

D2.6: Evaluation results of the third functional prototype and updated requirements

Dissemination level: Public

Document type: Report

Version: 1.0.0

Date: November 29, 2019



This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement #769553. This result only reflects the author's view and the EU is not responsible for any use that may be made of the information it contains.

Document Details

Project Number	769553
Project title	Council of Coaches
Title of deliverable	Evaluation of third functional prototype and updated requirements
Due date of deliverable	August 31, 2019
Work package	WP2
Author(s)	Mattienne van der Kamp (RRD), Tessa Beinema (RRD), Marijke Broekhuis (RRD), Harm op den Akker (RRD), Dominic de Franco (UDun), Alison Pease (UDun), Sita Ramchandra Kotnis (DBT), Frederikke Ellemose Vad (DBT), Eskil Welan (DBT)
Reviewer(s)	Harm op den Akker (RRD), Jorien van Loon (CMC), Konstantina Kostopoulou (iSPRINT)
Approved by	Coordinator
Dissemination level	PU: Public
Document type	Report
Total number of pages	67

Partners

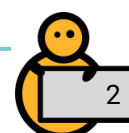
- University of Twente – Centre for Monitoring and Coaching (CMC)
- Roessingh Research and Development (RRD)
- Danish Board of Technology Foundation (DBT)
- Sorbonne University (SU)
- University of Dundee (UDun)
- Universitat Politècnica de València, Grupa SABIEN (UPV)
- Innovation Sprint (iSPRINT)

Abstract

This deliverable reports on the evaluation of the third functional prototype. It describes the setup of the third evaluation in respect to evaluation one and two, provides an overview of the third functional demonstrator, presents the detailed design and methodology of the third evaluation, reports the results of the evaluation and reflects on results by providing additional updated requirements.

Table of Contents

1	Introduction.....	7
2	Objectives	8
3	Overview of the third functional demonstrator.....	9
3.1	Recap: Results from the first and second evaluation	9
3.1.1	The first evaluation	9
3.1.2	The second evaluation.....	10
3.2	The third functional demonstrator	11
3.2.1	The Main Menu	11
3.2.2	Register account	11
3.2.3	Baseline questions.....	12
3.2.4	Coach selection.....	13
3.2.5	Living room.....	14
3.2.6	Introduction by Coda.....	15
3.2.7	Interaction and introduction to coaches	15
3.2.8	Coaching content	16
3.2.9	Interaction with widgets	17
4	Methods of evaluation of the third functional prototype.....	19
4.1.1	Participants	19
4.1.2	Ethics & Recruitment	20
4.1.3	Study procedure	20
4.1.4	The evaluation tasks.....	23
4.1.5	Preparation of tablet/iPad screen capture.....	26
4.1.6	Revoke Access to RRD activity coach on Fitbit.....	30
4.1.7	Interview guidebook.....	32
4.1.8	Questionnaire of quantitative assessment of usability and user acceptance.....	33
4.2	Data analysis.....	36
4.2.1	Usability issues	36
4.2.2	Task metrics.....	36
4.2.3	Questionnaire results.....	37
4.3	Methodology Denmark.....	38
4.3.1	Recruitment of test participants	38
4.3.2	Changes to the test design.....	38
4.3.3	The setup of the focus group.....	39
4.3.4	The focus group session layout.....	40
5	Results.....	41
5.1	Demographics.....	41
5.2	Usability issues.....	42
5.2.1	Navigation & structure	44



5.2.2	Content & Information	45
5.2.3	Design & Presentation	47
5.2.4	Other	48
5.3	Task metrics	49
5.3.1	Task 1: Account creation.....	50
5.3.2	Task 2: Free exploring.....	51
5.3.3	Task 3: Connect Fitbit.....	52
5.3.4	Task 4: Find performed steps	53
5.3.5	Task 5: Personalised dinner recipe.....	54
5.3.6	Discussion and recommendations	55
5.4	Questionnaire results	55
5.4.1	The Netherlands + Scotland.....	55
5.4.2	Denmark	56
5.5	Discussion.....	57
6	Conclusion and updated requirements.....	58
6.1	Updated requirements.....	58
6.1.1	Functions and Events	58
6.1.2	Interaction and Usability.....	59
6.1.3	Content and Structure.....	61
6.1.4	Style and Aesthetics	62
7	Bibliography	64

List of figures

Figure 1: The Main Menu of COUCH.....	11
Figure 2: Council of Coaches first step of creating a new account.....	12
Figure 3: Example of baseline questions during the account creation process of Council of Coaches.....	12
Figure 4: Example of the in-take questions used for the automatic coach recommender.....	13
Figure 5: The coach selection screen on which the users can select their coaches.....	14
Figure 6: The main living room user interface for Council of Coaches, where the coaches live.....	14
Figure 7: Coda's introduction round.....	15
Figure 8: Coach specific introduction dialogues.....	15
Figure 9: Example of coaching content dialogue with Olivia, the Physical Activity coach.....	16
Figure 10: Activity Book widget that Olivia (Physical Activity coach) can use to show the user's data.....	17
Figure 11: The Radio widget with simple channel select, volume, and play and stop controls.....	18
Figure 12: Screen Recorder app in Google Play Store.....	27
Figure 13: Settings android device - show touches.....	28
Figure 14: Settings of the screen capture app.....	29
Figure 15: Screen Capture app recording control. Use top button to record screen capture video.....	29
Figure 16: Control centre of iPad settings to include Screen Recorder functionality.....	30
Figure 17: iPad Screen Recorder icon.....	30
Figure 18: Fitbit page.....	31
Figure 19: Fitbit settings page.....	31
Figure 20: Authorized applications page Fitbit.....	32
Figure 21: Overview of the number of unique usability issues per task.....	43
Figure 22: Overview of the number of unique usability issues per category.....	43
Figure 23: Performance metrics radar chart for task 1: account creation.....	50
Figure 24: Performance metrics radar chart for task 3: Connect Fitbit.....	52
Figure 25: Performance metrics radar chart for task 4: find steps Saturday.....	53
Figure 26: Performance metrics radar chart for task 5: Get personalized diner recipe.....	54

List of tables

Table 1: Overview of the third evaluation study in deliverable 2.6	7
Table 2: Council of Coaches evaluation cycles.	8
Table 3: Overview of all coaches in the Council of Coaches.	16
Table 4: Schematic study procedure for the third evaluation.	21
Table 5: Overview of the tasks.	23
Table 6: Relevant aspects of task one.	24
Table 7: Relevant aspects of task two.....	24
Table 8: Relevant aspects of task three.	25
Table 9: Relevant aspects of task four.....	25
Table 10: Relevant aspects of task five.....	26
Table 11: After Scenario Questionnaire.	26
Table 12: Interview guidebook.	32
Table 13: The System Usability Scale (SUS).	34
Table 14: The Technology Acceptance Model (TAM).	34
Table 15: The System Usability Scale (SUS) scoring table. Expressed in grades, percentile range, adjectives, acceptability and net promotor scale (NPS) (Sauro, 2016) (Tullis & Albert, 2008).	37
Table 16: Focus group session layout.....	40
Table 17: Demographics of the 30 evaluation participants.	41
Table 18: Exemplary usability issues in category Navigation & Structure.	44
Table 19: Exemplary usability issues in category Content & Information.....	46
Table 20: Exemplary usability issues in category Design & Presentation.....	48
Table 21: Relative filling area radar charts (figure 21 till 24).	49
Table 22: Performance metrics for task 1: Account creation.....	50
Table 23: Performance metrics for task 2: Free exploring.....	51
Table 24: Performance metrics for task 3: Connect Fitbit.	52
Table 25: Performance metrics for task 4: Find steps Saturday.	53
Table 26: Performance metrics for task 5: Get personalized dinner recipe.	54
Table 27: Overview of the Technology acceptance model outcomes (NL+SC).	56
Table 28: Overview of the Technology acceptance model outcomes (DK).	56

Symbols, abbreviations and acronyms

CMC	Centre for Monitoring and Coaching
COUCH	Council of Coaches
CP	Chronic Pain
D	Deliverable
DBT	Danish Board of Technology Foundation
DK	Denmark
DM2	Diabetes Mellitus Type-2
EC	European Commission
ISPRINT	Innovation Sprint
M	Month
MS	Milestone
NL	the Netherlands
RRD	Roessingh Research and Development
SU	Sorbonne University
SUS	System Usability Scale
TAM	Technology Acceptance Model
UDun	University of Dundee
UK	United Kingdom
UPV	Universitat Politècnica de València
UT	University of Twente
WP	Work Package

1 Introduction

This deliverable describes the execution and outcomes of the evaluation study of the third functional demonstrator of the Council of Coaches system. This third functional demonstrator of Council of Coaches was originally released in May, 2019 but has received additional updates before the execution of this evaluation. In Overview of the third functional demonstrator, we describe the updated version of this third functional demonstrator and the changes we have made to the system based on the previous evaluation studies reported in D2.4 (Beinema, et al., 2018) and 2.5 (Beinema, et al., 2019).

Methods of evaluation of the third functional prototype describes the study methodology of this evaluation. This third evaluation primarily focuses on eliciting usability issues for further improvement of the system to the technical prototype, with specific attention to (1) the new design and new aspects in the demonstrator (radio, activity book & robot account page), (2) the coaching content which the coaches provide to the users, and (3) on task performance of commonly performed actions within the Council of Coaches. The third evaluation study resembles the second evaluation in applying a retrospective think-aloud and eliciting usability issues. However, the protocol for this evaluation was extended by evaluating task performance metrics and evaluating technology acceptance and willingness to pay. This usability evaluation study was performed in all three evaluation sites (Enschede, the Netherlands; Dundee, United Kingdom; and Copenhagen, Denmark).

In Results 5, the results of the evaluation study are described. A total of 30 participants took part in this study. A total of 395 usability issues were found across all participants. Deduplication of usability issues that occurred among two or more participants, resulted in a total of 247 unique usability issues: 77 positive usability points and 170 negative usability issues. The majority of the negative issues were minor issues that have a low impact on user performance. In total, there were only 23 serious and 8 critical issues compared to 24 serious and 16 critical issues in the previous evaluation.

Table 1 below provides a brief overview of the study (method, setting and participants).

Table 1: Overview of the third evaluation study in deliverable 2.6

Study	Method	Setting	N	Participants < 54	Participants > 55	Participants with health conditions (DM2, CP)
Evaluation of the third functional prototype	Retrospective think-aloud + Task performance metrics + System usability and technology acceptance.	Lab-setting	30	1	29	DM2 (8) CP (5) DM2+CP (2)

We conclude this deliverable with additional requirements for the Council of Coaches system that stem from the evaluation study in this deliverable. These requirements will be used to further improve the system for the third and final usability evaluation of the third functional demonstrator.

2 Objectives

The objective of this deliverable is to describe the evaluation of the third functional demonstrator of the Council of Coaches system. In Council of Coaches, there are four “official” cycles of demonstrator releases followed by evaluations, as depicted in Table 2 below. Each of these cycles will result in a report on the evaluation results and updated requirements. The first and second round has been completed; see D2.4 (Beinema, et al., 2018) and D2.5 (Beinema, et al., 2019). This deliverable (D2.6) describes the results of the third evaluation round.

Table 2: Council of Coaches evaluation cycles.

Council of Coaches Evaluation Cycles	
Cycle 1	
M9	Milestone 2: First Functional Prototype
M12	D2.4: Evaluation results of first functional prototype and updated requirements
Cycle 2	
M15	Milestone 3: Second Functional Prototype
M18	D2.5: Evaluation results of second functional prototype and updated requirements
Cycle 3	
M21	Milestone 4: Third Functional Prototype
M24	D2.6: Evaluation results of third functional prototype and updated requirements
Cycle 4	
M27	Milestone 5: Technical Prototype
M36	D7.7: Final Demonstration Results

	<i>Completed</i>
	<i>To be done</i>

3 Overview of the third functional demonstrator

The third evaluation for the Council of Coaches will evaluate the third functional prototype. This section provides a description of the third functional demonstrator, which can be found online at the following address:

<https://www.council-of-coaches.eu/beta/>

3.1 Recap: Results from the first and second evaluation

For each new functional demonstrator, we extend the scenario so that it becomes more complete and gets closer to our vision of the final Council of Coaches system. In the first functional demonstrator evaluation, the user met the coaches for the first time and they introduced themselves. The demonstrator provided us insights on the core elements of the Council: Would users accept multiple coaches? Would they like to have coaches who had personalities and background stories? Would the topics that the coaches would coach on be suitable for the target group?

The second functional demonstrator included tailored coach action strategies based on the motivation type of the user (a-motivation, extrinsic motivation and intrinsic motivation), based on a recent study of Velsen et al. (van Velsen, Broekhuis, Jansen-Kosterink, & op den Akker, 2019). The second demonstrator furthermore provided usability insights and updated requirement on the topics; Navigation & Structure, Content & Information, Design & Presentation and Other.

We built the third demonstrator version based on what we learned from the user requirements and stakeholder analysis processes, a tailored approach of the second demonstrator combining the usability evaluations. The next step within this development towards the final Council of Coaches prototype was to add more coaching content. Moreover, the scenario for the third functional demonstrator is to test acceptance and usability of more advanced features (agent animations and behaviours, interaction concepts). The description of the third demonstrator can be found in Section 3.2: The third functional demonstrator.

3.1.1 The first evaluation

The first evaluation, reported in D2.4 (Beinema, et al., 2018), which evaluated the first functional prototype and two additional demonstrators on the theme of acceptance, resulted in insights and recommendations on the topics of 'interaction with the technology', 'opinions on character design', 'appearance of the virtual coaches', and 'content of the coaching'. Two card sorting tasks also provided additional information on the users' preferences for topics and coach characteristics. Some important insights from the evaluation were:

- Interaction with technology:
 - The interface (e.g. buttons, speak bubbles) should be intuitive. Participants were sometimes wondering which coach was speaking.
 - Participants seemed to prefer interaction with multiple coaches and found that there was a game-like element to it.
- Opinions on character design:
 - Participants liked the background stories of the coaches, but the amounts of coaching content and social content should be carefully balanced.
- Appearance of the virtual coaches:
 - The participants preferred human-like coaches to the more neutral blob-men.
- Content of the coaching:
 - The language used in the dialogues should not be too difficult. Participants would have preferred more coaching content.

From the card sorting tasks we learned that users thought the diet and physical activity topics were important, but that they also appreciated the brain-training topic. We also learned that for character

traits they preferred humour and friendliness in addition to our confirmed assumption that they should be relatable experts or experience experts.

3.1.2 The second evaluation

The second evaluation, reported in deliverable D2.5 (Beinema, et al., 2019), was set up to evaluate the second functional demonstrator on two main topics. 1) *Do participants indicate that a coaching strategy, that is assumed to be more motivating for the participant's type of motivation, is indeed more motivating than a strategy that is assumed to be not motivating?* and 2) *What is the system usability and user experience of the demonstrator?*

On the first topic, evaluation 2 showed that the average ratings for all four strategies were quite close and this is reflected in the selection of the most preferred strategy by the participants, which does not show a significant preference. The overall results of the online experiment showed that most experiment participants are intrinsically motivated or fall within the so-called "dual motivation" category. This is in line with expectations, as non-motivated participants are unlikely to participate in this type of studies.

With regards to the usability study, overall the results of the retrospective think-aloud method, showed that the second Council of Coaches demonstrator was well understood and perceived to be usable by the test participants. Test participants showed high engagement with the demonstrator and the study in general, resulting in a large amount of recommendations and leading to additional requirements.

In the next subsection we will first describe the third functional demonstrator and the adjustments that we made to it to support the scenario for this third evaluation round.

3.2 The third functional demonstrator

3.2.1 The Main Menu

The Web interface of the demonstrator has a welcoming and familiar visual of a building entrance. It is an entrance to the “Council of Coaches” home. The functions available in this screen are: account creation or log in with an account that was created before, selection of preferred language, additional menu options, that give a “professional feel” of the application like the “Credits scroller”, a Patch Notes section, and access to the privacy statement (see Figure 1).

The login functionality is placed strategically in the centre of the image, on the doorway with the reference of an entry to the system (i.e. entering the Council of Coaches house) with the opening door, during log-in. The wall-paper that can be seen in the hallway is similar to the one which is found later on, during the baseline questions and of the living room with all the coaches. Therefore, the log-in screen offers the user a direct link to the living room where all coaches are located.



Figure 1: The Main Menu of COUCH.

3.2.2 Register account

The first time when users interact with the Council of Coaches application, they have to create an account, so that his preferences and information can be stored, and dialogues can be personalized. The account creation procedure serves a dual purpose, as it also introduces users to the COUCH system and learns them to interact with the dialogue system. Figure 2 shows the account creation interface with an example of the embedded tutorial (helper arrows, e.g. “Choose your response below...”).

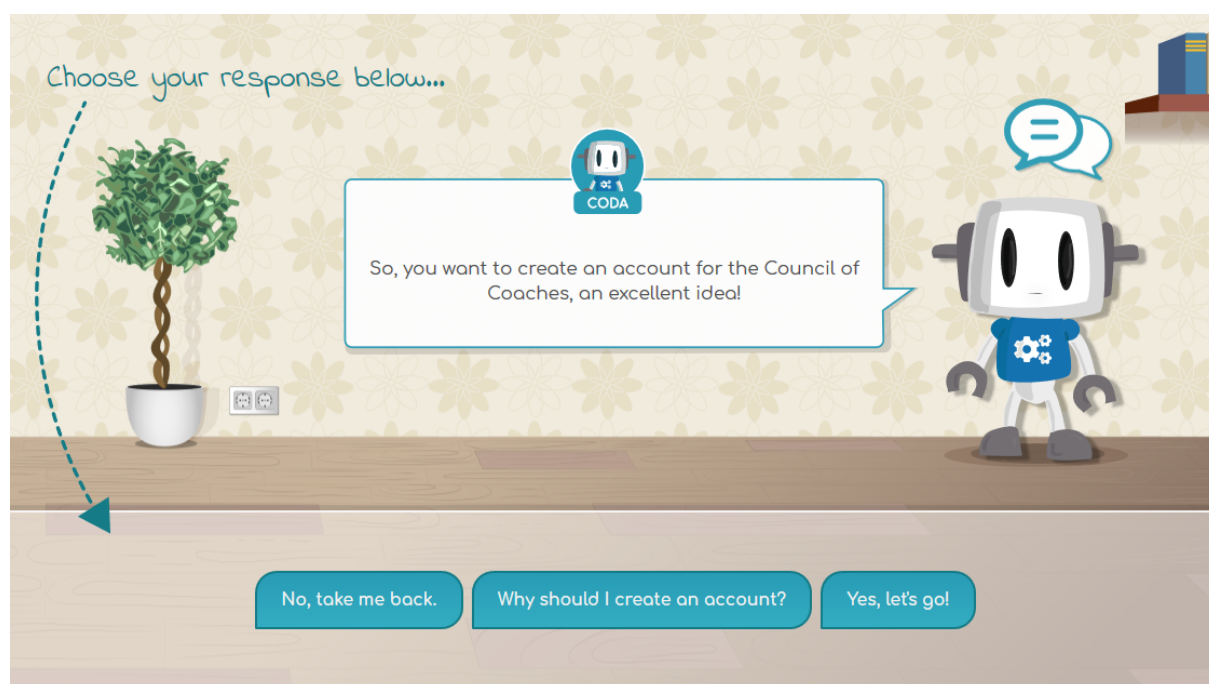


Figure 2: Council of Coaches first step of creating a new account.

3.2.3 Baseline questions

After registering the account information, the user will undergo a series of baseline questions, to enable the system to identify the type of user. These questions provide demographic variables like gender, age and level of education which are used to personalize the system (Figure 3). Furthermore, the baseline questions contain a 3-item health literacy questionnaire and four questions to get a rough estimate of technology savviness by asking whether the user uses a computer/laptop, tablet/iPad, smartphone or game console at home. These items are currently part of the system in order to acquire information relevant for the evaluation of the system, but may later be used to further personalize the interaction with the coaches.

