









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

THE SUMMARY OF THE 2st ROUND OF TESTING THE @HBS

AUTHORS

Summary report edited by: Brigita Vrabič Kek

Country reports:

Finland: Anna Niemelä

Netherlands: Deirdre Giesen, Stefan Theunissen, Mart Nyholt

Slovenia: Mateja Zgonec

project number

847257 — 2018-NL-INNOV

27 February 2020

summary

This report is a summary of the second usability test round of the @HBS and includes results of the testing and recommendations for improvements and proposals for the next version of the app.

Contents

. Introduction		3
Sun	nmary of main findings	4
2.1	Communication and motivation strategy	4
2.2	App as tested	4
Des	scription of testing	6
3.1	Testing period	6
3.2	Staff involved	6
3.3	Recruitment method and results	6
3.4	Summary of main characteristics of test respondents	6
3.5	Description of test	7
3.6	Materials evaluated with respondents	7
3.7	App tasks tested	7
Det	ailed results	9
4.1	First response after opening app (Full Observation Group only)	9
4.2	General evaluation after using the app	9
4.3	Response behaviour during week (Retrospective Group only, NL and SI)	10
4.4	Functionalities respondents miss in the app	11
4.5	Main findings per task	12
4.6	List of technical issues (bugs) found in the test	20
4.7	Communication, motivation and overall data collection strategy (NL nad SI)	22
4.8	Consent and data linkage	23
Rec	commendations for future development	25
5.1	Most important/critical ones (from the viewpoint of technical issues and/or dat	a quality)25
5.2	Recommendations to improve user experience (not critical for production)	28
	Sun 2.1 2.2 Des 3.1 3.2 3.3 3.4 3.5 3.6 3.7 Det 4.1 4.2 4.3 4.4 4.5 4.6 4.7 4.8 Rec 5.1	Summary of main findings 2.1 Communication and motivation strategy 2.2 App as tested Description of testing 3.1 Testing period 3.2 Staff involved 3.3 Recruitment method and results 3.4 Summary of main characteristics of test respondents 3.5 Description of test 3.6 Materials evaluated with respondents 3.7 App tasks tested Detailed results 4.1 First response after opening app (Full Observation Group only) 4.2 General evaluation after using the app 4.3 Response behaviour during week (Retrospective Group only, NL and SI) 4.4 Functionalities respondents miss in the app 4.5 Main findings per task 4.6 List of technical issues (bugs) found in the test 4.7 Communication, motivation and overall data collection strategy (NL nad SI) 4.8 Consent and data linkage Recommendations for future development 5.1 Most important/critical ones (from the viewpoint of technical issues and/or data)

Attachments

- 1 Protocol @HBS test round 2
- 2 Translation data linkage information with screenshots test NL

1. Introduction

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, special characters, search list). The @HBS project is a co-operation of ONS, Stat Austria, Stat Finland, Stat Netherlands, Stat Slovenia and the University of Essex. 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 and feedback). The testing was conducted in Finland, Slovenia and the Netherlands. The University of Essex and ONS reviewed the test plan and the test results.

The Household Budget Survey (HBS) app is available in both in the Google Play Store and iOS App Store (app store). Upon opening, the app users choose the country. Available versions are UK, Suomi, The Netherlands and Slovenia.

This report is a summary of the second usability test round and includes results of the testing and recommendations for improvements and proposals for the next versions of the app.

The test interviews were conducted in the period from 13th to 30th January 2020 according to common guidelines and templates for all relevant testing materials (detailed explanations are available in the document: Protocol @HBS test round 2). The staff involved in the interviews were mainly methodologists, accompanied by survey manager, UX/UI designer, intern or some other colleagues.

The overall goal of the test was to evaluate both the recruitment & motivation strategies/materials and the app and partly improved functionalities of the app. The main goal of test round two was to evaluate the changes in the app and the new brochure. Additionally – if interview time allowed - feedback was collected regarding the explanation of data linkage. This linkage is not yet implemented in the app, but a draft description of how this might work was available during the tests.

Regarding the app, the goal of the test was to explore:

- How easy or difficult is it for respondents to find the app, install it and login?
- How well does the app together with the materials provided (letter, brochure, detailed instruction document) 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?

2. Summary of main findings

2.1 Communication and motivation strategy

In general, the communication and motivation strategy as tested seems to work well. In this round of testing, we ensured respondents explanation on how to use the app within the app, which is an important and appreciated contribution. It seems that a combination of a short letter in paper form and dynamic walk through tutorial could meet respondents' expectations. Even though there were some technical problems with the tutorial, this turned out to be a promising addition.

Some improvements should be made. Most importantly, it should be made clearer that the main purpose of the app is data collection for statistical purposes. In addition, it should be communicated better that expenses of other households members should also be included.

2.2 App as tested

Overall impressions of the app were positive and the app was considered clear, simple and visually nice. It works well and is perceived as user friendly by the test respondents, even for those who do not use apps.

Part of the test respondents installed the app at home and used it for a week. All could use the app independently. For almost all using the app to document their expenses only took a few minutes a day.

The current app is clearly improved over the previous version, especially the flow for manual product entry. Still, the tests have yielded many ideas for further improvements. Some of these should be fixed before the app can be put in real data collection.

The app is very intuitive especially for younger respondents. Everything looks very easy. They use the app in a way "learning by doing", not by reading the instructions.

Some clear improvements since the first test round:

- The new manual entering screen is more intuitive e.g. since "Add" and "Complete" are on separated screens.
- The editing function has improved and the risk of deleting single items by mistake is clearly smaller, as well as the risk of closing a partly entered expense without saving.
- The automatic tutorial has brought added value but needs to be refined.
- The main usability problems of editing the price noticed in the first test round have been solved.
- Finding a shop by name seems to go smoother than using restricted shop categories as in the first test round (FI).
- The improved product list has made the responding task easier (FI).

