investigate how respondents interact with a questionnaire or an app and their user experience. The common feature of these methods is that we examine in detail how respondents in the target group for the Household Budget Survey reason or act. Therefore, we also call these information-rich techniques. However, qualitative methods do not tell us how widespread the reactions to a certain solution are. Qualitative methods can be very valuable in the development of questionnaires because they provide the opportunity to:

⁸ See the description of the process model for statistical production here: <https://www.ssb.no/omssb/kvalitet-i-offisiell-statistikk/prosessmodell-for-statistikkproduksjon>.

⁹ We use the term "user testing" as a collective term for various methods to test how easy a system is to use to accomplish certain tasks for users. As such, we include usability testing in user testing.

1. Gain deeper insight into the respondent's perspective. We can better understand respondents' thoughts, opinions, and experiences related to the questions and identify any challenges or misunderstandings.
2. Discover possible problems or weaknesses in the questionnaire. Respondents can provide feedback on whether the questions are confusing, unclear, inappropriate, or difficult to answer. This helps identify areas that need improvement or revision.
3. Test the questionnaire's functionality and usability. This includes evaluating visual design, instructions, navigation, and functionality.
4. Iteratively improve the questionnaire. Based on feedback from test respondents, we can make adjustments, revise questions, and make changes to optimize the questionnaire. This ensures that the final questionnaire is tailored to the respondents' needs and understanding.
5. Improve the reliability of the collected data. By gaining insight into respondents' perceptions and interpretations of the questions, we can avoid misunderstandings or misinterpretations that could affect data quality.

In summary, qualitative testing methods contribute to a thorough, user-oriented, and quality-assured development of questionnaires. They help identify potential issues, improve functionality, and optimize questions to achieve reliable and valid data¹⁰. Where possible, we have also referred to quantitative data to confirm or contest findings or hypotheses from the qualitative work.

Here, we will briefly describe the methods and types of tests we have used most in the development of the measurement instrument for the Household Budget Survey¹¹. It is important to note that different testing methods can overlap or be used together. For example, cognitive interviews will often also have elements of user testing, and vice versa. Specific test designs are described in more detail for each user testing round in Chapter 3.

Expert Evaluation

When there is a draft questionnaire, such as for the Household Budget Survey, we start by evaluating it based on established guidelines for questionnaires, web design, and user interaction. This work is done by experts in questionnaire design, user interaction (UI), and user experience (UX). We typically distinguish between experience-based evaluations and evaluations based on a structured setup.¹²

For the Household Budget Survey, we primarily used available insights and recommendations. This is not what is commonly referred to as an "expert evaluation" but rather refers to the analysis done by subject matter experts. Unlike user testing, which is normally conducted among potential respondents, this analysis is based on experience, and real users are not involved.

We used at least two questionnaire methodologists and often also experts in web design and UI and UX when conducting expert evaluations in the development work. After individually reviewing the questionnaire, we came together to compile our common observations into one report. Compared to user testing, such an evaluation is a more cost-effective way to identify weaknesses in a

¹⁰ «Praktisk brukertesting», SSB (2006).

¹¹ For a more detailed overview of various types of tests, we refer to SSB's handbook for "Praktisk brukertesting", SSB (2006).

¹² Such structured coding systems often rely on frameworks like Forsyth's 0-Questionnaire Review (Forsyth 1996) or Snijkers' Compact Expert Questionnaire Appraisal Coding System (Snijkers 2013) for their analyses.

questionnaire or measurement instrument. The results of such evaluations can contribute to reducing the respondent burden, increasing user-friendliness, and improving data quality.

Cognitive interviews

Cognitive interviews (Ericsson, K. A., 1984) are a qualitative user testing method used to gain insight into the question-answer process, as shown in Figure 2.1. In these interviews, we focus on how people process information. We try to uncover respondents' perception and cognition from the moment they read/hear a question until they have answered it or completed a survey. The focus is on how respondents think to arrive at an answer, in other words, the response process.

We used this technique to gain insight into the user journey and the user's ability to use the app as intended. We aimed to understand how most people think about their household consumption, what constitutes a natural grouping and reporting of types of consumption, how respondents interpret questions and interaction in the app, and how they retrieve and evaluate their answers and complete challenging tasks.

Usability testing

Usability testing is aimed at assessing how user-friendly and understandable a solution is for performing specific tasks. Usability tests are more technical compared to traditional cognitive tests. In usability tests, we examine the interaction between respondents and the solution, including the interaction between text and graphics, functionality, and navigation. Various types of usability tests can be used to identify the risk of measurement errors and gain insights into how to improve the user experience. The different types of tests primarily vary in terms of their degree of structuredness. At one end of the scale, there are explorative tests that are investigative and unstructured. These are suitable for discovering unforeseen issues. At the other end of the scale, there are validation tests. These are structured and intended to be the final rehearsal. Between these two extremes, we find, for example, comparison tests, where two versions of the same question or solution are tested against each other. The choice of test type depends on what needs to be tested and where in the development process one is.

To assess how an app or solution performs in different user situations or scenarios, we have often used scenario testing. This method is based on various user stories that describe a situation in which the user needs to use a digital tool to solve these scenarios. We use this type of usability testing to describe the interaction between the user and the solution and to evaluate the user experience and ensure usability. Such testing can identify errors and challenges and ensure that the solution works as intended. Scenario testing of prototypes is a cost-effective way to test solutions but has its limitations as neither the test situation nor the solution is authentic. This can affect both the use and the experience during the tests.¹³

For the Household Budget Survey, we began with exploratory tests to gain insights into the subject matter and respondents. When sketches and/or functionality were developed, we used scenario testing and comparison tests, such as A/B or split tests, to compare two versions of the same solution. We also conducted evaluation tests after the pilot survey and after the fieldwork to examine how the survey functioned.

Implementation

The user testing process described were conducted by experienced moderators. The tests were conducted as digital remote tests, primarily through video calls¹⁴. This approach was adopted as an

¹³ You can find more information about scenario testing and A/B testing in "Praktisk brukertesting" from 2006, Chapter 4.4.5.

¹⁴ This, because our scope for action was limited during the development period due to the Covid-19 pandemic. (Between March 12, 2020, and April 2022, Norway was under lockdown, see

<https://www.regjeringen.no/no/tema/Koronasituasjonen/tidslinje-koronaviruset/id2692402/>)

alternative to in-person testing, especially during the pandemic. During these remote tests, participants were asked to share their screens and use their mouse to point to what they were looking at, reading, or trying to solve. They were also encouraged to "think out loud" while completing tasks.

In user testing of technical solutions, active listening¹⁵ and planned follow-up questions or probing are often allowed immediately after each element or functionality is tested. This approach aims to facilitate a smoother dialogue and allows testers to recall the considerations and evaluations they made during the process. This approach can provide deeper insights into the user experience.

The testing method used saved time and resources and enabled the recruitment of participants from across the country. The tests were conducted either as one-on-one interviews or as focus group discussions with multiple participants simultaneously, following an interview guide that could be more or less open or structured. The person conducting the tests is referred to as the moderator or test leader, whose role is to lead the conversation without influencing the answers or outcomes. An observer often participated as well, observing the interaction between the moderator and the test participant, taking notes from the conversation/group, and contributing to the analysis.

In some tests, we utilized sketches, prototypes, or different versions of questionnaires that participants were required to use to complete various scenarios or tasks. Tasks ranged from concrete tasks, such as comparing different user interfaces in A/B tests, to more exploratory tasks aimed at exploring conceptual understanding and gaining in-depth information. Through observation and feedback collection, we were able to identify patterns and gain in-depth insights into participants' interaction with and experience of the solutions.

In agile development processes, time frames are often short due to need to control technical development and resources. This means that the time for qualitative testing must be reduced compared to what would be ideal for qualitative testing. "Quick tests" have often been delivered, allowing for the testing of many versions of elements in the measurement instrument. Development work has thus been characterized by exploration, trial, and error, which has contributed to increased confidence in the chosen solutions in the app. When reducing the testing time, certain elements can be omitted or simplified. For example, the role of an observer or recorder may be eliminated, and recorded notes may be in bullet point form rather than detailed transcripts. The interview guide or the length of the test setup can also be shortened. Priority was given to which topics and target groups are most important to investigate. Sometimes, the test setup relied more on the "thinking aloud" technique with probing during the process than a separate retrospective review. To streamline the process, summaries of analyses were presented in oral debriefings, and summary reports with bullet points in PowerPoint were delivered.

Sample and recruitment

Selection of test participants for usability testing is characterized by its strategic approach. This means that we typically invite participants who have a good understanding of the subject matter or those whom we believe will encounter the most challenges when using the web app, in order to learn from their experiences. This approach differs from random sampling, which aims to represent

¹⁵ This means that we allow for immediate follow-up on what we refer to as "back channels," such as exclamations or in what we call "verbal communication" from the test participants. This implies that the moderator can follow up immediately with questions like "Can you elaborate?" or "Could you explain what you expected?" instead of addressing it later in a retrospective review guided by the moderator. In active listening, the moderator should still avoid interrupting and influencing the test participants.

an entire population. In qualitative interviews, we often recruit individuals who represent presumed problem groups or different "personas."¹⁶

Recruitment

Recruitment primarily occurred through networks and social media advertising. In the early stages of the process, some participants were recruited from a list of individuals who had volunteered to be contacted by SSB for similar purposes. Test appointments were confirmed through emails, and we experienced minimal dropouts during the tests.

Incentives

All test participants (not employed by SSB) received a universal gift card as a token of appreciation for their participation. Informed consent was obtained from all participants, emphasizing that participation was voluntary, and they could withdraw from the interview at any time.

Ethical considerations

Regarding ethical considerations, user testing for the Household Budget Survey was not considered particularly sensitive, as it did not focus on highly sensitive activities. Therefore, no specific ethical considerations were made in relation to these tests.

Documentation

Documentation for all tests was in the form of audio and video recordings, with the consent of the participants. Some tests were conducted with both a moderator and an observer/notetaker. The moderator (and observers) recorded various observations during the interview, and all notes made during the interview were collected. Observations and interpretations were distinguished in the notes from the interview.

Recordings from all tests were reviewed by at least one, often two, of the moderators after the test. This allowed the moderator to view the interview from an analytical perspective and ensured that important details from the interviews were not overlooked. The analysis was based on the recording as it is the only primary source of data from the interview. All notes from the tests must be considered secondary sources, as these notes have already been "filtered" through the moderator.

Data processing and security measures

Data processing and security measures were implemented. Documents containing personally identifiable information were stored separately from documents containing substantive data from the tests. Documents containing substantive data from the tests did not contain information that could identify the test participants. Personally identifiable material was deleted when the tests and analyses were completed. All other materials, including audio and video recordings, were deleted following the information provided to the participants.

2.3. Analysis methods

Systematization of data from the tests

Data from each individual interview were entered into an Excel data file specifically designed for this purpose. The data was organized in a way that facilitated comparison across tests and questions/form elements.

¹⁶ A "persona" is a detailed description of the profile of a fictional person or user, often accompanied by a user story, to help the service developer gain a better understanding of how the service or product is used.

The data were recorded with a level of detail that made it possible to understand what happened during the interviews for both the moderator and the rest of the testing team who were not present during the interviews. The interviews are not fully transcribed. In cases of doubt, we used the recordings to re-examine what was actually said and done.

Methodological challenges

In development of a new electronic data collection solution for the Household Budget Survey there were three specific challenges, in addition to the general goal of reducing respondent burden:

1. Transition to a new data collection mode, i.e., from paper to web app,
2. the use of a proxy interview to assess the consumption behaviour of other household members through a reference person for the household, and
3. respondents' ability to remember behaviour and details of purchases from 12 months ago.

First and foremost, the transition to a new data collection mode posed methodological challenges. This concerns the shift from paper diaries (PAPI) to web/app diaries (CAWI), where we were concerned with both comparability across time and mode, and whether an electronic data collection channel that requires "digital language"¹⁷ would be inclusive or exclusive for different target groups.

Furthermore, mapping the total consumption of households can be demanding. Primarily because the survey has multiple parts/sequences, and because keeping a diary over a period can be perceived as challenging to complete. In particular, we note that it is difficult to get everyone in the household to keep a diary of their ongoing expenses. It is equally difficult to have his reported correctly by the reference person on behalf of others in the household, known as "proxy" reporting. In addition to ongoing and regular or fixed expenses¹⁸ for a reference week, participants are also required to record major purchases made in the past year in order for us to map the household's annual consumption.

A particular challenge with questions about behaviour is respondents' ability to remember and recall detailed and accurate information about what they actually purchased. Memory problems can lead to systematic underreporting of specific purchases. A methodological challenge has been to come up with a form design that helps respondents with interpretation, recall, evaluation, and reporting of precise consumption information.

