

ALFRED

Personal Interactive Assistant for Independent Living and Active Ageing



D8.2.1.2 Piloting & Validation I: Individual Usability

Deliverable Lead: NFE

Contributing Partners: CHA, ESE

Delivery date: 09/2016

Dissemination level: Public

Version 1.0

This deliverable presents the results of all the iterative evaluations with end users in the ALFRED project. It gives an overview of the final success rates and satisfaction of end users with the usability of ALFRED. It is the public version.



Document Status	
Deliverable Lead	Nina van der Vaart, National Foundation for the Elderly
Internal Reviewer 1	Michael Krummen, Ascora
Internal Reviewer 2	Jorge Doménech, AITEX
Type	Deliverable
Work Package	WP8 Piloting & Validation
ID	D8.2.1.2 Piloting & Validation I: Individual Usability, Public version
Due Date	30.09.2016
Delivery Date	30.09.2016
Status	For Approval

Note

This deliverable is subject to final acceptance by the European Commission.

Disclaimer

The views represented in this document only reflect the views of the authors and not the views of the European Union. The European Union is not liable for any use that may be made of the information contained in this document.

Furthermore, the information is provided “as is” and no guarantee or warranty is given that the information is fit for any particular purpose. The user of the information uses it at its sole risk and liability.

Project Partners

 <p>Ascora GmbH, Germany</p>	 <p>Atos Spain sau, Spain</p>
 <p>Worldline, Spain</p>	 <p>Charité - Universitätsmedizin Berlin - Department of Geriatrics, Germany</p>
 <p>Asociacion de Investigacion de la Industria Textil, Spain</p>	 <p>Technische Universität Darmstadt, Germany</p>
 <p>National Foundation for the Elderly, The Netherlands</p>	 <p>Talkamatic AB, Sweden</p>
 <p>E-Seniors, France</p>	 <p>TIE Nederland N.V., The Netherlands</p>
 <p>IESE Business School, Spain</p>	

Executive Summary

This D8.2.1.2 is a follow up of D8.2.1 and D8.1.2 and contains a complete overview of the end user involvement in Iterative Evaluation Cycles 2 to 5. This deliverable is simultaneously released with D8.2.2.2 which is a confidential deliverable. This deliverable contains summarized information on the final results of the Iterative Evaluations to protect possible business opportunities of ALFRED. For the complete version, please consult D8.2.2.2.

The approach of ALFRED is to involve target groups continuously in the development of the system through Iterative Evaluations. The aim of this approach is to provide to system developers continuous recommendations from end users in order to obtain final results that are easy to use by older people and are adapted to their requirements and needs. The first Iterative Evaluation in M12 (reported on in D8.1.2) of the project was focused, among others, on the WoZ methodology to obtain more insight on voice interaction and older people.

This document continues this approach by reflecting the results of Iterative Evaluation Cycles 2, 3, 4 and 5.

Cycle 2 and 3 focus on different technical components of the ALFRED system. The results were analysed and reported back to the technical partners. Through a set of simple tables with recommendations, technical partners are informed on problems and necessary improvements to adapt to target group needs. The recommendations are ranked in priorities based on the amount of incidences during the usability sessions.

In the second Iterative Cycle the ALFREDO Marketplace, the Dance with ALFRED app and the ALFRED T-Shirt were tested with a total of 12 Test Persons (TPs). The ALFREDO Marketplace received a good user rating in general, the main issues were related to navigational and data entry problems. Most participants enjoyed playing the “Dance with ALFRED” game and especially had fun with the multiplayer modus. On the other hand, more challenging levels and a better introduction on how to play the game were desired. The main problems with the T-shirt was the position of the sensor unit in front of the chest. This and the thick material caused a discomfort and lead to excessive sweating of the users.

The third Iterative Cycle evaluated was the User Profile Editor, the (improved) ALFRED T-Shirt and the connected health app, the Dancicians game and the (improved) ALFREDO marketplace. A total of 25 TP’s were involved. The User Profile Editor was considered easy to use, but the majority of the users experience inconsistencies in the user interface. Suggestions are given to improve usability. The ALFRED T-shirt was received positive as users liked the idea to check their health status, but doubts were raised on data sharing. Different recommendations are given to improve usability and improve data transparency. The Dancicians game (before Dance with ALFRED) remains difficult for users. Although many improvements were made compared to Cycle 2, there are still some usability issues that must be tackled. Finally the ALFREDO marketplace received very good ratings and only minor recommendations.