Identified issues:

- Manual registration of receipts requires additional effort. Respondents usually do not want to take the complicated "path" to get to the goal: completing the diary as soon as possible with as little effort as possible. The function of taking pictures of receipts is easy

- and what is important to the respondents not time consuming. The respondents often preferred this option to enter expenses.
- One important issue for discussion is the use of search lists for the manual entry of products. This data entry process seems sometimes to discourage an accurate entry of the product description and category. The searching list for products is vague and sometimes misleading.
- The categories provided after choosing "Not found?" are incoherent and need further development.
- As noticed in the test round 1, the compulsion to choose only one product category to describe scanned receipt causes frustration and uncertainty.
- Some respondents expected that by clicking on the arrow in the search field for shop/product, would be displayed the full list of shops/products. But they returned to "Add receipt" screen (see section 4.5: "Manual registration of a receipt").
- There are some technical issues with the presentation of amounts entered via receipts in the overview screen. In Slovenia it seemed iOS respondents saw only manually entered expenses in the "Overview screen". In the NL, it seemed that there were some inconsistent problems with the first receipt entered in both Android and iOS, but that most amounts entered via receipts were correctly included in the overview screen.
- Discount entry on a product/service remained misunderstood.
- The automatic tutorial with many pages, laborious navigation and confusing highlighted buttons causes annoyance and disturbs the fluent responding process. "i" is not noticeable and it is not placed in the same place on every screen. "Skip" is not noticeable and it is placed on the left side of the screen (it should be on the right side of the screen).
- Completing a day with (out) expenses is not intuitive (a small number of respondents knew how to complete the day and they did not even know they had to do it.).
- Several respondents failed to notice the function quantity (consequently, it was not used).
- When deleting the entire receipt (by mistake) the warning message disappears too quickly (it is not noticeable).
- Notifications: Unusual time picker for Android respondents. For some respondents on Android setting the time was difficult.
- Thirty-day reporting period on "Calendar screen" should be country specific element (depends on the duration of completion of the diary).
- Receipts that are untypical like A4 sized or electronic cause uncertainty.

3. Description of testing

3.1 Testing period

The test interviews were carried out from 13th to 30th January 2020.

3.2 Staff involved

The staff involved in the interviews were mostly a methodologist, a UX/UI designer, and an intern, depending on the organizational aspects of such a work within each participating NSI. While observing, the observant already began to write down important actions, quotes, or found bugs/issues from the interview. In addition, the observers could ask additional questions at the end of the interview.

3.3 Recruitment method and results

Test respondents were selected based on different socio-demographic characteristics and other important characteristic for this testing. We divided them into two groups - namely, those who independently installed and used the app for one week or more at home (retrospective interview group), and those who installed the app and did tasks with the app while being observed at NSI (full observation group). Test respondents from both groups were interviewed in person.

In total eight individual interviews were carried out per participating NSI. Respondents received an incentive (gift card or a practical gift).

3.4 Summary of main characteristics of test respondents

In all three counties, there were 24 respondents altogether. Among them, there were more women (13) than men (11).

۸۵۵	distribution	of roc	nandanta
Age	uistributioi	1 01 163	ponuents.

Age class	Number of respondents		
1.80 0.000	FI	NL	SLO
18-34	3	3	4
35-54	2	3	3
55 or older	3	2	1

On general, they were well educated mostly with tertiary education. They were mostly in employment, but respondents were also students, housewives/men and retired persons. Most of them were shoppers or at least they shop with someone else, but there were also some non-shoppers.

For comprehensive testing, we selected respondents with different types of phones and mobile operating systems to participate. Among them 10 had iOS and the others Android operating systems.

One of the main characteristics of test respondents was also evaluation of their smart phone skills. This was done during the interview with questions about respondents' ability to use

or perform particular skills. We rated them with low, average, high. The majority of them got the rate "high".

Evaluation of smart phone skills of the respondents:

Rate	Number of respondents
High	18
Average	5
Low	1

3.5 Description of test

Almost all (except one) interviews took place at NSI premises. Interviews were conducted using the internationally coordinated topic list (see Attachment 1: Protocol @HBS test round 2). At the beginning of the interview, respondents were briefly told about how a test interview will proceed, the purpose of the interview and provided other relevant information related to the interviews. The interviewees were informed to think aloud while using the app. The tasks and evaluation questions were included in the interview protocol and asked during the test. Also, the usual probing techniques were used whenever it was needed.

All interviews were recorded by using "Mr Tappy" camera device and screen. The interviews took 40 to 90 minutes. All recordings were made with the written permission of the respondents.

3.6 Materials evaluated with respondents

In the test we evaluated the following materials:

- The national version of the HBS app
- Advance letter with leaflet

In addition:

NL:

- Text explaining data linkage with screenshots of a prototype design for data linkage of a bank account (all respondents) – see Attachment 2: Translation data linkage information with screenshots test NL for an English version App tasks tested
- Clickable prototype of data linkage with bank accounts

FI: — Description of possible data linkage solutions in future development

3.7 App tasks tested

- Installing the app
- Logging in
- Look at a walk through tutorial
- Starting entry of an expense via a date in the calendar
- Starting entry of an expense via the + button
- Manual registration of a receipt

- Taking a picture of a receipt
- Manual registration of a receipt with 2 or more identical items
- Manual registration of discount on a product
- Manual registration of a discount on a total receipt
- Change price of a manually registered expense
- Delete one item in a list of more than one manually registered items
- Complete a day with expenses
- Completing a day without expenses
- Use of Calendar screen
- Use of Spending screen
- Use of Overview screen
- Use of Settings screen
- NL: tried prototype for linking bank data

4. Detailed results

4.1 First response after opening app (Full Observation Group only)

After logging in the walkthrough tutorial opened automatically. The tutorial caused some confusion. Many respondents had difficulties with navigation by swiping, swiping through the screens is not so easy — as not all parts of the tutorial can be swiped. Several respondents also interpreted the highlighted +-button in the instructions as the actual entry button and repeatedly tried to tap the button in order to start registering the expenses. After closing the instructions, almost all respondents managed to find the entry functions quickly and got started without larger difficulties. Many respondents did not notice "SKIP" or "i" button.

One respondent who was quite a beginner with applications needed some guidance. Unlike the test round 1, this time only one respondent started the registration via the calendar. This change might be related to the automatic instructions as the respondent is not immediately able to examine the first page independently.

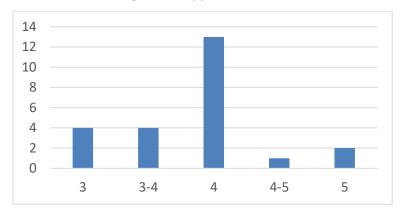
One respondent, who had adjusted her phone to have a large size letter, really got stuck in the tutorial. Because of the adjusted letter size crucial information on how to use and navigate the tutorial was not displayed.

