



Office for
National Statistics



Statistics Finland



University of Essex

@HBS >An app-assisted approach for the Household Budget Survey

Testing Materials Test August- September 2019

Deirdre Giesen, Stefan Theunissen, Mart Nyholt – Statistics Netherlands
Brigita Vrabič, Mateja Zgonec- Statistics Slovenia
Marjaana Järvensinu, Anna Niemelä – Statistics Finland

project number 847257 — 2018-NL-INNOV

28 August 2019

summary This report documents the international templates for the testing materials for the first usability test round of the @HBS project. The three testing countries (Finland, the Netherlands and Slovenia) will use these templates for the development of the national documents. This report is a further specification of the Test plan @HBS 2019 and includes a) recruitment goals b) recruitment screening list c) information material for test respondents d) example of a consent form for test respondents e) test protocol and f) reporting template.

Content

1. Background and purpose of this report	4
2. General design test round 1	4
2.1 Scope of app tested	4
2.2 Goal of the test	5
2.3 Overall structure of the test	5
2.4 To be evaluated in the test	6
2.5 Interviewing and observation during tests	7
2.6 Time frame test round 1	7
3. Recruitment	7
4. Communication and motivation strategy	9
5. Information for test respondents prior to test interview	9
5.1 All respondents	9
5.2 Full observation group	9
5.3 Retrospective testing group	9
6. Full Observation Group - protocol content	10
6.1 Materials needed to be brought to the test:	10
6.2 Introduction (10 minutes)	10
6.3 Discussion/test of contact/motivation strategy and materials (10 minutes)	11
6.4 Installing the app (5 minutes)	11
6.5 First opening of the app (5 minutes)	12
6.6 Using the app (40 minutes)	12
6.6.1 Spontaneous use	12
6.6.2 Evaluation spontaneous use	14
6.6.3 Scripted use observation & evaluation	14
6.7 Assessing quality of scanned receipts (5 minutes)	15
6.8 Evaluation communication materials and overall data collection strategy (10 minutes)	15
6.9 Rounding off (5 minutes)	16
7. Retrospective interview group – protocol content	16
7.1 Materials needed to be brought to the test:	16
7.2 Introduction (10 minutes)	16
7.3 Evaluation of experiences during independent test (20 minutes)	19
7.3.1 General impression	19
7.3.2 Task evaluation	19
7.3.3 Evaluation of communication materials	20

7.4	Scripted tasks (10 minutes)	20
7.4.1	Scripted use observation & evaluation	20
7.5	If time allows: Assessing quality of scanned receipts (5 minutes)	20
7.6	If time allows: HBS other issues (5 minutes)	20
7.7	Rounding off (5 minutes)	21
8.	National reporting template	21
9.	List of digital attachments	22
9.1	Attachment 1 Recruitment questionnaire template	22
9.2	Attachment 2 Content of Excel document for recruitment	22
9.3	Attachment 3 Manual for installation and use of app as used in NL	22
9.4	Attachment 4 Form for comments of respondents retrospective interview group	22
9.5	Attachment 5 Example of consent form	22
9.6	Attachment 6 Checklist for scanned receipts	22
9.7	Attachment 7 National reporting template	22

1. Background and purpose of this report

The @HBS project develops an app-assisted approach towards data collection for the household budget survey (HBS). This app contains both generic and country-specific components (e.g. expenditure classification). ONS, Stat Austria, Stat Finland, Stat Netherlands, Stat Slovenia and the University of Essex co-operate in the @HBS project. Part of the project is performing small scale cognitive and usability testing of both the app and the accompanying recruitment and motivation strategy and materials (e.g. letter, incentives, feedback). The testing will be conducted in Finland, Slovenia and the Netherlands. The University of Essex and ONS will review the test plan and the test results. Also, Statistics Austria has volunteered to provide feedback on the test plan.

This report is a further specification of the first usability test round as described in the “Test plan @HBS 2019”, This report includes

- a) recruitment goals
- b) recruitment screening list
- c) information material for test respondents
- d) example of consent form for test respondents
- e) test protocol and
- f) reporting template.

The purpose of this document is to describe guidelines and templates for all relevant testing materials. The three testing countries (Finland, the Netherlands and Slovenia) will use these for the development of the national documents. Materials will be adapted to national requirements. Test protocols and reporting templates may be adapted during testing.

2. General design test round 1

2.1 Scope of app tested

The Household Budget Survey (HBS) app is available in both in the Google play store and iOS app store. Upon opening the app users choose the country. Available are now UK, Suomi, Nederland, Slovenija.