2.4. Strengths and weaknesses of user testing

To ensure that the questionnaire functions in line with its intentions, one must first be aware of and adhere to established best practices, or "Current Best Methods" (CBM). Secondly, the questionnaire should be tested on actual respondents. Respondents' thoughts and behaviours are too complex to predict how a questionnaire will perform without exposing it to those who will use it. Testing is the only way to ensure that the questions in the questionnaire effectively communicate what we intend to respondents (Campanelli 2008).

¹⁷ By "digital language" we indicate that the user needs "language" or experience with digital technology to participate in a digital survey. That is to understand how a PWA or app works and how to use it. An able user must have the skills to find, interpret, and evaluate conventional design and content in an app, and to interact with digital information or the app's user interface.

¹⁸ The Household Budget Survey distinguishes between ongoing and fixed expenses. Ongoing expenses are usually food and household items, clothes, hobbies, holidays and free time, or other expenses that you manage yourself. The amount is not predetermined or agreed in advance, as it is for fixed expenses. Fixed expenses can be, for example, housing expenses (electricity, rent, communal expenses, etc.), interest and instalments on loans, transport expenses, public charges, insurance, mobile expenses, nursery expenses and the like.

User testing based on qualitative methods such as observation, interviews, or interpretation of human actions and expressions often provides opportunities for multiple valid alternative interpretations and rarely yields a "neutral" or "objective" answer or finding. Several factors can influence the quality of user testing. For example, test participants may provide socially acceptable responses or behaviour, or the testing situation may affect them differently than a real situation would. The use of incentives and recruitment strategies can also result in more motivated test respondents than ordinary respondents randomly selected to participate in our surveys. Therefore, it is important to distinguish between when we investigate a phenomenon in a test situation versus a real situation, such as when we conducted follow-up interviews with participants in the pilot or standard survey for the Household Budget Survey. When participants are aware that they are participating in a test, this is different from when they are actually participating in a real survey and can affect their motivation and behaviour.

The testing situation can affect test participants in several ways, both positively and negatively. The advantages of being in a testing situation include increased attention and fewer distractions. Test participants may be more attentive and engaged in tasks when they are aware of being observed and evaluated. They may also experience fewer distractions in a controlled testing environment, which can help improve their performance. However, the testing situation can also have some disadvantages. Test participants may experience performance anxiety, which can negatively affect their performance. They may also behave unnaturally or exaggerate their behaviour to create a specific impression on the testers, leading to incorrect or misleading data. The observer effect can also influence participants' behaviour, as they may be affected by the attention and presence of the testers.

Well-established and carefully designed procedures with experienced moderators are essential to ensure the quality of qualitative observations and findings. We have also used method triangulation by gathering observations from interviewees and service responses, as well as quantitative data sources such as interview, process, and survey data, whenever possible. Using data from multiple sources has helped strengthen or weaken observations from our qualitative work.

3. Tests

To take into account the perspectives of the Household Budget Survey target group, extensive testing was carried out in multiple rounds of iterative testing and evaluation. The test series began with gaining insights to understand the data collection process and describing possible user scenarios and user journeys. In this phase, we primarily relied on in-house test participants. The idea was to identify and rectify the most elementary issues before involving "external" participants outside of SSB. The work then transitioned to testing various prototypes, using both in-house and external participants. In the final stages of the development process, usability tests were primarily conducted with participants outside of SSB, and the app was subjected to realistic testing. The final rounds of testing clearly emphasized usability testing of app functionality together with focus on the entire user journey and the cognitive response process.

Between each round of testing, those responsible for conducting user testing involved the entire HBS Team in analysis and evaluation meetings. Here, changes were discussed, the data collection solution was adjusted, and new tests were planned in an iterative development cycle where insights from one part of the process were used in the next, and so on. In this work, it was essential to be able to quickly identify potential problems, user challenges, or areas for improvement before data collection began. Due to resource constraints, much of the testing had shorter timelines for planning, execution, analysis, and reporting than what is described in the methodology in Chapter 2. Especially the initial rounds of testing can be described as "quick testing." In subsequent test rounds, we worked iteratively within the round, not just from one test round to another.

All user testing was conducted via Teams. One advantage of face-to-face or F2F¹⁹ online testing is that we could share screens and observe how test participants completed tasks in the app. At the same time, test participants may have felt that they could work quietly on tasks without someone observing them present in the room. This can have contributed to a more realistic reporting situation.

In table 3.1, we have compiled an overview of the user tests conducted for the Household Budget Survey app.

Please note that the Norwegian version of this report ([Brukertesting av ny forbruksundersøkelse \(ssb.no\)](https://www.ssb.no/brukertesting-av-ny-forbruksundersokelse)) includes detailed descriptions of all 9 test rounds conducted during development. In this English version, we have not included this content because an understanding of the Norwegian language and culture is a prerequisite to comprehend the analysis from each test.

In the following chapter, we will summarize the tests and discuss methodological challenges.

¹⁹ F2F is short for "Face-to-Face" and means that the interviewer and the interviewee or test person see each other. Whether this happens physically or online needs to be specified as well.

Table 3.1 Overview of conducted user tests

#	Name	Period	Number	Recruitment and target group	Length	Incentives
1	Usability test of the first version of a prototype	Week 45, 2020	9 interviews (Teams)	In-house sample via family and friends of SSB employees	45 min	No
2	Usability test of revised prototype	W50/51, 2020	7 interviews	Population sample: - one-person households - couples - multi-person households	45 min	Yes
3	Usability test of the first version of the app	W04/05, 2021	14 interviews	Population sample: 1. Outside the labour market 2. Seniors 3. Non-native speakers	45 min	Yes
4	Focus groups with pre-task	W11, 2021	4 focus group with a total of 19 participants	Recruitment via SSB's Facebook page: -2 gr adults -2 gr young	1,5-2 t	Yes
5	Comparison test of solutions for the home screen	14-15.04.21	11 interviews	In-house sample via family and friends of SSB employees -Colleagues outside the team, family or friends, aged 27-68 -Mix of household size	ca 45 min	No
6	Cognitive interviews with participants from the pilot	June, 2021	18 interviews total, 16 ind. + 1 couples	From pilot sample (started/completed): - Young people - Low education - Couples wo. children/children living at home - Multi-person households	ca 30 min	Yes
7	User test of the user journey	5-12.10.2021	12 interviews	Recruited via the SSB.no panel	45 min	Yes
8	Cognitive interviews before field start with groups hard to recruit or struggles to use the app	Week 48/2021	10 test persons a total in 7 interviews/ groups	SSB.no panel and FB & intranet adds: - Not Norwegian as first language - Multiple person HH (30-50 years) - Reduced vision - 70 years+	10-60 min	Yes
<i>Data collection period for the Household Budget Survey 2022: Week 01-52, 2022</i>						
9	Cognitive follow-up interviews after field start HBS 2022	Week 18-21/2022	25 interviews	From the sample of HBS 2022/ Interview status: - Not started - Started - Finished	5-30 min	Yes
SUM:			125 interviews			

3.1. Total sample for all test rounds

In total, over 9 test rounds, we conducted 105 user tests with 125 test participants²⁰. The test participants represent various subgroups of respondents. Please refer to Table 3.2 for an overview of the background information about the test participants.

Table 3.2 Distribution of Samples per round, overall and by demographics

Characteristics	User tests									Sum
	Test 1 First prototype W45/20	Test 2 Rev prototype W50/20	Test 3 First app W04-05/21	Test 4 Focus- groups W11/21	Test 5 A/B-test W15/21	Test 6 After Pilot July/21	Test 7 User journey W04-05/21	Test 8 Target groups W48/21	Test 9 After field start W18- 21/2022	
Total	9	7	14	19	11	18	12	10	25	125
Sex										
Man	3	2	9	7	5	8	7	5	10	56
Woman	6	5	5	12	5	10	5	5	15	68
Age										
17-24		1		4	5	1	1		4	16
25-34	3	3			4			1		32
35-49	2	1	7	4	1	7		1	10	12
40-59	2	1	1		1		4	3		20
50-69	1		3	5	1	9	7	2	8	28
70-79	1	1	3	6		1		3	3	18
80 years+										
Household size/type										
1 person HH	2	1	1		5					9
2+ persons HH	7	6	12		6					31
Adult couples without children living at home	2	3	5							10
Employment										
Student	1	1	2				1			5
At work	6	5	3				9			23
Outside of work			3							3
Retired/social security non-Oslo	3	1	7				2			13
Special target groups (recruitment/limited app use/accessibility need)										
Not Norwegian as first language			9					9		18
Reduced vision			2					2		4
70 years+			5					3		8
Outside of work			5							5
Assessment of own digital experience										
High			3							3
Medium			2							2
Low			9							9

As Table 3.2 shows, we achieved a diverse distribution in terms of gender, age, household size, and employment type. We also interviewed targeted groups that we anticipated would be challenging to recruit or may have difficulties using the app. We do not have information about education, geography, or household finances, but we have an impression that we have covered these aspects well too. Regarding location or geography, we note that conducting tests online contributed to a broader geographic representation than we would typically have resources for in face-to-face

²⁰ Technical testing done within the HBS team is not included here.

testing. Unfortunately, we have not succeeded in recruiting test persons with lower education levels to the degree planned. This is a general challenge in official statistics.

3.2. Documentation

For each test round documented in the Norwegian version of this report ([Brukertestning av ny forbruksundersøkelse \(ssb.no\)](#)), the following documentation has been archived:

- Recruitment plan
- Invitation to participate **Feil! Fant ikke referanseilden.**
- Interview guides **Feil! Fant ikke referanseilden.**
- Analysis file
- Report
- Recordings of team workshops when conducted

4. What we have learned

In this chapter, we begin with a summary of all the rounds of testing and what we have learned about the use of the web app in the new Household Budget Survey 2022. Following that, we discuss methodological challenges for this type of data collection.

4.1. Summary of the tests

User journey

Interviewer contact

When transitioning from interviewer-administrated data collection to self-administered data collection for a complex survey like the Household Budget Survey 2022, there is an increased need for effective survey communication throughout the user journey, from the initial contact to the final interaction with the selected participants. This includes recruitment, reminders, and follow-ups via phone calls, SMS, and email throughout the survey period until thank-you letters with gift vouchers are sent. In addition to the information provided in the email invitation, the app contains information about the survey in the welcome message upon first login, as well as in instructions and help texts. However, respondents spent very little time reading this information and instructions. An overall impression from all the tests is that respondents tend to skip such information if they can and rarely consult instructions. Therefore, we have observed that interviewer contact and information provided by interviewers are critical in recruitment and are crucial in explaining what needs to be done and guiding respondents when they encounter difficulties or get stuck in the app.

Login

Logging in from a web link received via SMS is easy for most participants. With a link in an SMS, the response burden is reduced since respondents do not have to search for the right app or download it from the App/Google Store. Respondents simply click on the link, log in with ID-porten or bankID, which practically all adult digital users are familiar with. When participants were informed by interviewers that they would receive an SMS with a link from SSB, most of them felt it was safe to open it. Younger participants believed they could assess the security of opening the link themselves, while some older participants were cautious about clicking on links and said that their practise was to be careful about this. Furthermore, most respondents were accustomed to using ID-porten and found logging in to be straightforward. We only encountered challenges with logging in using bankID among the oldest age groups. It requires further investigation to determine which age groups face difficulties with this type of login, but based on our experience, we can say that this is a clear challenge from the age of 80-85 and onwards.

Since the start of the survey, there has been an increase in warnings from government agencies and banks about not clicking on links. Therefore, it is reasonable to assume that the login solution used in this survey may not be best practice for SSB in the future. It may be necessary to explore alternative solutions to ensure representativeness in all age groups for future household budget survey or similar surveys. Additionally, we have observed a budding mistrust in the population towards technology, privacy, and data security, which should also be addressed with measures that make participation feel safe for everyone.

"Getting Started"

Communicating the survey task through the "getting started" instructions in the app is a significant challenge because the mobile format offers very little space for textual information. A consistent impression from the tests was that test participants quickly skimmed through the welcome page, information about what they needed to do, and tips on the home page about what was expected of them. This is also a common experience in tests of other surveys. Respondents felt there was too much text and were eager to get started with the task. They clicked determinedly to find out what

they needed to do. Their primary focus was on how to record their ongoing expenses in the reference week. When they couldn't figure out how to use the app or understand its functionality, our experience was that younger participants tried various answers or navigation options to move forward. The impression is that adult respondents made more efforts to figure out how to do it right and to provide correct answers. They sought help and technical support more frequently, primarily from those around them but also within the app and from SSB. It is clear that if respondents cannot figure out a task or a function, there is an increased risk that it may hinder them from starting or completing the survey.

During the survey, there were not many participants who spent time searching for or reading instructions, but they still expected instructions to be available and easy to find.