4.2 General evaluation after using the app

Overall impressions were positive. The app was considered clear, logical and simple, even easier to use than expected. A couple of respondents who were not very familiar with apps learned quickly to use the app. The layout and colors were found nice. The walkthrough tutorial with many pages and laborious navigation caused some annoyance. The automatic instructions were found disturbing, swiping did not function smoothly and navigation through numerous pages was considered laborious. The highlighted fields/buttons in instructions made another usability problem as the users often understood them as functional parts of the app and tried to tap them. Some critical feedback was also directed to product list and product categories which were in some cases regarded as insufficient.

The app made a great impression and grabbed the attention of test respondents. One respondent was positively surprised that he had an option taking a picture of receipt. Some respondents had expected more functionality of taking a picture, i.e. automatic capture of at least some parts of the information of the receipt. In general, respondents were pleasantly surprised at the ease of the use of the app.

Distribution of ratings of the app:

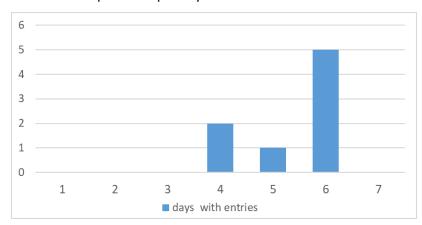


Additional question in NL - How would describe the purpose of this app? gave interesting results: most respondents perceived the app as being dedicated to monitor their own expenses.

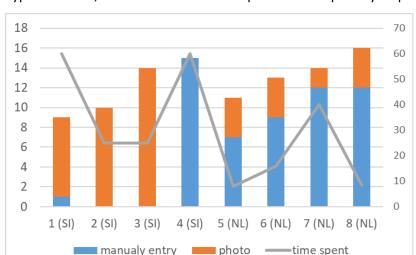
4.3 Response behaviour during week (Retrospective Group only, NL and SI)

This task was done only in NL and SI, with four respondents in each country in one-week observation period. No one from retrospective group needed help during independent using the app. The observation period was one week; the respondents made entries on 4 to 6 days, depending on their shopping activities, majority of then on 6 days.

Number of respondents per days with entries:



The majority of respondents used both ways of entering the receipts, 2 of them only took pictures and one respondent chose only manual entry. The time spent for these activities was very diverse, depending on the number of items pre receipt and the number of receipts. Several respondents said that they preferred to make the entries immediately after a purchase. For this, it was annoying for one respondent that the camera function in the app does not support flash. Because of this, she could not always make a picture of a receipt.



Type of entries, number of entered receipts and time spent by respondent:

4.4 Functionalities respondents miss in the app

Respondents mentioned the following functionalities:

Entry of expenses:

- Instant feedback on the quality of a photo; alert if the photo does not meet quality requirements or confirmation if it does.
- Option to zoom in pictures of receipts to check content / readability
- Automatic processing of parts of the receipts, such as the total amount. This would save time.
- An option to choose several days in the date entry in order to entry repeated identical expenses (like a daily lunch or coffee) in one go.
- An option to scan only the barcodes of receipts
- An option to add categories (not just products), or to modify the categories of insight screen.
- An option to use categories in manual registration
- An option to search products by brand
- Option to upload a digital receipt/bill, preferably in a way that you can easily remove personal information in the bill.
- Option to enter amount lower than one euro without the 0 (so ,89 instead of 0,89).
- Option to change the currency for if you buy things abroad.
- Option copying total price from scanned receipt to field "Total price".
- Option to monitor/enter income
- "Ok" button to confirm the selection of filter.
- Option to navigate with the Android back button

Screens in the app:

- Detailed "Overview" from scanned receipts. An option to enter more than one product category to a scanned receipt.
- Option creating/adding their own categories in the "Overview" screen.
- Infographics in "Overview "screen (for example fork/knife because they are easier to read and understand).
- An additional Start screen / home screen to give more overview over the app
- Provide an addition of the amount spent each week/month in the spending screen.

- Option in the overview screen to see which entries are related to the amounts shown.
- Insights: An option to compare one's own consumption e.g. with average consumption or certain risk limits (alcohol, tobacco, fat).

Additional functionalities:

- Virtual assistant to chat with to get help with use of the app
- Add some motivational messages in the app, comparable to for example My Fitness pall, for example: "Thank you for entering the expenses for your first day in the survey". "Great, you have already completed expenses for 7 days now".
- Option using the app after reporting period (for example respondent would like to see how much does he spend on a fuel per month/year for his own budget planning).
- "Complete a day" button on the "Spending" screen.
- A question type "reminder" for completing a day with expenses.
- For completing a day without expenses: It should be written in the middle on the screen something in a sense:" I have not had expenses this day". Checkbox (which must be seen all the time that respondent must confirm) should be put in front of this text.
- "Ok" button to confirm the reminder clock (when you set the time).
- Tutorial should contain infographics.
- FAQs.
- Introduction movie.
- Elements of gamification in the app (the app should be a bit more like a game, not an obligation).
- A searchable help
- Option to swipe to next and previous day in the day overview that is shown when you select a day in the calendar.
- Option to navigate with back button of Android phone.
- More information on the purpose of the data collection and data privacy; a link to the data privacy statement.
- Optional additional security for opening the app (e.g. using face recognition)

4.5 Main findings per task

Task description	Summary test results
Installing the app	 Finding and installing the app goes smoothly for most respondents. Issues found: Why is the name of the app in English? The logo of the app in the letter helps to recognize the correct app in the app store. QR code from the letter: in Slovenian test it was confusing for the respondents who did not realize they had to install the app first. They immediately used the QR and consequently found out it contains only username and password but not the whole process for installing the app (even though it was all explained in the letter). This problem did not occur in other countries. It is hard to say if this is a coincidence or related to how the letters were structured. However, there will always be respondents who do not read well and immediately start scanning a QR code. It would

Task description	Summary test results
	therefore be ideal if the same QR code could be used for installing the app and logging in.
Logging in	Safety issues: one would expect to be able to protect access to the app with for example face ID, if access to the app is not secured this would be a reason not to give access to banking data
Look at a walk through tutorial	Several users brought out that the instructions as such are useful but they should be short, and not compulsory to swipe through or skip in every screen: Generally, for most users the automatic tutorial in every screen was disturbing or annoying. Swiping does not work well - response is very slow. All respondents had problems with swiping. They did not know how to swipe - move to another screen trough tutorial (contradiction between arrows & finger pointing at right while the user was supposed to swipe left to proceed). Sometimes this caused frustration and even angry reactions. Some respondents were convinced that swiping was not working. Highlighted fields and buttons: Several users tried repeatedly to tap them since they interpreted them as functional parts of the app. Several respondents did not notice "i" on each screen and it is not in the same place on every screen (confusing). "Skip" is not immediately noticeable and it is placed on wrong side — it should be on the right side of the screen and it should be replaced with an icon "x". Tutorial does not show correctly on a phone for which the respondent has set larger letter size. Possible solutions: Respondents would prefer starting the use by themselves and then looking for instructions if needed Automatic tutorial is as a good solution but only in the first screen of the app. FAQs and short introduction movie would be helpful. Inclusion of infographics in the tutorial to make it more practical and attractive to see.
Starting entry of an expense via a date in the calendar	No problems, works fine.
Starting entry of an expense via the + button	No problems, works fine.
Manual registration of a receipt	 Understanding response task: At the beginning of entering the words of products it is not immediately understandable which level of detail is required (i.e. item level)