Cycle 4 and 5 focus on two versions of the integrated ALFRED prototype. In these cycles the same methods were used as in Cycle 2 and 3, but the definition of the task success

D8.2.2 Piloting & Validation I: Individual Usability	Document Version: 1.0	Date: 2016-09-23	Status: For Approval	Page: 4 / 43
http://www.alfred.eu/		Copyright © ALFRED Project Consortium. All Rights Reserved. Grant Agreement No.: 611218		

rate was added. These simplified success rates provide a general picture of how the ALFRED system and different apps are working. As we are working with a small number of observations and rough estimate of partial success scores they do not give a complete picture, but an indication on the improvement needed to make the ALFRED system work properly. Instead of reporting the findings and recommendations in tables to the partners, the end user partners worked directly in Gitlab, so that all feedback could be taken up more quickly and effectively.

The fourth Iterative Cycle, which was conducted in May 2016 proved to be rather challenging due to technical problems in the integration of the ALFRED system. A total of 10 TPs participated in an individual session where they performed 14 tasks. The overall success rate was 36,79%. Looking at the overall satisfaction rate this quite low. Feedback and recommendations were collected continuously by end user partners and taken up and implemented by the technical partners. Commands were extended to offer end users a wider range of interaction.

The fifth Iterative Cycle was implemented additionally in a short cycle of improvement on the integrated version of the ALFRED system. The sessions were implemented previously to the start of the pilot with the same TPs. A total of 19 TPs were involved. The success rates improved considerably to 72.9%. This was also reflected by the PSSUQ with a slight improvement in overall satisfaction.

Based on the PSSUQ ratings it can be concluded that the assumptions for the project were correct and the results are good and valid. When compared to the ambitious KPI's previously defined it cannot be unnoticed that they have been only partially met. Finally the Iterative Evaluations show that early prototype testing with users is very valuable for the prototypes as the developers are able to obtain user recommendations and start adapting their solution at an early stage of the development. This early testing helps also the end-user partners to plan and define the ideal final validation (pilot) test settings, as the researchers are very much involved in the development and have seen the earlier versions of the system components. The pilots, as the technical development, have been developed in an iterative process adjusting to the reality of each moment.

D8.2.2 Piloting & Validation I: Individual Usability	Document Version: 1.0	Date: 2016-09-23	Status: For Approval	Page: 5 / 43
http://www.alfred.eu/		Copyright © ALFRED Project Consortium. All Rights Reserved. Grant Agreement No.: 611218		

Table of Contents

1	Introduction	9
1.1	ALFRED Project Overview	9
1.2	Deliverable Purpose, Scope and Context	10
1.3	Document Status and Target Audience	10
1.4	Abbreviations and Glossary	10
1.5	Document Structure	10
2	Key Findings from Iterative Evaluations	11
2.1	Iterative Evaluation Cycle 2	11
2.1.1	Summary	11
2.1.2	Participants	12
2.1.3	Evaluation Tasks and Results	13
2.1.3.1	ALFRED Sensor T-shirt	13
2.1.3.2	ALFREDO Marketplace	14
2.1.3.3	Dance with ALFRED	14
2.1.4	ASQ Ratings	15
2.1.5	PSSUQ Ratings	16
2.2	Iterative Evaluations Cycle 3	17
2.2.1	Summary	18
2.2.2	Participants	18
2.2.3	Evaluation Tasks and Results	19
2.2.3.1	User Profile Editor	19
2.2.3.2	Sensor T-shirt	20
2.2.3.3	Dancicians	20
2.2.3.4	The ALFREDO Marketplace	21
2.2.4	PSSUQ Rating	21
3	Key findings from Iterative Evaluations on the Integrated Prototype	23
3.1	Integrated version of ALFRED	23
3.2	Methodology	25
3.2.1	Overall methodology	25
3.2.2	Success Rate	25
3.2.3	Analysis	26
3.3	Iterative Evaluation Cycle 4	26
3.3.1	Summary	26
3.3.2	Participants	27
3.3.3	Evaluation Tasks and Results	28
3.3.3.1	Findings and recommendations	28
3.3.3.2	PSSUQ ratings	28
3.4	Iterative Evaluations Cycle 5	29
3.4.1	Summary	29
3.4.2	Participants	29
3.4.3	Drop out	30