The main functionalities of the app tested in this test round include

- An overview page with a calendar with a 30 day reporting period and an overview of incentives earned.
- Registration of expenses per day by either:
 - Scanning of a receipt per shop and manually providing information about type of shop, total price of the receipt and the types of products/services.
 - Manual registration per shop of prices of all purchased products/services; shops and products/services can be either automatically coded by finding them in search lists or coded manually.
- A chronological list of all entered expenses
- A visualization of all types of expenses (per category and over time)
- Settings enabling notifications

Note: For this test round the back-end of the app is not developed far enough to actually collect and store data other than on the respondents' phone. For the November/December test round, the back-end should be developed to the extent that is possible actually collect data entered. For the next test round the app will also be extended with the option to use geolocation to help remember expenses and consent for big data linkage.

2.2 Goal of the test

The overall goal of the test is to evaluate both the recruitment & motivation strategies/materials and the app, as adapted for Finland, the Netherlands and Slovenia .

Regarding the recruitment & motivation strategy the goal of the test is to explore:

- How do respondents spontaneously interact with the recruitment & motivation materials (e.g. do they notice all materials, do they only look briefly at the material or actually read?).
- How well can respondents understand the recruitment & motivation materials?
- How do respondents evaluate the recruitment & motivation materials (e.g. how do respondents judge the attractiveness, relevance, clarity).
- What are factors that would motivate respondents to participate, what are barriers for them to participate ?

Regarding the app the goal of the test is to explore:

- How easy or difficult is it for respondents to find and install the app?
- How well does the app communicate the response task: e.g., meets the respondents' needs for understanding the required task?
- How well does the app facilitate performing the response task (usability of using the app, e.g. meets the respondents' needs for completing the required tasks)?
- How do respondents perceive the task of responding via the app compared to alternatives (e.g. a paper questionnaire, a web questionnaire, collecting paper receipts).

For any issues found regarding the questions above, the tests will explore what the problems were and why they occurred.

2.3 Overall structure of the test

The general approach of the test is:

- Make respondents perform the tasks that are essential for the use of the app, in a way that is as realistic as possible and evaluate these tasks with them. In each test we will evaluate effectiveness (are respondents able to perform tasks), efficiency (to what extent can they do so quickly / easily) and satisfaction (how do the respondents perceive the app in general and specific tasks within the app).
- In each country we test as much as possible in a similar way and document findings in a template that allows easy comparison of findings in the three countries.

Each country will recruit 12 test respondents who are not familiar with the project and/or the HBS survey and split the test respondents over two groups:

- Group 1 "full observation group": respondents who install the app and perform several tasks with the app (based on their own receipts brought to the test and vignettes) while being observed and videotaped. After observation, the respondents will be interviewed. Observation and interview will take about 1.5 hour.

- Group 2 “retrospective interview group”: respondents who are provided with materials to independently install the app at home, use it for at least three days and will then be interviewed in person. After the retrospective interview they will also be asked to perform some of the same tasks as group 1, will be observed and videotaped doing this and these tasks will also be evaluated in an interview. The respondents are assumed to spend about half an hour prior to the interview installing the app and entering some expenditures, the retrospective interview will take about 1 hour.

From the group 1 observations and interviews we expect to get detailed and recorded information of how people interact with the app. From the group 2 respondents we expect to receive more realistic user experience feedback of people using the app independently, in real life situations, with real expenditures and over several days.

All interviews and observations may be done in the lab or somewhere else (e.g. the respondents home). This choice will depend on the preferences and possibilities of the researchers and respondents. Recommended setting is a place where respondent and interviewer will not be disturbed by others.

2.4 To be evaluated in the test

- a. Proposed communication strategy for an app assisted approach for the HBS
- b. App assisted data collection as a method for the HBS
- c. Installation of the app (and related instruction materials)
- d. To test most important features of the app we aim to test the following tasks (most tasks cannot be tested for each respondent as testing time will not allow that, but we aim to test each task for each country for at least a few respondents)
 - o at least a few real expenses made by respondents’ household (for each respondent)
 - o expenditures over more than one day
 - o a simple short receipt (1 item) – both manually and scanned
 - o a short receipt with different items – both manually and scanned
 - o a long receipt (8 items or more)
 - o duplicate products (e.g. three identical cans of soup) – manually
 - o products with a discount – manually
 - o discount on the overall receipt – manually sea
 - o adding a shop or product/service not found in the search list
 - o an online purchase
 - o a purchase abroad
 - o completing a day with expenses
 - o competing a day without expenses
 - o deleting an entry
 - o changing an entry
 - o using Overview page (calendar etc.)
 - o using Expenses overview
 - o using Insights
 - o using Settings

2.5 Interviewing and observation during tests

General instruction for interviewer and/or observer:

- Be neutral. Only show respondent that you are interested in what they are doing or saying, do not evaluate anything they do or say (so not: that seemed to go easy, excellent!).
- If respondent asks for help/explanation/confirmation; do not help too quickly, first ask: what would you do if I was not here?
- If respondents get to the point they would stop or call a help desk you can help the respondent.