App's user interface

Our goal is that the visual design, navigation, and interaction, or the user interface, of the app should be self-explanatory for all users and not exclude anyone. These are high demands to place on an app for a complex survey with multiple components, proxy interviews, and diary entries for a reference period. User tests have revealed several challenges in interpretation and assessment, but our impression is that most users figured out how things worked and found the app's user-friendliness to be good. Even participants with less digital experience generally navigated through it fairly well, although they might not have utilized or benefited from all the technical features, such as pinning to the home screen, changing settings, scanning receipts, etc. Individuals with less digital experience more frequently encountered technical issues, which, in turn, could lead to dropout. *Therefore, our experience is that the app's user interface cannot replace the dialogue with interviewers.*

Respondents' expectations

The test persons we have spoken to were positive about using a web app, and most of them used it on their mobile devices. Regarding the user interface, they expected it to be conventional, intuitive, and efficient. They wanted to recognize functionality and find things where they were used to finding them, without repetitive questions and tasks. They expected the app to use interactions that guide them through it correctly and with as few clicks as possible. These are things that users take for granted should work well. Along with an interface that provides an intuitive understanding of the task at hand, this is the key to a good user experience. From our work with user testing, it is clear that the app's user interface alone has challenges in conveying what SSB expects respondents to do in the app. We tried to compensate for this with the use of interviewers, targeted survey communication (in addition to the app) throughout the process and offering technical support to those who need it.

Universal design

The requirements for universal design in the new Equality and Anti-Discrimination Act were only applicable after the HBS app was developed. Therefore, the app does not meet all universal design requirements but adheres to SSB's previous design standard. Deficiencies in this regard also emerged in the tests we conducted with individuals requiring universal design accommodations. Several of the user interface requirements can be easily adapted in the HBS app in a possible future iteration. However, the user group with needs for universal design is diverse, and SSB needs to better understand the various needs and how to accommodate them.

Response process

Respondents

Regarding the response process, we note that respondents who are responsible for their household's finances, or have experience with bill payments and household purchases, quickly understood what they needed to do. For this group, it was important that they felt the survey covered the entire household's consumption. They reacted if they perceived significant omissions

and lost trust in SSB's ability to provide accurate statistics on total consumption. In a population survey, we also have individuals who may have varying levels of language proficiency and some who struggle more than others to understand the task and interpret questions. To accommodate those who do not have Norwegian as their first language, we added English as an option alongside Norwegian written standard languages *Bokmål* and *Nynorsk* in the app²¹.

Understanding the task

The response process involves comprehension, retrieval, judgement, and formulation of response. For many respondents, the biggest challenge was understanding what information SSB required, both in terms of information from other household members and the desired level of detail. This was unclear for too many when they opened the app. Many had not grasped how detailed their reporting should be, that all consumption for everyone in the household for one week should be included. Not everyone was aware that the response burden could be reduced by gathering receipts and taking pictures of them for automatic entry instead of manually entering them. In some cases, they were aware of this but lacked receipts from the store. A few were unsure whether they should both scan and manually record expenses.

Learning from the system?

As we mentioned, instructions were often not seen or understood. Respondents often jumped into the task, expecting it to resemble a questionnaire and that they would figure it out through trial and error. And many did just that; they learned from the app or the interviewers that contacted them during the reference week. Those who did not learn, or did not "learn" as much, were at risk of dropping out of the survey. Some were afraid to try new things and dared not test scanning but chose manual registration because it was familiar and seemed manageable.

Some concepts (both terminology and what is included and excluded) were unclear and open to interpretation, such as the labels of open text fields in the diary section where one should report "store/place," (where something was bought) "I bought," (what was bought) and "cost" In the diary section. Not everyone was sure what type of expenses to report in the diary section and what to report in the questionnaire section. Some reported fixed costs in both the diary and the questionnaire, which was not the intention²². Not everyone had the same level of experience or familiarity with conventions for navigation and interaction in the app. These challenges in interpretation and assessment, combined with not all respondents reporting as diligently for others in the household, aggregating purchase items, not reporting immediately after shopping, or not remembering well, contribute to a reduction in data quality.

Memory aid

Reporting of one's own behaviour is most accurate when it occurs as soon as possible after the behaviour, and conversely, the longer the time between behaviour and reporting, the less precise it becomes. From the cognitive interviews, we saw that respondents had challenges with memory when it came to remembering all purchases or exact amounts or periods or remembering details about various costs and periodicities over the past 12 months. Few checked with their partners or other sources like contracts, online banking, invoices, or the like. When the retrieval of information is difficult the assessment and formulation of answers can become demanding and result in imprecise data.

²¹ English in the app is an assistance to respondents that is not proficient in Norwegian, but not optimal, as the search keywords for store and product names in the app are still in Norwegian.

²² Fixed costs were moved from manual registration in the diary section to the questionnaire section to simplify reporting for the respondents.

Quality of responses

Diverse interpretations, difficult retrieval, and assessment of what should be included and how accurately participants report contribute to uncertainty regarding the validity and comparability of the data. However, we do not have the impression that respondents themselves perceive their inaccurate responses as problematic or as providing an incorrect picture of their consumption. On the contrary, they believed that what they had entered should provide a good picture. This may be related to the earlier mentioned challenge in clearly communicating in the app what information SSB needs to create accurate statistics. We should also not disregard the possibility that the interaction with the moderator may lead the respondent to defend themselves or provide socially acceptable answers.

Recording practices

Reporting device

Most participants chose to respond to the survey using their mobile phones. Some used tablets or PCs, and some switched between several types of digital devices. Younger participants used only mobile phones, while it was clear that older participants preferred larger screens and larger text. Seniors are accustomed to performing similar tasks, such as online banking, on larger screens than mobile phones. This may be a temporary phenomenon, and we might eventually see older individuals also prefer mobile screens, or the younger generation may start preferring larger screens as they age.

Reporting behaviour and memory

From the tests, we observed that few participants scanned or recorded their purchases in the app immediately after shopping. Such a practice, as mentioned, would have been advantageous for the quality of the reported information. Participants collected receipts and recorded them in batches when they had a calm moment. This could affect what they remember to include. Many also forgot to get physical receipts from the store or did not have routines for keeping them, often resulting in manual entry of purchases. The impression is that many participants then combined items into categories and simplified things so that not all the details were included as required by SSB. This could be a strategy of "satisficing," i.e., taking shortcuts to meet the minimum requirements for the response. As we have pointed out, it did not seem particularly common for participants to seek sources or aids to remember, even though they were encouraged to do so. However, among older participants, several reported checking online banking, invoices, with their partner, or other sources to remember what they had purchased. Most people based their reporting on what they remembered there and then and did not interpret it as SSB needing more than that.

Scanning or manual registration

Many participants did not understand SSB's information about the advantages of scanning receipts. Those who explored and tested the scanning feature figured out on their own that it was most efficient for longer receipts. Some were uncomfortable with new technology like scanning and chose manual registration because it felt safe, and they assumed it was the simplest option. Both of these practices are unfortunate since we know that details often are lost when purchases are manually entered. This is especially true for number of grocery purchases (at the product item level), which make up 60-70% of the reporting during the reference week.

Reviewing own responses

Respondents made limited efforts to review and edit their data, entered either manually or by scanning in the app. Some wondered why they were sent to an overview of what they had registered. They found review and editing unnecessary and did not realize they had a verification task. Many thought the overview was fine or useful, but our impression is that the expense overview was not used to verify the completeness of the information, as intended.

The diligence with which respondents reported for other household members varied. It seems that younger individuals had less patience and a lower sense of responsibility for this task, often opting for the easy way out. They took shortcuts by answering in a way that met the minimum requirements to proceed through the app. We got the impression that younger participants omitted purchases or grouped items into categories more often than older participants. Older participants were more meticulous, sought help, and checked sources. Inaccurate reporting and the absence of scanning may be more related to respondents not having the concentration and endurance for such a complex reporting task as the Household Budget Survey requires, and it may be a strategy to get through it.

Finishing the survey

The possibilities offered by technology have led users to expect everything to happen quickly and in "real-time," i.e. immediately. Storing and confirming that information has been submitted, as SSB requests, can today be perceived as outdated. Along with complex survey communication, this may have contributed to many not realizing they needed to confirm in the app that they had completed the survey. Follow-ups through SMS and/or interviews were crucial for many respondents to confirm that they had completed the survey and included everything and had finished.

User experience

Complex experience

Throughout all the tests, we have seen that it is difficult to convey detailed information to respondents about the task and how to do it using general survey communication, interviewers, and the app. User tests have shown that the app is easy to use for most people, but the app's user interface alone struggles to convey the task and user journey, as well as what is expected of respondents. Lack of clear communication of the task creates uncertainty and can make the survey feel more challenging than necessary. This is especially evident as not everyone thinks that scanning receipts is the easiest method. The understanding of the importance of scanning receipts instead of manually entering purchases is critical for several reasons. Firstly, it can reduce both the perceived and actual response burden for respondents. Secondly, scanning receipts ensures a level of detail and accuracy in the data that is difficult to achieve with manually entered information.

Task interpretation and participation

Furthermore, we know that when the task becomes unclear or complex and seems difficult, many people hesitate to take the time to participate. It is therefore important to keep the response burden low to ensure participation and that respondents complete the entire survey. We have done this by optimizing survey communication, the app's user interface, reducing the reference period to one week, moving fixed costs from the diary section to the questionnaire section, and reducing the length of the questionnaire somewhat. However, it is clear that more can be done to simplify the tasks and the time spent on the survey, such as using more registry data, further reducing the questionnaire, and providing better universal design improving accessibility.

Interactive messages

The experience of tips and SMS messages from SSB was positive and can be strengthened. Several younger respondents requested more interaction and feedback from the app during the reference week, such as messages that pop up with positive messages like "Wow, you're halfway through the week! Thanks for your help!" with a thumbs-up emoji. Technology allows us to tailor such motivation to different demographic groups. Younger and digitally experienced users expected us to use this. For the Household Budget Survey, the opportunity has been underutilized in app, partly due to time and resource constraints and because we chose to develop a web app, which currently has limitations in this area. Going forward, the opportunity to "give something back" to the user along with more interactive user guidance is very interesting. How such "feedback" affects the user and data quality is something we need to study further.

A consistent request from respondents throughout the tests has been that they wanted and expected participation in the Household Budget Survey to provide them with a (better) overview of their own consumption. For many, this is a key motivation for participating. They expect us to leverage the possibilities of technology and provide data back to them in the app, such as an overview of what they have registered as expenses, for example, in the form of a pie chart based on their own data; compared to others; or previous years, etc. Many respondents expect to see this continuously in the app, as they do in many other apps. Such functionality has not been utilized in the new Household Budget Survey, both for resource reasons and because we know little about the effect of informing respondents about the conditions the survey aims to measure as objectively as possible.

Clear and unambiguous message

It is important that all parts of the survey communicate the same message. This includes survey communication, questions and answer fields in the app, graphics and functionality, and what interviewers say. A good user experience requires clear communication about what respondents should do and that it is easy to navigate and use the app. Some groups, such as older individuals, more often encountered technological barriers. This can be both an expression of being less experienced users of digital technology and also that they have more respect for the task and care more about doing things correctly. It is clear that technical issues must be addressed, and extra support must be provided to those who need it.

4.2. Methodological challenges

One of the main objectives of developing the new web app-based Household Budget Survey was to reduce the response burden for respondents. In assessing the extent to which we have achieved this goal, we must apply the respondent perspective (see chapter. 2.1, page 10). How do various respondent groups experience the new Household Budget Survey? Furthermore, a requirement was to continue to be able to compare over time in Norway and with other countries conducting the same survey. These considerations have significantly influenced the development work. The balance between maintaining established practices for measuring consumption in Norwegian households and embracing the new opportunities offered by web-based data collection with a web app has been of central importance. The transition to data collection in a web app has provided opportunities for some parts of the sample, while creating limitations for others. The development of the new Household Budget Survey also meant that parts of the survey are now self-administered, as opposed to the previous interview-administered survey. The reference period for the new Household Budget Survey was halved from 14 to 7 days, which represented a significant reduction in response burden. This change partly resulted from strong impressions from user testing regarding high perceived response burden.

Web-based data collection in a web app offers both opportunities and limitations. Data collection in an app requires that respondents have some technical skills and an understanding of digital or interactive communication. Therefore, it has been crucial to ensure that the new Household Budget Survey is user-friendly. In development, we have specifically focused on issues relevant to people we assumed might have challenges participating in the survey or using the app, such as older respondents or those with limited technical skills, individuals with disabilities, younger respondents, and people whose first language is not Norwegian. During the development process, we gained valuable insights into how different parts of the sample handled the new Household Budget Survey on the web app. Our approach has been qualitative and therefore not suitable for general conclusions on size and weight of usability issues. However, we have identified potential problems and clear patterns and trends after observing and interviewing over a hundred test participants.