Task description **Summary test results** There are no clear instructions yet on how to deal with student house (whether to treat them as a household or not). Expenses when respondent did not have a receipt. Precaution and prevention of misunderstanding: if you do not receive receipt, you simply do not enter these expenses into the app. Respondents were wondering why the difference in requested information between the two entry options (when taking a picture of receipts, there is no specific product information). user hesitated to add the product name herself because she was unsure how the added product impacts on the app and wondered if it causes "disorder to the system". Response burden overall: Many respondents said manually entry is time consuming and difficult. One respondent said that he would enter manually only curvy short receipts. Far too much, work to provide details at such a detailed level, not understanding why this information is necessary, because there is no personal interest to keep track of expenses at such a detailed level. Additional burden to enter items that are not found in the list. Entering process: Entering shop, date, adding items they can find in the search lists and price works intuitively for most respondents. Respondents find it positive that so many items /shops can be found in the search list. However, the automatic instructions sometimes caused misunderstandings. Several users swiped forward, did not recognize they were still in the tutorial and tried to tap the highlighted "Enter shop" field in the instructions to start entering. Some dissatisfaction was directed to the product list that some users found insufficient. If an item they type in is not found, most respondents seem to try to type in a different description. For example if typing "light cracker" does not yield a good result, then "cracker". When entering shop/products on the "Add receipt" screen the entire list of shops/products is not visible immediately. They did not know that they have to start typing in the search field; by clicking the arrow in the search field (← "Find shop"/"Find product/service") they could not come on whole list of shops/product as they expected - they returned to "Add receipt" screen. The search result list is not always in a proper alphabetical order, which in some cases slowed finding the right product. Besides, the

Task description	Summary test results	
	 algorithm does not recognize all relevant products comprehensively. Respondents find it strange that list with suggested items contains typing /spelling errors, this looks unprofessional. Missed the item rounding "afronding" to enter extra 2 cents for rounding. This illustrates a dominant strategy when entering items manually from a receipt, respondents want to enter literally, what they see on the receipt. Respondents want to make sure that the totals of the receipt they enter manually match the total of what is entered in the app. This seems like a micro award for the task. Receipts only in a mobile app of a store chain: how to easily register long electronic receipts. 	
	 Display of category of item: The category can confuse or disturb respondents, the result is wrong classification of the product (coffee in a restaurant as coffee/coffee (first hit in search list) instead of coffee/catering industry, socks are entered as women socks instead of men's socks. 	
	Adding a product/shop A few respondents do not find this option by themselves, most do.	
	 Lists of recent / frequent items Some respondents seem to like this, others do not use it. Results from entering receipts should not be shown as recent/frequent item) 	
	Online/abroad Works OK.	
	Warning message: If a respondent leaves the manual entry field without completing a warning is shown "You are about to close this transaction, if you continue this transaction will be deleted". One respondent was a bit confused by this message and was afraid she would delete all data previously entered. Also, the word 'transaction' may be not so clear as it is not used elsewhere in the app.	
Taking a picture of a receipt	 This function is quite easy to use. Some respondents were a bit disappointed with the lack of functionality of taking a picture; they had hoped that the app would at least read automatically the total of the receipt. Respondent thought it is not possible to make a correct picture of a very long receipt, therefore decides this must be entered manually. 	

Task description	Summary test results
	 Tips for taking a picture /quality picture Tips for taking a picture seem to help a bit (e.g. not choosing a white background). However, several respondents took pictures of wrinkled receipts, ignoring instructions to make sure the receipt is as flat as possible. Interestingly, even when respondents still remembered the instructions like keeping the receipt flat they did not think about following them when taking pictures.
	 Entering details for picture goes smoothly with the following remarks: Calendar starts with Sunday, Monday would be more logical and also consistent with Calendar screen) Several respondent find it difficult to choose a category, categories are not so clear (where to put curling iron, pest control)
	- Difficult to find a category when receipt has multiple types of items - the "Overview" screen could not show a "true picture" of the expenditure.
	 There is no check on choosing a category; picture can be added with choosing category iOS respondent noticed that the expenses were not shown in the "Overview" screen when taking the picture, then she tried to enter receipts manually. Manually entered expenses were displayed on the "Overview" screen. If respondents only paid part of the receipt, they cannot enter that.
	 UX taking picture: Sometimes picture is shown in landscape, most of the time portrait. Not clear why, this disturbs respondents. Also seem to make receipts even less readable for checking. Logo shown when taking a picture is Android, looks strange on iPhone Not possible to zoom in on pictures taken. Respondent missed the flash option, wanted to take a picture in the car, directly after shopping in the evening. Multiple pictures of a long receipt: this is not possible and so takes the picture from a further distance. After several entries (about 3 pictures and 1 manual entry) app crashes when taking a picture. After restart no problems.
Manual registration of a receipt with 2 or more identical items	Most respondents easily find option to enter multiple identical items.
Manual registration of discount on a product	Very problematic; difference between "discount on total receipt" and "discount on product" not clear for most respondents, despite the fact that the letter contained brief and simple instructions on how to enter both types of discount (SI version). Unfortunately, the