General approach while observing respondent performing tasks:

- Ask respondent to perform tasks as much as possible as they normally would if they were using the app in a real survey.
- Ask them to think aloud. If needed probe with “what are you thinking?”, “what are you looking at?”, “please keep telling me what you are doing”. Occasionally give feedback like “mmm, mmm” or “thank you” to indicate their comments are heard and appreciated. Do not overburden respondent with too many probes. Not every respondent is comfortable thinking aloud and probing them to do so may interfere with the response task.
- Observe which steps respondent takes to perform their tasks.
- Note any behavior that shows a suboptimal flow of the task performance (e.g. tapping several times on a button, looking for help, trying the wrong button).
- If respondents look for help/additional information observe where they look for this (e.g. in the recruitment materials, in the app – where specifically in the app?)
- Note (non)verbal responses of the respondent indicating emotions (e.g. surprise, frustration).

2.6 Time frame test round 1

	What	Who	Deadline
1.	Testing period	Stat Finland, Stat Netherlands, SORS	August 26-Sept 13
2.	National test reports ready	Stat Finland, Stat Netherlands, SORS	September 25
3.	Preparing draft summary report overall test results	SORS	October 7
4.	Feedback on draft summary report	All partners	October 14
5.	Finalization summary report test round 1	SORS, Stat Finland, Stat Netherlands	October 28

3. Recruitment

Test respondents for the usability tests will be recruited to reflect the target population for the @HBS app and various types of possible response /usability problems.

During the recruitment the following background characteristics will be collected from the respondents. This is needed to check the composition of the test groups and for analysis and reporting. See attachment 1 for a template of a questionnaire that can be used during recruitment. See electronic attachment 2 for an Excel template to register recruited respondents.

1. Contact information
2. Age
3. Gender
4. Work status
5. Highest current or attained educational level
6. Number of persons in household
7. Number and ages of children in household
8. If main shopping is done by respondent, respondent and others or by others
9. Extent to which respondent usually keeps track of expenditures (rather well, a bit, hardly or not at all)
10. Type of phone
11. Use of phone
12. Ever installed app on phone
13. How easy or difficult to install app

Recruitment goals :

- a. Only respondents who are used to using smartphones and apps (if no experience it is not very likely that they would use the app for HBS and they may have usability issues that are not typical for the target population; however experience with installing apps is not necessarily required, as some groups of users may typically not install apps themselves).
- b. For both group 1 and group 2 we should have at least:
 - both iOS users and Android users.
 - A single household, a multi-person household
- c. We also aim to get for each group:
 - Some variation in ability to install and use apps
 - Both people who normally keep track of their spending and people who do not
 - A person who is the main shopper for the family and someone who is not the main shopper
 - A family with teenagers (because of issues of keeping track of their expenses)
 - A family with young children (because these typically are very time pressured)
 - Some variation in educational level
 - Both men and women
 - Different age groups (≥ 18)

Respondents can be recruited via various methods as practical for the NSI, for example flyers in supermarkets, ads, or recontacting respondents who have participated in a NSI survey before and have allowed the NSI to contact them again. People with knowledge of statistical data collection, the HBS survey or app development should be excluded as test respondents.

Each country documents their specific recruitment strategies and results in the country specific test report.

4. Communication and motivation strategy

The communication and motivation strategy evaluated in the test will be country specific, adapted to national field work requirements. For the Netherlands for example, in the test we will provide the respondents with a draft advance letter as proposed for the real implementation of an app assisted, a leaflet about the HBS from the last HBS implementation and a short booklet with instruction on how to download the app (including name, QR code and link for both appstores) , explanation of how the app works and a list of FAQs (see attachment 3). For a next test round and for implementation this information will be made available on a dedicated website. In the test report we will describe the communication and motivation strategy and materials as evaluated in the tests for each country.