Some of the observations about where the biggest problems in the web app occur were made in the early tests. These impressions were further reinforced throughout the development process. We have gained a more and more solid impression of how severe these problems were as the testing series progressed and through follow-up interviews with participants in the final survey.

Elderly respondents

One of the main challenges in transitioning to a self-administered web app for data collection has been addressing the needs of all participants in the sample. While most people in Norway now have access to smartphones, technical skills vary, often correlated with age and work-related factors. Elderly respondents often have less experience, are more anxious about making mistakes, have trouble with small font sizes, and may be sceptical about clicking links and downloading and using mobile apps. However, older respondents tend to have a better understanding of key concepts in the survey compared to younger respondents. It is particularly older respondents who indicate a preference for responding on a larger screen, such as a computer or tablet, instead of a mobile app. They also take the time to prepare and complete the survey conscientiously. Many older respondents are motivated to participate in the Household Budget Survey for the sake of societal benefit.

The transition to data collection with a web app can mean that some parts of the sample may not have as good access to the data collection tool as others. Effective use of a web app-based data collection solution requires technical skills that are not evenly distributed in the population. A central objective of this project was to offer participants a more user-friendly solution than before. However, there is a risk that the transition to a self-administered web app survey for the Household Budget Survey could contribute to digital exclusion for those with limited access to digital devices or less experience with digital technology. It has been crucial to consider the needs of all parts of the sample regardless of digital competence in the development of the new Household Budget Survey.

The results from user testing have provided valuable insights into potential challenges regarding the recruitment of elderly respondents to the survey. For example, we have observed that older respondents are more motivated for accurate reporting and are more likely to double-check sources for correct information compared to younger respondents. When they encounter challenges using a data collection mode they are not proficient in, their motivation to participate can decline. Many are reluctant to try because they don't believe they will succeed or don't want to learn new skills, and they give up quickly because they find it foreign and complicated. If we cannot offer an alternative to the web app for this group, we are concerned about our ability to include them in the survey. We are considering various ways to address this issue at SSB, including the possibility of offering a paper survey or an interview-administered survey.

Elderly respondents may sometimes have difficulty understanding certain concepts related to mobile technology, and it takes them some time to become familiar with the functionality. Dropout analyses in other web surveys show that dropout rates among the elderly increase with age. This suggests that web surveys may underrepresent older respondents compared to telephone surveys.

Reference person for the household

At the opposite end of the spectrum, we have young respondents with good technical skills who are supposed to answer questions on behalf of their household. The Household Budget Survey is a household survey in which one person, the reference person, is selected as the representative for the entire household. The result of this sampling method is that the sample can include respondents who are not responsible for the household's finances or purchases. The challenge for these respondents is that they do not always have access to the same information as the person(s) who are financially responsible in the household. This raises methodological issues related to both proxy interviews, where someone answers on behalf of others, and coverage regarding which

expenses are included and which are omitted. It is particularly important to note that it is challenging to get everyone in the household to accurately record ongoing expenses or to ensure that this information is correctly reported by the reference person in the household.

The choice of who answers on behalf of the household can impact the quality and accuracy of the answers. And it is a demanding sampling system to involve one respondent in a multi-person household to obtaining information from other members of the household. In many ways, this process can be described as an internal data collection within the household. As a result, the risk of measurement errors increases when participants rely on information from others. This can lead to incomplete or inaccurate responses.

Young respondents

Young respondents often have good technical skills and feel comfortable using the new Household Budget Survey app. Their confidence and competence often lead them to choose to "trial and error" rather than spending time preparing by reading instructions or other information, which our impression is that older respondents do more often. Our interpretations from the usability tests suggest that young respondents are more inclined to answer in ways that meet only the most basic requirements of the app. There is little evidence that they spend more time than necessary obtaining information from others in the household, verifying registered information, or correcting any errors. We have already mentioned that proxy interviews in general can increase the risk of measurement errors. This concern increases if the reference person is among the youngest in the household.

The youngest participants often had a limited understanding of the scope of the task, including the duration of the registration period and the requirements for detail. This may be partly because they felt that the email invitation they received was too long and that the link in the SMS was the most important. As a result, they often did not read the information in the email very carefully.

In interviews, young respondents reported that the web app was user-friendly. They also have high expectations that the web app must be efficient and have good functionality. The youngest participants preferred to use their mobile phones for reporting. A common impression of this group of respondents is that they do not have a routine for collecting and keeping receipts, so they often forgot to do so. This led to widespread manual recording of expenses. They were unsure of whether they had finished, thought that the web form was too long and challenging, and felt that the time spent increased the response burden.

Young respondents expressed a desire for simpler language and more colloquial phrasing where possible. Taking pictures and potentially retrieving information on a computer was considered more of a hassle for the younger respondents compared to older respondents.

People with disabilities

The transition to web-based data collection can have both advantages and disadvantages for people with disabilities. Around 20-25% of the population in Norway report having some form of disability. If large parts of these groups struggle to use the web app, they constitute a significant portion of the population and therefore the sample for the new Household Budget Survey. It is therefore necessary to consider people with disabilities to ensure inclusion and representativeness in data collection. This includes adaptations of text formatting and text quantity for small screens, as well as a solution that provides an overview of the entire task and easy navigation between the parts of the instrument.

The transition to digital platforms brings new interactive possibilities, such as touch screens and swiping. It is important to understand how these changes affect respondents and data quality. Our

user tests have not provided sufficient insight into challenges related to people with disabilities. However, we assume that they may encounter challenges that could have been addressed by interviewers in the past.

To promote participation among people with disabilities, several measures were implemented in accordance with the requirements for universal design (UD). We particularly considered the needs related to visual impairments. This includes options for magnification within the app, desires for horizontal screen display (which we were unable to provide), bold fonts, clear contrasts, and user-friendliness for people with visual impairments or those who prefer audio as a medium. When repeating the Household Budget Survey, SSB will be obligated to follow the UD requirements. This may affect the visual and interactive tools we can use. From other surveys, we have learned that several UD measures will also be beneficial for other groups. However, it is necessary to balance universal design with the principles of good practice in questionnaire design, as these considerations can sometimes be in opposition to each other.

Non-native speakers

The user test series included people whose first language is not Norwegian because we wanted insights into whether there could be language-related challenges related to key concepts in the survey. Experiences from other surveys also show that it can be challenging to motivate this group to participate in SSB's surveys. We experienced greater challenges related to trust in government agencies like SSB and concerns about privacy and surveillance than for others. It is necessary to convey that participation in the survey contributes to something significant, regardless of linguistic or cultural background.

Challenges for all target groups

Self-administration

A significant change from the old to the new Household Budget Survey is the increased use of interviewers' resources for follow-up and motivation, rather than traditional interview activities. In the old Household Budget Survey diary keeping and the final interview were administered by interviewers, while in the new survey, this has become self-administered in the Household Budget Survey app. The transition from interviewer-administered to self-administered data collection transfers responsibility and initiative from SSB to the respondents. Previously, respondents could provide oral information when contacted by the interviewer, but now they must take the initiative to enter information in the digital questionnaire themselves. Therefore, respondents must interpret how to answer the questions on their own to a greater extent.

Reduced contact with interviewers during the response process can negatively affect the reporting process itself and places higher demands on clear questions to minimize the risk of misunderstandings, dropout, and incomplete responses. At the same time, interviewer contact is important for recruiting and motivating respondents to complete the survey. Reduced interviewer contact can, however, also have a positive effect on reporting, as respondents may feel less pressured to give socially accepted answers and have more control over the progress of their responses.

Machine quality controls

In an app, the system or the user interface between the app and the user can offer many opportunities to support and guide the user. This is especially important since interviewers are no longer present to monitor and ensure the quality of the responses. Instead, the quality depends on the design of the survey. From web surveys, users are familiar with the use of controls, such as "soft" alerts that can be ignored or "hard" error messages that must be addressed. With the use of an app, we also have an entirely new language with its conventions for symbols, navigation, human-machine interaction with guidance, and prompting the user to complete tasks as intended. The app

requires that specific fields must be answered, and this is communicated visually and/or with warnings about what needs to be done to proceed.

It is also possible to trigger messages through user actions to motivate or "nudge" the user to use the app as expected and guide them through the survey. At the time the Household Budget Survey app was developed, such possibilities in the technical solution were limited. However, we could use keyword lists and have the app suggest store and product names from a searchable list that automatically appears with suggestions when the respondent begins typing. The app can also sum up manually entered purchases, so the user does not have to do it. It can alert the user to illogical answers and ask the user to check if they are correct. Some interactions in the app, the user can choose to read and/or follow themselves, while others must be actively dismissed. The questions, guidance, and controls that the user receives throughout the survey can be limited to only those respondents to whom it is relevant, thereby optimizing the app's user interface for each individual.

The user interface, or human-machine interaction, can replace some of the interviewer's guiding role in data collection, but it is not human and is not perceived as equally relevant or useful and user-friendly as an interviewer. It is also important to balance the use of alerts, tips, and controls. Widespread use can negatively affect respondents' experience of the data collection instrument.

The use of new technology going forward

In the development of the Household Budget Survey app, we have seen that the possibilities for designing solutions have increased dramatically with the use of smartphones and smart technology. We have access to OCR (Optical Character Recognition), GPS, and other sensors in mobile phones, which enable us to move towards more passive data collection, reducing the tasks for respondents significantly. Additionally, today's technology offers greater opportunities to customize the user interface in apps for different user groups, making them even "smarter." From web surveys, we are familiar with drop-down menus with lists of answers. Now we can use search functions that suggest answers to the user. With the help of machine learning and increased data capacity, we have unique opportunities to streamline the data collection process and improve guidance in the measurement instrument.

However, such opportunities do not come without the risk of new sources of error. We need more information on how access to digital devices, different operating systems, and screen sizes will affect respondents' responses. What about different user interfaces in different systems and human-machine interaction—what significance will they have? Will passive receipt scanning provide sufficient data quality? We have seen that active confirmation of the scan does not necessarily mean that respondents have actually checked the content. Therefore, it is important going forward to investigate the effects of new technology and assess to what extent it is appropriate to fully utilize all the possibilities.

Strengths and weaknesses of the development process

A total of 9 rounds of testing, with a total of 125 user tests, have provided us with a solid foundation for making well-founded recommendations for improvements to the recruitment process, survey communication, and the measurement instrument for HBS. Strategic recruitment of test participants, varied testing methods, and an iterative process with the involvement of many different professional fields have particularly contributed to a successful development process.

Through the user tests, we have gained valuable insights and critical in-depth knowledge about how the measurement instrument actually functions when interacting with respondents. It has also provided us with knowledge about unexpected issues in the app, the perceived response burden on respondents, and the response process of target groups for the developed solution.

Qualitative studies do not provide information about the extent of the findings we observe. However, qualitative analysis can identify possible issues and provide valuable learnings about users and app use that is invaluable in a development process. One of the significant strengths of the documented test series here is the solid data material. We have experienced that observations made early in the development process have evolved into hypotheses that gained more weight as we applied other methods, spoke with more test participants, and had issues or hypothesis enlightened from multiple perspectives. Thus, observations and learning from this development process are robust.

The pandemic situation made remote testing easier, as a large part of the Norwegian population started using Teams and Zoom, becoming accustomed to video conferences and document sharing in such meetings. The fact that our HBS Team already worked in an agile team on Teams also made online testing easier for us.

Further work

Data collection for the Household Budget Survey was completed at the end of 2022, and work on analysis, production, as well as the publication of statistics, will be completed during 2023. In this work, Statistics Norway (SSB) will closely examine the data from the survey to gain even deeper insights into how the survey setup and measurement instrument have functioned for different groups and validate the findings from the user tests.

The User testing team will compile what we have learned from the Household Budget Survey and the Time Use Survey, which developed a similar app in 2022-23, regarding respondents' interaction with a diary app. This information will be used to inform the development of other SSB surveys, and to share our experiences with other statistical agencies.