Task description	Summary test results
	respondents did not read it. Only one respondent has successfully completed this task, because she read the letter very carefully.
	Several respondents were surprised when the task was given (they did not know how/where to enter discount; they were guessing and trying out they needed to do).
	They tried to enter discount on a product by entering discount on the total of the receipt and/or by switching on "discount/return/etc.". When they switched on "discount/return/etc.", they were surprised because they saw only the negative amount.
Manual registration of a discount on a total receipt	 Logic of entering discount not clear: Respondent warned us that in some cases discount on a total receipt is not discount on everything item in the receipt because there is no discount on certain products. Respondent suggested that we should rename a "Discount" field to "Discount on total receipt". Cannot delete discount entered.
Change price of a manually registered expense	 Works fine, but after changing an item, the order is changed. This is confusing as when people check their entry they want to keep the same order of the receipt. Some respondents try to edit an item by double tapping the item in the expense overview.
Delete one item in a list of more than one manually registered items	The function is easily noticed and easy to use. Compared with test round 1 the editing function has improved and the risk of deleting single items by mistake is clearly smaller.
items	 Deleting complete entry: If scrolling is needed to see the edit/delete option this is not found by some respondents. Respondent want an extra check question if they really want to delete something. One respondent deleted the entire receipt by mistake. The warning message disappeared too quickly (it was not noticeable).
Complete a day with expenses	 No new findings compared with test round 1: The remaining usability issue is that there is no sign or hint where to find this function. Besides, the purpose of completing a day was unclear to respondents. If completing a day is a crucial task to do in the HBS data collection, it should be communicated clearly. Some respondents wondered what is the reason that days should be completed and what would be the consequences except for changing the color of a day in the calendar. Some respondents were not sure if they could still enter/change data if a day has been completed.

Task description	Summary test results
Completing a day without expenses	Possible solutions: - Function "Complete a day" should be on the "Spending" screen (for example, it could be a button "Complete" at the end of "Spending" screen). - One respondent found the "Filter expenses" function, by mistake, on the "Spending" screen. There he also searched for function "complete a day". - One respondent proposed to put a question type "reminder" in the sense: "Have you received any other receipt today?" or "Would you like to complete your day?" Respondents from retrospective group did not have any problems; full observation group respondents do not see the option to complete a day.
	Possible solutions: - Put in the middle on the screen something in a sense:" I have not had expenses this day". Checkbox (which must be seen all the time - that respondent must confirm) should be put in front of this text.
Use of calendar screen (other findings than listed above)	Clear, simple. Issues found: A few respondents wondered why the calendar contained a thirty-day reporting period. Respondent does not see this screen as a starting / home screen. Respondent does not understand how the award is calculated. Respondent mentions that he does not recognize the reporting week in the app. This may have to do with the fact that the app marks a 30 day period (with a light shaded blue), whereas in the leaflet a one week reporting period is mentioned. Layout of top not consistent with other screens ("i" is here in upper right corner instead of left and not title of screen). The users made some development proposals: Add year to Calendar screen Add more information on purpose of the app One user noticed that in the calendar, it is possible to scroll down to the next month but moving back is only possible by using
Use of Spending screen (other findings than listed above)	arrows. - A year (next to a month) should also be written on the calendar. The page was found clear, logical. It demands a lot of scrolling if there are many expenses. Issues found: - Respondent would like to see totals per week/month here. - Several respondents mention that they expect the oldest expenses to be on the top (now the ordered is the other way around).

Task description	Summary test results
	 One respondent filtered the expenses on this screen. He wanted to confirm his decision of filter (would expect button to confirm filter (for example "OK"). Country specific element: there were no special letters displayed on this screen (e.g. "č", "š", "t", etc.)
Use of Overview screen	 Screen was generally evaluated as nice, informative screen, very interesting, useful and motivating. Issues found: Full titles of categories are not visible. Many categories are not logical, for example pest control was registered as "Furnishing, household, equipment" and writing material is listed under the main header "recreation and culture". These non-intuitive categories make the overview less useful for respondents. Some respondents want to be able to link the amount spent in a category in the overview to entries they have made. This is not always clear (also has to do with the fact they find the names of the categories not always logical).
	 If overview contains data of several household members, it would be useful to be able to distinguish spending per person/household members. Some technical problems with display of this screen. The chart is shown only for a few seconds and then disappears. Also seems to be missing some entries. Later – after restarting of app - it looks OK. Respondents expect/want to discounts taken in to account in the overview. The category prostitution / narcotics surprises respondents, even causes embarrassment.
	 Development proposals: An option to compare one's own consumption e.g. with average consumption. Infographics in this screen (for example fork/knife) is more readable and understandable.
Use of Settings screen	 Generally is fine. Issues found: A few respondents were wondering why is option to change the language. One Android respondent said that the time picker is unusual (he was surprised). He didn't know immediately how to set the time. One would like to change currency here, would be practical if he would be on holiday.

Task description	Summary test results
Registration of identical expenses on several days (FI only)	The duplication function is easy to find and use. Those who did not use it, registered all under one expense by adding the number of products or using mental calculation.
	Most users would still prefer choosing multiple dates from the calendar when entering expenses like daily lunch/coffee. If the duplication is made through the Expenses screen, the duplicated new expense registered automatically to the ongoing day, but if the duplication is made through the calendar screen, the duplicated expense is registered to the same date as the original expense. This confused respondents, as they did not always find the new expense. Besides, in some cases, the latest expense with scanned receipt in the Expenses screen opens after the duplication.

4.6 List of technical issues (bugs) found in the test

Issue nr	Description
1	 Mobile phone issues: Installing the app: did not open in Samsung Galaxy J4+ Telephone with large letters walk through screen not displayed correctly. Possible solution: do not support dynamic text in the app? App crashes when after taking picture (only once, before and after no problems) Android respondents are used to use the back-button of their smartphone. They could not navigate through the app with this button.
2	 Manual registration: An odd html-alert was displayed when entering the price starting with a comma. The search result list is not in a proper alphabetical order. See Appendix, screenshot 1. Special letters - Not displaying special characters: Missing special letters (when searching shops/products that start with a special character; also letter "ž" in the middle in the word is removed). If we typed a product that starts (or have special letter in the middle of word) with the "ž", "č", "š" letters (e.g., Žakelj) and then select "Not Found", then the product name was displayed without that first letter.
3	 Tutorial: Not all area of walk through tutorial is swipe able (text field not swipe able?) Walk through tutorials screen Settings: part of text on Contact information not displayed.
4	 Calendar screen: Switching between months in calendar: Switching back from February to January via arrow did not work for some respondents in the beginning, but now seems fixed The name of the month is only partly visible, partly covered by the StatFin logo. See Appendix, screenshot 2.