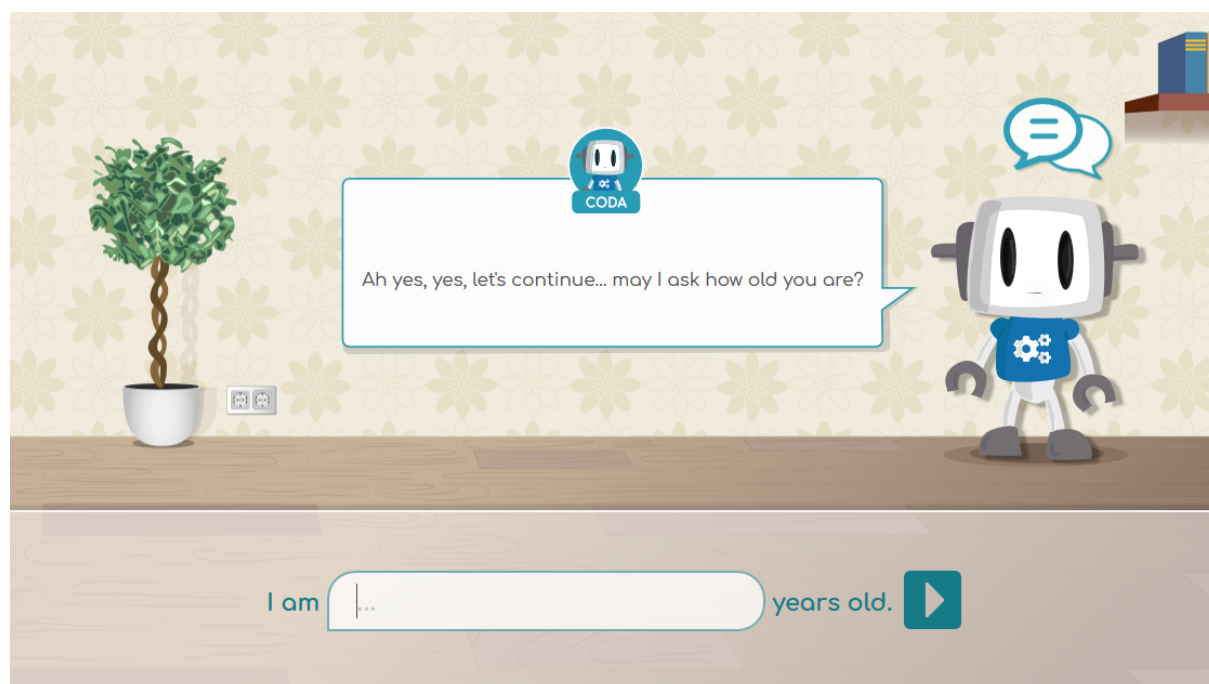


Figure 3: Example of baseline questions during the account creation process of Council of Coaches.

3.2.4 Coach selection

The third demonstrator can help users to select their “Council of Coaches”. This functionality consists of 2 phases;

- 1) Coda will ask whether or not the user is diagnosed with Diabetes Mellitus type 2 (DM2) or chronic pain (CP). If the answer is yes to the DM2, the diabetes coach Katarzyna will be automatically selected. If the answer is yes to CP, the chronic pain coach Rasmus will be automatically selected. If the answer is no to these questions the accompanying coaches cannot be selected.
- 2) Coda will ask whether the user wants to get help with the selection of the coaches or to select the coaches themselves. When the user indicates to want help, he will be taken through a 2-item questionnaire (self-perception and state of change) regarding the four coaching domains (Physical activity, Nutrition, Social, Cognition) (Figure 4). Based on the given answers the system determines the most appropriate coaches for this participant and displays this in the coach selection screen (Figure 5). The user can get more information of each coach by clicking on the individual portraits. The pre-set automatic selection of coaches can be altered using the checkboxes below the portraits. In case the user indicates not to want help with the selection of the coaches, he/she will immediately be directed to the coach selection screen.

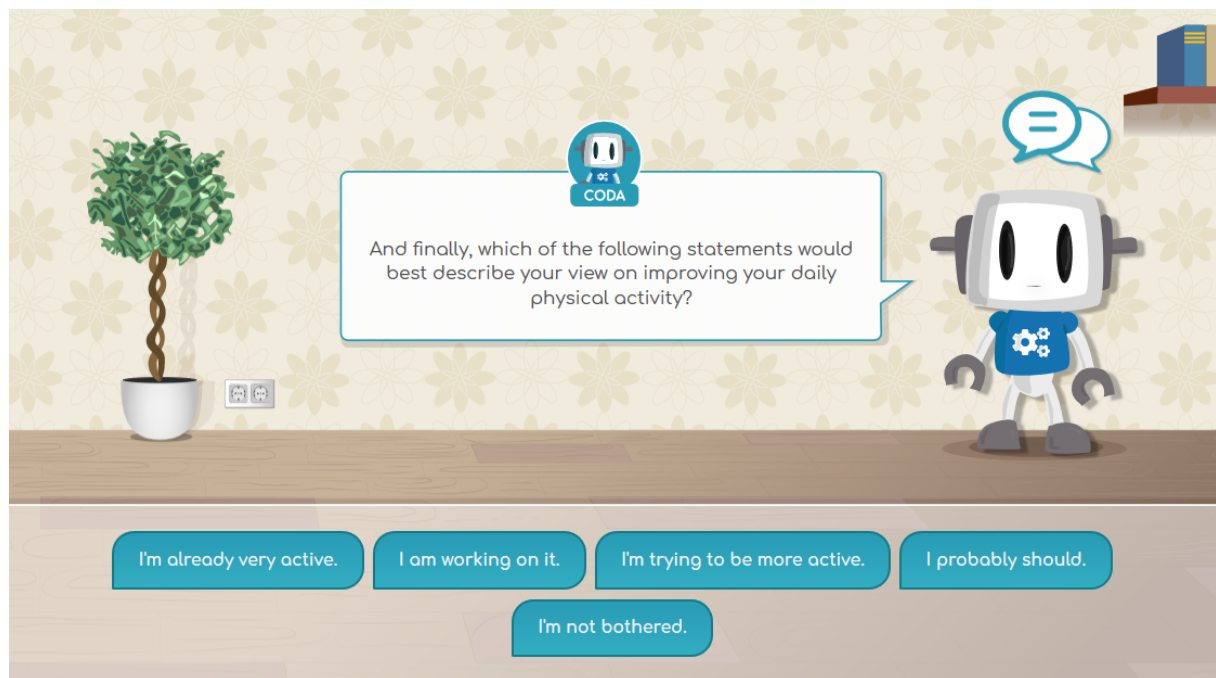


Figure 4: Example of the in-take questions used for the automatic coach recommender.

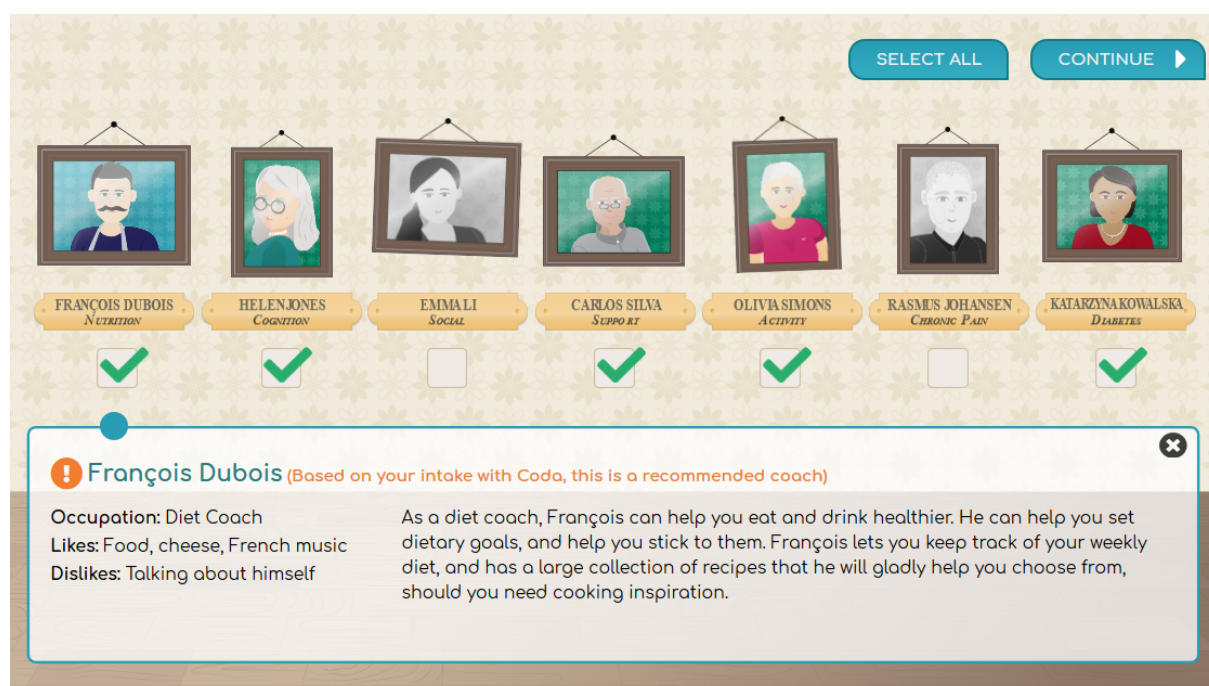


Figure 5: The coach selection screen on which the users can select their coaches.

3.2.5 Living room

After the coach selection, users proceed to the living room with the coaches. Each coach has his/her own place within the living room based on their expertise (Figure 6). Two of the most identifiable coaches are the nutrition coach “Francois” in the kitchen and the activity coach “Olivia” behind the bicycle ergometer.



Figure 6: The main living room user interface for Council of Coaches, where the coaches live.

3.2.6 Introduction by Coda

The Council of Coaches experience starts with Coda showing the user around, introducing the user to every coach and explaining their purpose (Figure 7).

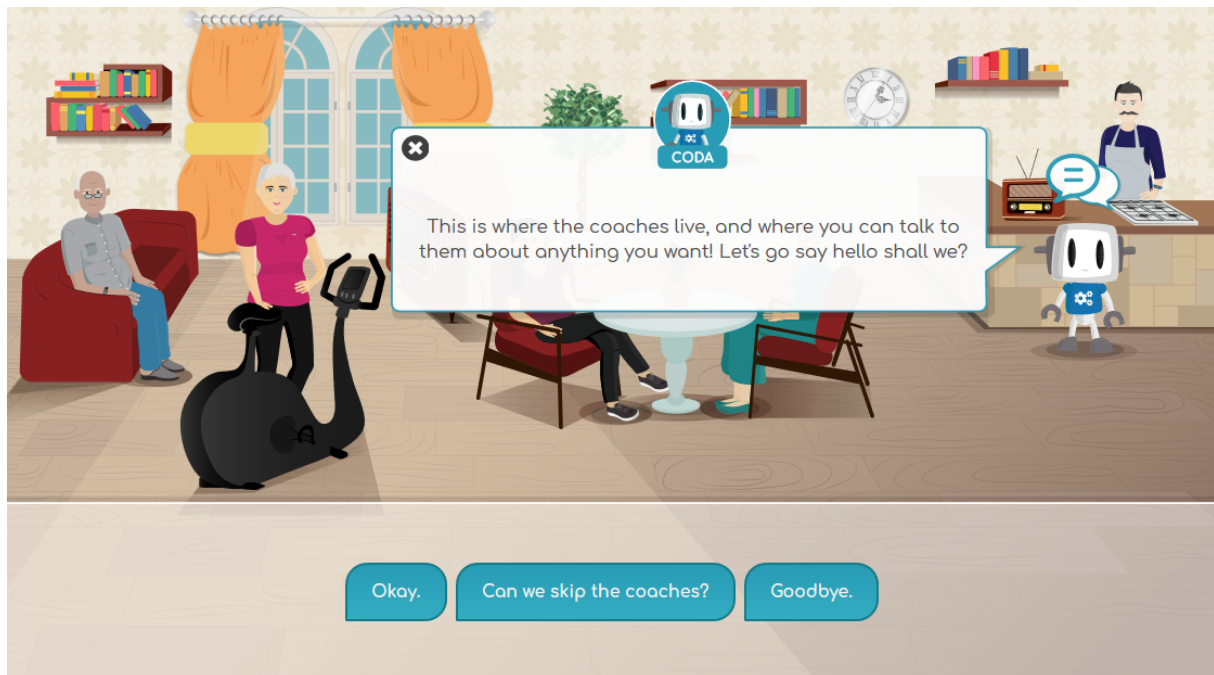


Figure 7: Coda's introduction round.

3.2.7 Interaction and introduction to coaches

Users can determine themselves which coaches to interact with. The first personal dialogue with each coach consists of a short introduction and a personal story or domain relation question (i.e. François asks the user whether he/she likes to cook) (Figure 8).

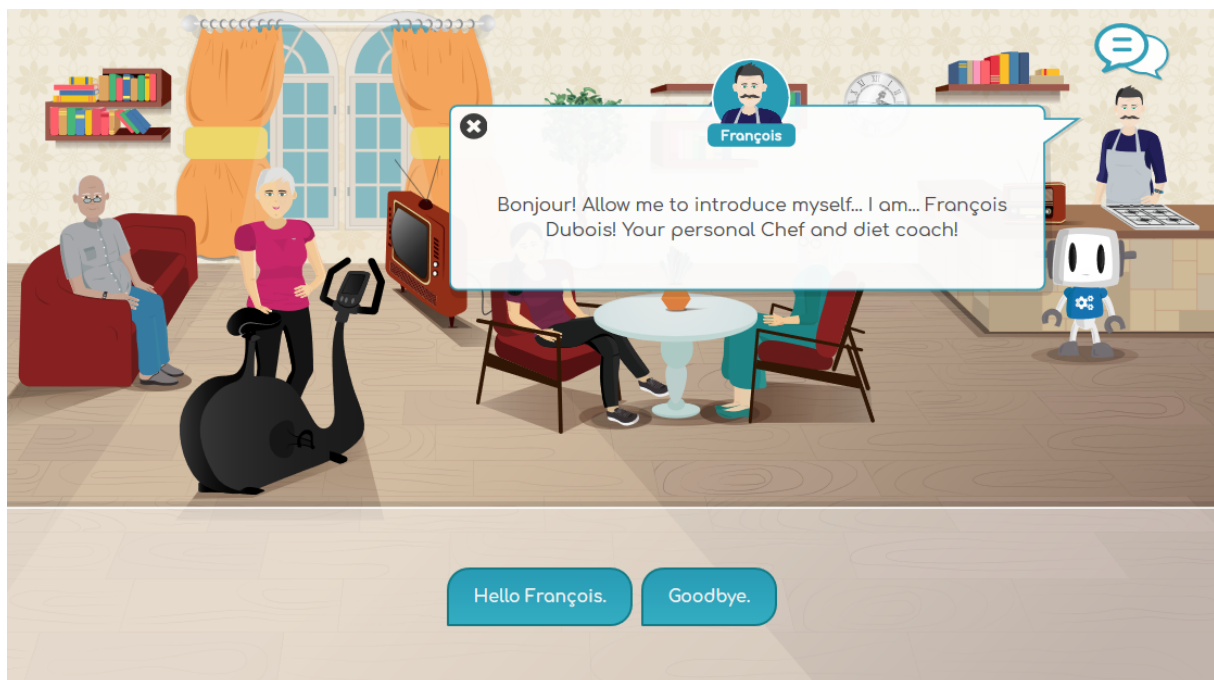


Figure 8: Coach specific introduction dialogues.

3.2.8 Coaching content

The council of coaches consists of a variety of coaches, each with different fields of expertise. Table 3 shows an overview of the coaches and their role within the Council of Coaches. More information about the coach content can be found in D3.4: Final coaching actions and content. When starting a conversation with one of the coaches there is the option to have a social conversation but also to proceed to specific coaching topics. As an example, the coaching content of Olivia the physical activity coach consists of various options (Figure 9). When the user selects “Let’s talk about coaching” Olivia suggest to discuss the activity progress in the activity book, the goals or the coaching strategy. Furthermore, Olivia can provide any advises or talk about the activity tracker and help to connect such device to their Council of Coaches account.

Table 3: Overview of all coaches in the Council of Coaches.

Role	Name	Nationality	Gender	Age
Physical Activity Coach	Olivia Simons	Dutch	Female	52
Nutrition Coach	François Dubois	French	Male	45
Social Coach	Emma Li	American	Female	28
Cognitive Coach	Helen Jones	British	Female	64
Peer Support	Carlos Silva	Portuguese	Male	67
Chronic Pain Coach	Rasmus Johansen	Danish	Male	33
Diabetes Coach	Katarzyna Kowalska	Polish	Female	45
Assistant	Coda	-	-	-

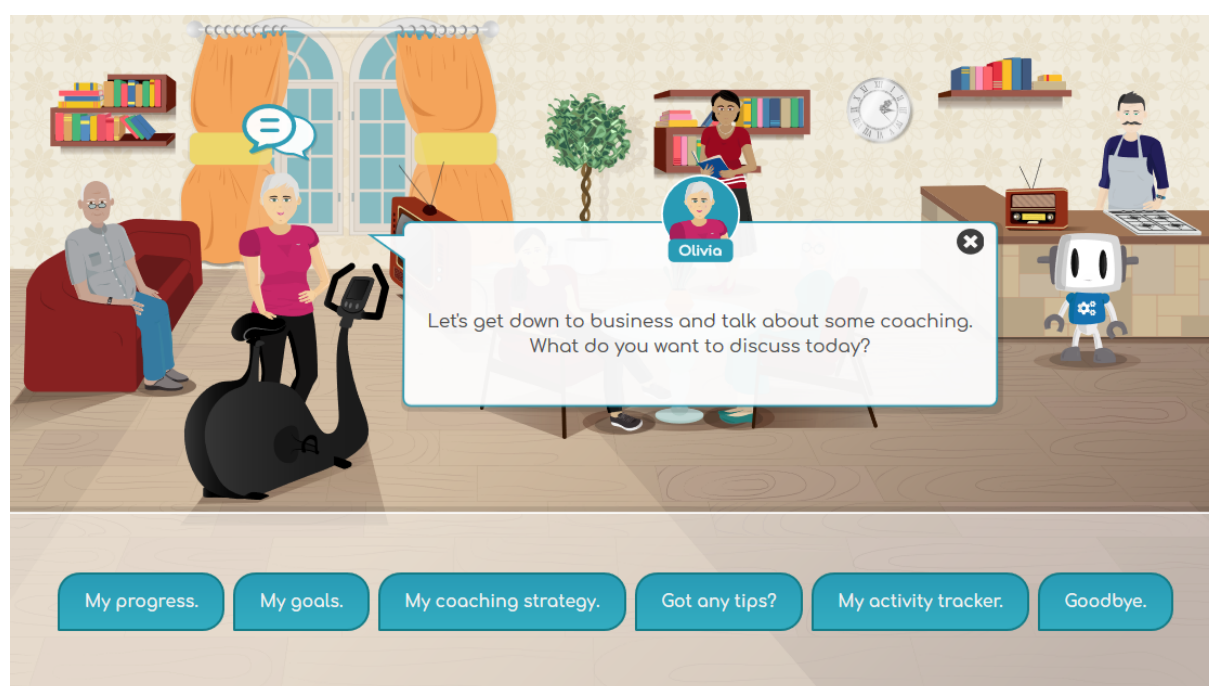


Figure 9: Example of coaching content dialogue with Olivia, the Physical Activity coach.

3.2.9 Interaction with widgets

Next to the coaching dialogue language content (D3.4: Final coaching actions and content), the third demonstrator contains several new elements (called widgets) to transfer content and/or improve user experience. Different widgets were conceptualized, including i.e. a physical activity book that Olivia can use to show details about the user's physical activity behaviour – these features are tightly connected to the content of the coaches. This widget was present in the third demonstrator and showed the number of steps per day from the previous week of activity (Figure 10).