5. Information for test respondents prior to test interview

5.1 All respondents

All respondents will be briefed in general about the purpose and general design of the test during recruitment (in the Netherlands this will be done via email and telephone). During this briefing also the appointment is made for the interview.

5.2 Full observation group

This group will receive a letter and/or email with the following content

- confirming time and place of the appointment for the interview
- mention interview will take about 1,5 hours
- reminder to bring
 - their charged phone
 - if needed password to install apps
 - at least five receipts of recent expenses made by their household (more receipts are welcome!)
- confirmation that data will be treated confidentially
- name and contact details of contact person in case they have any questions or need to reschedule
- If applicable; confirmation (if applicable) of the promised incentive and information on how to reach the test location

5.3 Retrospective testing group

The retrospective testing group will be briefed about the test during recruitment. During this briefing also the appointment is made for the interview in the lab. Installing the app and the three reporting days should be as close as possible to the interview date.

Following up on the recruitment, respondents from the retrospective testing group will receive the following information via email and/or paper mail:

- A letter confirming what has been discussed on the phone :
 - install the app at or before date X

- register all expenses of their household for dates XYZ [tell them on the phone if this is too much work to try to do as much as possible]
- collect as many receipts as possible of the expenses of their household during XYZ and bring these to the interview
- request to write down some short notes on their experiences installing and using the app in the provided template and bring this to the interview
- date, time and location of the interview (as close as possible after Z)
- contact person and contact details if any problems or questions
- if relevant: promise of incentive as agreed; Information on how to reach the test location
- Attachment with a form for comments for respondents; a template for reporting any experiences installing and using the app (attachment 4).
- Attachment with info on how to install and use the app (appendix 3).

6. Full Observation Group - protocol content

6.1 Materials needed to be brought to the test:

- a. Protocol for interviewer and (if present) for extra observer
- b. Form with information on respondent collected during recruitment
- c. If in the office; Wi-Fi code for respondent
- d. Consent form for respondent (note: the consent form should allow to make a recording and to share this within the international project team – see attachment 5 for example of consent form used in the Netherlands).
- e. Recording equipment
- f. Communication materials that are tested (e.g. advance letter, manual for respondent).
- g. Vignettes with receipts (these will serve as back up receipts in case respondents did not bring enough receipts to cover all types of expenditures we want to test).
- h. Checklist to assess quality of scans (see attachment 6).
- i. An Android and iOS test phone that are fully charged, updated and that do not have the app installed yet and search history for browser erased.
- j. If relevant: gift card

6.2 Introduction (10 minutes)

- Introduce all persons present (also if observing in another room)
- Thank respondent for help with test
- Explain that for a European project a new way for data collection is being developed and that this test is intended to get feedback on how this works. d
- Assure confidentiality
- Make sure the respondent understands we are not testing the respondent but the materials and that there are no wrong or right answers. All comments are welcome, both positive and negative.
- Explain structure of the test:
 - First we want you to look at the letter / walk you through the way would normally contact respondents for this survey [country specific intro of approach and motivation strategy/materials].
 - Next we would like you to try out the app.

- Finally we will ask some additional questions
- This will take about 1,5 hours.
- Ask for consent. If respondent refuses to be video recorded see if at least an audio recording is possible and/or only recording the screen and hands of the respondent.
- Preferably by other researcher while the interviewer continues: Check receipts respondent has brought to the test. Check to which extent the receipts have the following characteristics (note: one receipt may score on more than one characteristic):
 - i. at least two receipts with different dates
 - ii. simple and short receipt with only 1 item (without special characteristics such as online/abroad/discount)
 - iii. short receipt with about 3 different items (with or without special characteristics)
 - iv. long receipt with many (>8) items (with or without special characteristics)
 - v. receipt 2 or more identical items
 - vi. receipt with discount on a product/service
 - vii. receipt with discount on the total of the receipt
 - viii. online expenditure
 - ix. expenditure abroad

Select a number of receipts from the receipts the respondent has brought to the test to cover as much as possible the different characteristics (at least three). Note which characteristics are missing and select the prepared vignettes that are covering these characteristics.

- If in the office: ask if respondent wants to use WiFi
- If at respondents home: discuss where in the home the test can best be done, if needed ask to change to a more quiet room, to sit at a table, turn down background music etc.
- Ask if respondent wants to put phone on 'do not disturb' to prevent that private messages will show during the test.
- Document which type of phone is used and which operating system.
- Ask if respondent has any questions before proceeding

6.3 Discussion/test of contact/motivation strategy and materials (10 minutes)

- Show or discuss with respondent how they would be recruited for a real implementation of an app assisted HBS (e.g. give respondent the advance letter in an envelope and/or explain contact strategy).
- How would you probably respond to such a strategy. Why?
- If a letter is available:
 - Assume you would open the envelope, please check this letter as you normally would. Observe how respondent reads the letter.
 - How would you normally respond to this letter? Why?
 - Here also discuss, if applicable, who would open a letter addressed to the household in their household.