Glossary and abbreviations

- **A/B or split test** is a method for comparing two versions of the same technical solution or data collection mode to assess which one provides the best data quality.
- **COICOP classification** is a European standard classification for personal consumption that the Household Budget Survey should report according to.
- **Data collection solution** encompasses all the components of a survey (f. ex. diary plus web questionnaire) and the respective data collection methods (like online, telephone, F2F, or paper). For the Household Budget Survey, this includes both the recruitment interview by telephone interview and the diary and the questionnaire sections in the HBS app, and the underlying data collection system.
- **Interviewer-administered data collection**, in contrast to self-administered, is when the interviewer administers or guides the execution of the survey for the individual respondent. This means that the interviewer handles the measurement instrument and interprets and clarifies questions from the respondent.
- **Measurement instrument**. This is the questionnaire or the part of the data collection solution that respondents use to report their data in, e.g., the HBS app.
- **Mouse-over function** means that when you move the mouse or cursor over an element on the screen, the element's function is activated. That is, a new element or window can appear, in the form of text or an image.
- **OCR** stands for Optical Character Recognition, which is using a smartphone or tablet to take pictures of receipts and software that interprets the image into characters that a system can analyse and code mechanically.
- **Onboarding** is a term used in web development to describe how the user is welcomed and introduced to or provided with information about what the solution is intended for, how it is used, and what rules apply. In our case, this would include information about the purpose of the survey, what respondents should do in the app, privacy, how their data is processed, and which statistics data will be included in, etc.
- **Prototype** is a design sketch with an authentic user interface that also shows the user's interaction with the solution. For testing purposes during the development of an application, the entire solution or parts of it can be developed as a prototype.
- **Proxy interview** means that one person or interviewee answer on behalf of another, or answer as a substitute, such as when a parent or guardian answers for a minor or someone who is unable to answer on their own. For the Household Budget Survey, this can be one of the adults in the household who is expected to know the household economy best. F. ex. the one that normally pays most of the bills or do most of the shopping etc.
- **PWA** is a Progressive Web App that opens and runs from an internet browser and works on all devices, not just on a specific platform like iOS or Android. To the user it appears to look like and work like an actual (or "native") app. A PWA is often referred to (orally) as a web app.
- **Reference period** for the Household Budget Survey is the period for which respondents are required to keep a diary of all household expenses. Previously, this was 14 days. Due to response burden and response rate considerations, the period was reduced to 7 days just before the Household Budget Survey started. Therefore, we often refer to this as the reference week.
- **Reference person** for the Household Budget Survey is the person in the household selected to report for the entire household. The selection is based on who in the household is assumed to provide the best answers to questions about the household's finances.
- **Response burden**. The term refers to the effort required of respondents to answer. Often referred to as either perceived response burden or time burden.
- **Response process model** is based on the work of Tourangeau and others to describe the cognitive process respondents go through to provide a response to each question in a survey (Tourangeau et al., 2000).

- **Respondents** are those who participate in surveys. Several other terms can be used for those who participate in surveys, such as "sample," "interview subjects" (IO), or "users."
- **Satisficing** is a cognitive process where the respondent uses mental shortcuts to formulate their answers quickly and get through the form rapidly, choosing the solution they perceive as sufficient to complete a task. Getting through the app and completing the survey may weigh more heavily than whether the answer is correct or detailed.
- **Scenario testing.** This method is based on different user stories that describe a situation in which the user needs to use a digital tool to solve given tasks.
- **Self-administered data collection**, in contrast to interviewer administered, means that instead of the interviewer, the respondent must keep track of everything that needs to be done and taken care of. For the Household Budget Survey, this means that the respondent must download/open the form themselves, collect receipts, scan/fill out answers, and confirm that they are finished in the app-based solution.
- **Smart survey** is data collection that combines survey data and digital data, often in the same device, such as with a mobile app. In this way, digital behavioural data, or as in the case of the Household Budget Survey, scanning of receipts from the digital device, can be combined with self-reported responses from the respondents. Such digital behavioural data is based on sensor data and is often referred to as smart technology, see the description below. The idea is that digital behavioural data should be recorded without the involvement of respondents.
- **Smart technology** is technology that utilizes sensors such as location, movement, camera, etc., found in "smart" or digital devices like smartphones, etc. For the Household Budget Survey, it involves the use of the camera for optical scanning or reading.
- **Survey communication** refer to all communication with respondents outside the actual app, such as invitation letters, recruitment calls, SMS messages, etc.
- **Test persons** are those who participate in user tests. Several other terms can be used for those who participate in tests, such as "test respondents", "users", "sample," "interview subjects" (IO). Note that in some cases, the test participants in the user tests we describe may also have been actual respondents in the Household Budget Survey.
- **UI** stands for "user interaction" or user interaction with technical solutions.
- **UX** stands for "user experience" or user experience with technical solutions.
- **User interface** is the app's visual design and how the system and the user communicate or interact with each other. Often referred to as "human-computer interaction."
- **User journey** refers to the user's or respondents' experience of participating in the survey from beginning to end: From receiving the invitation, through being contacted by one of our interviewers, to opening the web link to the survey, logging in, and answering and completing all parts of the survey. Those who do not wish to participate, do not start, or start but do not complete will have a different user journey than those who complete.
- **Web app** is the most commonly used term for a PWA (Progressive Web App). It is an app that opens and runs from an internet browser and works on all devices and screen sizes, not just on a specific platform like iOS or Android. To the user it appears to look like and work like an actual (or "native") app.

References

- Boren, T. and Ramey, J. (2000) Thinking aloud: reconciling theory and practice, in IEEE Transactions on Professional Communication, vol. 43, no. 3, pp. 261-278, Sept. 2000, doi: 10.1109/47.867942.
- Brinck, Tom et al (2002): Usability for the Web: Designing Web Sites That Work. San Francisco, Calif.: Morgan Kaufmann
- Brukertesting av ny forbruksundersøkelse. Dokumentasjonsrapport 2023. Statistics Norway's publications, from: [Brukertesting av ny forbruksundersøkelse \(ssb.no\)](https://www.ssb.no/brukertesting)
- Campanelli, P., De Leeuw, E., Hox, J. and Dillman, D. (2008): International handbook of survey methodology.
- Campanelli, P., Nicolaas, G., Jäckle, A., Lynn, P., Hope, S., Blake, M., Gray, M., (2011): A classification of question characteristics relevant to measurement error and consequently important for mixed mode questionnaire design, not published: http://www.websm.org/uploadi/editor/1364221156Campanelli_et_al_2011_A_classification_of_question_characteristics_relevant_to_measurement_error
- Confluence/prosjektdokumentasjon for utvikling av ny forbruksundersøkelse (kun tilgjengelig internt i SSB): <https://statistics-norway.atlassian.net/wiki/spaces/TID/overview?homepagelid=1788936261>
- Mick Couper (2013): Surveys on Mobile Devices: Opportunities and Challenges. <https://www.ncrm.ac.uk/documents/Mick-Couper.pdf>
- Croskerry, P. (2002). Achieving quality in clinical decision making: cognitive strategies and detection of bias. Academic Emergency Medicine, 9(11), 1184-1204.
- Dillmann, Don A., Jolene D. Smyth and L. M. Christian (2014): Internet, mail and mixed-mode surveys. The tailored Design Method. Wiley: New Jersey.
- Ericsson, K. A., & Simon, H. A. (1984): Protocol analysis: Verbal reports as data. The MIT Press.
- Forsyths Questionnaire Review Coding System, 1996.
- Forbruksundersøkelsen 2012. Dokumentasjonsrapport. Statistics Norway's publications, from: https://www.ssb.no/inntekt-og-forbruk/artikler-og-publikasjoner/_attachment/169278?_ts=144f30b31a8
- Handbook of Recommended Practices for Questionnaire Development and Testing in the European System (2006).
- HETUS: Harmonised European Time Use Survey. <https://ec.europa.eu/eurostat/documents/3859598/11597606/KS-GQ-20-011-EN-N.pdf/2567be02-f395-f1d0-d64d-d375192d6f10?t=1607360062000>
- Lessler, J. og Forsyth, B. (1996): A Coding System for appraising Questionnaires. I N. Schwartz and S. Sudman (eds.): Answering Questions: Methodology for determining Cognitive and Communicative processes in Survey Research. San Francisco, Jossey-Bass.
- Loranger, H. (2016, October 23). The Negativity Bias in User Experience. Retrieved February 24, 2018, from <https://www.nngroup.com/articles/negativity-bias-ux/>
- Nielsen, J. (2000). Why you only need to test with 5 users.
- Om Forbruksundersøkelsen, <https://www.ssb.no/innrapportering/Forbruk>
- Documentation report from the pilot (for the Household Budget Survey 2022): Statistics Norway's publications 2021.

<https://statistics-norway.atlassian.net/wiki/spaces/F/pages/2487779329/Rapport+fra+piloten#Utvalg>

Praktisk brukertesting (2006): Statistisk sentralbyrås håndbøker. SSH 87.

https://www.ssb.no/a/histstat/ssh/ssh_87.pdf

Rothgeb, J., Willis, G. and Forsyth, B. (2007): Questionnaire pretesting methods: Do different techniques and different organizations produce similar results? *Bulletin of Sociological Methodology*, volume 96, page 5—31, Sage Publications Sage CA

Råd for utvikling og utforming av webskjema (2004): Statistisk sentralbyrås håndbøker.

https://www.ssb.no/a/histstat/ssh/ssh_81.pdf

Sauro, J. (2012). 9 Biases in Usability Testing. Retrieved February 23, 2018, from:

<https://measuringu.com/ut-bias/>

Shafir, E., Simonson, I., & Tversky, A. (1993). Reason-based choice. *Cognition*, 49(1-2), 11-36.

Snijkers, G., Haraldsen, G. & Jones, J. (2013). Planning the survey. *Designing and conducting business surveys journal*, 127-163.

Tourangeau, R., Rips, L. J., Rasinski, K. (2000), *The psychology of survey response*.

Willis, Gordon and H. Boeije (2013): The Cognitive Interviewing Reporting Framework (CIRF). Towards the Harmonization of Cognitive Testing Reports. *Methodology* 2013; Vol. 9(3):87–95.

Weichbroth, P. (2020). Usability of mobile applications: a systematic literature study. *IEEE Access*, 8, 55563-55577.

Zhang, Li-Chun (2012). Topics of statistical theory for register-based statistics and data integration. *Statistica neerlandica*. Vol. 66, Issue1, Special Issue: Methodological Challenges of Register-based Research, page 41-63.

Appendix A: The HBS team

The work on the new Household Budget Survey was divided between three project groups or teams: D1 Transaction Data, D2 Survey Solution, and D3 Statistical Production. User testing was included in and related to the D2 Survey Solution team. The team at large in D2 Survey Solution was called the HBS (or "Household Budget Survey") Team. See the composition of the whole Household Budget Team (D2) in Table A.1.

Table A.1 Participant in the HBS Team - DP2: Surveyløsning

Name	Section	Active role int the team
Kristin Egge-Hoveid	Section for Income and Living Conditions	Product Owner and Project Manager for the new Household Budget Survey
Esben Berg Aasgaard	Section for Business Communication	Designer (Sketching App & Testing and Advisory)
Ole Bredesen-Vestby	Section for IT architecture	System developer
Yassin El Barkani	Section for Data platform	System developer
Dag Blakstad	Section for Data platform	System developer
Eiliv Læg Reid	Section for Data platform	System developer
Prabakar Venkataraman	Section for Data platform	Adviser
Peter Fløgstad	Seksjon for Quality and group management	Front-end developer
Bjørn Frode Kvernstuen	Seksjon for Data platform	System developer
Aina Holmøy	Section for Data collection/Personal Surveys	Responsible project leader data collection HBS
Gezim Seferi	Section for Data collection/Personal Surveys	Recruitment, Moderator Assistant, and Advisory
Karianne Lund	Section for Methodology/Instrument Development and User Testing	Responsible for Planning and Conducting User Tests and Moderator
Nina Berg	Section for Methodology/Instrument Development and User Testing	Responsible for Planning and Conducting User Tests and Moderator

Appendix B: Illustration of the HBS app

This is the last version of the app. It was used from the start of the field period and tested in test 9: "Cognitive follow up interviews after the fielding of the survey".