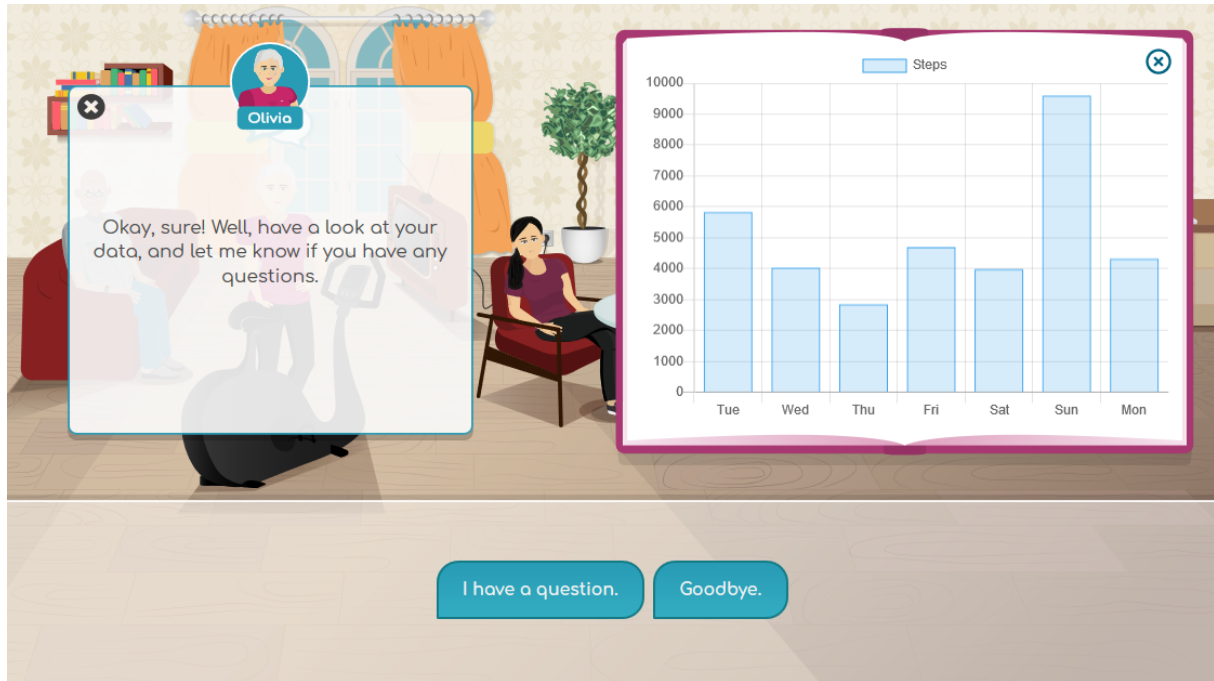


Figure 10: Activity Book widget that Olivia (Physical Activity coach) can use to show the user's data.

Moreover, a radio widget was created to break the silence and offer some background music to the users (Figure 11). The living room design lends itself perfectly for this type of additions (a little old-fashioned radio that is a fitting decoration). This old radio has four channels that the user can choose from, with Jazz and Blues on channel 1, Classical Music on channel 2, different French songs on channel 3 (Francois loves it), and Rock & Roll on channel 4.

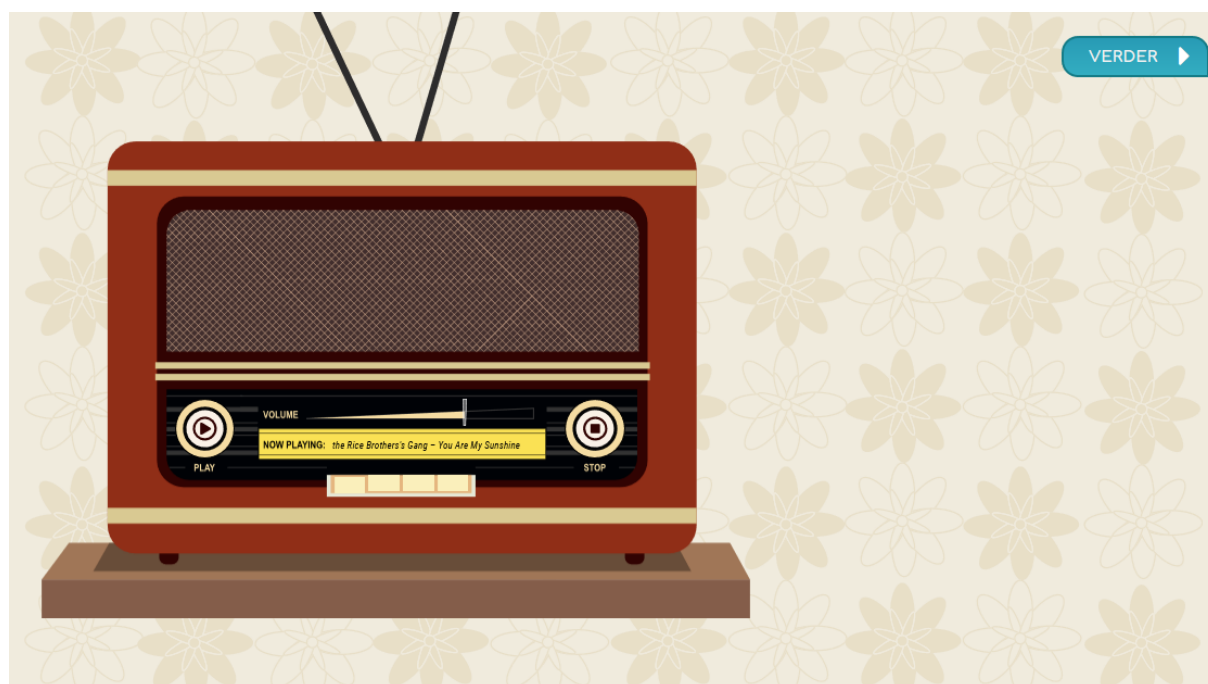


Figure 11: The Radio widget with simple channel select, volume, and play and stop controls.

4 Methods of evaluation of the third functional prototype

In this section we will describe the methods and aims of this usability and task performance evaluation. We decided to perform the evaluation on all three sites involved in the requirements gathering process (Roessingh Research and Development (RRD), Danish Board of Technology (DBT), and the University of Dundee (UDun), to take into account potential cultural differences.

As the Council of Coaches system is maturing, it is important for this third demonstrator to examine how well potential end-users perform common tasks within the system and how end-users currently perceive the usability of the system and evaluate the user experience. A high and positive usability and user experience evaluation can improve satisfaction of a system, which in turn can heighten perceived intention-to-use (Belanche, Casaló, & Guinalíu, 2012). The difference between usability and user experience is that the first focuses on the functional elements of a system, can a user efficiently and satisfactory complete tasks, while the latter focuses on the interaction the user has with the system, how the user feels about the system (Bevan, 2009).

This evaluation study had three main goals:

- 1) Identifying usability problems
- 2) Benchmarking the usability
- 3) Identifying the user experience and intention to use

Regarding the first goal, a retrospective think-aloud protocol was conducted. Qualitative data collection was to elicit usability issues is generally accepted as the 'golden standard' in usability evaluation studies. The study set up was based on the second evaluation study D2.5 (Beinema, et al., 2019). In addition to this protocol, task metrics were measured. Because this third demonstrator has new aspects (e.g. homepage, Fitbit integration, new coaching content), we wanted to know how potential users evaluate these domains. Therefore, users will be given several tasks to perform in these specific domains. Task metrics such as task completion, task completion time and task satisfaction are an indicator of the effectiveness and efficiency of user-system interaction in these domains. The results of both the retrospective think-aloud method and task metrics help to elicit new usability issues and additional user requirements.

Because this is the last iteration round before the final version is developed, we wanted to benchmark the usability of the Council of Coaches system. For this purpose, participants filled out the System Usability Scale (Brooke, 1996) after completing the retrospective think aloud. The SUS in combination with task metrics provides good insights in the usability of eHealth systems.

Finally, in this research, the external variables of the technology acceptance model (4.1.8) are used as experience domains. In literature, multiple constructs were found for user experience of eHealth services such as aesthetics (Lavie & Tractinsky, 2004), enjoyment (van der Heijden, 2004), privacy (Heinz, Tuch, Seckler, Opwis, & Forde, 2014) and trust (McKnight, Carter, Thatcher, & Clay, 2011), (Wang & Emurian, 2005). These are mostly measured quantitatively through questionnaires and online surveys. Combining these with questionnaires about acceptance, perceived usefulness, ease of use and willingness to pay will provide us the complete picture of all factors explaining the intention to use the Council of Coaches system.

4.1.1 Participants

Thirty participants will be recruited to participate in this evaluation, 10 from each research site (RRD, DBT, UDun). The participants need to meet the target group of Council of Coaches, according to the following inclusion criteria:

- Participants having one or more of the following conditions: Chronic pain, Diabetes type 2 (at least 18 years old).
- Or participants of at least 55 years old.
- Native speaker or C2 level based on European framework.

- Willing to provide informed consent

4.1.2 Ethics & Recruitment

Considering the nature of this research (usability evaluation of prototype in a voluntary session), this study does not require formal (medical) ethical approval. Older adults (aged 55 or over) will be approached by email or telephone to inform them about the study and offer the opportunity to ask questions. If they indicate to be interested to participate, they receive a recruitment email with the formal information letter and consent form attached. Voluntary participants are scheduled for a two-hour session at the research site of the project partner. If participants are unfit to travel to this location, the study can be done on location but this is not preferred. At the start of this session the participants will be informed orally about the study and are asked for written informed consent. Only after informed consent is received, the study continues and only the data from those consenting will be included. All data will be anonymized.

4.1.3 Study procedure

This usability evaluation is set up to simulate and evaluate real-life use. Two strategies are used to reflect real-life use, namely free exploration and task performance of commonly performed actions. Both task performance metrics and usability issues are tracked while participants go through the system by performing five consecutive tasks (4.1.4). The approximate length of each session is 1.5 to 2 hours.

Prior to the start of the evaluation, each participant fills out an informed consent form. While the participant interacts with the demonstrator, one moderator is present to coordinate the execution of the task according to the evaluation guidebook and to start the screen capture and audio recording.

After task one (the account creation) the COUCH demonstrator will lead the participant through a series of demographic baseline questions (gender, age, level of education etc.), which will be stored for the purpose of this evaluation. Afterwards the user can choose to select their own coaches or to proceed with an intake (motivation and state of change questions) to help select the coaches. From there on the participants were asked to freely explore the Council of Coaches interface (task 2) and to fulfil the last three tasks (task 3 to 5).

After the user interaction, an interview (retrospective think-aloud) will be held with the participant. During this interview, all the phases of the user interaction will be discussed task by task. The interview guidebook (4.1.7) provides guidance on where to ask for during the interview.

After the interview participant will be asked to fill out the system usability scale (SUS) questionnaire and the technology acceptance model (TAM) questionnaire. These questionnaires can be found in 4.1.8. Participants will receive full explanation about the purpose of the study and the next steps to be taken at the end of the study session.

The full study procedure is schematically displayed in Table 4.

To summarize, the data that will be collected per participant during the study is the following:

- The email-address used in the system for creating an account (these are provided by the moderators and have the following format: *RRDCOUCH[subjectnumber]@RRD.nl* / *DUNDEECOUCH[subjectnumber]@DUNDEE.uk* / *TEKNOCOUCH[subjectnumber]@TEKNO.dk*).
- Interactions with the system (i.e. responses to the agents, answers to demographic questions and coach selections) will be stored by the system automatically.
- Task performance metrics for every task (task completion, completion time, task satisfaction, steps, errors)
- A screencapture video of the user interacting with the system on a tablet, which records the screen, the user's screen touches, and audio through a microphone. (We use the "Screen Recorder" app.)
- An audio recording of the post-interaction retrospective think-aloud interview.
- SUS and TAM questionnaire results.

Table 4: Schematic study procedure for the third evaluation.

No	Phase	Activity	Explanation	Materials	Time
1	Introduction	Welcome	Welcome and explanation of study. Introduction to Council of Coaches system.	Coffee/tea	5
		Obtain informed consent and permission for audio recording	Participant fills out an informed consent form on paper.	Informed consent form	3
2	User interaction with system	Introduction and explanation retrospective think-aloud.	"During the following 30 minutes you will receive a series of tasks one by one. Perform these tasks, during your tasks you are allowed to make any comments as you wish" End of introduction: Moderator turns on screen capture software on tablet + additional audio recorder	Screen capture software (4.1.5)	2
		Task 1: Account creation	"Create an account and continue to the council of coaches" Task starts on the home page. https://council-of-coaches.eu/beta/ The moderator selected the right language prior to the start of the task. <i>Register with:</i> <i>e-mail:</i> <i>For the Netherlands:</i> - RRDCOUCH[subjectnumber]@RRD.nl <i>For Scotland:</i> - DUNDEECOUCH[subjectnumber]@DUNDEE.uk <i>For Denmark:</i> - TEKNOCOUCH[subjectnumber]@TEKNO.dk <i>Password: 123456</i> Task ends when the participant pushes the finish button when Coda confirms to have set up the account. Moderator: tracks time on task Moderator fills in if respondent completes task within the time limit or not. If task is not completed after 5 minutes, moderator stops the task.	Stopwatch (for time on task) See start and end of task 1 in (4.1.4) See Task metrics (4.2.2)	5
			Participant fills out After-scenario questionnaire for task 1	After-Scenario Questionnaire (4.1.4.1).	1
			The participant now gets the time to go through the baseline questions and coach selection menu.		5
		Task 2: Free exploration	"Freely explore the options and functionalities of the system within 10 minutes". Task start after entering the living room for the first time when coda introduces the living room. Task ends when 10minutes are over OR when the participant indicates he or she wants to continue with the next task. <i><u>Moderator: Make sure that all introduction dialogues of the coaches are settled, in case the task is completed without talking to all coaches.</u></i>	Stopwatch (for time on task) See start and end of task 2 (4.1.4) See Task metrics (4.2.2)	10

			Participant fills out After-scenario questionnaire for task 2	After-Scenario Questionnaire (4.1.4.1).	1
		Task 3: Connect fitbit	<p>"Connect the fitbit activity tracker account with the COUCH account"</p> <p>Task starts in living room, <u>without</u> an ongoing coach conversation. Provide the participant with the login information.</p> <p>Username: couch01@activity-coach.com</p> <p>Password: couch01!</p> <p>Task ends when the the data sharing with fitbit is allowed by pushing the allow button and you return to living room.</p> <p>Moderator: tracks time on task</p> <p>Moderator fills in if respondent completes task within the time limit or not.</p> <p>If task is not completed after 5 minutes, moderator stops the task.</p>	<p>Stopwatch (for time on task)</p> <p>See start and end of task 3 (4.1.4)</p> <p>See Task metrics (4.2.2)</p>	5
			Participant fills out After-scenario questionnaire for task 3	After-Scenario Questionnaire (4.1.4.1).	1
		Task 4: Steps	<p>"Find out how many steps were performed last Saturday"</p> <p>Task start in living room, <u>without</u> an ongoing coach conversation.</p> <p>Task ends when in the activity book shows the step graph over the week and the participant found the amount of steps on Saturday.</p> <p>Moderator: tracks time on task</p> <p>Moderator fills in if respondent completes task within the time limit or not and the amount of steps that were indicated by the participant.</p> <p>If task is not completed after 5 minutes, moderator stops the task.</p>	<p>Stopwatch (for time on task)</p> <p>See start and end of task 4 (4.1.4)</p> <p>See Task metrics (4.2.2)</p>	5
			Participant fills out After-scenario questionnaire for task 4	After-Scenario Questionnaire (4.1.4.1).	1
		Task 5: Recipe	<p>"Receive your personalized diner recipe"</p> <p>Task start in living room, <u>without</u> an ongoing coach conversation.</p> <p>Task ends when Francois provides the personalized recipe.</p> <p>Moderator: tracks time on task</p> <p>Moderator fills in if respondent completes task within the time limit or not.</p> <p>If task is not completed after 5 minutes, moderator stops the task.</p>	<p>Stopwatch (for time on task)</p> <p>See start and end of task 5 (4.1.4)</p> <p>See Task metrics (4.2.2)</p>	5
			Participant fills out After-scenario questionnaire for task 5	After-Scenario Questionnaire (4.1.4.1).	1
		End of user interaction	Moderator turns off screen capture software.		
(2.5)	Break		Ask the participant if he/she likes to have a small break.		(10)
3	Retrospective Think-aloud	Start retrospective interview	Explain retrospective interview to participant: together you will watch the actions taken in a video. Ask them to		2

			comment on their actions and verbalize the thoughts that they had during these phases.		
			Per phase (each task), the participant will be asked to comment on his/her actions during that phase while rewatching his actions using screencapture software. After rewatching each phase, the participant will be asked some general questions, on what went well/wrong and his/her thoughts and feelings during this phase.	Start playing screencapture software from beginning. Pause after each task. See interview guidebook (4.1.7)	40
4	Closure		Participant fills in the exit interview questionnaires SUS +TAM. Moderator turns off audiorecorder. Full explanation study's purpose + explain about next steps.	See questionnaires SUS and the TAM (4.1.8)	8
5	Reset Fitbit account	Reset Fitbit account	Moderator: Go to fitbit.com > dashboard > settings > applications: and revoke access to the RRD activity coach. Moderator: Log out of fitbit.com Now the same fitbit account can be used for the next participant.	Revoke access information (4.1.6)	-
			Total time		110 (120) min

4.1.4 The evaluation tasks

Five relevant actions which take the user through the council of coaches system were evaluated. Table 5 shows an overview of these five tasks. Task 2 consists of free exploring of COUCH after entering the living room, as the introduction of the coaches requires almost 5 minutes and we would also like to evaluate the other interactions.

Table 5: Overview of the tasks.

Task	Description of task	Module	Duration
Task 1	"Create an account and continue to the council of coaches."	Home page	5 minutes
Task 2	"Freely explore the options and functionalities of the system within 10 minutes."	Living room	10 minutes
Task 3	"Connect the fitbit activity tracker account with the COUCH account."	Activity coach	5 minutes
Task 4	"Find out how many steps were performed last Saturday."	Activity coach	5 minutes
Task 5	"Receive your personalized diner recipe."	Nutrition coach	5 minutes

- **Task 1: Account creation.**

Participants are asked to perform the following task: "Create an account and finish the task by pressing the continue button". Task one takes the participant through the home-page and the account creation procedure. This is an essential task for user in real-life to get even started with the Council of Coaches system. Relevant aspects of this task which we aim to get more information about are listed in Table 6.

Table 6: Relevant aspects of task one.

Category	Task 1: Relevant aspects
Interaction	<ul style="list-style-type: none"> Is it easy for the participants to create an account? Do the participants understand that they have to enter their personal information to create an account? Do the participants get that they cannot click on the description itself? Do the participants understand the description added by arrows?
Content	<ul style="list-style-type: none"> Do the participants understand the meaning of the account? Do the participants get that they have to enter the provided email-address? Do the participants like a robot (Coda)? Are the participants content with a robot?
Design	<ul style="list-style-type: none"> How do participants judge the design of the home screen? Are the participants able to read the font of the descriptions added by arrows? Do the participants notice the description?

○ **Task 2: Free exploration of multi-coach interaction in living room**

Participants are asked to “freely explore the options and functionalities of the system within 10 minutes”. During this phase the participant will be introduced to the coaches, interacts with coaches of their choice, go through coach content of several coaches and might interact with some of the new widgets in the demonstrator (radio, clock, activity book). To evaluate real time use it is very insightful to let the participants freely explore the Council of Coaches system from the first moment they entered the living room and observe which route they take to get to know the system. Relevant aspects of this task which we aim to get more information about are listed in Table 7.

Table 7: Relevant aspects of task two.

Category	Task 2: Relevant aspects
Interaction	<ul style="list-style-type: none"> Do the participants understand how to get to the different attributes (radio, activity book, etc.)? Does participant understand how to start a dialogue with one of the coaches? Do the participants get how to stop a dialogue? Do the participants get how to go back?
Content	<ul style="list-style-type: none"> Do participants like to give personal information? Do the participants like the introduction dialogues? Does the participant have coach preferences? Do the participant get that Coda is in the room to assist if they need help?
Design	<ul style="list-style-type: none"> Do the participants like the design of the living room? Do the participants miss anything in the current version? What would they like to see? Are the participants able to read the dialogue boxes?