6.4 Installing the app (5 minutes)

- Now assume you want to participate, how would you proceed? [goal: test if materials provided are clear enough to help respondent find and install the app]
- When the respondent starts to use phone make sure that phone use is recorded.

- Interviewer stays in the room with the respondent, preferably observing from a short distance on a laptop.
- Observe how respondent tries to install app. Code this as:
 - 1) respondent completed the task independently without difficulties and errors [=ideal flow, efficient and effective]
 - 2) respondent completed the task independently but had some difficulties and/or made some minor errors
 - 3) respondent completed the task independently but with great difficulty and/or great errors
 - 4) respondent could only complete task with help
 - 5) respondent could not complete the task, even with help
- Make notes of any difficulties encountered by respondent
- Make notes of any (non)verbal communication indicating emotions about the task

6.5 First opening of the app (5 minutes)

- Observe how respondent proceeds after opening the app .
- Encourage respondent to act as he/she normally would.
- Note what respondent does and any issues that disturb an optimal flow
- Make notes of any (non)verbal communication indicating emotions about the app or response task

6.6 Using the app (40 minutes)

6.6.1 Spontaneous use

After the respondent has looked at the app and maybe has spontaneously tried out some functionalities explain:

- Please use these receipts to enter your expenditures in the app (hand over selection of receipts in random order). Please do as you would normally do if you were really using the app.
- Let respondent at least use two or three own receipts first before vignettes are introduced. Score which tasks have been tried/done and how this went (see table Scoring of task performance).
- If needed (see above): after entering own receipts provide vignette receipts so respondent experiences receipts with as many of all characteristics as listed under 6.2 as time allows. Hand over vignette 1. First please take your time to read. Do you have any questions? [Note: it is important that we give respondents time to process and understand the vignette, if needed walk them through each line of the receipt; to prevent issues with understanding the vignette to interfere with issues using the app] How you would register this expense in the app?. Please remember to think aloud. (if needed: to be repeated with several vignettes)
- Code for each of tasks if and how respondent performed as in table below.

Table 1 Scoring of task performance full observation group

1= completed task independently without difficulties and errors [=ideal flow, efficient and effective]; 2= completed task independently but had some difficulties and/or made some minor errors; 3= completed the task independently but with great difficulty and/or great errors; 4= could only complete task with help; 5= could not complete the task, even with help.

Task nr	Task	Performed at least once in test? 0=no 1=yes	Task performance (if task performed more than once and with different results note both and explain in notes)	Notes e.g. types of issues, emotions, learning curve, different findings for manual / scanning etc.
1.	Installing the app		1 – 2 - 3 – 4 – 5	
2.	Manual registration of a simple receipt with one item only (without online, abroad, discount)		1 – 2 - 3 – 4 – 5	
3.	Scan of a simple receipt with one item only		1 – 2 - 3 – 4 – 5	
4.	Manual registration of a short receipt with about three different items		1 – 2 - 3 – 4 – 5	
5.	Scan of a short receipt with about three different items)		1 – 2 - 3 – 4 – 5	
6.	Spontaneous registration of a long receipt (if R chooses manual registration you may stop them after 3 or 4 items to save time).		1 – 2 - 3 – 4 – 5	
7.	Scan of long a receipt		1 – 2 - 3 – 4 – 5	
8.	Manual registration of a receipt with 2 or more identical items		1 – 2 - 3 – 4 – 5	
9.	Manual registration of discount on a product		1 – 2 - 3 – 4 – 5	
10.	Manual registration of discount on a receipt		1 – 2 - 3 – 4 – 5	
11.	Adding a shop or product/service not found in the search list		1 – 2 - 3 – 4 – 5	
12.	Manual registration of expenditure abroad		1 – 2 - 3 – 4 – 5	
13.	Manual registration of online expenditure		1 – 2 - 3 – 4 – 5	

Task nr	Task	Performed at least once in test? 0=no 1=yes	Task performance (if task performed more than once and with different results note both and explain in notes)	Notes e.g. types of issues, emotions, learning curve, different findings for manual / scanning etc.
14.	Change price of a manually registered expense		1 – 2 - 3 – 4 – 5	
15.	Change price of scanned receipt		1 – 2 - 3 – 4 – 5	
16.	Delete one item in a list of more than one manually registered items		1 – 2 - 3 – 4 – 5	
17.	Add an item to an manually entered transaction		1 – 2 - 3 – 4 – 5	
18.	Completing a day with expenses		1 – 2 - 3 – 4 – 5	
19.	Competing a day without expenses		1 – 2 - 3 – 4 – 5	
20.	Removing the app		1 – 2 - 3 – 4 – 5	

6.6.2 Evaluation spontaneous use

Thank you! What is your overall impression of using this app ?