3.4.4	Evaluation Tasks and Results.....	30
3.4.4.1	Findings and Recommendations	30
3.4.4.2	PSSUQ Rating.....	31
4	Results from the Usability Perspective.....	32
4.1	Overall results on the Iterative Cycles	32
4.1.1	PSSUQ Rating	32
4.2	KPI's of the Usability Perspective	33
4.2.1	KPI 1: Positive Feedback in the Validation Groups.....	34
4.2.2	KPI 2: Improvement of Positive Rating in Each Feedback Loop.....	34
4.2.3	KPI 3: Rise of PSSUQ Satisfaction Levels.....	34
4.2.4	KPI 4: Success Rates	35
5	Conclusion	36

List of Figures, Tables and Listings

List of tables

Table 1: Test Participants in Cycle 2	11
Table 2: ASQ Ratings in Cycle 2	15
Table 3: PSSUQ Ratings Cycle 2	17
Table 4: Test Participants in Cycle 3	18
Table 5: PSSUQ Ratings Cycle 3	22
Table 6: Apps of the Integrated ALFRED Prototype	23
Table 7: TPs Evaluation Cycle 4	27
Table 8: TP Experience in Use of Technology	27
Table 9: PSSUQ ratings.....	28
Table 10: TPs Evaluation Cycle	29
Table 11: TP Experience in Use of Technology	30
Table 12: Use of Mobile Device	30
Table 13: PSSUQ Ratings Cycle 4 and 5	31
Table 14: KPIs of the Usability Perspective	34

List of Figures

Figure 1: ALFRED Personal Assistant	24
Figure 2: Gitlab Findings and Recommendations	26
Figure 3: Iterative Evaluation Cycles	32
Figure 4: Overall PSSUQ Ratings.....	33

List of Annexes

Annex 1: Overview Test Participants Cycle 2	37
Annex 2: Overview Test Participants Cycle 3	38
Annex 3: Test Persons Evaluation Cycle 4	40
Annex 4 Test Persons Evaluation Cycle 5	41

1 Introduction

ALFRED – Personal Interactive Assistant for Independent Living and Active Ageing – is a project funded by the Seventh Framework Programme of the European Commission under Grant Agreement No. 611218. It will allow elderly people to live longer at their own homes with the possibility to act independently and to actively participate in society by providing the technological foundation for an ecosystem consisting out of four pillars:

- **User-Driven Interaction Assistant** to allow older people to “talk” to ALFRED and to ask questions or define commands in order to solve day-to-day problems.
- **Personalized Social Inclusion** by suggesting social events to older people, considering his interests and his social environment.
- A more **Effective & Personalized Care** by allowing medical staff or carer to access vital signs of older people monitored by (wearable) sensors.
- **Physical & Cognitive Impairments Prevention** by incorporating serious gaming to improve the physical and cognitive condition by offering games and quests to older people.

1.1 ALFRED Project Overview

One of the major problems today is the increasing isolation of older people, who do not actively participate in society either because of missing social interactions or because of age-related impairments (physical or cognitive). ALFRED will allow overcoming this problem with an interactive virtual butler for older people, which is fully voice controlled.

The ALFRED project is wrapped around the following very clear main objectives:

- Empowering people with age related dependencies to live independently for longer by delivering a virtual butler with seamless support for tasks in and outside the home. The virtual butler ALFRED will have a very high end-user acceptance by using a fully voice controlled and non-technical environment.
- Prevailing age-related physical and cognitive impairments with the help of personalized, serious games.
- Fostering active participation in society for the ageing population by suggesting and managing events and social contacts.
- Improved care process through direct access to vital signs for carers and other medical staff as well as alerting in case of emergencies. The data is collected by unobtrusive wearable sensors monitoring the vital signs of older people.

To achieve its goals, the project ALFRED conducts original research and applies technologies from the fields of Ubiquitous Computing, Big Data, Serious Gaming, the Semantic Web, Cyber Physical Systems, the Internet of Things, the Internet of Services, and Human-Computer Interaction. For more information, please refer to the project website at <http://www.alfred.eu>.