○ **Task 3: Connect the Fitbit account to the Council of coaches.**

Participants are asked to “Connect the Fitbit activity tracker account with the COUCH account”. During this task they have to interact with the activity coach, Olivia, to get to the coaching content regarding activity trackers. During this dialogue Olivia will explain the purpose of monitoring activity with an activity tracker and suggest to connect a Fitbit device. After a

complete explanation on how to connect the account, the participant is automatically sent to the Fitbit website to log in (with provided account information) and allow COUCH access to the activity data. This task could be seen as quite a technical task for the target population and therefore very relevant to evaluate, as it is commonly known that use-errors can lead to drop outs in technology use. Furthermore, within the Council of Coaches activity coaching is an important domain. Therefore, participants will be encouraged to use and connect activity monitors so that the COUCH system can provide specific and personal activity coaching. Relevant aspects of this task which we aim to get more information about are listed in Table 8.

Table 8: Relevant aspects of task three.

Category	Task 3: Relevant aspects
Interaction	<ul style="list-style-type: none"> Is it easy for the participants to connect the Fitbit? Do the participants understand that they have to enter their Fitbit account to connect this account? Do the participants get that they will be send to the Fitbit webpage to log-in and allow access? Do the participants understand to go to the activity coach to connect the Fitbit?
Content	<ul style="list-style-type: none"> Do the participants mind the English on the Fitbit page? Do the participants understand the explanation of the connection steps? Do the participants understand why to connect an activity tracker?
Design	

○ **Task 4: View the activity book and retrieve how many steps were performed on Saturday.**

Participants are asked to “Find out how many steps were performed last Saturday”. To achieve this, they have to speak to Olivia, the activity coach, again. Now the system is connected to an activity tracker in task 3, the participant can go to “My Progress” under the central coaching content dialogue of Olivia to open the activity book. The activity book is one of the key features of the activity coach to review the personal activity data and the use for personalized coaching strategies and goal setting, indicating its importance to evaluate. Relevant aspects of this task which we aim to get more information about are listed in Table 9.

Table 9: Relevant aspects of task four.

Category	Task 4: Relevant aspects
Interaction	<ul style="list-style-type: none"> Do the participant get how to get to the activity book?
Content	<ul style="list-style-type: none"> Do the participants understand which data is presented in the activity book? Do the participants miss specific information?
Design	<ul style="list-style-type: none"> Do the participants understand the illustrated data/graphics? Do the participants like the design of the activity book? Are the participants able to read the text? (font, font size).

○ **Task 5: A personalised dinner recipe.**

Participants are asked to “Receive your personalized diner recipe”. During this task they have to interact with Francois, the nutrition coach. Within his coaching dialogue he will take the participant through a food intake conversation about the participant’s meal preferences. Afterwards, there is the option to discuss a recipe with Francois after he confirmed some of the food preferences. The basis of nutrition coaching lies managing a healthy balanced diet.

Therefore, next to informing about healthy diet, providing personalized recipes is a direct way of coaching the participant and therefore a valuable aspect of the Council of Coaches system. Relevant aspects of this task which we aim to get more information about are listed in Table 10.

Table 10: Relevant aspects of task five.

Category	Task 5: Relevant aspects
Interaction	<ul style="list-style-type: none"> Is it easy for the participants to create an account? Do the participants understand that they have to enter their personal information to create an account? Do the participants get that they cannot click on the description itself? Do the participants understand the description added by arrows?
Content	<ul style="list-style-type: none"> Do the participants understand the meaning of the account? Do the participants get that they have to enter the provided email-address? Do the participants like a robot (Coda)? Are the participants content with a robot?
Design	<ul style="list-style-type: none"> How do participants judge the design of the home-screen? Are the participants able to read the font of the descriptions added by arrows? Do the participants notice the description?

4.1.4.1 After scenario questionnaire

After each task the participant is asked to fill in the After Scenario Questionnaire (ASQ) (Lewis, 1991) (Table 11). This questionnaire features three questions, post-task. The ASQ is commonly used, and research has supported that it “has acceptable psychometric properties of reliability, sensitivity, and concurrent validity, and may be used with confidence in other, similar usability studies.”

Table 11: After Scenario Questionnaire.

Item	Answer options	
	Strongly disagree	Strongly agree
1. I am satisfied with the ease of completing the task	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
	1 2 3 4 5 6 7	
2. I am satisfied with the amount of time it took to complete the task	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
	1 2 3 4 5 6 7	
3. I am satisfied with the support information (online help, messages, documentation) when completing the task	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
	1 2 3 4 5 6 7	

4.1.5 Preparation of tablet/iPad screen capture

4.1.5.1 Tablet screen capture (Screen Recorder: AppSmartz)

A screen capture video of the user interacting with the system will be made on a tablet, which records the screen, the user's screen touches, and audio through a microphone. We use the “Screen Recorder” app, which can be found in the Google Play Store (Figure 12).

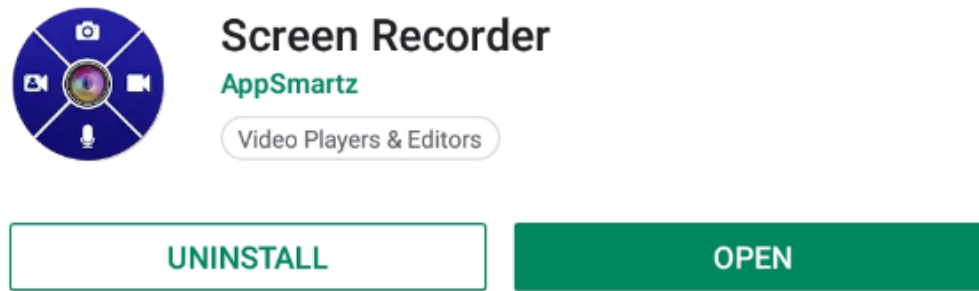


Figure 12: Screen Recorder app in Google Play Store.

After installing the application, the android device has to be set-up to “Show touches”. This can be achieved by performing the following actions:

- 1) Open Settings on the Android device.
- 2) Scroll to the very bottom of the menu and select ‘About phone’.
- 3) Scroll to the very bottom of the menu and select ‘Build number’ repeatedly. After the 7th or 8th click you should see a message telling you that you are a developer.
- 4) Click back to return to the main Settings menu.
- 5) There should be a new ‘Developer Options’ option above to ‘About phone’ option. Select ‘Developer Options’.
- 6) Under the ‘Input’ heading there is a ‘Show touches’ option. Selecting this will show all touch events on the screen (Figure 13).

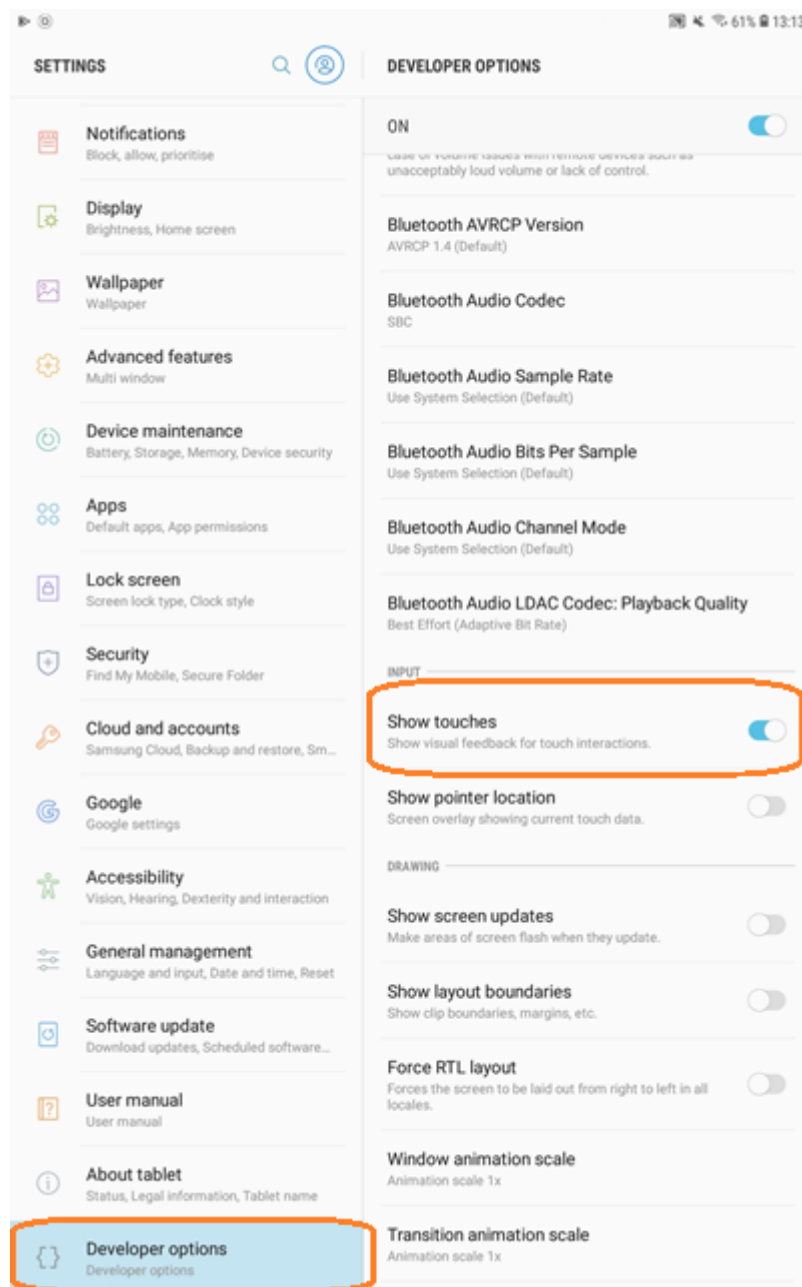


Figure 13: Settings android device - show touches.

The first time the Screen Capture app is opened accept all permissions and go to the settings. Use the settings as displayed in Figure 14 (hide recording controls>on / lock to stop recording>off / watermark>off / record audio>on). The screen capture recording can now be started with the control panel on the screen, use the top video icon to start recording, Figure 15 (now the recording control will disappear and when the countdown is finished the recording starts). To stop the recording swipe down the notification tray of your android device and click on the screen capture notification.

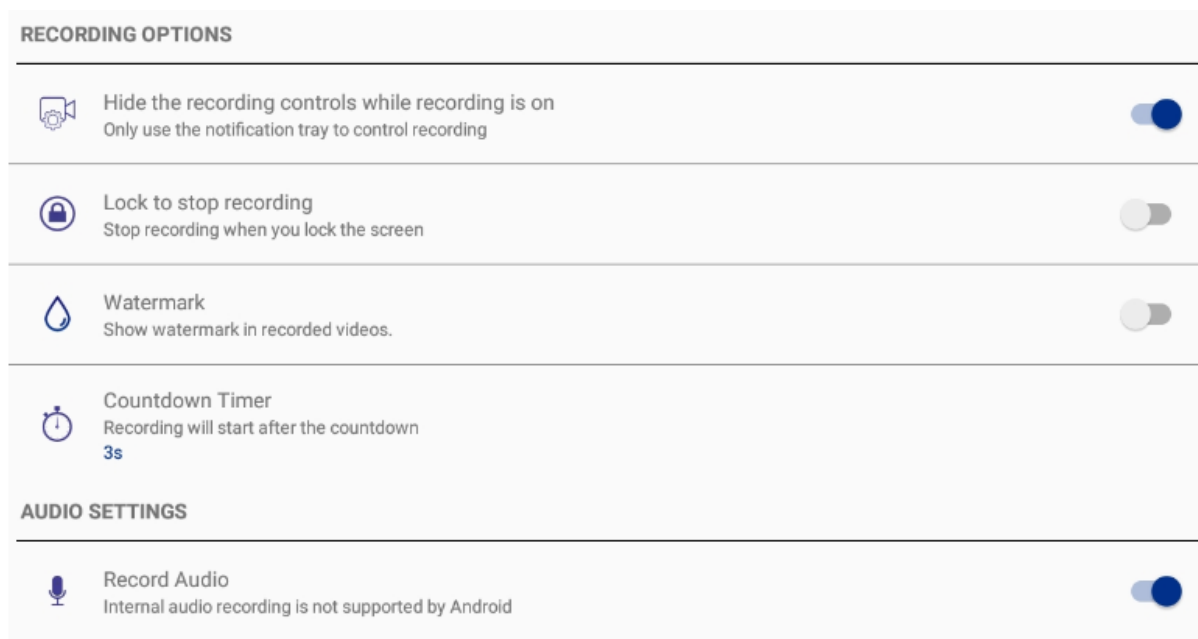


Figure 14: Settings of the screen capture app.

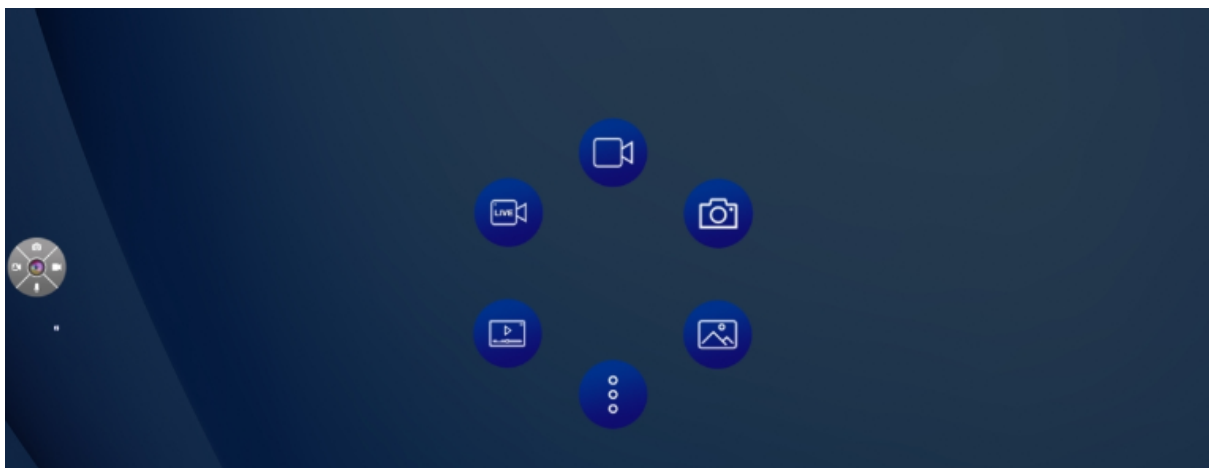


Figure 15: Screen Capture app recording control. Use top button to record screen capture video.

4.1.5.2 iPad screen capture (Build in iPad app)

Apple iPads have a build in screen capture function. To activate this option;

- 1) Go to the iPad settings > Control Center > Customize Controls.
- 2) Include the Screen Recording by tapping the plus sign (Figure 16).
- 3) Swipe down from the right upper corner of the screen to open the control panel.
- 4) Press deeply on the screen recording Icon (Figure 17).
- 5) Tap on the microphone to activate simultaneous sound recording

The screen recorder icon can now be found within the control panel and is ready to record both visual and audio output of the iPad. Unfortunately, the iPad has no “show touches” option. To start a screen recording just tap on the screen recording icon in the control panel. Tap on the red status bar to stop the screen recording. The recorded video can be found in the Photo’s app.

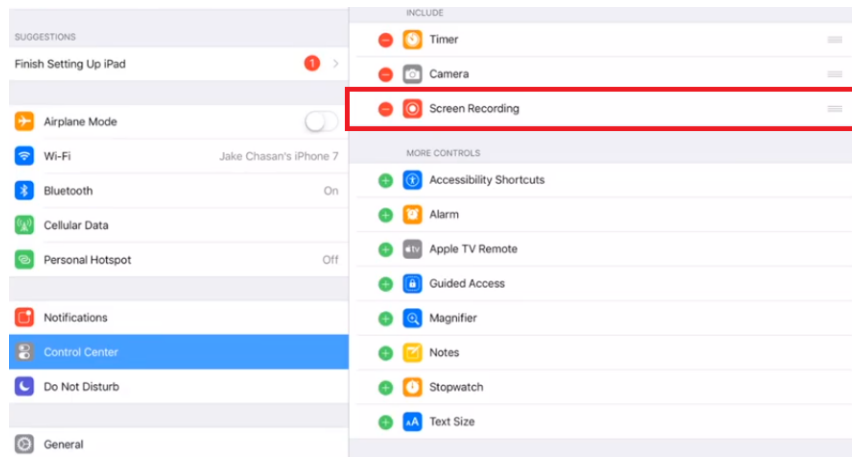


Figure 16: Control centre of iPad settings to include Screen Recorder functionality.

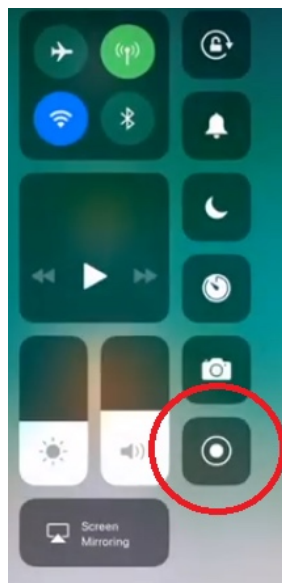


Figure 17: iPad Screen Recorder icon.

4.1.6 Revoke Access to RRD activity coach on Fitbit.

To be able to re-use the same Fitbit account for the purpose of evaluation three, the access to RRD activity coach has to be revoked after each session. The description on how to revoke this is described below;

- 1) Go to fitbit.com > (log in with couch01@activity-coach.com and password: couch01! If you're not logged in automatically) > Go to Dashboard (top left) > Go to settings (top right) (Figure 18: Fitbit page.).

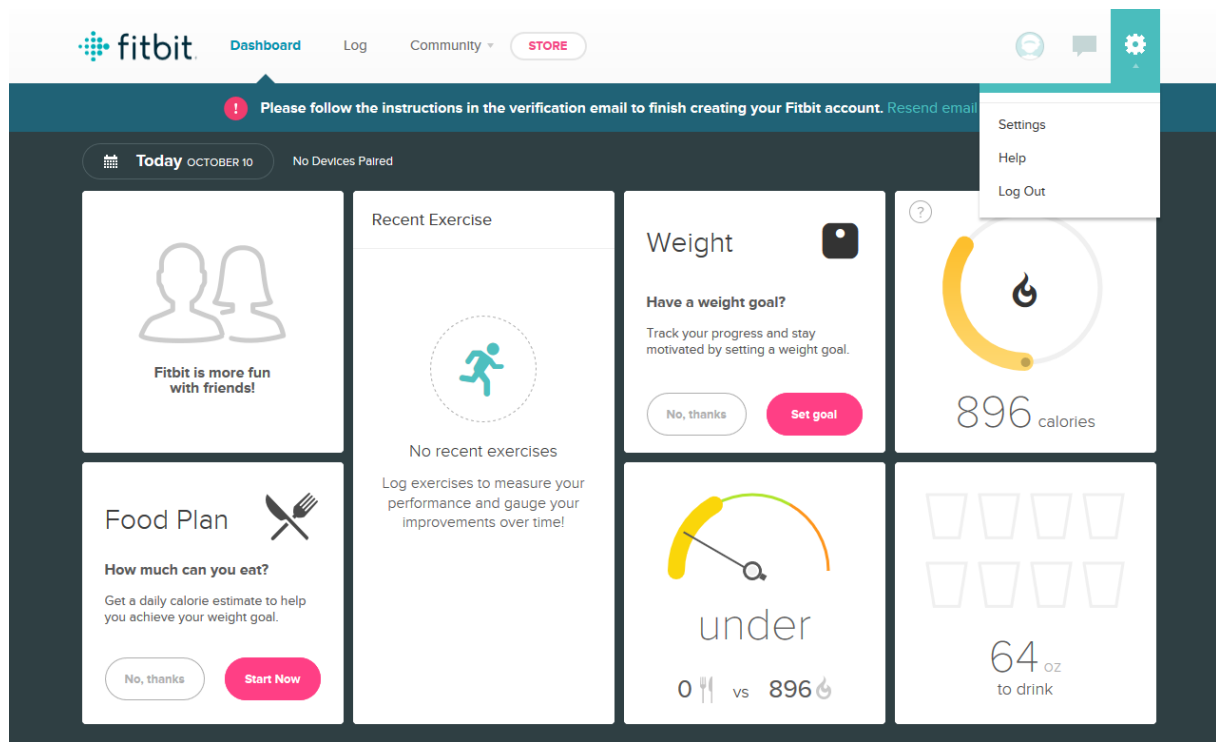


Figure 18: Fitbit page.

2) Click on Applications (left menu, second on the bottom) (Figure 19).