In app stores apps can be rated with a five star system. How many stars would you give this app?

Why would you give this rating?

To what extent was this app and the tasks involved what you expected when you installed the app?

Was there anything that surprised you? Anything that was easier or more difficult than you expected?

Where there (other) things you liked about this app?

Where there (other) things you did not like about the app?

Are there any functionalities you miss?

[Note: first evaluation with respondent about spontaneous use should be done before scripted use; by giving respondents scripted tasks we will affect how they perceive the app (e.g. making them aware of functionalities they may not have noticed).]

6.6.3 Scripted use observation & evaluation

Check which of the tasks the respondent has not yet performed and select as many as them as time of this part of the test allows. It will probably not be possible to test all tasks with all respondents in each country. We will try to make sure that at least all tasks are tested with some respondents. Also try to give tasks in a way that each respondent at least:

- Enters a receipt manually. (“So far you scanned all receipts. What if you had an expense without a receipt? For example, what if you bought flowers in a flower stall for 25 euro’s. How would you enter this?)
- Scans a receipt (“So far you have manually entered all expenses. Did you see there is another way to enter them? Could you please try that with one of the receipts?”)
- Sees and discusses Expenses overview (What would you expect here? Please have a look. What do you think?)
- Sees and discusses Insights (What would you expect here? Please have a look. What do you think? Make sure respondent looks both at ‘by category’ and ‘over time’.)
- Sees and discusses Settings (What would you expect here? Please have a look. What do you think?)

After each scripted task, the task can be briefly evaluated with the respondent.

6.7 Assessing quality of scanned receipts (5 minutes)

Randomly select three scanned receipts and assess the quality according to the checklist. If short on time this can be skipped and may be done afterwards based on the video recordings of the respondents’ phone.

6.8 Evaluation communication materials and overall data collection strategy (10 minutes)

- a) Review communication materials (letter, instruction you tube movies...) / strategy with the respondent. Goal is to get an impression of how useful the materials are for the respondent, which parts are particularly helpful and if and how the materials should be improved.
- b) Discuss where respondents would expect the information about the response task to be – it could be in the recruitment materials, within the app as a series of screens they go through the first time they log in, or within the app as an FAQ section, or on a project related web site – what would respondent expectations be about how they can access this information?
- c) Discuss alternative ways for HBS data collection (e.g. paper questionnaire, web questionnaire, collecting receipts), which data collections method(s) would the respondent prefer, why?
- d) Here also discuss the use of an app / smartphone for data collection: would only an app be OK for respondents, or would they also like a PC version? Why?
- e) Would they like other household members to also have access to the app from their own phones? How would they feel about app in which all expenses entered by all household members can be visible (but can also be hidden from other household members)?
- f) Would they like the app to have the option to register and show geolocation data information to help them remember places for possible expenses and what they did on a particular day?
- g) How would they feel about the option to give permission to
 - use scanner data based on time/location
 - use bank transaction data
 - use data from some loyalty cards

- h) If not discussed yet spontaneously or during recruitment: to what extent and how does respondent usually keep track of expenses? (precisely, somewhat, not; methods used)

6.9 Rounding off (5 minutes)

Anything else you would like to share with us about this app or how we plan to approach respondents?

Thank for help. If applicable: Give gift card

Ask if we may re-contact respondent for additional questions or a new test if needed.

Ask if we may keep (some) of their receipts (make sure any personal information from collected receipts is made invisible).

If applicable: remind respondent to switch off "do not disturb".