Issue	
nr	Description
5	Duplicating an expense:
	- If the duplication is made through the Expenses screen, the duplicated new expense registered automatically to the ongoing day, but if the duplication is
	made through the calendar screen, the duplicated expense is registered to the
	same date as the original expense. The logic should be similar in both cases.
	Besides, in some tests, the latest expense with scanned receipt in the Expenses
	screen opens after the duplication. This confused respondents as they did not
	always find the new entered expense.
6	Adding a shop or product via the "not found?" option:
	- Dots and capitals are not kept. For example: eBay.com is entered by respondent,
	but the app registers this as Ebaycom.
	- After choosing "Not found" - button, only category "Other" is displayed in the list.
7	Information on receipt:
	- If not all fields are completed (shop, price and/or category) it is not possible to
	select "done". If respondents forget a field, they do not see how to proceed. We
	would need to show an error message that not all fields are completed if someone
	tries to select "done" for an incomplete field. Maybe it is also good to show the
	missing fields with a red line instead of a black line. For manual entry there is
	already a message "Information incomplete".
	- Cannot delete discount on total receipt
	- The calendar language was English in the receipt scanning entry (Finnish in the manual entry)
8	Scanning of a receipt:
	- In some cases, the scanned receipt is displayed in a horizontal position. This was
	already noticed in the test round 1. See Appendix, screenshot 3.
	- All rows are not fully visible. See Appendix, screenshot 4.
	- It is possible to leave a product group empty, save an expense and move on even
	if 'Ready' button is grey. This was already noticed in the test round 1.
	- The orientation of one photo of a Jumbo receipt was changed while originally shot
	in the right orientation. The app itself turned a receipt (not in all cases).
9	Overview screen:
	- Not all expenses that were registered via a picture were displayed in the insight.
	Inconsistent findings in the tests on how often and how this occurred.
	- for 1 respondent the pie chart first is only shown for a second and then not visible, also seemed not correct. After (later) crash and restart it worked fine.
	e u con contra de la contra del la
	 Full titles of some categories not visible the last/first numbers in the graph are cut off - whole expenses should be visible.
10	Settings:
	- Language: If you change the language the product names and categories remain
	in the original language. This has the strange effect that the app than has two
	different languages.
11	Insight & Spending:
	- The time filter cannot be set if expenses are entered to the year 2020. If expenses
	are entered to 2019, the time filter functions.
	- Missing dot after date on "Spending" screen. For example: Sreda, 22. januar

Issue nr	Description
12	Expenses & entry screen:
	- In one respondent's phone (iPhone 8), the price was not fully visible in entry
	screen and the expenses screen. See Appendix, screenshot 5.

4.7 Communication, motivation and overall data collection strategy (NL nad SI)

4.7.1 Evaluation of materials / strategy discussed in test

NL

Letter, login info and leaflet

- Letter only tested by full observation respondents. They are positive about the letter. Clear, not too long.
- Gift card noticed and appreciated
- Several respondents mention that they like the promised incentive
- One respondent comments that he feels a bit obligated by the card (but does not seem to mind too much). The same respondent says that an additional 30 euro for participating in the survey is quite a lot of money. This makes him wonder if participation would take more time than he initially thought.
- Showing the logo of the app in the letter helps respondents to quickly select the correct app in the app store.
- When reading the letter one respondent wonders if he can only enter expenses if he still has the receipt, later in the app discovers that he can also enter manually.

Letter and leaflet raise expectation that app also monitors income

- Some respondent indicate that they would expect and like the functionality to also enter income. The leaflet uses the term "budget survey", this raises the expectation that the app can monitor both income and expenses.
- One respondent does not understand why research into spending does not measure income. This respondent refers to the content of the letter that implies that this type of data is collected. The letter starts with "What does life cost in the Netherlands? Can people make ends meet? What part of our income goes to food, to clothing, to housing? What is the impact of our financial situation on our health and well-being? Statistics Netherlands research these important questions once every five years. "

SI

We slightly adapted test protocol according to our own practice. For example, we explained current contact strategy at the beginning of the test. The current practice of HBS is to send an advance letter and the leaflet via post. The advance letter does not contain information that respondents will have to keep a paper diary. At the end of CAPI interview interviewers have to persuade respondents to keep paper diary.

- Letter

We prepared short "additional" letter that will be given to the respondents personally by the interviewer. The respondents from full observation group received "additional" letter at the beginning of the test. The respondents from retrospective group received "additional" letter before the test. The majority of the respondents looked briefly at the "additional" letter. Actually, they read "keywords" such as the name of the app, username and password. We noticed that the respondents did not look at the "additional" letter after logging in (unless they had a dilemma while entering expenses).

As mentioned in Chapter 3. 6 (task 1), the first version of the letter contained a QR code (QR code was for logging in, not for installing). After a few tests, we removed the QR code from the letter because it disrupted optimal flow. The problem was that the respondents had to manually find the app on the app store anyway, and then they had the option of manually entering their username and password or scanning the QR code. Respondents thought that scanning the QR code would ensure them full access to the app.

- A walk through tutorial

With a walk through tutorial, we are on the right track. Some improvements need to be made (see Chapter 4). For example, in the "Settings" screen, under "i" should be placed "whole" tutorial (option to see a whole tutorial at the same time). And of course, "i" should be easier to find.

4.7.2 App use and alternative methods for HBS data collection (Full Observation Group only)

NL

Not systematically discussed. But:

- Several respondents spontaneously say that an app is a convenient way of keeping track of expenses as you can do that on the go.
- One respondent mentions spontaneously he likes to use new technologies as one of the reasons for being willing to participate in HBS with app data collection.
- Another respondent with whom we have discussed the possibility of using interviewers says she would not like a visit from an interviewer.

SI

We were interested in which, according to the respondent's opinion and experiences with this app, would be the most appropriate mode to collect data for HBS survey or how they would like to participate in this survey.

4.8 Consent and data linkage

Note: not all parts of data linkage discussed with all respondents.

This topic was briefly discussed with six respondents in Slovenia, seven respondents in Finland and 4-8 respondents (depending on the data instance) in the Netherlands. Sharing geo-location was discussed as relevant feature.

4.8.1 Linking with banking data

As in the test round 1, allowing the application to import bank transaction data split the group in half. For some it appeared as one way to ease reporting expenses but for the others suggestion was intrusive and not acceptable. Interestingly, several respondents who would be willing to link data questioned the added value of doing so. Banking data are not always very informative and would still require editing to make sense. This may take more time than

just entering an expense directly. On a more general level, adding additional data sources to the app may make reporting expenses more complicated.

When discussing the use of banking data for reporting fixed expenses such as rent, insurances et cetera, respondents noted that some of these expenses are paid quarterly or annually. To detect these expenses consent for linking banking data should allow to look back up to a year (90 days is the default for PSD2 regulation – this may probably be changed by explicit consent though).