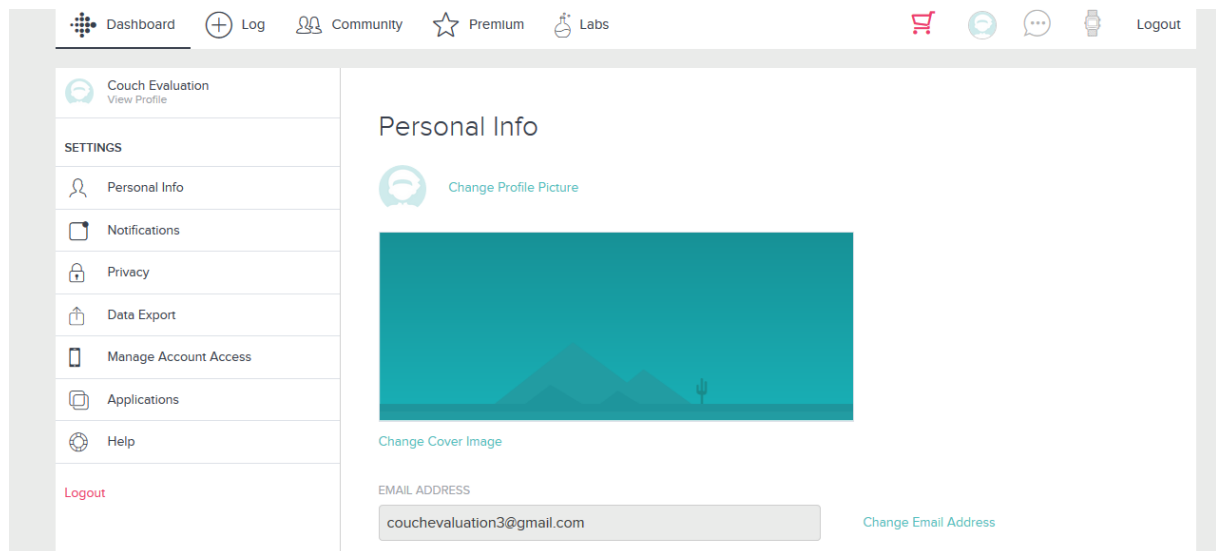


Figure 19: Fitbit settings page.

3) Click on the pink "Revoke Access" > Confirm > and log out afterward (top right) (Figure 20).

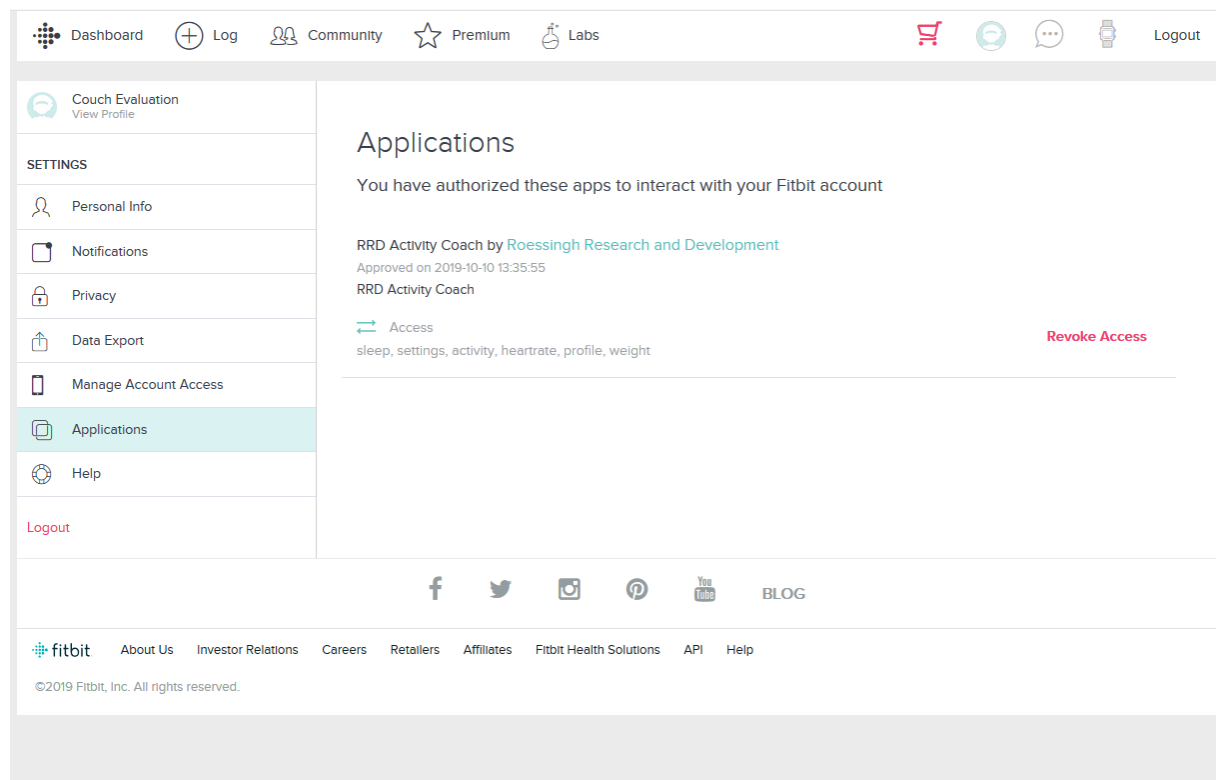


Figure 20: Authorized applications page Fitbit.

4.1.7 Interview guidebook

The interview guidebook (Table 12) gives guidance for the moderator. We will use the same protocol as used in deliverable 2.5 (Beinema, et al., 2019), but now focused on several specific tasks within the system, which are expected to be commonly used in real-life use. During the interview the moderator does not intervene, unless the participant remains quiet for a specified amount of time. The moderator will remind the participant to verbalize his thoughts. After each task the moderator evaluates all elements of the task with open non-directive questions. When the participant indicates they have covered all the issues for this task the moderator can proceed to the next task. After evaluating all tasks, the participant will be asked whether he or she has something else to mention on using the Council of Coaches system.

Table 12: Interview guidebook.

Part		Text
Introduction		You just finished using the Council of Coaches system, during which we recorded your actions. Next, we would like watch these recordings together and ask you to comment on your actions during this re-watch. After each task, we will have a short interview.
Materials		<ul style="list-style-type: none"> Turn on audio recorder
Task 1	<i>During rewatch</i>	<ul style="list-style-type: none"> User comments on his/her actions Moderator asks participant to comment if participant is quiet
	<i>After rewatch</i>	<ul style="list-style-type: none"> What went well? What went wrong? Could you describe your thoughts when you were performing these actions? Could you describe your feelings when you were performing these actions? Did you understand the information provided?
Task 2	<i>During rewatch</i>	<ul style="list-style-type: none"> User comments on his/her actions Moderator asks participant to comment if participant is quiet
	<i>After rewatch</i>	<ul style="list-style-type: none"> What went well? What went wrong? Could you describe your thoughts when you were performing these actions? Could you describe your feelings when you were performing these actions?

		<ul style="list-style-type: none"> ▪ Did you understand the information provided?
Task 3	<i>During rewatch</i>	<ul style="list-style-type: none"> ▪ User comments on his/her actions ▪ Moderator asks participant to comment if participant is quiet
	<i>After rewatch</i>	<ul style="list-style-type: none"> ▪ What went well? ▪ What went wrong? ▪ Could you describe your thoughts when you were performing these actions? ▪ Could you describe your feelings when you were performing these actions? ▪ Did you understand the information provided?
Task 4	<i>During rewatch</i>	<ul style="list-style-type: none"> ▪ User comments on his/her actions ▪ Moderator asks participant to comment if participant is quiet
	<i>After rewatch</i>	<ul style="list-style-type: none"> ▪ What went well? ▪ What went wrong? ▪ Could you describe your thoughts when you were performing these actions? ▪ Could you describe your feelings when you were performing these actions? ▪ Did you understand the information provided?
Task 5	<i>During rewatch</i>	<ul style="list-style-type: none"> ▪ User comments on his/her actions ▪ Moderator asks participant to comment if participant is quiet
	<i>After rewatch</i>	<ul style="list-style-type: none"> ▪ What went well? ▪ What went wrong? ▪ Could you describe your thoughts when you were performing these actions? ▪ Could you describe your feelings when you were performing these actions? ▪ Did you understand the information provided?
Closing	<i>After rewatch</i>	<ul style="list-style-type: none"> ▪ Is there anything else you want to mention about the Council of Coaches system that you've tested today?

4.1.8 Questionnaire of quantitative assessment of usability and user acceptance

The usability of the third demonstrator is assessed with performance metrics and the System Usability Scale (Brooke, 1996). The SUS presented ten statements about the perceived usability of the application. Participants could indicate on a scale from 0 to 4 to what extent the presented statements were true for them. The SUS score ranges from 0 to 100 (low and high usability, respectively), for more details about the interpretation of SUS score see Section 4.2.3. We will use a Dutch and Danish translation of the original English version of the SUS scale, as no validated translated version is available.

Table 13: The System Usability Scale (SUS).

SUS		Strongly disagree				Strongly agree
1	I think that I would like to use this system frequently	1	2	3	4	5
2	I found the system unnecessarily complex	1	2	3	4	5
3	I thought the system was easy to use	1	2	3	4	5
4	I think that I would need the support of a technical person to be able to use this system	1	2	3	4	5
5	I found the various functions in this system were well integrated	1	2	3	4	5
6	I thought there was too much inconsistency in this system	1	2	3	4	5
7	I would imagine that most people would learn to use this system very quickly	1	2	3	4	5
8	I found the system very cumbersome to use	1	2	3	4	5
9	I felt very confident using the System	1	2	3	4	5
10	I needed to learn a lot of things before I could get going with this system	1	2	3	4	5

End-user experience of the third COUCH demonstrator is assessed by means of a questionnaire with summated rating scales (on a 7-point scale), based upon the Technology Acceptance Model (TAM) (Davis, Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology, 1989). TAM originates from the 1980s and has been used numerous times to assess and explain the acceptance of new technology. We expanded TAM with factors that have been found to shape the user experience of mHealth technology: enjoyment (Crutzen, et al., 2011), aesthetics (Baumel & Muench, 2016), control (Jameson, 2007), trust in the technology (van Velsen L., van der Geest, van de Wijngaert, van den Berg, & Steehouder, 2015) and willingness to pay. We hypothesize that these factors affect the core factors of TAM that explain the intention to use (perceived usefulness and ease of use). We will use a Dutch and Danish translation of the original English version of the TAM, as no validated translated version is available.

Table 14: The Technology Acceptance Model (TAM).

Original item		Source	COUCH app item (7point scale)
Enjoyment*: “The extent to which the activity of using the computer is perceived to be enjoyable in its own right, apart from any performance consequences that may be anticipated” (Davis, Bagozzi, & Warshaw, Extrinsic and Intrinsic Motivation to Use Computers in the Workplace, 1992)			
1	[XXX] was [enjoyable – disgusting]	(van der Heijden, 2004)	The COUCH app was [enjoyable – disgusting]
2	[XXX] was [exciting – dull]		The COUCH app was [exciting – dull]
3	[XXX] was [pleasant – unpleasant]		The COUCH app was [pleasant – unpleasant]
4	[XXX] was [interesting – boring]		The COUCH app was [interesting – boring]
Aesthetics: “Visual beauty or the study of natural and pleasing (or aesthetic) computer-based environments” (Jennings, 2000)			

1	[XXX] looks clean	(Lavie & Tractinsky, 2004)	The COUCH app looks clean
2	[XXX] looks clear		The COUCH app looks clear
3	[XXX] looks pleasant		The COUCH app looks pleasant
4	[XXX] looks symmetrical		The COUCH app looks well balanced
5	[XXX] looks aesthetic		The COUCH app looks pretty
6	[XXX] looks original		The COUCH app looks original
7	[XXX] looks sophisticated		The COUCH app looks sophisticated
8	[XXX] looks fascinating		The COUCH app looks fascinating
9	[XXX] looks creative		The COUCH app looks creative
10	[XXX] uses special effects		The COUCH app uses special effects
Control: “The belief that the user can choose to bring about or prevent particular actions or states of the system” (Jameson, 2007)			
1	I have a lot of control over what I can do on [XXX]	(van Velsen L. , van der Geest, van de Wijngaert, van den Berg, & Steehouder, 2015)	I have a lot control over what I can do on the COUCH app
2	On [XXX] you can choose freely what you want to see		On the COUCH app you can choose freely what you want to see
3	I can determine for myself what happens on [XXX]		I can determine for myself what happens on the COUCH app
Trust in technology: “The belief that a technology has protective legal or technological structures (e.g., encryption) that assure that business can be conducted in a safe and secure manner” (McKnight, Choudhury, & Kacmar, 2002)			
1	The security on [XXX] does not set my mind at rest	(McKnight, Choudhury, & Kacmar, 2002)	The security on the COUCH app does not set my mind at rest
2	The law and security technology protect me well against problems with [XXX]		The law and security technology protect me well against problems with the COUCH app
3	Your personal data are protected well when you use [XXX]		Your personal data are protected well when you use the COUCH app
4	[XXX] is not safe		The COUCH app is not safe
Perceived usefulness			
1	Using the [XXX] helps me to understand my physical condition		Using the COUCH app helps me to understand my physical condition
2	Using the [XXX] improves my physical condition		Using the COUCH app improves my physical condition
Ease of use			
1	My interaction with [XXX] is clear and understandable	(Venkatesh & Davis, 2000)	It is clear and understandable how I can work with the COUCH app

2	Interacting with [XXX] does not require a lot of my mental effort		I do not have to think hard when working with the COUCH app
3	I find [XXX] to be easy to use		I find the COUCH app easy to use
4	I find it easy to get [XXX] to do what I want it to do		I find it easy to get the COUCH app to do what I want it to do
Intention to use: "A person's intention to use a technology once it is available to him or her"			
1	If [XXX] would be available for me, I would definitely use it	(van Velsen L. , van der Geest, van de Wijngaert, van den Berg, & Steehouder, 2015)	If the COUCH app would be available for me, I would definitely use is
2	I would recommend [XXX] to others		I would recommend the COUCH app to others
3	I hope that [XXX] becomes available to me		I hope that the COUCH app becomes available to me
Willingness to pay: "A person's intention to use a technology once it is available to him or her"			
1	I would pay for [XXX] if it becomes available to me.		I would pay for [XXX] if it becomes available to me.
2	I would spend a maximum of euro per month to be able to use the COUCH.		

4.2 Data analysis

4.2.1 Usability issues

Usability issues will be identified using the following process:

1. One researcher identifies all errors in the think-aloud transcripts and observational notes;
2. A second researcher also examines this dataset. Discrepancies will be solved and the first researcher re-analyses the full data set with this final list.
3. The first researcher creates an overview of usability issues by grouping similar errors into one usability issue (e.g., recurring errors from clicking on non-clickable elements were grouped as 'the user has difficulty distinguishing clickable from non-clickable elements in the interface').
4. The second researcher examines this usability issue overview.
5. The researchers discuss discrepancies and created a final overview;
6. The first researcher awarded each usability issue with a severity score (minor, serious, or critical), following a procedure from (Duh, Tan, & Chen, 2006)
 - a. Minor issue: Occurred infrequently among the participants and/or the problem only increased task completion time slightly;
 - b. Serious issue: Occurred frequently among the participants and/or the problem severely increased task completion time;
 - c. Critical issue: Occurred when all participants had the same problem and/or the problem prevented participants from completing tasks
7. The severity ratings were verified by the second researcher.
8. The usability issues were converted into categorical code (0 = negative, 1 = positive) to allow for statistical analyses.

4.2.2 Task metrics

Task metrics were used to evaluate task performance. All five tasks will be evaluated on;

- Task completion (yes/no)
- Time to complete task (min)
- Task satisfaction (After Scenario Questionnaire (ASQ))

- Number of steps per task
- Number of errors per task

The time to complete the task and the number of steps to perform a task were averaged over all participants that did succeed in performing the task. The other metrics were averaged over all participants.

All metrics will be scaled from 0-100% and visualized in a radar chart. All parameters are scaled from 0 to 1, with 0 being the centre and 1 the outline of the radar chart; Task completion (%), Completion time (1-seconds to complete task/300), Task satisfaction (ASQ score/7), Steps (1/(Steps done/Steps needed)) and Errors (1/(1+amount if errors/2)). A completely filled plot therefore indicates best possible performance.

4.2.3 Questionnaire results

System usability score (SUS)

To obtain the final SUS score, the following calculations have to be made;

- For every odd-numbered question, subtract 1 from the score (X-1),
- For every even-numbered question, subtract the score from 5 (5-X),
- Sum the just calculated scores from all 10 questions,
- Multiply the total with 2.5.

This results in a score between 0 and 100 (low and high usability, respectively). The scores can be interpreted according to general guideline below (Table 15). We will use descriptive statistics to explore the SUS scores of the participant and evaluate the net promotor score (NPS).

Table 15: The System Usability Scale (SUS) scoring table. Expressed in grades, percentile range, adjectives, acceptability and net promotor scale (NPS) (Sauro, 2016) (Tullis & Albert, 2008).

Grade	SUS	Percentile range	Adjective	Acceptable	NPS
A+	84.1-100	96-100	Best Imaginable	Acceptable	Promoter
A	80.8-84.0	90-95	Excellent	Acceptable	Promoter
A-	78.9-80.7	85-89		Acceptable	Promoter
B+	77.2-78.8	80-84		Acceptable	Promoter
B	74.1 – 77.1	70 – 79		Acceptable	Passive
B-	72.6 – 74.0	65 – 69		Acceptable	Passive
C+	71.1 – 72.5	60 – 64	Good	Acceptable	Passive
C	65.0 – 71.0	41 – 59		Marginal	Passive
C-	62.7 – 64.9	35 – 40		Marginal	Passive
D	51.7 – 62.6	15 – 34	OK	Marginal	Detractor
F	25.1 – 51.6	2– 14	Poor	Not Acceptable	Detractor
F	0-25	0-1.9	Worst Imaginable	Not Acceptable	Detractor

Technology acceptance model (TAM)

The technology acceptance will be scored in the following categories; enjoyment, aesthetics, control, trust, perceived usefulness, ease of use and intention to use. The scores of the accompanying questions (Table 14) will be averaged per category and scaled to a 0-100% scale.

4.3 Methodology Denmark

4.3.1 Recruitment of test participants

Due to the low population of the test (10 test subjects per site), we decided on a qualitative approach to our test. Due to the selection criteria of the participants (Diabetes/Chronic Pain patients OR older adults of at least 55 years old) we also needed to test the test design, as we do not have in house experience with software tests on older adults. Additionally, we wished to test the software on participants without prior knowledge in order to pass 'The Mom Test'¹. Specifically, we needed to explore whether the usability issues with IT hardware and software would cloud the content testing.

Several activity centres for senior citizens were contacted in order to reach a broad selection of older adults/seniors. An initial test of the test design was set up with the General Manager of one of these centers, as he had expert knowledge of this age group and was 55 years of age himself. He went through the test as described in 4.1.3. The major feedback from this test contained the following:

- The seniors at the centre will not be not comfortable with tablets, especially as the text on screen got very small and hard to read.
- He experienced difficulties with creating a user. As the seniors generally are not comfortable with new software, he predicted that minor difficulties like that would put off a lot of the seniors from the beginning.
- The test was too long, especially for the elder seniors (the average age of the seniors at the center is 75).
- We would need to define much more precisely, what the purpose of the software was, and at the same time what the purpose of the test was.
- Task 3 would be very hard, as most of the seniors would need a much more thorough introduction to online activity trackers.
- He could recommend a handful of seniors who would ace the usability part of the test, but they would be far from representative for the age group.

4.3.2 Changes to the test design

Based on the feedback from the General Manager, the test design changed to a series of focus groups instead of individual tests, and to do the tests on computers instead of tablets. This allowed us to:

- Heighten the possibility of getting participants to sign up by making the test a social event.
- Minimize the initial usability issues the tablets could have caused.
- Minimize the uncertainty of using unknown software by grouping the test subjects, so that they could support one another.
- Minimizing the 'Mom Factor' as the participants talked about the product to each other instead of just talking to us.
- Focus on getting feedback on the content and overall impression of the software, instead of focusing on process.