Figure B 1 Home screen HBS-app

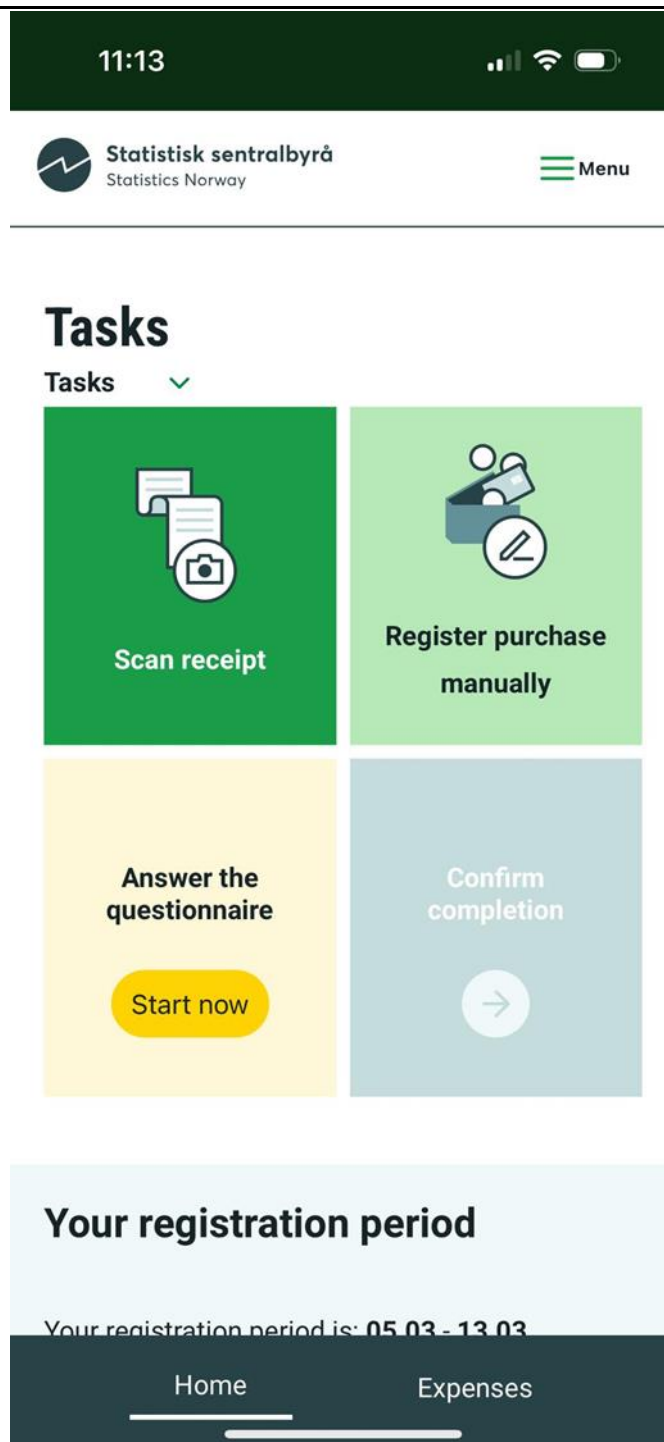


Figure B 2 Open app (PWA) in an internet browser and add to start screen (like a native app)

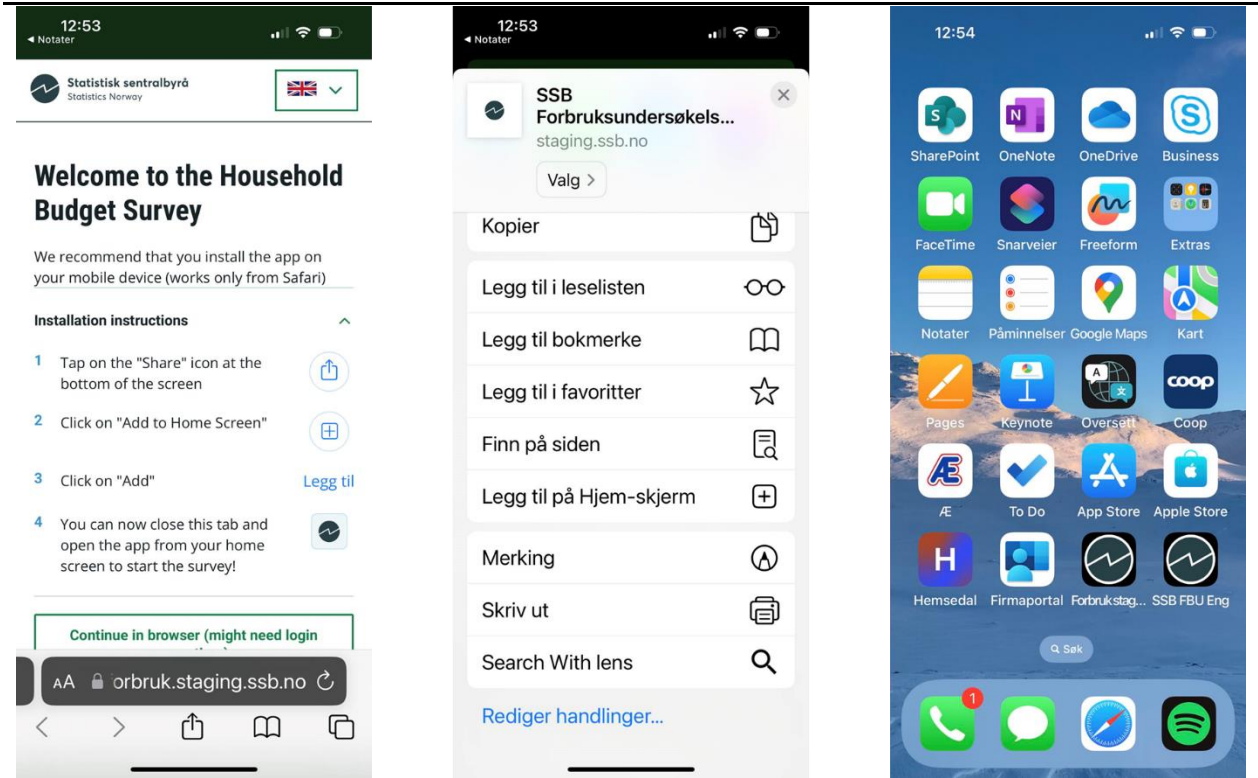


Figure B 3 Login once with the national ID-porten for secure logon

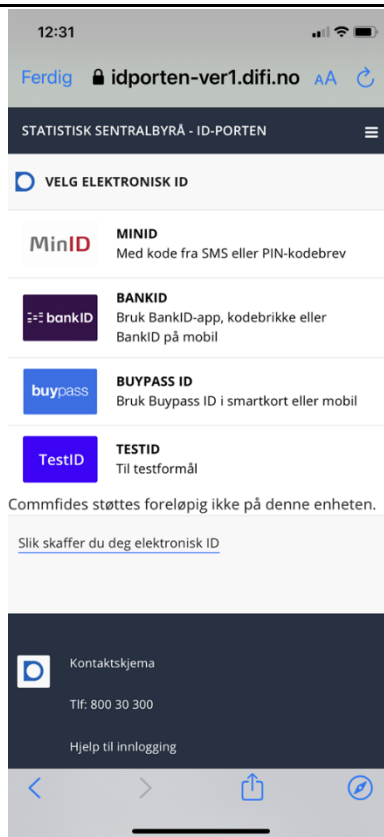


Figure B 4 Onboarding or “get started” instructions with consent

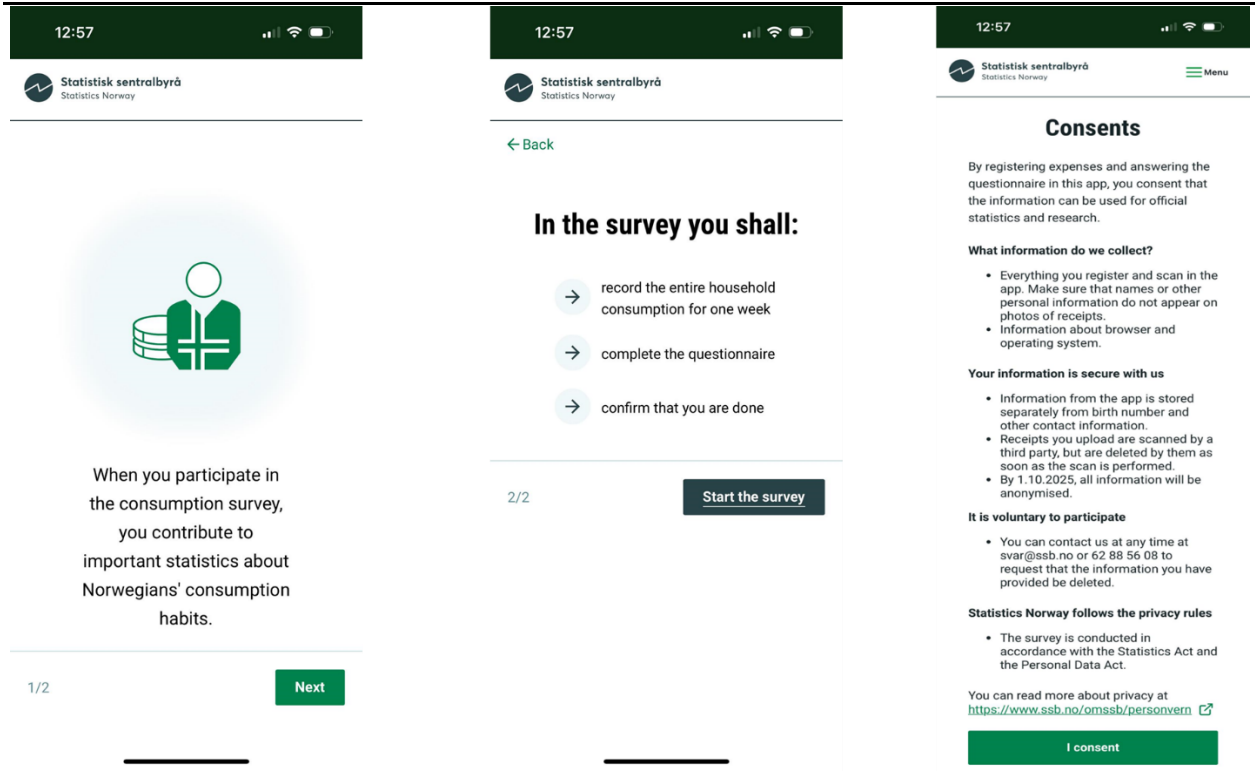


Figure B 5 Popup tutorial when respondent login (one time – can be found under Help later)

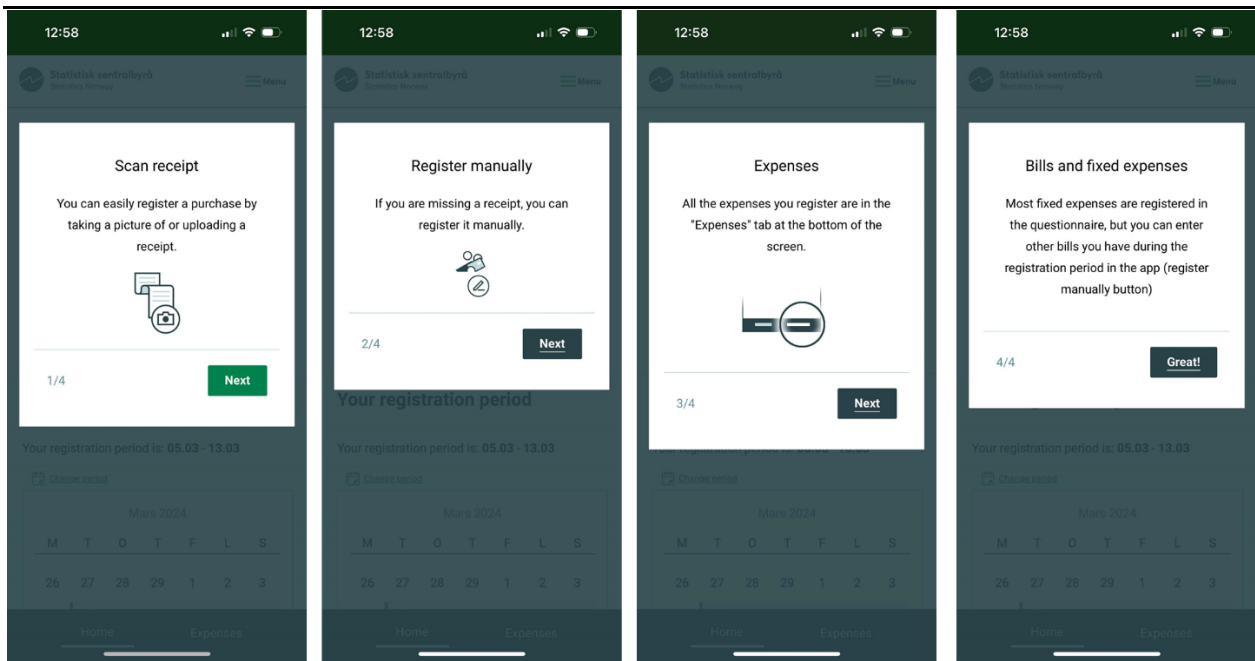


Figure B 6 Home screen with and without drop-down menus open for task and for calendar