NL Prototype:

The flow of the prototype works well. It is clear what to do and what they see.

Small usability issue: respondent does not immediately see he can tap the words "no linkage", tries tapping the header "linkage" first (reconsider design if linkage button).

Remarks /questions from respondents:

- I want to check the data that has been retrieved from my bank account
- For which period / for how long would they retrieve data
- If I disconnect the data linkage, I would expect that the data retrieved from my bank account is removed from the app (R2)
- Why both include Paypall and checking account? Expenses from Paypall are also registered in your checking account. Oh.. maybe because there is more information in your Paypall account? "

4.8.2 Sharing geo location

The results ran parallel to test round 1. From most respondents who were asked how they feel about the use of geo location were skeptical about it. In the context of the HBS, they did not see the usefulness of this information and/or they were not themselves willing to permit geo location of their mobile phone as it violates one's privacy. Several respondents emphasized that they generally not want to share their location to any other applications either and expressed the concerns for battery life duration.

Maybe younger respondents are more willing to give permission to keep track of their location.

4.8.3 Linking with scanner data (NL and FI)

Compared with banking data linking, the scanner data option was evaluated more positively as it did not provoke similar privacy-related doubts. It was considered similar to loyalty card data and thus something that people use to share anyway with store chains. Generally, the scanner data option was not easy to understand and communicate. The process and benefits did not come clear based on the short information document.

4.8.4 Other possible data sources discussed with respondents

FI: The possibility to link loyalty card data was discussed shortly. This option was the most familiar to respondents and all except one assumed they could give consent to link the data to the app.

5. Recommendations for future development

Number of crucial issues will have to be addresses and carefully implemented before the @HBS app is integrated in the "real" HBS survey. That especially applies for all countries that will like to implement the app into their own infrastructure. The concerns are:

- How to proceed after the project (from testing results it is clear that app needs improvement)? When/how to get final version of the app with all promised features and country specific elements?
- Who will set up priorities?
- Can Stat Netherlands continue support of the app? If yes; to what extent? How/When?

5.1 Most important/critical ones (from the viewpoint of technical issues and/or data quality)

Development of very important parts of the app (which are not visible to respondents), fix bugs, simplification and modification of some functions of the app is necessary:

- Developed app back-end in order to evaluate how suitable the scanned receipts are for further processing like OCR, and to evaluate the requirements the app will set to the national data systems as well.
- Developed of the **receipt classification pipeline** (preprocessing, OCR, text processing and classification).
- The automatic **walkthrough tutorial** does not provide a good user experience as it is now. The instructions must be made lighter and less compulsory/automatic.
- The categories provided after choosing "Not found?" when entering the shop or product are incoherent and must be developed. The determination of this list should be more logical and stable, e.g. to permanently include the most common shop/product categories + "additional".
- Developed FAQs.
- Developed short introduction movie.
- Developed solution for a household data collection to form a comprehensive assessment of the app.

5.1.1 Communication

- Make it more explicit in the communication that the main purpose of the app is **data collection** for a survey. One way to do this is by adding a little video that starts automatically that stresses the purpose of the app: data collection for the statistical agency (even with a clear explanatory text the respondent don't seem to read or remember the response task good enough. clear illustrations and animations tend to stay better in people's memory than plane text.) Another way to stress that this app is a data collection tool may be to have a short questionnaire included in the process of starting the use of the app.
- Make it very explicit in the communication that expenses from **all household members** should be collected and what should be counted as a household members (e.g. student houses).

- Display both the name and the logo of the app in the information, to facilitate finding

5.1.2 Walkthrough tutorial

- Make text also swipe able.
- Change the swipe icon in one that explains **the right swipe direction**. Use an animation if possible.
- Make swipe response speed instant.
- There could be e.g. an automatic tutorial only on the first page and the rest of instructions would be accessible through i-buttons.
- "i" should be noticeable (bigger) and placed in the same place on every screen.
- Option to see "whole" walk through tutorial (not separated on each screen), at least on the "Settings" screen.
- **Shorter text** in tutorial.
- "X" instead of "Skip" and it should be placed be on the right side of the screen.
- Change the tutorial texts so that there is no risk of people thinking you can use the explained buttons while in the tutorial. Also consider to add a title "explanation" to the numbering each tutorial screen (e.g. explanation 1/5)?

5.1.3 Entering products

a) Reconsider **the use of the search lists** for entering products/categories and stores. It may be better for data quality to redesign this part of the app using the logic that is currently used in the Dutch online budget survey. In this flow respondents type in the product in an open field, based on this input either one category is offered (that can be changed by the respondent), the respondent has to choose between several categories, or the respondent has to make several hierarchical selections to choose a category – entering a new category is always possible.

The main issues with current search list design are:

- If a product /store that the respondent wants to enter is not offered in the search list, several respondents choose or try to come up with an alternative description that they can find in the list. For example: curling iron become curling set, post-its become note book, Christmas socks are changed in socks, riding school becomes riding school shop. This may affect the data quality and in the long run may cause the lists of products to become outdated. Some respondents do not see that they can enter a new product if they cannot find it in the search lists. Others are aware of the option, but seem to prefer to use the search list.
- Several respondents seem unaware of the categories displayed below the products, so they choose the first option of socks or coffee, without checking if these are men socks or coffee in catering.
- The search lists (as tested in the Netherlands) is that it contains spelling and typing errors, which looks very unprofessional.
- Of course, forcing respondents to fully type in product/shop names and explicitly selecting categories will increase the response burden. However, it may be worth to do this given the expected gain in data quality.
- b) The product search algorithm needs further development. The current search function has improved since test round 1 but some incoherence still appears. The search result