Due to the changes in the evaluation design it was not possible to:

- Get screen captures of the user interactions.
- Comply with the retrospective think-aloud design, as the participants commented the tasks while doing them. But being a group to solve the task might have helped with the sensory overload of thinking aloud while doing it.

¹ Never ask your mom whether your project is a good idea. She will always say yes. Get someone who do not know you to test your product without them knowing it is yours – then you get an honest opinion to further develop your product on. *The Mom Test*, Rob Fitzgerald, <http://momtestbook.com/>

4.3.3 The setup of the focus group

The test was framed as a 1-hour workshop, in which the participants could try out the Council of Coaches. There were 2 workshops at the activity centre for seniors where we tested the design with the General Manager, and 1 workshop at another activity centre for seniors. In total 12 participant did participate (see 0). The setup of an individual session went as follows:

- Each workshop started with a short presentation of ourselves and the project. We presented ourselves as impartial test planners, minimizing the fear of seeming impolite when criticizing.
- After the introduction, the participants were divided into groups of 1-3 participants, allowing them to choose groups if they so wished. Each group got a computer, and each participant was asked to fill out a consent form.
- We then started the tasks, allowing each group to continue with the next task when finished with the previous one. In between each task we asked each group to comment on the task. During the first task, we collected the baseline questions from each participant.
- Each group was sound-recorded while doing the tasks.
- After 50 minutes we stopped the test, collected comments on the task at hand and asked the participants to fill out the SUS and TAM questionnaires before finishing the workshop.

4.3.4 The focus group session layout.

The different focus groups are called sessions, and below in Table 16 the layout of participants in each session is presented.

Table 16: Focus group session layout.

Session 1	Session 2	Session 3
<ul style="list-style-type: none">▪ Test user 1▪ Test user 2	<ul style="list-style-type: none">▪ Test user 3▪ Test user 4▪ Test user 5▪ Test user 7▪ Test user 8	<ul style="list-style-type: none">▪ Test user 6▪ Test user 9▪ Test user 10▪ Test user 11▪ Test user 12

5 Results

5.1 Demographics

A total of 30 participants evaluated the third functional demonstrator. Each research site contributed a fair share of participants, ten in Enschede, NL; twelve participants in Copenhagen, DK; eight participants in Dundee, UK. Table 17 shows an overview of the demographics of all participants. Most participants were men (57%) and the average age was 69 years. 33% of the participants had diabetes type 2 and 23% of the participants answered to have chronic pain. Health literacy was higher in Scotland (3.7) compared to the Dutch participants (2.9) ($p < 0.01$). From the participants 83% uses a computer/laptop, 60% a tablet, 80% a smartphone and 3% uses a game console at home, which indicates that the technology use within this population was quite high. Some participants reported physical impairments that might have influenced their performances during this evaluation such as; dry fingers (P03), slow understanding due to previous brain infarction (P05), reduced eyesight (P08) and pain in the hands due to neuropathy (P23/P24).

Table 17: Demographics of the 30 evaluation participants.

Country	ID	Gender	Age	Diabetes mellitus type 2	Chronic pain	Health literacy	Education
The Netherlands	01	M	64	N	N	3.33	Higher vocational education
	02	M	60	N	Y	2.00	Higher vocational education
	03	M	57	N	N	2.67	Higher vocational education
	04	F	61	Y	Y	3.00	Lower vocational education
	05	M	73	N	Y	2.33	Vocational education
	06	M	71	N	N	3.33	Higher vocational education
	07	M	67	N	N	3.00	Higher vocational education
	08	M	78	N	N	3.00	Lower vocational education
	09	F	62	N	N	3.33	Vocational education
	10	M	79	N	N	3.00	Lower vocational education
Denmark	11	M	78	N	N	-	Vocational education
	12	M	68	N	Y	-	Higher vocational education
	13	M	75	N	N	-	Vocational education
	14	F	64	Y	N	-	Vocational education
	15	F	87	Y	N	-	Vocational education
	16	F	74	N	N	-	Vocational education
	17	F	87	N	N	-	Basic

	18	M	82	N	N	-	Vocational education
	19	F	70	N	N	-	Vocational education
	20	F	68	N	Y	-	Basic
	21	F	79	N	N	-	Basic
	22	F	81	N	Y	-	Vocational education
Scotland	23	M	71	Y	Y	3.33	Vocational education
	24	F	63	Y	N	4.00	Higher vocational education
	25	M	56	Y	N	2.67	Vocational education
	26	F	51	Y	N	3.67	Vocational education
	27	M	64	Y	N	4.00	Lower vocational education
	28	M	56	Y	N	4.00	Higher vocational education
	29	M	63	N	N	3.67	Higher vocational education
	30	F	63	Y	N	4.00	Higher vocational education

5.2 Usability issues

A total of 395 usability issues were found across all participants. Deduplication of usability issues that occurred among two or more participants, resulted in a total of 247 unique usability issues: Of these, 77 were positive usability points and 170 negative usability issues (139 minor, 23 serious and 8 critical issues). Since the majority of these issues are minor, meaning they only slightly hindered task completion time or only occurred to a few participants, the usability of the third version of the Council of Coaches system can be considered quite good.

Serous and critical issues have a much larger impact on task completion and task completion time. Serious issues severely hamper task completion and critical issues prevent a user from completing tasks (Duh, Tan, & Chen, 2006). Inspection of the usability issues per task reveal that task 2 yielded most unique usability issues (14 positive and 27 negative). Most serious and critical issues were found in task 3 and 1, which reflects the task completion percentages seen in Section 5.3. Task 5 revealed a relatively high amount (14) of positive points against just 1 serious and 1 critical issue. Figure 21 shows an overview of the unique usability issues per task.

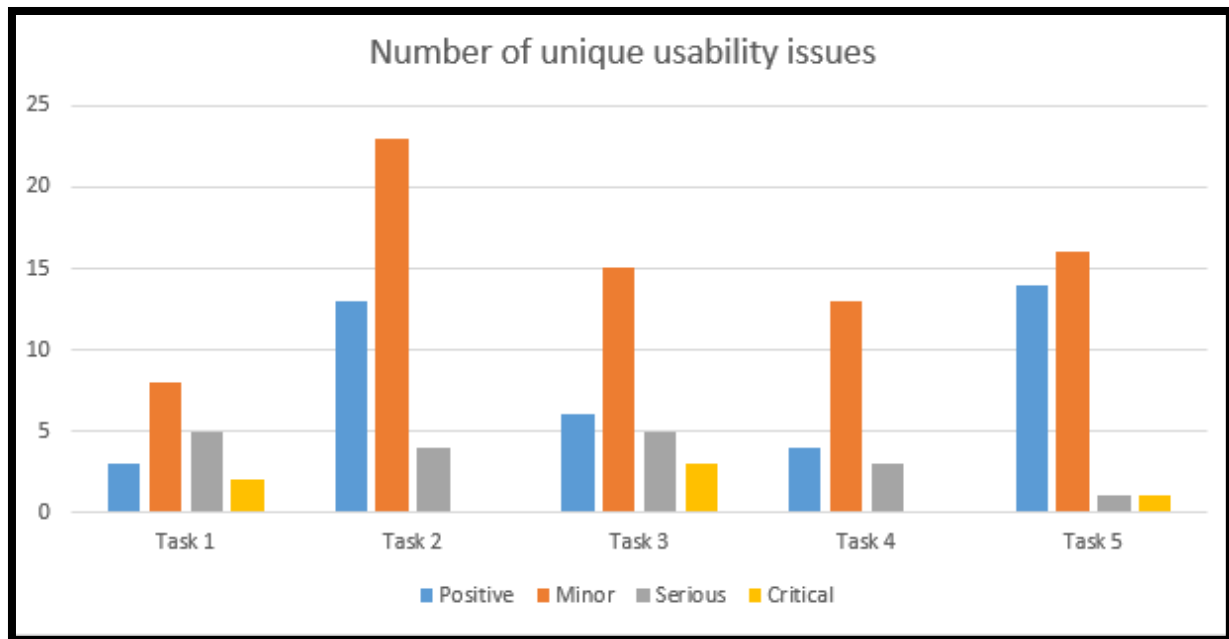


Figure 21: Overview of the number of unique usability issues per task.

The usability issues were then categorized using the framework of van der Geest (2004), who distinguishes four categories: (1) Navigation & Structure, (2) Content & Information, (3) Design & Presentation and (4) Other. Figure 22 provides an overview of the number of both positive and negative unique usability issues found per category. It shows that the most critical issues were found within the category Navigation & Structure (6 critical issues). The other 2 critical issues were found within the Design & Presentation category. No critical issues were found in the categories Content and Other. In the next sections, each category will be discussed in more detail.

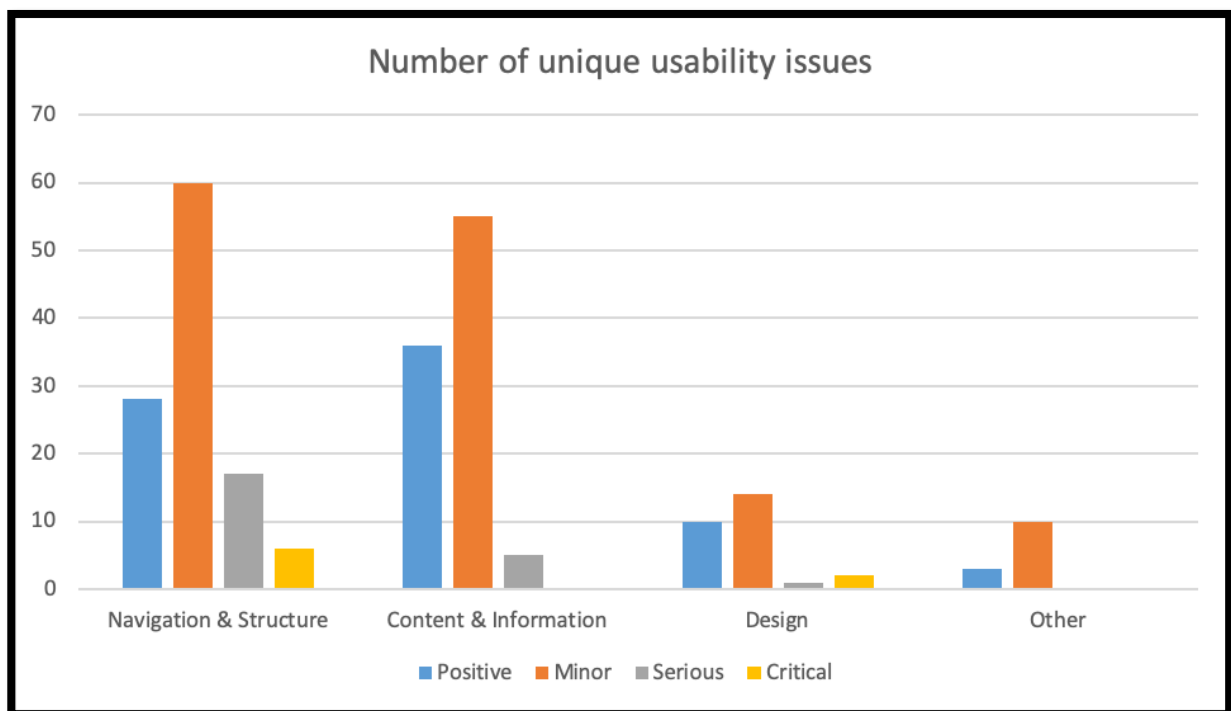


Figure 22: Overview of the number of unique usability issues per category.

5.2.1 Navigation & structure

A total of 111 unique usability issues were identified in the Navigation & Structure category. Of all these issues, there were 28 positive points and 83 negative issues (of which 60 were minor issues, 17 serious, and 6 critical).

Examples of these issues that were found, can be viewed in Table 18. In general, participants were quite positive regarding the navigation during the account creation and baseline questions of Coda. However, within the interaction with the coaches and execution of the coach specific tasks (3, 4 and 5) a relatively high amount of serious and critical issues regarding navigation & structure were identified.

Many participants struggled to connect the Fitbit (which is confirmed with the task completion of 17% in section 0). First of all, it was not clear for eight participants to talk to Olivia to get the activity tracker connected. Some thought they first had to logout from COUCH or that Coda could help, as he is a robot with technical knowhow. Another issue within the Fitbit connection process were the system errors during the actual connection itself. Six users were redirected to the Council of Coaches home-page and asked to login again. In some other tests the users were redirected to the living room, however the connection was not successfully made.

Another main issue which really impeded the execution of the tasks was a slow response of the system to a click of the user, when a medium-spec tablet was used. Therefore, most of the following test were performed with a high-spec iPad, reducing its impact drastically.

The interaction of participant 23 revealed a system error during the state of change questions within the introduction conversations with coda; “the app stopped working”.

The last critical issue was found during the recipe conversation of Francois. Users were confused when the dialogue ends after the food intake instead of providing the recipe, resulting in different strategies to proceed. 1) Go on in the conversation with Francois and ask again for the recipe, and eventually (after confirming food preferences) get the recipe, 2) Go on with the conversation with Francois, but try out all other coaching content dialogue, as they already walked through the recipe dialogue. 3) Close the conversation with Francois and check out other options, resulting in failure to achieve the task within 5 minutes.

A striking serious issue during the free-exploring phase, where participants were introduced to the main room of the COUCH system for the first time, was that a lot of users indicated that after the introduction dialogues it was not clear for them to click on the coaches to interact.

Regarding the overall system interaction and navigation, participants furthermore mentioned they sometimes would like to use a return button to go back in the conversation to alter a given answer.

Next to these serious and critical issues participants also provided a lot of positive and minor issues. For example, many participants thought it was easy to use the system. One participant (P29) mentions that he likes that you can go through the system ‘at one’s own pace’. This user does not feel rushed to continue with the conversation and indicated to like the interactive nature of the application.

Table 18: Exemplary usability issues in category Navigation & Structure.

Severity		Example of usability issue	Participants
Positive		The user indicates that the account creation steps were easy.	P09, P10, P23, P24, P25, P26, P27, P28, P29, P30
		The user does find it clear to interact with Francois (the diet coach) for a recipe.	P03, P09, P10, P26, P29
		The user liked that you could go at your own pace.	P29
Negative	Minor	The user prefers the explanation of the activity book before showing data.	P04

	Serious	The user likes to have a return button in the dialogues.	P09, P10, P25, P28, P30
		After the introduction dialogues it was not clear for the user to click on the coaches to interact.	P03, P04, P05, P06, P07, P08, P25, P27
		The user had problems reading when the tablet screen resizes to fit in the keyboard.	P01, P02, P03, P04, P23
	Critical	The user was confused when the dialogue ends after the food intake instead of providing the recipe.	P01, P02, P04, P06, P07, P09, P10, P25, P26, P27, P30
		The user indicates it was not clear to first chat with Olivia the activity coach to connect the Fitbit.	P03, P05, P06, P07, P08, P10, P23, P24
		Despite the correct set of actions to connect the Fitbit account, the system responds with returning to home screen of COUCH instead of continuing dialogue in living room.	P03, P04, P11, P28, P29, P30

5.2.1.1 Recommendations for improvement

There are some serious and critical negative issues found in this category. The following requirements are made to further improve the system:

- The Fitbit connection system coding should be checked, to prevent failing connections and redirections to the home-page (D2.6-F2).
- The recipe dialogue of Francois should end with the actual recipe (D2.6-I7).
- When introduced to the coaches in the living room, information should be provided more clearly to the user on how to interact with the system (D2.6-I2).
- The system should include an option 'go back', during the conversation with the virtual coach(es) (D2.6-I8).
- The system should make it clearer to the user to interact with Olivia if they have to connect an activity tracker (D2.6-I4).
- Effort should be taken to optimize system performance, to prevent slow system responses and enable proper system use all tablets/iPads (D2.6-F3).

See Section 6, for a complete and detailed overview of the additional requirements that are derived from this study.

5.2.2 Content & Information

A total of 95 unique usability issues were identified in the Content & Information category. Of all these issues, there were 35 positive points and 60 negative issues (of which 55 were minor issues, 5 serious, and 0 critical).

Examples of these issues that were found, can be viewed in Table 19. In general, participants were positive regarding the content provided by COUCH. They understand the added value of coaching on the physical activity, nutrition, social and cognition domains. Some indicated that they already use health apps for one of these domains and like to integrate those with a system like COUCH. So, as P11 indicated "With COUCH, there's a good opportunity to gain more knowledge on specific topics you find interesting". However, the participants were a bit divided about how the content was provided to them. Some users get impatient during the dialogues because of the large number of repetitions and the amount of small talk. While others really like the small-talk and the funny interruptions.

No critical issues were found in the content & information domain. The minor and serious content related issues were found throughout all tasks. For example, a minor issue was noticed in the baseline questions where a user joked/indicated "Nowadays, there should be more gender options here!". Furthermore, during the coach selection process users that choose to select their coaches themselves were a bit confused by the follow up dialogue which states "because you did not finish the intake, we can't make a suggestion for suitable coaches for you". This direct sentence evokes the feeling that they

did something wrong, which was experienced as annoying. Articulating the sentence differently might help solve this serious issue. Moreover, during the coach selection phase two users indicated that they did find it weird that they could not select both the diabetes coach as well as the chronic pain coach, after all they had both disorders.

During the free exploring phase many participants enjoyed the dialogue content of the coaches (P08, P09, P24, P28) and only positive notes were made about the multifaceted approach of the coaching. The radio functionality was not often found by the participants themselves ($\pm 10\%$). Most of the participants indicated the radio looks nice (“it looks magnificent, I just had such radio in the past.”), but they also indicated to not really mind the radio functionality as they don’t often use the radio or they have another device/app for this. A serious issue indicated during this task by a user with heart disease, was the fact that Olivia introduces herself as a coach that will advise you based on monitored activity and will help you with an activity program. The participant indicated that the system should be careful with this. For example, in patients with heart or lung diseases this could result in a big risk of (severe) adverse events without detailed information on the capabilities of these individual patients. This participant advised to include a screening question in the dialogues to ask for specific (medical) limitations and risks concerning physical activity.

The content issues found for the task 4 and 5, concerning the connection and data of the activity tracker, revealed that participants do find that the explanation of the connection consists of a lot of steps and that the actual connection page is not understandable due to the English content (for Dutch and Danish users). They indicated that visual support during the explanation would be an option to help them.

After the connection participants do indicate to prefer more detailed information about their activities in the activity book. For example, the division between their cycling, swimming and walking minutes. Moreover, they indicated to want to view data from longer ago, by swiping through the book.

Finally, in task five, most participants indicated that confirming the most important food preferences before getting the recipe is good (“you are very precise about that, it is good for some people to ask this again”). However, after getting the actual recipe some participants were a bit disappointed that it did not match their indicated food preferences. If they indicated to prefer meat, they had liked a recipe with meat.

Table 19: Exemplary usability issues in category Content & Information.

Severity		Example of usability issue	Participants
Positive		The user indicates it is good to confirm the food preferences, before providing the recipe.	P06, P03, P09
		The user does find specific parts of the dialogue content funny.	P02, P08, P09, P25
		The user likes the idea of a coach selection screen for getting to know all coaches.	P10
Negative	Minor	The user does not mind the radio functionality	P05, P06, P07, P08
		The user prefers to see more detailed information about their activities performed.	P06, P07
		The user indicates that during the food intake it is important to ask for specific preferences/restriction for salt.	P04, P06
	Serious	The user indicates to not understand The Fitbit connection page as it has English content.	P01, P04, P05, P06, P08
		The user feels confused about the fact that the system states “to not have finished the intake” when they choose the option to select their own coaches.	P05, P07, P09.