Rate the smart phone skills of the respondent (1 low level of competence, 2=average level of competence, 3=high level of competence)

7. Retrospective interview group – protocol content

7.1 Materials needed to be brought to the test:

- a. Protocol for interviewer and (if present) for extra observer
- b. Form with information on respondent collected during recruitment and testing phase (any help needed during independent testing?)
- c. If in the office; Wi-Fi code for respondent
- d. Consent form for respondent (note: the consent form should allow to make a recording and to share this within the international project team).
- e. Recording equipment
- f. Communication materials that are tested (e.g. advance letter, manual for respondent).
- g. Vignettes with receipts (these will serve as back up receipts in case respondents did not bring enough receipts to cover all types of expenditures we want to test).
- h. Checklist to assess quality of scans (see attachment 6).
- i. An Android and iOS test phone that are fully charged, updated and that do not have the app installed yet and search history for browser erased.
- j. If relevant: gift card

7.2 Introduction (10 minutes)

- Introduce all persons present (also if observing in another room)
- Thank respondent for help with test
- Explain that for a European project a new way for data collection is being developed and that this test is intended to get feedback on how this works.
- Assure confidentiality
- Make sure the respondent understands we are not testing the respondent but the materials and that there are no wrong or right answers. All comments are welcome, both positive and negative.
- Explain structure of the test:
 - o First we want to evaluate the respondents experiences with the app and the accompanying communication materials.

- Next we will ask the respondent to try some specific tasks with the app.
- Finally we will ask some additional questions
- This will take about 1 hour
- Ask for consent. If respondent refuses to be video recorded see if at least an audio recording is possible and/or only recording the screen and hands of the respondent.
- (Preferably by other researcher while the interviewer continues:) Check receipts respondent has brought to the test. Check to which extent the receipts have the following characteristics (note: one receipt may score on more than one characteristic):
 - i. at least two receipts with different dates
 - ii. simple and short receipt with only 1 item (without special characteristics such as online/abroad/discount)
 - iii. short receipt with about 3 different items (with or without special characteristics)
 - iv. long receipt with many (>8) items (with or without special characteristics)
 - v. receipt 2 or more identical items
 - vi. receipt with discount on a product/service
 - vii. receipt with discount on the total of the receipt
 - viii. online expenditure
 - ix. expenditure abroad

Note which characteristics are missing and select the prepared vignettes that are covering these characteristics.

- If in the office: ask if respondent wants to use WiFi
- If at respondents home: discuss where in the home the test can best be done, if needed ask to change to a more quiet room, to sit at a table, turn down background music etc.
- Ask if respondent wants to put phone on 'do not disturb' to prevent that private messages will show during the test.
- Document which type of phone is used and which operating system.
- Ask if respondent has any questions before proceeding

Table 2 Scoring of task performance retrospective interview group

1= completed task independently without difficulties and errors [=ideal flow, efficient and effective]; 2= completed task independently but had some difficulties and/or made some minor errors; 3= completed the task independently but with great difficulty and/or great errors; 4= could only complete task with help; 5= could not complete the task, even with help.

Task nr	Task	Performed during independent test at home? 0=no 1=yes	Performed while observed? 0=no 1=yes	Task performance (if task performed more than once and with different results note both and explain in notes)	Notes e.g. types of issues, emotions, learning curve, different findings for manual / scanning etc.
1.	Installing the app			1 - 2 - 3 - 4 - 5	
2.	Manual registration of a simple receipt with one item only (without online, abroad, discount)			1 - 2 - 3 - 4 - 5	
3.	Scan of a simple receipt with one item only			1 - 2 - 3 - 4 - 5	
4.	Manual registration of a short receipt with about three different items			1 - 2 - 3 - 4 - 5	
5.	Scan of a short receipt with about three different items)			1 - 2 - 3 - 4 - 5	
6.	Spontaneous registration of a long receipt (if R chooses manual registration you may stop them after 3 or 4 items to save time).			1 - 2 - 3 - 4 - 5	
7.	Scan of long a receipt			1 - 2 - 3 - 4 - 5	
8.	Manual registration of a receipt with 2 or more identical items			1 - 2 - 3 - 4 - 5	
9.	Manual registration of discount on a product			1 - 2 - 3 - 4 - 5	
10.	Manual registration of discount on a receipt			1 - 2 - 3 - 4 - 5	

11.	Adding a shop or product/service not found in the search list			1 – 2 – 3 – 4 – 5	
12.	Manual registration of expenditure abroad			1 – 2 – 3 – 4 – 5	
13.	Manual registration of online expenditure			1 – 2 – 3 – 4 – 5	
14.	Change price of a manually registered expense			1 – 2 – 3 – 4 – 5	
15.	Change price of scanned receipt			1 – 2 – 3 – 4 – 5	
16.	Delete one item in a list of more than one manually registered items			1 – 2 – 3 – 4 – 5	
17.	Add an item to an manually entered transaction			1 – 2 – 3 – 4 – 5	
18.	Completing a day with expenses			1 – 2 – 3 – 4 – 5	
19.	Competing a day without expenses			1 – 2 – 3 – 4 – 5	
20.	Removing the app			1 – 2 – 3 – 4 – 5	

7.3 Evaluation of experiences during independent test (20 minutes)

7.3.1 General impression

How did it go?