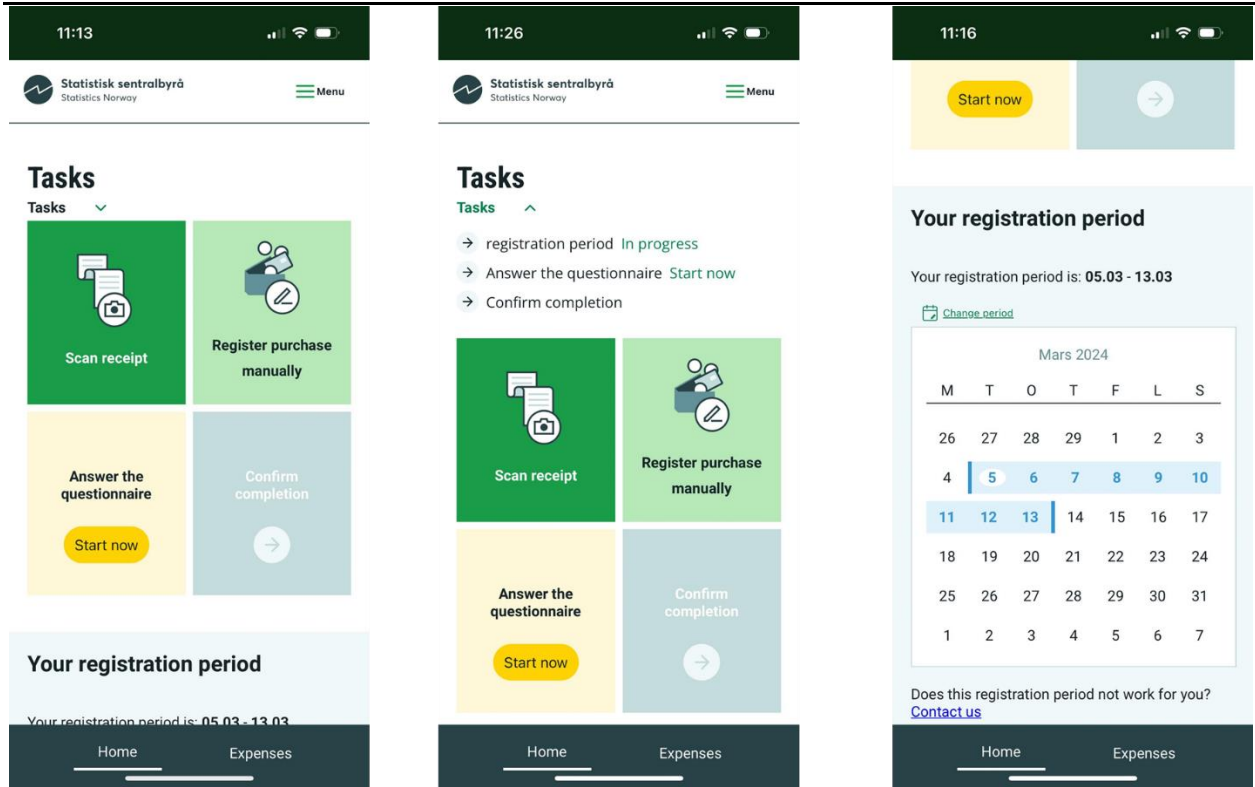


Figure B 7 Receipt scanning

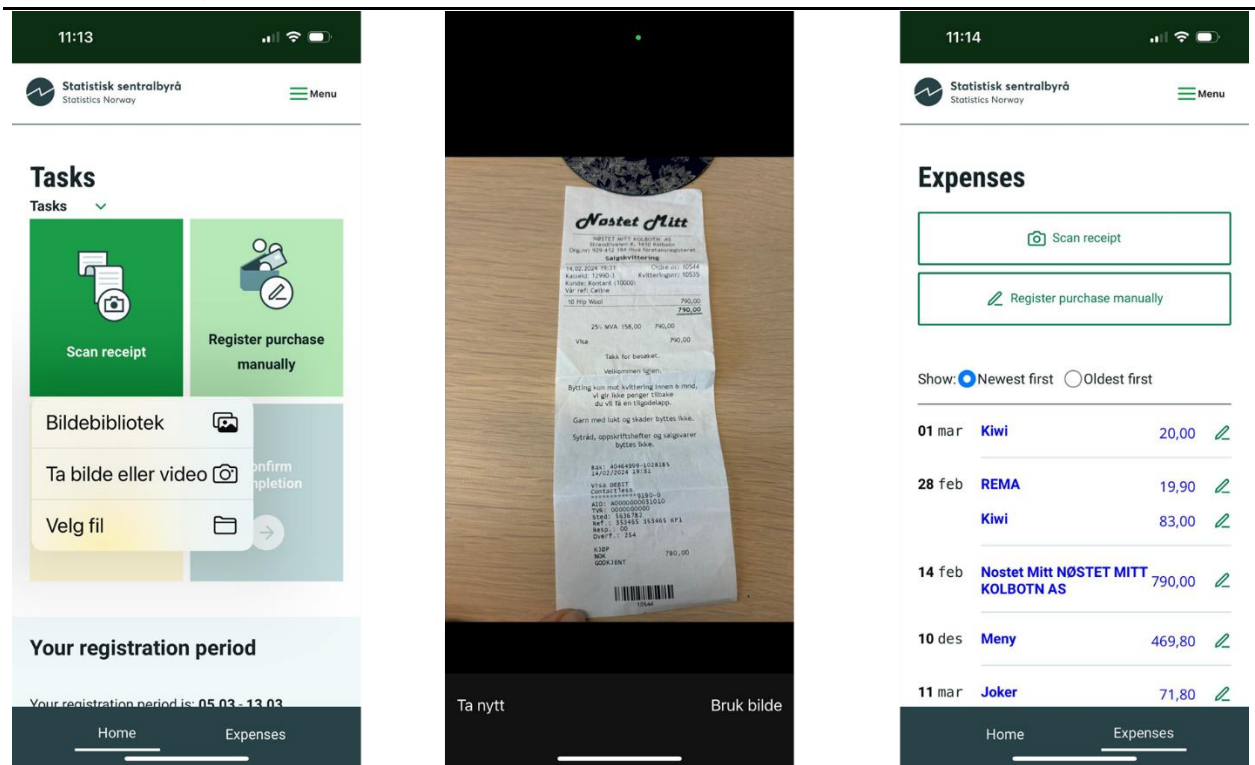


Figure B 8 Manual registration of purchase, part 1

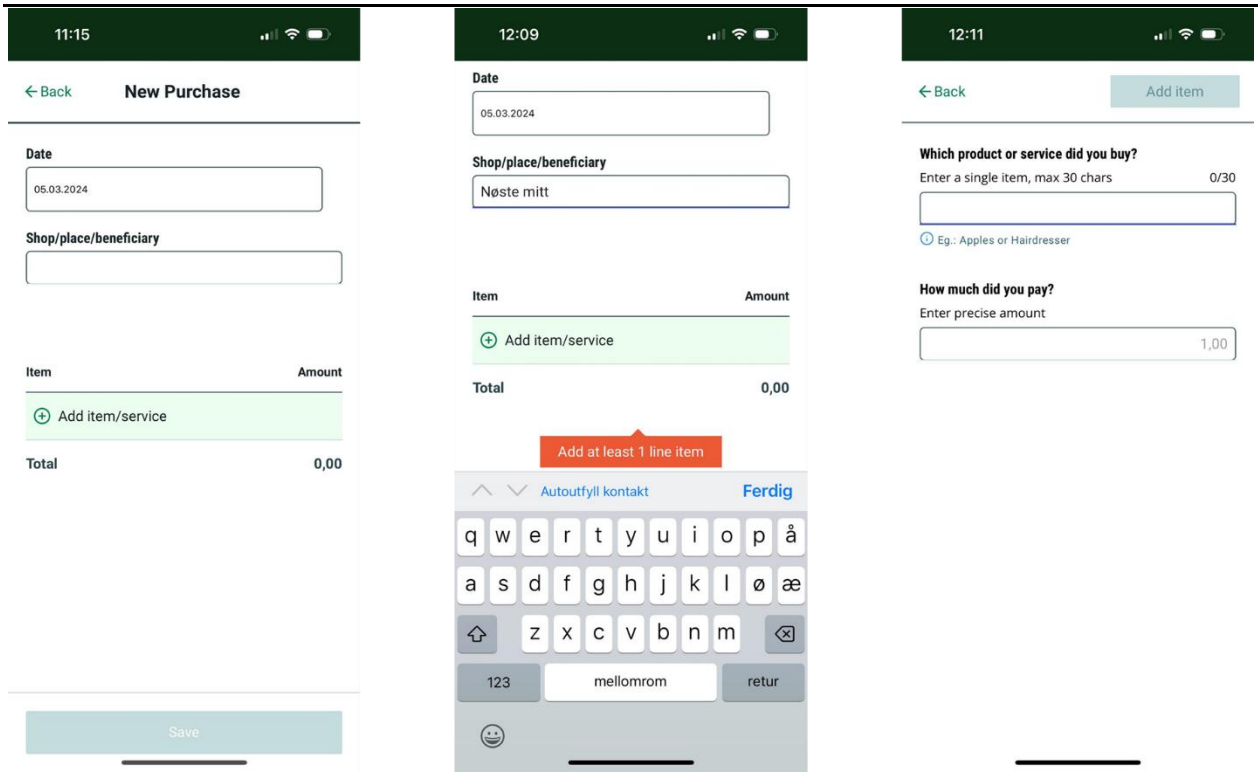


Figure B 9 Manual registration of purchase, part 2

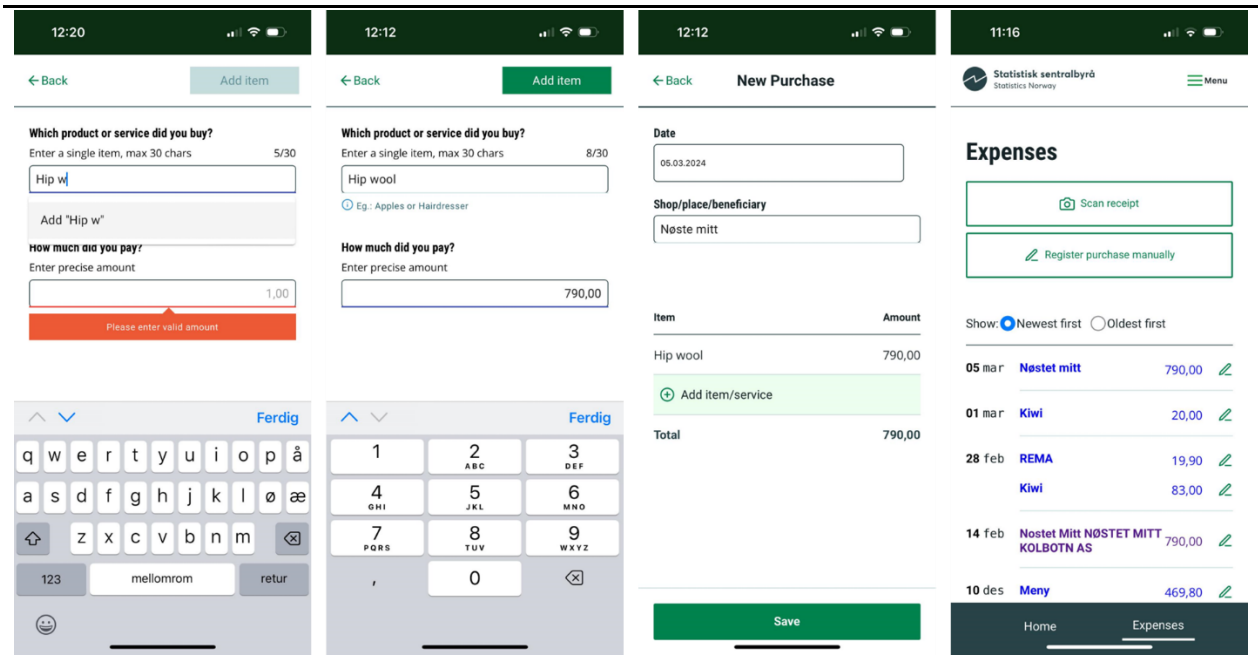


Figure B 10 Optional edit of scan/manual registration screen

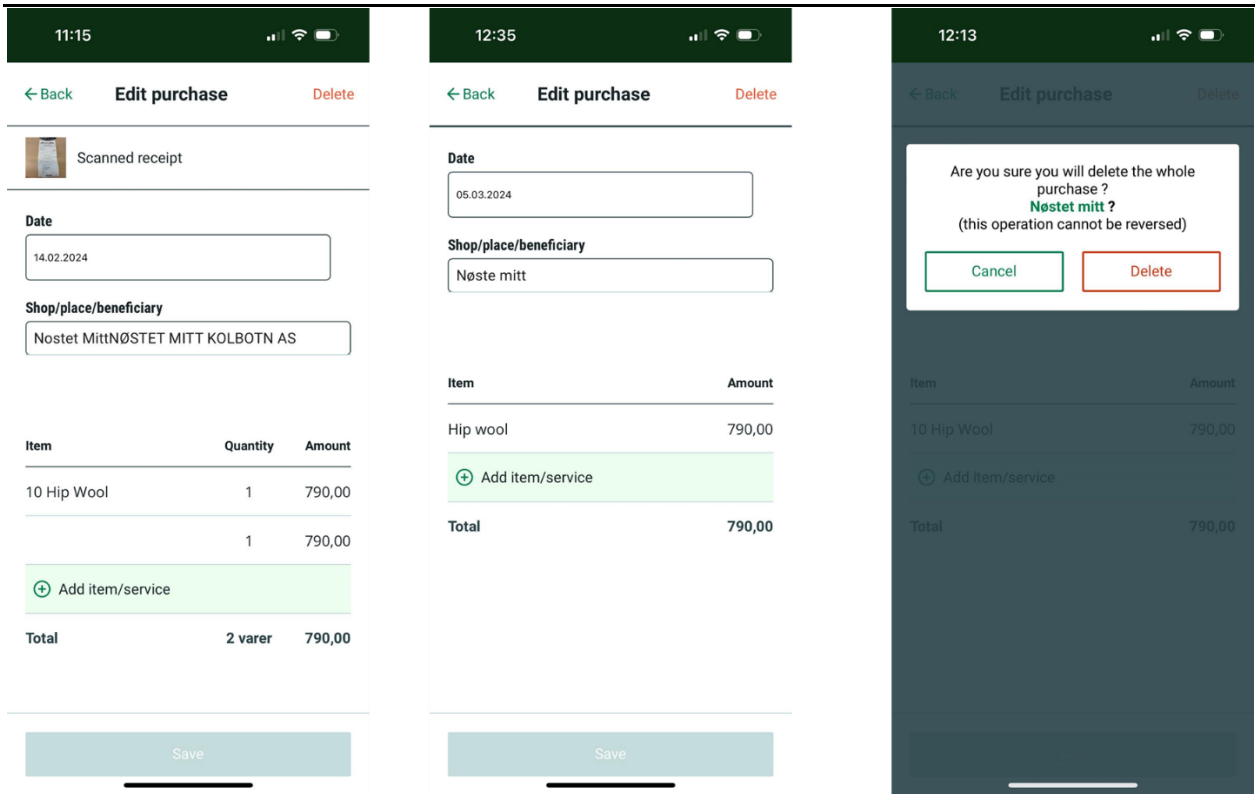


Figure B 11 Overview of registered expenses

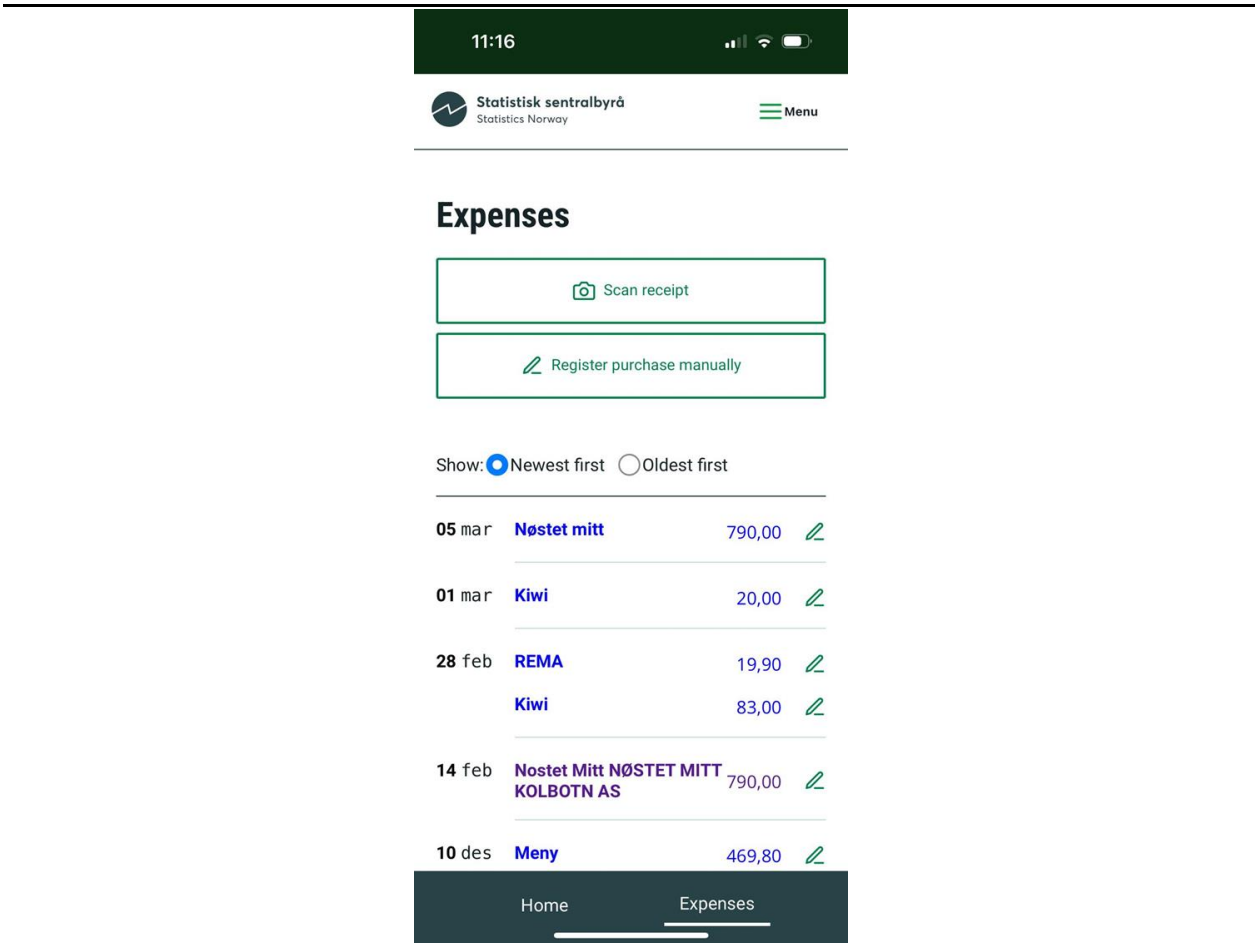


Figure B 12 Intro text for the questionnaire section of the app

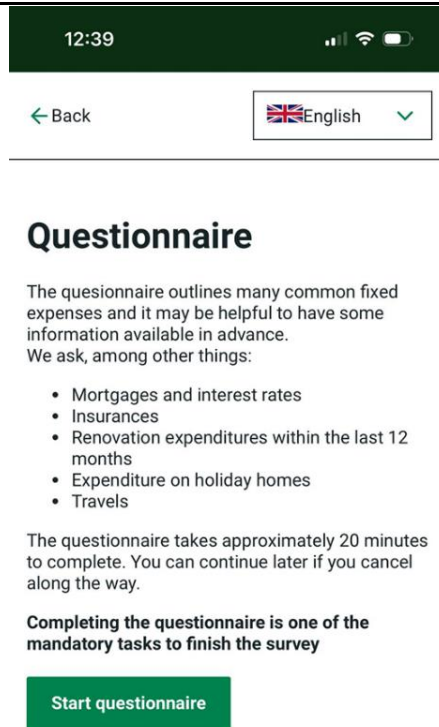
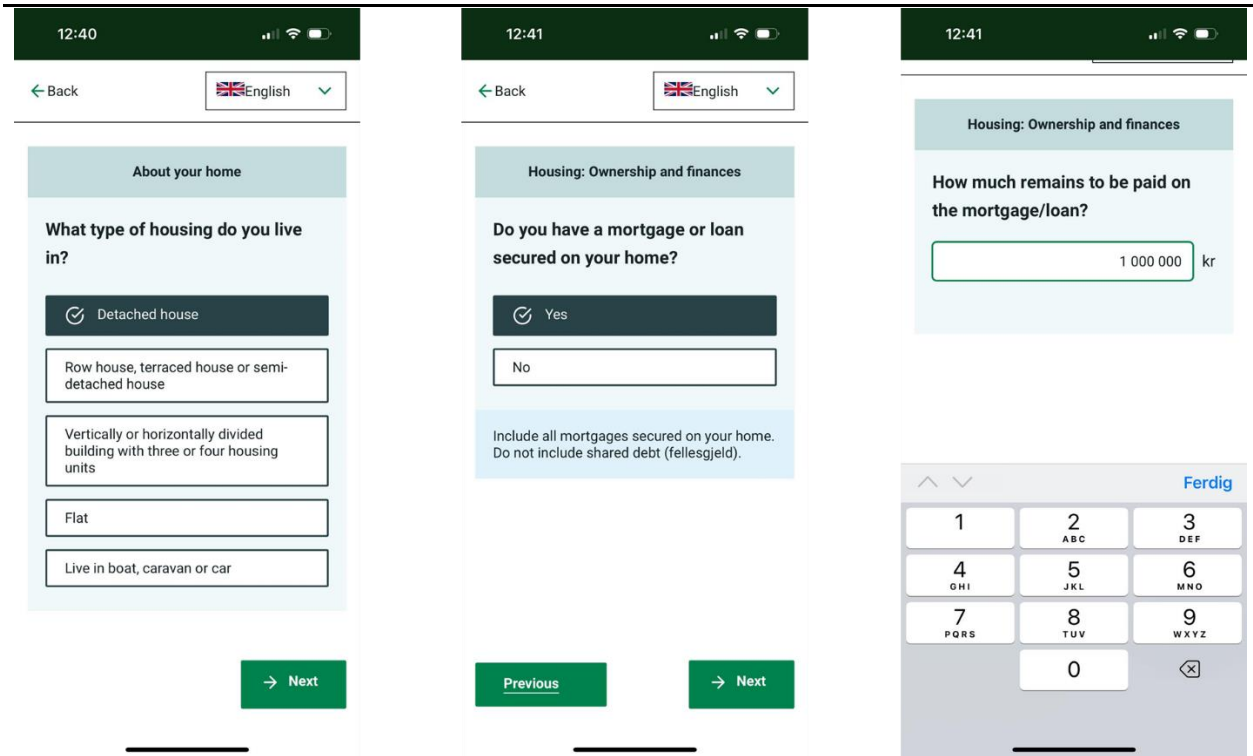


Figure B 13 The layout of random questions in the questionnaire section of the app



Appendix C: Reasons for non-response HBS 2022

Table C.1 Reasons for non-response for the HBS 2022 (preliminary figures)

Non-response Reasons HBS 2022	Number	Percent
Finished	3 525	30 %
Not started	3 259	27 %
Hard refusal, threats	2 346	20 %
Login information sent	767	6 %
Do not wish to participate	443	4 %
Started	276	2 %
Do not contact	251	2 %
Language problems	232	2 %
No internet, computer etc.	169	1 %
Long-term illness, impairment, mental health inhibited	136	1 %
Not time	131	1 %
Illness/death in the family, other unforeseen event	88	1 %
Other reason for withdrawal	79	1 %
Temporarily absent due to holiday etc.	48	0 %
Only phone number missing/wrong phone number	30	0 %
Resident abroad (6 months or more)	28	0 %
Others refuse for IO	28	0 %
Recruited	24	0 %
Temporarily absent due to school/work	20	0 %
IO resides at an institution	13	0 %
Only e-mail is missing	6	0 %
Does not give consent	4	0 %
Death	3	0 %