		The user wants to have both diabetes and chronic pain coach.	P04, P28
		The user indicates Olivia should also ask for other conditions before advising an activity program, as this could be a big risk for i.e. heart patients.	P03
		The recipe does not match the selected food preferences.	P06, P23, P24, P25, P26, P27, P28
	Critical	-	-

5.2.2.1 Recommendations for improvement

There are some serious negative issues found in this category. The following requirements are made to further improve the system:

- Personalize the amount of small talk, funny interactions and fun facts. (D2.6-F4)
- Change the dialogue content after selecting to choose your own coaches to a less direct way of saying “are you sure you want to select your coaches yourself? (D2.6-C1)
- Enable the option to select both the diabetes and chronic pain coach. (D2.6-F5)
- Include screening questions in Olivia’s dialogue for specific (medical) limitations and risks concerning physical activity. (D2.6-C2)
- Include visual support during the explanation of the connection of the Fitbit. (D2.6-C3)
- Provide more (optional) detailed information within the activity book (i.e. types of activities, history of activity by swiping through book). (D2.6-C4)
- Expand the number of recipes so that a better match to the food preferences of the users can be found. (D2.6-C5)

See Section 6 for a complete and detailed overview of the additional requirements that are derived from this study.

5.2.3 Design & Presentation

A total of 27 unique usability issues were identified in the Design & Presentation category. Of all these issues, there were 10 positive points and 17 negative issues (of which 14 were minor issues, 1 serious, and 2 critical).

Examples of these issues that were found, can be viewed in

Table 20. In general, there were few issues reported concerning the design and presentation of the Council of Coaches system. The critical issues were found only within task one. The fact that the create account button did not stand out, resulted in a failure of task one for two participants. The login screen was more clearly present according to them, which made the participants focus on filling in their email/password there instead. Another critical issue was the fact that it was unclear for participants to press the button next to the text entry after the text entry, which increased the time to complete the task eminently. The button is highlighted before filling in the data, however due to scaling/displacement when opening the keypad this is not visible anymore.

The participants were very positive about the design of the living room and the coaches, but indicated to prefer a visual label of the coach's expertise as expressed by P23 with "I think it would be easier if there was effectively a label on the coaches saying what they specialise in."

Finally, a commonly argued issue was the fact that the speech bubble of Coda (or other coaches) block some of the other coaches. Sometimes even the coaches that are being talked about become blocked, like when Helen is introduced as the woman at the left side of the table, while she is not visible due to the speech bubble.

Table 20: Exemplary usability issues in category Design & Presentation.

Severity		Example of usability issue	Participants
Positive		The COUCH interface looks pretty.	P01, P03, P05, P10, P25, P30
		The user does find the activity steps overview graph clear.	P02, P06, P09
		The user recognises the Fitbit login page from his/her own use.	P02
Negative	Minor	The user indicates that coaches should be visually labelled with their expertise.	P20, P23, P30
		The user indicates that the colour differences between a touched button and untouched button is too small.	P01
	Serious	The user indicates that the coaches (i.e. Helen) are not visible when introduced by Coda, due to the size and location of the speech bubble.	P01, P03, P06, P09
	Critical	The user indicates that the create account button does not stand out.	P03, P23
		Participant doesn't know they've to press the button next to the text entry when entering email address.	P23

5.2.3.1 Recommendations for improvement

There are some serious and critical negative issues found in this category. The following requirements are made to further improve the system:

- Make the “create account” button stand out for the first time the COUCH system is used (D2.6-S1).
- Better highlight the next button next to the text entry fields (D2.6-S2).
- Include a visual label of each coach’s expertise (D2.6-S3).
- Prevent that coaches that are talked about are not visible due to the speech bubble of another coach (D2.6-S4).

See Conclusion and updated requirements, for a complete and detailed overview of the additional requirements that are derived from this study.

5.2.4 Other

A total of 13 unique usability issues were identified in the Design & Presentation category. Of all these issues, there were 3 positive points and 10 negative issues (of which 10 were minor issues, 0 serious, and 0 critical).

Some of the issues classified as “other” concerns more the definitions of the task instead of system related issues. For example, one participant didn’t understand task five, as he thought he had to search for a medicine prescription instead of a food recipe (as in Dutch “recept” has two definitions).

A participant (P05) indicated that the system was more understandable compared to previous versions of the demonstrator and liked the progress that was made. Participants also indicated that the application could help people who are lonely. A final general minor negative issue was given by P02, who indicated that the COUCH system required thorough reading to use the system.

5.2.4.1 Recommendations for improvement

No serious or critical negative issues were found in this category. Moreover, no specific requirements follow from the issues found in this section.

5.3 Task metrics

This subsection describes the task performance matrices for each of the five tasks. The results of this section are based on the Dutch participants (Participant ID 01 to 10) and the Scottish participants (Participant ID 23 to 30) combined, as Denmark was not able to retrieve task metrics data as described in Methodology Denmark. Task 1 and 5 had the best task performance based on the areas of the radar charts (Table 21). Task 3 and 4 scored the lowest on task performance. In the subsections 5.2.1 till 5.2.5 we will go through the detailed task performance per task.

Table 21: Relative filling area radar charts (figure 21 till 24).

Task number	Area (relative % filling)
Task 1	57%
Task 2	N.A.
Task 3	16%
Task 4	28%
Task 5	45%

5.3.1 Task 1: Account creation

In general, the account creation task went quite well for most participants. 83% of the participants did successfully complete their task. The participants who did not complete the task within 5 minutes, tried to login multiple times instead of creating an account. Moreover, there was some variation in the time to complete the task, which was mainly dependent on whether participants had difficulty with typing in their email/password. When the task completion time variable was split into 2 groups based on whether the participant uses a tablet at home, we found that participant that do use a tablet at home perform task 1 on average in 110 seconds versus 128 seconds in participants without tablet use at home, a not significant difference ($p=0.26$).

The participants that completed the task on average did not need much additional steps. The errors that were made were limited to typing errors in the email/password ($n=5$) and trying to login instead of create an account ($n=4$). One participant was also curious for the purpose of creation an account and therefore pressed “why should I make an account?”.

The task scored high on the ASQ score, indicating the participants were satisfied with the ease of the task, the time it took and the amount of support they got.

Details of the performance metrics can be found in Table 22 and in the visual radar chart (Figure 23).

Table 22: Performance metrics for task 1: Account creation.

Task number	ID	Task completion (%)	Task completion time (sec)	Task satisfaction (ASQ)	Steps to complete (Steps done/Steps needed to complete task)	Errors
Task 1	M (SD)	83%	115 (44)	5.8 (2.1)	1.1 (0.2)	0.8 (1.6)
	95%-CI		93-138	4.8-6.7	1.0-1.2	0.1-1.6

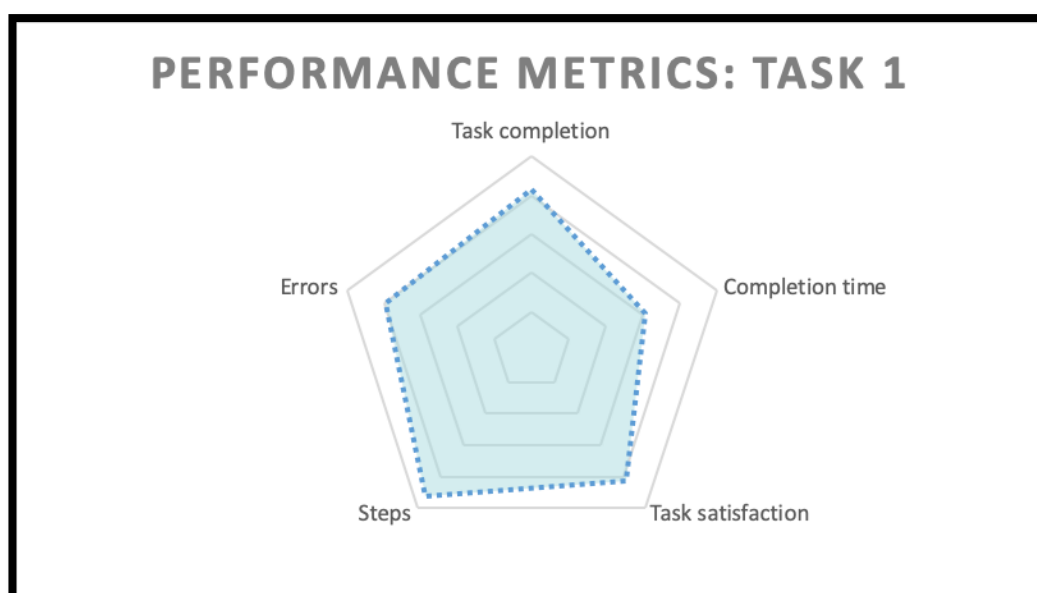


Figure 23: Performance metrics radar chart for task 1: account creation.

5.3.1.1 Coach selection

After the account creation the participants could continue to perform the baseline questions and proceed to the coach selection. 21% of the participants indicated to make their own coach selection, versus 79% who proceeded with the state of change questions to be automatically helped in the coach selection by the COUCH system.

In the coach selection screen 79% of the participants inspected at least one of the portraits to get more information. Only 18% of the participants, that received their personalized coach selection based on the intake questions, altered the coach selection within the coach selection screen. Although it should be noted that it is not known how many of the participants knew that they could alter their coach selection. One participant remarked to want to alter his predetermined council of coaches but did not understand how to do that.

5.3.2 Task 2: Free exploring

Participants took on average 5 minutes and 24 seconds and used 48 steps to explore the living room with all the coaches. The two main errors that were made within the free exploring were; 1) Some participants (P02/P03) clicked on one the coaches during the introduction dialogues of another coach, nothing happened while they expected that the other coach would start talking. 2) When the introduction dialogues are finished it is often not clear for the participants (P03/P06/P09/P10) how to continue, which results in the fact that they randomly click within the living room, repeat the introduction, press the return button of the browser or even logout. The random clicking resulted in the fact that one participant found the radio functionality by himself (P10). For the other participants it was not self-evident to find this functionality.

The task scored high on the ASQ score, indicating the participants were satisfied with the task. If we investigate the questions in more detail it became clear that within the task satisfaction the question about the time it took to complete the task was answered the most negatively, indicating some participants (2 even scored with a 1 out of 7) does find the introduction too comprehensive.

Details of the performance metrics can be found in Table 23.

Table 23: Performance metrics for task 2: Free exploring.

Task number	ID	Task completion (%)	Task completion time (sec)	Task satisfaction (ASQ)	Steps performed	Errors
Task 2	M (SD)	N.A.	324 (153)	5.8 (1.7)	48 (18)	1.1 (2.1)
	95%-CI		253-395	5.0-6.6	40-57	0.1-2.1

5.3.3 Task 3: Connect Fitbit

Task 3 can be described as the toughest task of this evaluation, as quoted from P04 “You make it very difficult for me this time”. Only 17% of the participants managed to successfully connect the Fitbit (or at least went through all the steps correctly, as in a couple of cases a system error occurred after clicking “allow all”, which is described in more detail in Navigation & structure) within the time deadline of 5 minutes. The fact that a lot of participants did not succeed lays in the fact that they had no clue to go to the activity coach. Many participants first tried to logout, go directly to fitbit.com in the browser, go to the browser settings or speak to coda. These actions took so much time or resulted in the fact that the subjects were not able to return to the COUCH demonstrator, so that the task failed. When these participants were provided with a hint afterwards, that they should talk with one of the coaches to connect the Fitbit, many of them still succeeded, indicating that better navigation towards Olivia in case of activity sensor connection could contribute to better task completion.

The participants that did complete the task did this quite quick (on average 2 minutes 53 seconds) without much additional steps (on average 23 instead 22 that are minimally needed to complete the task).

Details of the performance metrics can be found in Table 24 and in the visual radar chart (Figure 24).

Table 24: Performance metrics for task 3: Connect Fitbit.

Task number	ID	Task completion (%)	Task completion time (sec)	Task satisfaction (ASQ)	Steps to complete (Steps done/Steps needed to complete task)	Errors
Task 3	M (SD)	17%	173 (9)	4.0 (2.2)	1.0 (0.1)	4.9 (4.6)
	95%-CI		163-183	3.0-5.0	1.0-1.1	2.8-7.1

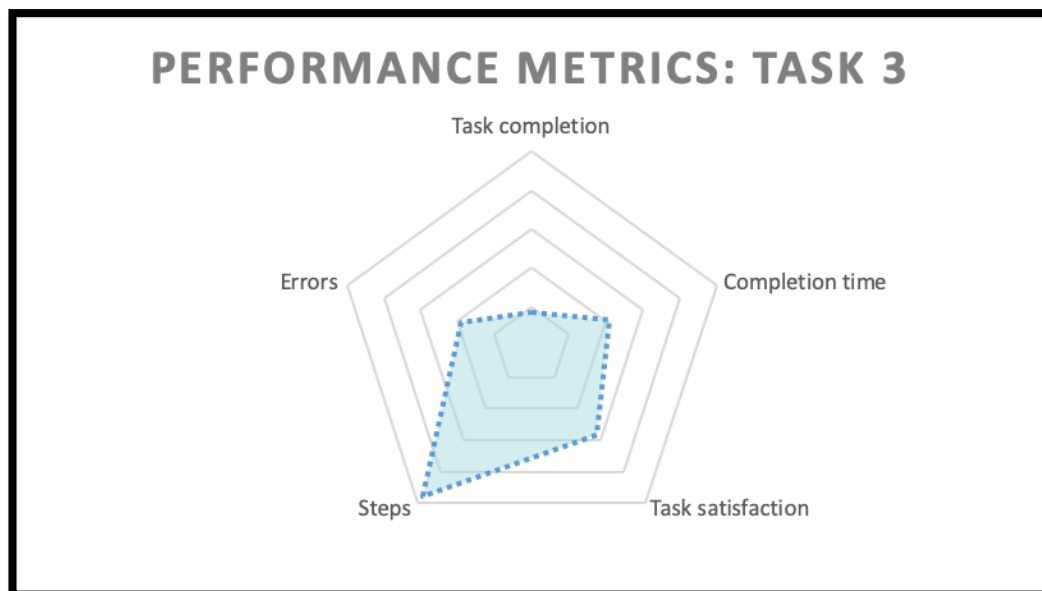


Figure 24: Performance metrics radar chart for task 3: Connect Fitbit.

5.3.4 Task 4: Find performed steps

After performing task 3 it was obvious for most participants that they needed Olivia the activity coach again for task 4: “Find out how many steps were performed on Saturday”. This results in a task completion rate of 93% and a quick task completion time (1 minute 49 seconds on average), which could even be quicker as participants did on average 2.4 times more steps than needed and made on average 2.8 errors during the task.

A variety of errors were made during this task. First of all, the most common error (in 77% of the participants) was to go to the “activity tracker” dialogue instead of “my progress” to review the data from the activity tracker. Also, the other dialogue options of Olivia were often tried. Therefore, it could be concluded that “my progress” is not directly indicative for viewing the activity data. 38% of the participants went through the explanation of the activity book before providing the number of steps. One participant (P23) did select every other coach before going to the activity coach.

Furthermore 29% of the participant also clicked on the bar to review the exact number of steps the other 71% did read the steps from the graph-axis.

Details of the performance metrics can be found in Table 25 and in the visual radar chart (Figure 25).

Table 25: Performance metrics for task 4: Find steps Saturday.

Task number	ID	Task completion (%)	Task completion time (sec)	Task satisfaction (ASQ)	Steps to complete (Steps done/Steps needed to complete task)	Errors
Task 4	M (SD)	93%	109 (61)	4.8 (1.5)	2.4 (1.4)	2.8 (2.9)
	95%-CI		76-142	4.0-5.6	1.6-3.2	1.3-4.3

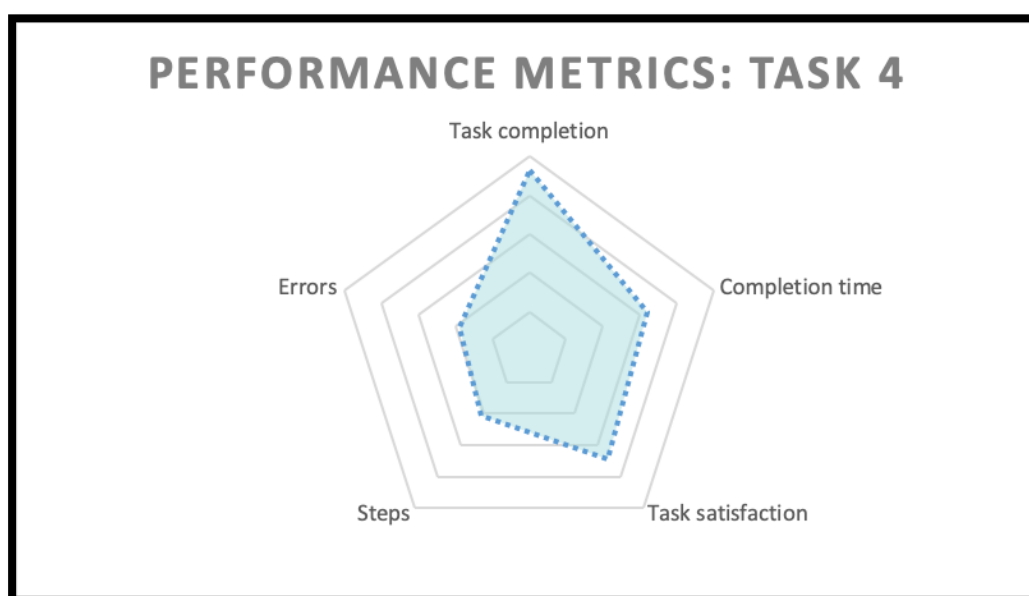


Figure 25: Performance metrics radar chart for task 4: find steps Saturday.

5.3.5 Task 5: Personalised dinner recipe

Task 5 was successfully performed in 89% of the participants. The overall task performances were good, as indicated by the area in the radar chart (Figure 26). Two participants (P02 and P04) did not go directly to Francois (one did go to Katarzyna (the diabetes coach) instead). However, most of the participants (89%) did directly chat with Francois to go to the “recipe” dialogue. One participant (P03) first did the food preferences via the “food intake” dialogue before going to the “recipe dialogue”, because as he said “before I click on recipes, I have to indicate my preference, otherwise I would not receive a personalized recipe”. For the other participants the “food intake” dialogue was still being run through, as the recipe dialogue automatically linked to the food intake. However, after finishing the food intake questions many participants were confused. “Huh, why don’t I get my recipe?”. This confusion led to multiple following errors such as talking to other coaches and/or selecting the other dialogue options of Francois.

Details of the performance metrics can be found in Table 26 and in the visual radar chart (Figure 26).

Table 26: Performance metrics for task 5: Get personalized dinner recipe.

Task number	ID	Task completion (%)	Task completion time (sec)	Task satisfaction (ASQ)	Steps to complete (Steps done/Steps needed to complete task)	Errors
Task 4	M (SD)	89%	169 (56)	5.2 (1.5)	1.1 (0.1)	1.3 (1.6)
	95%-CI		142-197	4.5-5.9	1.0-1.2	0.5-2.0

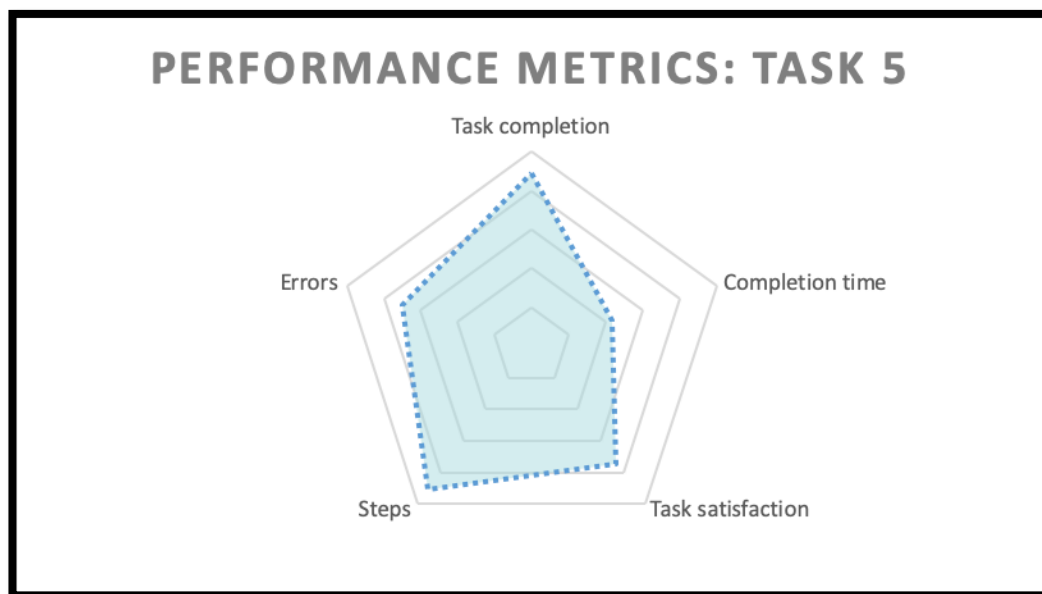


Figure 26: Performance metrics radar chart for task 5: Get personalized dinner recipe.