What is your overall impression?

To what extent was this app and the tasks involved what you expected when you installed the app?

Was there anything that surprised you? Anything that was easier or more difficult than you expected?

In app stores apps can be rated with a five star system. How many stars would you give this app?

Why would you give this rating?

Where there (other) things you liked about this app?

Where there (other) things you did not like about the app?

7.3.2 Task evaluation

Ask respondent to show you the expenses registered and any notes they made on the template provided.

Try to assess / reconstruct how these tasks were performed (e.g. both assess how easy it was to respond to do the task and try to find out if the task was performed correctly). Code results as in table 2

7.3.3 Evaluation of communication materials

If not yet discussed: did the respondent use any of the materials provided? Why? If materials used: what did the respondent think about it?

Discuss where respondents would expect the information about the response task to be – it could be in the recruitment materials, within the app as a series of screens they go through the first time they log in, or within the app as an FAQ section, or on a project related web site – what would respondent expectations be about how they can access this information?

7.4 Scripted tasks (10 minutes)

7.4.1 Scripted use observation & evaluation

Check in table which of the tasks the respondent has not performed and select as many as them as time of this part of the test allows. It will probably not be possible to test all tasks with all respondents. We will try to make sure that at least all tasks are tested with some respondents. Also try to give tasks in a way that each respondent at least:

- Enters a receipt manually. (“So far you scanned all receipts. What if you had an expense without a receipt?”)
- Scans a receipt (“So far you have manually entered all expenses. Did you see there is another way to enter them? Could you please try that?”)
- Sees and discusses Expenses overview (What would you expect here? Please have a look. What do you think?)
- Sees and discusses Insights (What would you expect here? Please have a look. What do you think?)
- Sees and discusses Settings (What would you expect here? Please have a look. What do you think?)

After each scripted task, the task can be briefly evaluated with the respondent.

7.5 If time allows: Assessing quality of scanned receipts (5 minutes)

Randomly select three scanned receipts and assess the quality according to the checklist.

7.6 If time allows: HBS other issues (5 minutes)

Discuss with the following priority

- a) Would they like to the app to offer geodata information to help them remember their day and places for possible?
- b) How would they feel about the option to give permission to
 - o use scanner data based on time/location
 - o use bank transaction data
 - o use data from some loyalty cards
- c) alternative ways for HBS data collection (e.g. paper questionnaire, web questionnaire, collecting receipts), which data collections method(s) would the respondent prefer, why?

- d) Here also discuss the use of an app / smartphone for data collection: would only an app be OK for respondents, or would they also like a tablet or PC version? Why?
- e) Would they like other household members to be able to enter own expenses? How would they feel about app in which all expenses entered by all household members can be visible (but can also be made private).

7.7 Rounding off (5 minutes)

If not discussed yet spontaneously or during recruitment: to what extent and how does respondent usually keeps track of expenses? (precisely, somewhat, not; methods used)

We are now approaching the end of this test. Is there anything else you would like to share with us about this app or how we plan to approach respondents?

Thank for help.

Ask if we may recontact respondent for additional questions if needed.

Give giftcard, if needed make respondent sign for receiving give card.

If applicable: remind respondent to switch off "do not disturb".

Rate the smart phone skills of the respondent.

8. National reporting template

Each country will summarize their testing procedures and findings in a national report. The national reports will be synthesized in one overall test report. For a draft reporting template see attachment 7. It will be further developed in the course of testing, as needs for reporting may change with the test findings.

9. List of digital attachments

Click on the attachment button (paper clip icon) on the upper left to display attachments.

9.1 Attachment 1 Recruitment questionnaire template

9.2 Attachment 2 Content of Excel document for recruitment

9.3 Attachment 3 Manual for installation and use of app as used in NL

This is rough translation of Dutch manual .

9.4 Attachment 4 Form for comments of respondents retrospective interview group

To be used to write down experiences of installing and using the app during the days respondents independently test the app at home prior to interview.

9.5 Attachment 5 Example of consent form

9.6 Attachment 6 Checklist for scanned receipts

9.7 Attachment 7 National reporting template

The reporting template is partly based on reporting templates as developed for previous European testing projects by Statistics Denmark and Statistics Finland.