Other reasons for leaving - specify	2	0 %
Capacity problems, illness etc. at interviews	2	0 %
Missing both phone number and email	1	0 %

Source: The household Budget Survey 2022, Statistics Norway

List of figures

Figure 1.1	The development process for the app.....	8
Figure 2.1	The response process model adjusted.....	11
Figure 2.2	The process model for development of measurement instruments	13
Figure B 1	Home screen HBS-app	40
Figure B 2	Open app (PWA) in an internet browser and add to start screen (like a native app)	41
Figure B 3	Login once with the national ID-porten for secure logon	41
Figure B 4	Onboarding or “get started” instructions with consent.....	42
Figure B 5	Popup tutorial when respondent login (one time – can be found under Help later).....	42
Figure B 6	Home screen with and without drop-down menus open for task and for calendar.....	43
Figure B 7	Receipt scanning	43
Figure B 8	Manual registration of purchase, part 1	44
Figure B 9	Manual registration of purchase, part 2	44
Figure B 10	Optional edit of scan/manual registration screen	45
Figure B 11	Overview of registered expenses	45
Figure B 12	Intro text for the questionnaire section of the app	46
Figure B 13	The layout of random questions in the questionnaire section of the app	46

List of tables

Table 1.1	Participants in the user testing team for the Household Budget Survey.....	9
Table 3.1	Overview of conducted user tests	21
Table 3.2	Distribution of Samples per round, overall and by demographics.....	22
Table A.1	Participant in the HBS Team - DP2: Surveyløsning.....	39
Table C.1	Reasons for non-response for the HBS 2022 (preliminary figures)	47