D8.2.2 Piloting & Validation I: Individual Usability	Document Version: 1.0	Date: 2016-09-23	Status: For Approval	Page: 9 / 43
http://www.alfred.eu/		Copyright © ALFRED Project Consortium. All Rights Reserved. Grant Agreement No.: 611218		

1.2 Deliverable Purpose, Scope and Context

This deliverable gives the results of the second and third iterative user evaluations that were made on different ALFRED components with older end users. It gives a set of recommendations in simple tables that are used by the technical partners to improve these components. The deliverable proceeds with detailing the final iterative evaluation on the Integrated ALFRED system within the usability perspective and plans the three pilots. Pilot 1 in the Netherlands and pilot 3 in France will run the pilot on the Integrated ALFRED system and Pilot 2 in Germany will focus on the health aspects of ALFRED with specialized medical staff.

1.3 Document Status and Target Audience

This document is listed in the Description of Work (DoW) as “confidential”, as it provides confidential information business components of ALFRED that are key to exploitation and can therefore not be used by external parties.

1.4 Abbreviations and Glossary

A definition of common terms and roles related to the realization of ALFRED as well as a list of abbreviations is available in the supplementary document “Supplement: Abbreviations and Glossary”, which is provided in addition to this deliverable.

Further information can be found at <http://www.alfred.eu>.

1.5 Document Structure

The following Chapter 2 describes the results of the second iterative evaluation which was in the spring of 2015 with older end users in Germany, France and the Netherlands. The evaluation included different ALFRED components, such as the ALFREDO market place, the Dancicians game and the sensor T-shirt. Based on this, a set of recommendations was developed that was communicated back to the technical partners. The third evaluation sessions took place in November and December of 2015 in the same countries and also included different ALFRED components, including notably the profile editor and the health monitoring app. The results of these evaluations are also provided to the technical partners as recommendations for the integrated ALFRED system.

Following up on D8.1.2, chapter 3 then proceeds to describe the detailed pilot methodology. It first describes the pilot methodology in Pilot 3 France and Pilot 1 the Netherlands, where the integrated version will first be tested in a final usability session to define the success rate of the tasks and then will be tested in the homes of older end users. The final section will describe the methodology for the German pilot which will focus on health aspects of the ALFRED system.

D8.2.2 Piloting & Validation I: Individual Usability	Document Version: 1.0	Date: 2016-09-23	Status: For Approval	Page: 10 / 43
http://www.alfred.eu/		Copyright © ALFRED Project Consortium. All Rights Reserved. Grant Agreement No.: 611218		

2 Key Findings from Iterative Evaluations

Within the ALFRED project the development of prototypes is accompanied by iterative evaluations. These evaluations are a continuous process to ensure that project results can be easily used by older people. Test Participants (TPs), of all target user groups are recruited by the end user partners (France, Germany and the Netherlands).

The first Iterative evaluation was in M12 and results were reported in D8.1.2. The process of Iterative Evaluations went on and this section will give the results for Cycle 2 (in Spring 2015) and Cycle 3 (in Fall 2015).

2.1 Iterative Evaluation Cycle 2

The aim is to give the key findings of cycle 2 user studies, organised in spring 2015. The details of the used methodologies can be found in D8.1.2 and in the cycle 2 workbook (see Annex 1). All data from the three countries was collected in a common Excel file and then analysed by the responsible end-user partners.

2.1.1 Summary

This section contains the results of the second prototype evaluation, including the ALFREDO marketplace, the game “Dance with ALFRED” and the ALFRED Sensor T-shirt. The After Scenario Questionnaire (ASQ) and Post-Study System Usability Questionnaire (PSSUQ) were used to evaluate the usefulness and usability of the currently available prototypes with older adults.

Altogether 12 (9 females, 3 males) seniors participated in the testing. Table 1 shows general information about the Test Participants in Cycle 2.

Table 1: Test Participants in Cycle 2

	Netherlands (NFE)	France (ESE)	Germany (CHA)	Total
Number of participants	6	3	3	12
Number of Female and Male participants	4(F), 2(M))	2 (F), 1 (M)	3 (F)	9(F), 3(M)
Average age of participants	74	70.3	73	72.4

The users had to perform 4 tasks with the ALFREDO Marketplace. 2 tasks and 6 questions had to be answered for the dancing game and finally 2 tasks and 9 questions needed to be answered for the Sensor T-shirt.