- list is not in a proper alphabetical order and the search does not recognize all relevant products comprehensively.
- c) Redesign the name of categories to be more meaningful to respondents. Both for the data entry process as for understanding the overview.
- d) Give automatic feedback if respondent tries to manually enter general descriptions such as "groceries" or "toiletries" ("please describe each item bought, so bread, milk, etc. instead of groceries).
- e) **Reconsider design of discount**. There seems no ideal option to facilitate the smooth manual entry of receipts that typically show the original prize of items and various types of discounts that may be related to one specific item, a subset of items or the total receipt. Maybe the following approach will work best (if also not perfect for everybody and every situation):
 - i. Instruct respondents that for items with discount they should report the price paid;
 - ii. Offer a calculation help button in the field for manually entering product price respondents can use this if they have to calculate the real price paid.
 - iii. Keep the option to enter negative amounts for example for refunds (but remove the word 'discount' here) respondents need to be able to enter their total receipt so they can check totals of what is entered and what they see on the receipt.
 - iv. Keep the option to enter discount on the total receipt, but only in euro's (not %). Entering % is not useful as in most cases the amount of discount is known.
 - v. If respondents enter discount on the total receipt, deduct a proportional discount on the items entered for the data shown in the overview field.
 - vi. In this design, the most problematic discount to deal with for respondents is discount over a subset of items on their receipt. Calculating the actual price per product may be difficult for such discounts. We would expect some respondents to be willing to make this effort (and maybe make some small mistakes in the process), others to take shortcuts that result in larger mistakes, e.g. putting all discount on one product or entering discount on a subset of products as a discount on the total receipt.
 - vii. We could rename a "Discount" field to "Discount on total receipt".
 - viii. Function "Discount on a product" should be removed.
- f) The arrow in the search field <-"Find shop"/"Find products" should be placed outside the search field.
- g) Function "Complete a day" should be on the "Spending" screen (for example, it could be a button "Complete" at the end of "Spending" screen).
- h) Options of categories and words prostitution and drugs (narcotics) on the "Overview" screen should be removed.

5.1.4 Warnings

Always **add a warning** when respondent tries to delete something. The warning message should not disappear quickly.

5.1.5 Entering details for picture:

- If a respondent tries to select "done " before all fields are completed, give a warning
 "information not complete" (like for manual entry is already implemented) (note to
 include check on choosing a category)
- Allow multiple categories
- The app itself should not turn a receipt.
- Scanned receipts should have a "zoom" function.

5.1.6 Country specific elements

- In the case of scanned the receipt the (sub)category of product on the "Overview" screen should be appropriately named (e.g. not "Receiving food and beverage"; not "Receiving transport").
- **Reporting period** should be adapted to national fieldwork requirements (for the Slovenian version of the app two weeks reporting period is needed).
- Inserted a dot after date on "Spending" screen. For example: Sreda, 22. januar
- **Special letters** "c","č","s", "š","z", "ž" should be working.
- In every case it should be displayed the full name of the product (for example, product that starts with the "ž", "č", "š" letters (e.g., Žakelj) and then select "Not Found", then the product name was displayed without that first letter).
- Option "change the language" on the "Settings" screen can be removed.

5.2 Recommendations to improve user experience (not critical for production)

Design of app in app store:

- Make name of statistical office directly recognizable in the app store.
- The name of the app when searching in the app store in national language.

Communication of data linkage

- If data linkage is offered in/with the app, carefully consider where and how to communicate this. Stressing sensitive options in the first communication may deter respondents who do not see / feel that allowing data linkage is optional.
- Invest in further research into the usefulness of linking to other data sources in the app. Some of the test results indicate that it is challenging to explain data linkage ant it may actually complicate the use of the app to use data from different sources.

Communication of response task

- Explain in communication that it is useful to keep/ask receipts for the research, but that it is not necessary.
- **Provide direct feedback on quality of picture.** Since receipt quality is a crucial factor in terms of data collection and if the aim is to bring the OCR data from the receipt back to the app and insights, a quality confirmation would bring reliability.
- Explain that "completing a day" is mainly intended as a memory support for the respondents and that one can still add, change and deleted items from completed days.
- Consider to integrate motivational messages in the app.
- Consider alternative solutions instead of forcing to choose only one product category when expenses are entered by receipt scanning and purchases belong to different

categories. Also, add an instruction for these cases. This might be solved when the real scanning and sending functions are developed into the app.

Language used in app and communication materials

Review use of language in the app for consistency. Make a table of all words used for the various screens / functionalities and check these are consistent in all screens/warning/help info etc. For example, check if and where "transaction" is used.

Layout calendar screen

Make location of "i" and use of title consistent with other screens.

Entering date of expense (both manual and picture entry):

- Week should start on Monday.

Lists of frequent/recent items in manual entry screen

- Results from entering receipts (in Dutch:"Bon voeding en drank") should not be shown as recent/frequent item.

Processing pictures of receipts

- Develop OCR for (parts of) receipts (this would probably both increase user experience and data quality)
- Move location of "other" to the last option and add open text field.
- Develop functionality to indicate if some part of receipt should be ignored (because for example paid for by others).
- Also, have an option to upload an online invoice or voucher. Especially because in the future more and more purchases will be made online. If possible, include in this option an easy way to make personal information on receipts invisible.

Completing a day

- If respondents want to complete a day without expenses make it an explicit choice to indicate that this day they had no expenses
- If a day is completed, indicate below "Day already completed" a text indicating that they can still enter, change and delete expenses.
- If a completed day is modified, deactivate "day completed" so that respondents have to explicitly complete a modified day

Spending screen

- Move buttons "delete/edit/duplicate" up so that they are always visible after opening an entry.
- Give warning "would you like to proceed" if an entry is deleted.
- Consider to make it possible to edit an item by directly tapping it in the spending screen lists.
- Add an option to choose several days in the date entry for registering repeated identical expenses (like a daily lunch or coffee) in one go. Add similar logic to the registration date of duplicated expenses regardless of the entry channel (Expenses/Calendar)

Overview screen

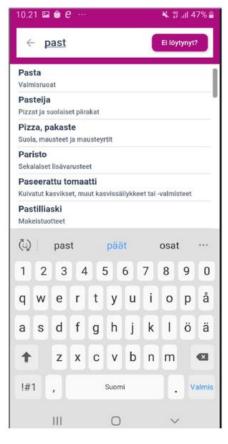
- Rename categories to make them more logical and useful to respondents. Current Dutch categories often are not intuitively linkable to expenses made.
- Deduct discounts from the amounts shown. This may be most problematic when processing OCR data from receipts.

Prototype linking banking data (only tested in NL)

- Reconsider design linkage button to make it more obvious that this part should be selected.

APPENDIX

Screenshot 1



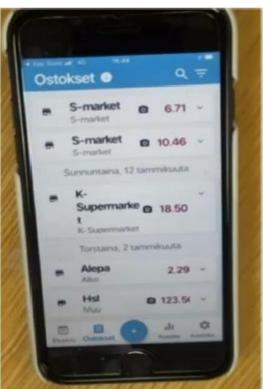
Screenshot 2



Screenshot 3



Screenshot 4



APPENDIX

Screenshot 5