5.3.6 Discussion and recommendations

In general, the task performance results showed that whether a task is successfully completed is primarily dependent on whether the first step of the task has been carried out properly and quickly (i.e. click account creation instead of trying to login or go to activity coach to connect Fitbit instead of directly go to Fitbit website). It is therefore very important to make the navigation towards the right initiating start-button/coach-dialogue very clear.

Next to this general discussion point, the following specific requirements were extracted from the task metrics results to further improve the system:

- Make the “create account” button stand out for the first time the COUCH system is used. (D2.6-S1)
- Add an indication in the coach selection screen that the coach choice can be altered. (D2.6-I1).
- When introduced to the coaches in the living room, information should be provided more clear to the user on how to interact with the system. (D2.6-I2)
- The system should provide the user with information on how to proceed after the introduction of coaches if there is no interaction for a while or when random clicking in the living room is recognised (D2.6-I3).
- The system should provide the option for an extensive or short introduction of the coaches (D2.6-F1).
- The system should make it clearer to the user to interact with Olivia if they have to connect an activity tracker. (D2.6-I4)
- Both the dialogue-path “my progress” as “my activity tracker” should lead towards the activity data. (D2.6-I5).
- The diabetes coach (Katarzyna) should have an option to refer to the nutrition coach Francois for food related questions and recipes (D2.6-I6).
- The recipe dialogue of Francois should end with the actual recipe (D2.6-I7).

See Conclusion and updated requirements, for a complete and detailed overview of the additional requirements that have been derived from this study.

5.4 Questionnaire results

We divided the analysis of the system usability and technology acceptance in 2 sections; one section evaluating the system usability and technology acceptance in the Netherlands and Scotland and one section evaluating the results of Denmark. This choice was made because the outcomes did significantly differ and were expected to be mainly due to the difference in methodology instead of being technology dependent.

5.4.1 The Netherlands + Scotland

5.4.1.1 System usability

The participants in the Netherlands and Scotland (n=18) scored the third demonstrator of the Council of Coaches with an average system usability score of 73.9 (± 17.9), indicating a good acceptability (B-category). The SUS scores are indicative for the likability of participant to recommend the product to a friend or colleague (net promotor scale). The average SUS score of the third demonstrator is indicative that it is not a promotor nor a detractor, so it has a passive promotional effect.

5.4.1.2 Technology Acceptance Model

As shown below in Table 27 the COUCH demonstrator passes on every aspect with at least a score of >50% for the Dutch and Scottish participants, indicating a good technology acceptance. It scored the lowest on willingness to pay and perceived usefulness. The highest scores were found in the domain intention to use and enjoyment. The largest variation between the participants was also found in the willingness to pay category.

Table 27: Overview of the Technology acceptance model outcomes (NL+SC).

ID	Enjoyment	Aesthetics	Control	Trust	Perceived usefulness	Ease of use	Intention to use	Willingness to pay
M (SD)	78.2 (18.9)	70.8 (17.8)	59.9 (21.6)	64.8 (14.2)	58.3 (20.6)	66.7 (18.7)	74.7 (21.4)	55 (28.4)
95%-CI	69.5- 87.0	62.6- 79.1	49.9- 69.9	58.3- 71.4	48.8- 67.9	57.1- 74.4	64.8- 84.6	37.4- 72.6

5.4.2 Denmark

5.4.2.1 System usability

The participants in Denmark (n=12) scored the third demonstrator of the Council of Coaches with an average system usability score of 50.6 (± 4.5), indicating a not acceptable acceptability. The SUS scores are indicative for the likability of participant to recommend the product to a friend or colleague (net promotor scale). The average SUS score of the third demonstrator is indicative that it is a detractor, so participants are not particularly thrilled by the service. They, with all likelihood, won't purchase again it and could potentially damage the reputation through negative word of mouth.

5.4.2.2 Technology Acceptance Model

The results of the TAM questionnaire in Denmark showed different results compared to the TAM results of the Netherlands and Scotland (Table 28). In Denmark the questions were answered more extremely. The elements "Ease of use" and "Control" scored very high (resp. 80.4 and 90.9), while "intention to use" and "willingness to pay" scored low (resp. 36.9 and 12.1). The largest variation between the participants was seen in the intention to use questions.

Table 28: Overview of the Technology acceptance model outcomes (DK).

ID	Enjoyment	Aesthetics	Control	Trust	Perceived usefulness	Ease of use	Intention to use	Willingness to pay
M (SD)	62.5 (30.3)	71.0 (17.8)	90.9 (11.5)	51.2 (14.8)	47.0 (28.9)	80.4 (14.6)	36.9 (29.5)	12.1 (23.7)
95%-CI	42.7- 79.6	62.6- 79.1	83.0- 97.7	40.2- 62.1	26.9- 64.1	69.6- 91.2	16.4- 54.3	-4.3-26.1

5.5 Discussion

The third functional demonstrator was evaluated to elicit usability issues from the end-user's interaction with the system. We will shortly discuss the main insights below and refer to Section 6: Conclusion and updated requirements to reveal how we can convert these issues to requirements for improvement of the system.

The results of the retrospective think-aloud process showed that there was a relative high number of positive points and that a low number of critical usability issues. Even more striking is that the amount of critical issues is reduced compared to D2.5 (Beinema, et al., 2019). Furthermore, multiple participants from the participants that also took part in the previous evaluation rounds indicated that they liked the progress which the Council of Coaches system has made.

Most serious and critical usability issues were found in the category Navigation & Structure. The third functional demonstrator changed a lot from the previous versions with regards to navigation & structure, included a lot of new features, and contained much more information which should be easy to navigate through. Therefore, it was expected that this category revealed many new issues. With this third evaluation and all the issues provided by the thirty participants we expect to have collected all the necessary feedback and requirements to improve the system towards excellent usability.

The task metrics results (how efficiently and effectively can users perform a task within the system) show that whether a task is successfully completed is primarily dependent on whether the first step of the task has been carried out properly and quickly. It is therefore very important to make the navigation towards the right initiating start-button/coach-dialogue very clear.

The results of the system usability and technology acceptance were quite different between NL&SC versus DK. This was expected to be mainly due to the difference in methodology for the evaluation (i.e. the group design in which the opinion of a participant can be altered by the most convincing and present voice of another participant). The overall system usability is considered good in NL&SC. The overall technology acceptance is quite good for our intended target audience, for example at all three sites the participants rated the system high on enjoyment, aesthetics and ease of use. This gives confidence that if the recommendations are implemented, the system has great potential to be used

The quantitative results of this evaluation of the third functional demonstrator (a high-fidelity prototype) can be used as a benchmark for the later stages of the system development process and the final evaluation round.

6 Conclusion and updated requirements

This document describes the setup, realization and results of a usability evaluation study performed using the third Council of Coaches functional prototype. With regards to the usability study, the overall results show that the second Council of Coaches demonstrator was easy to use and enjoyed by the test participants. Test participants showed high engagement with the demonstrator and the study in general, resulting in a large amount of recommendations, leading to additional requirements. This information will be used in the coaching strategies and content design process, as well as in the user modelling and content generation in WP3.

This study showed a good usability of the third Council of Coaches functional demonstrator and can therefore be used to build on the final version of the Council of Coaches demonstrator.

6.1 Updated requirements

Below we summarize the set of updated requirements that stem from the usability evaluation performed.

6.1.1 Functions and Events

ID:	D2.6-F1
Requirement:	The system should provide the option for an extensive or short introduction of the coaches.
Source:	Evaluation of the third demonstrator.
Rationale:	Some users indicate that the introduction is too comprehensive.
Priority:	Could
History:	

ID:	D2.6-F2
Requirement:	The Fitbit connection system coding should be checked, to prevent failing connections and redirections to the home-page.
Source:	Evaluation of the third demonstrator.
Rationale:	The Fitbit connection was not always successful, while the users did perform the correct set of actions. If the system failed to connect it sometimes returned to the COUCH home page and sometimes to the living room (without starting the successful connection dialogue)
Priority:	Must
History:	

ID:	D2.6-F3
Requirement:	Effort could be taken to optimize system performance, to prevent slow system responses and enable proper system use all tablets/iPads.
Source:	Evaluation of the third demonstrator.
Rationale:	Many participants complained about the slow response to their clicking. Using high-spec tablets/iPads would help enable proper use for all devices.
Priority:	Could
History:	

ID:	D2.6-F4
Requirement:	Personalize the amount of small talk, funny interactions and fun facts.
Source:	Evaluation of the third demonstrator.
Rationale:	The opinions of users are divided. The one likes the extensive (funny) dialogue while others would like a more direct way of communication.
Priority:	Should
History:	

ID:	D2.6-F5
Requirement:	Enable the option to select both the diabetes and chronic pain coach.
Source:	Evaluation of the third demonstrator.
Rationale:	Users with diabetes and chronic pain indicated to want the option to select both coaches.
Priority:	Should
History:	

6.1.2 Interaction and Usability

ID:	D2.6-I1
Requirement:	Add an indication in the coach selection screen that the coach choice can be altered.
Source:	Evaluation of the third demonstrator.
Rationale:	Some users did not understand that the coach selection could be altered.
Priority:	Should
History:	

ID:	D2.6-I2
Requirement:	When introduced to the coaches in the living room, information should be provided more clearly to the user on how to interact with the system.
Source:	Evaluation of the third demonstrator.
Rationale:	Many (wrong clicking) errors were made during the first interactions in the living room.
Priority:	Could
History:	

ID:	D2.6-I3
Requirement:	The system should provide the user with information on how to proceed after the introduction of coaches if there is no interaction for a while or when random clicking in the living room is recognised
Source:	Evaluation of the third demonstrator.
Rationale:	Some users struggle to interact when the introduction dialogue of Coda is finished.
Priority:	Should
History:	

ID:	D2.6-I4
Requirement:	The system should make it clearer to the user to interact with Olivia if they have to connect an activity tracker.
Source:	Evaluation of the third demonstrator.
Rationale:	Most users did not have any clue to talk to the activity coach to connect the activity tracker. They thought they had to go to the Fitbit website, logout or go to Coda.
Priority:	Must
History:	

ID:	D2.6-I5
Requirement:	Both the dialogue-path “my progress” as “my activity tracker” should lead towards the activity data.
Source:	Evaluation of the third demonstrator.
Rationale:	Many users go to “my activity tracker” to review their data.
Priority:	Should
History:	

ID:	D2.6-I6
Requirement:	The diabetes coach (Katarzyna) should have an option to refer to the nutrition coach Francois for food related questions and recipes.
Source:	Evaluation of the third demonstrator.
Rationale:	Some users (with diabetes) did went first to Katarzyna for getting a recipe.
Priority:	Could
History:	

ID:	D2.6-I7
Requirement:	The recipe dialogue of Francois must end with the actual recipe
Source:	Evaluation of the third demonstrator.
Rationale:	Users are confused when the recipe dialogue ends without the recipe. From this point a lot of errors were made instead of again select the recipe dialogue.
Priority:	Must
History:	

ID:	D2.6-I8
Requirement:	The system could include an option 'go back', during the conversation with the virtual coach(es).
Source:	Evaluation of the third demonstrator.
Rationale:	Some participants indicate to want to make changes to their already provided answers.
Priority:	Could
History:	

6.1.3 Content and Structure

ID:	D2.6-C1
Requirement:	Change the dialogue content after selecting to choose your own coaches to a less direct way of saying "are you sure you want to select your coaches yourself?"
Source:	Evaluation of the third demonstrator.
Rationale:	The current sentence did evoke the feeling in users that they did something wrong.
Priority:	Should
History:	

ID:	D2.6-C2
Requirement:	Include screening questions in Olivia's dialogue for specific (medical) limitations and risks concerning physical activity.
Source:	Evaluation of the third demonstrator.
Rationale:	Activity programs can result in medical risks for the user when it is not known if a (medical) condition influences their ability to be physical active.
Priority:	Must
History:	

ID:	D2.6-C3
Requirement:	Include visual support during the explanation of the connection of the Fitbit.
Source:	Evaluation of the third demonstrator.
Rationale:	Users indicate the explanation contains a lot of steps to remember. Moreover, the Fitbit page contains English so users indicate to want clear examples of what actions to do.
Priority:	Should
History:	

ID:	D2.6-C4
Requirement:	Provide more (optional) detailed information within the activity book (i.e. types of activities, history of activity by swiping through book).
Source:	Evaluation of the third demonstrator.
Rationale:	Users liked the activity book, however would like to see more detailed information.
Priority:	Could
History:	

ID:	D2.6-C5
Requirement:	Expand the number of recipes so that a better match to the food preferences of the users can be found.
Source:	Evaluation of the third demonstrator.
Rationale:	Users indicate the recipe does not match their preferences.
Priority:	Should
History:	

6.1.4 Style and Aesthetics

ID:	D2.6-S1
Requirement:	Make the "create account" button stand out for the first time the COUCH system is used.
Source:	Evaluation of the third demonstrator.
Rationale:	The users have trouble finding this button, resulting in the fact that they try to login instead.
Priority:	Could
History:	

ID:	D2.6-S2
Requirement:	Better highlight the next button next to the text entry fields
Source:	Evaluation of the third demonstrator.
Rationale:	During account creation, the keypad pop-up changes scaling/displacement making it unclear for some participants how to continue after filling in email/password.
Priority:	Should
History:	

ID:	D2.6-S3
Requirement:	Include a visual label of each coach's expertise.
Source:	Evaluation of the third demonstrator.
Rationale:	Some users indicate that it would make it easier for them to find the right coach if each coach had a label of its expertise.
Priority:	Could
History:	

ID:	D2.6-S4
Requirement:	Prevent that coaches that are talked about are not visible due to the speech bubble of another
Source:	Evaluation of the third demonstrator.
Rationale:	Many users recognize the fact that coaches are not visible when they are talked about by another coach, which annoys users.
Priority:	Should
History:	

7 Bibliography

- Baumel, A., & Muench, F. (2016). Heuristic Evaluation of Ehealth Interventions: Establishing Standards That Relate to the Therapeutic Process Perspective. *JMIR Mental Health*, 3(1), e5.
- Beinema, T., Broekhuis, M., op den Akker, H., van Velsen, L., ter Stal, S., Pease, A., . . . van Loon, J. (2019). *D2.5: Evaluation of second functional prototype and updated requirements*. The Council of Coaches consortium.
- Beinema, T., Huizing, G., Op den Akker, H., Snaith, M., Fides Valero, A., Traver, V., . . . Konsolakis, K. (2018). *D3.1: Initial knowledge base design and coaching strategies*. The Council of Coaches Consortium.
- Beinema, T., Op den Akker, H., Broekhuis, M., Huizing, G., van Velsen, L., ter Stal, S., & van Loon, J. (2018). *D2.4: Evaluation results of first functional prototype and updated requirements*. The Council of Coaches Consortium.
- Belanche, D., Casaló, L., & Guinalíu, M. (2012). Website usability, consumer satisfaction and the intention to use a website: The moderating effect of perceived risk. *Journal of retailing and consumer services*, 124-132.
- Brooke, J. (1996). *SUS: A "quick and dirty" usability scale*. Taylor & Francis.
- Crutzen, R., de Nooijer, J., Brouwer, W., Oenema, A., Brug, J., & de Vries, N. (2011). Strategies to Facilitate Exposure to Internet-Delivered Health Behavior Change Interventions Aimed at Adolescents or Young Adults: A Systematic Review. *Health Education & Behavior*, 38(1), 49-62.
- Davis, F. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, 13(3), 319-340.
- Davis, F., Bagozzi, R., & Warshaw, P. (1992). Extrinsic and Intrinsic Motivation to Use Computers in the Workplace. *Journal of Applied Social Psychology*, 22(14), 1111-1132.
- Duh, H., Tan, G., & Chen, V. (2006). *Usability evaluation for mobile device: A comparison of laboratory and field tests*. (Mobile HCI 2006, September 12-15, 2006 ed.). Helsinki, Finland: Conference: Proceedings of the 8th Conference on Human-Computer Interaction with Mobile Devices and Services.
- Heinz, S., Tuch, A., Seckler, M., Opwis, K., & Forde, S. (2014). Trust and distrust on the web: User experiences and website characteristics. *Computers in Human Behavior*, 39-50.
- Jameson, A. (2007). Adaptive interfaces and agents. In J. Jacko, & A. Sears, *Human-computer interaction handbook* (pp. 433–458).
- Jennings, M. (2000). Theory and models for creating engaging and immersive ecommerce Websites. *SIGCPR '00 Proceedings of the 2000 ACM SIGCPR conference on Computer personnel research* (pp. 77-85). Chicago, Illinois, USA: ACM New York.
- Lavie, T., & Tractinsky, N. (2004). Assessing dimensions of perceived visual aesthetics of web sites. *International journal of human-computer studies*, 269-298.
- McKnight, D., Choudhury, V., & Kacmar, C. (2002). Developing and validating trust measures for e-Commerce: An integrative approach. *Information systems research*, 334-359.
- McKnight, H., Carter, M., Thatcher, J., & Clay, P. (2011). Trust in a Specific Technology : An Investigation of its Components and Measures. *ACM Transactions on Management Information Systems (TMIS)*.
- Reidsma, D., Bruijnes, M., Huizing, G., Klaassen, R., Konsolakis, K., Weusthof, M., . . . Simonetti, D. (2019). *D7.3: Second functional prototype*. The Council of Coaches Consortium.
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1), 68.

- Sauro, J. (2016). *Quantifying the user experience : practical statistics for user research*. Morgan Kaufmann.
- Tullis, T., & Albert, B. (2008). *Measuring the user experience: collecting, analyzing, and presenting usability metrics*. Burlington, M.A.: Elsevier.
- van der Heijden, J. (2004). User acceptance of hedonic information systems. *MIS Quarterly*, 695-704.
- van Haak, M., De Jong, M., & Schellens, P. (2003). Retrospective vs. concurrent think-aloud protocols: Testing the usability of an online library catalogue. *Behaviour and Information Technology*, 339-351.
- van Velsen, L., Broekhuis, M., Jansen-Kosterink, S., & op den Akker, H. (2019). Tailoring persuasive eHealth Strategies on the basis of personal motivation: A survey-based study among older adults to identify motivational profiles for leading a healthy life and to identify preferred persuasive features. *Journal of Medical Internet Research (JMIR)*, Submitted.
- van Velsen, L., van der Geest, T., & Klaassen, R. (2011, 7). Identifying Usability Issues for Personalization During Formative Evaluations: A Comparison of Three Methods. *International Journal of Human-Computer Interaction*, 27(7), 670-698.
- van Velsen, L., van der Geest, T., van de Wijngaert, L., van den Berg, S., & Steehouder, M. (2015). Personalization has a Price, Controllability is the Currency: Predictors for the Intention to use Personalized eGovernment Websites. *Journal of Organizational Computing and Electronic Commerce*, 25(1), 76-97.
- Venkatesh, V., & Davis, F. (2000). A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. *Management Science*, 46(2), 186-204.
- Wang, Y., & Emurian, H. (2005). An overview of online trust: Concepts, elements, and implications. *Computers in Human Behavior*, 105-125.

Acknowledgements



The Council of Coaches project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement #769553. This result only reflects the author's view and the EU is not responsible for any use that may be made of the information it contains.

Headings and titles in this document, as well as the Council of Coaches logo use the Comfortaa font, designed by Johan Aakerlund and Cyreal and licensed under the Open Font License².

Additional text in this document uses the Roboto font, designed by Christian Robertson and licensed under the Apache License, Version 2.0³.

The Council of Coaches logo and Blobmen graphics were *drawn freely* in Inkscape, licensed under the GNU General Public License⁴.

² Open Font License: http://scripts.sil.org/cms/scripts/page.php?site_id=nrsi&id=OFL_web

³ Apache License, Version 2.0: <http://www.apache.org/licenses/LICENSE-2.0>

⁴ Inkscape License Information: <https://inkscape.org/about/license/>