The ALFREDO Marketplace received a good user rating in general, the main issues were related to navigational and data entry problems. Most participants enjoyed playing the “Dance with ALFRED” game and especially had fun with the multiplayer modus. On the

D8.2.2 Piloting & Validation I: Individual Usability	Document Version: 1.0	Date: 2016-09-23	Status: For Approval	Page: 11 / 43
http://www.alfred.eu/		Copyright © ALFRED Project Consortium. All Rights Reserved. Grant Agreement No.: 611218		

other hand, more challenging levels and a better introduction on how to play the game were desired. Despite the necessary tight fit of the ALFRED Sensor T-shirt, the participants had no trouble when they were putting on the T-shirt, also the ALFRED logo on the shirt received positive remarks. The main problems with the T-shirt was the position of the sensor unit in front of the chest. This and the thick material caused a discomfort and lead to excessive sweating of the users.

In the second part of the test, the ALFREDO marketplace was tested, which has been done with four tasks. Moreover, each individual task was evaluated by the tester with the ASQ and the PSSUQ was filled in after the second part of the task sets. The PSSUQ consisted of 19 questions (one overall satisfaction question was added) divided by four sub-measures (Both the ASQ and the PSSUQ are fully described in D8.1. Annex A for the ASQ and Annex B for the PSSUQ questionnaires).

The third and final part of the test consisted of the game 'Dance with ALFRED'. After finishing these questions, participants were thanked for their help and the test was over. In the Netherlands and France, the average sessions took around 45 minutes. In Germany, sessions took 20 min longer, as the T-shirt was tested as well.

2.1.2 Participants

Test participants are recruited by the end user partners (France, Germany and the Netherlands), using national projects and services to reach them. All participants visited the end users offices to be part of the iterative evaluations. Participants were guided with a workbook (including an Informed Consent, see Annex 2).

During the intake information about the participants was gathered regarding their year of birth, gender, nationality, household income, marital status, living situation, education, employment, residential situation, self-rated health, vision, hearing, motor control and the self-rated mobility level.

In the intake questionnaire participants were also asked about their technological attitude and behaviour. 7 Test Persons (TP) had a mobile phone, 4 had an Android smartphone, and one had a senior smartphone. 8 participants had a positive attitude towards technology and 4 participants had a neutral attitude towards it.

4 participants had a lot of experience with technology, 4 considered themselves to be average and 4 participants said they had little experience with technology. This corresponded with the results to the question "*What is your use of ICT Tools (computer, smartphone, tablets?*", where 6 participants answered that they used technology on a daily base and 3 participants said they used technology on a weekly base ("*I check my e-mail every Tuesday, but that is enough for me*") and 2 used their mobile phone on rare occasions only. The online activities of the participants differ; checking e-mail and searching the web (public transport information, road directions, looking up the news) are the main activities. Participants with a smartphone also use the text messaging or WhatsApp. Other activities mentioned were online shopping and playing games.

Finally, as games were a part of this testing round, participants were also asked some questions related to this. Four participants loved to play games, two thought it was fun once in a while, three people had a neutral opinion about playing games, one participant did not really like it and two participants really did not like playing games. 10 participants played games sometimes and two indicated playing games very often. When asked what

D8.2.2 Piloting & Validation I: Individual Usability	Document Version: 1.0	Date: 2016-09-23	Status: For Approval	Page: 12 / 43
http://www.alfred.eu/		Copyright © ALFRED Project Consortium. All Rights Reserved. Grant Agreement No.: 611218		

games participants played (multiple answers were possible), eight participants mentioned card games, five mentioned board games, and three mentioned puzzles and one mentioned parlour games, either online or offline. In Annex 2 “Overview of Test Participants Cycle 2” the extended information on the TP’s can be found.

2.1.3 Evaluation Tasks and Results

This section contains the results of the second prototype evaluation, including the Sensor T-shirt, the ALFREDO Marketplace and the Dance with ALFRED Game. The specific tasks were defined together with the partners developing each component, once it was clear that which app prototypes and which of their functionalities were ready to be tested. Moreover, the technical partners also detailed the type of end-user feedback would be the most valuable.

The evaluation was performed with two test administrators guiding the session and one test person. One person guided the user through the questionnaires and questions, another observed, helped and monitored the different tasks that needed to be done.

Based on the findings, recommendations are made for improvements of each part of the ALFRED system that was tested. Recommendations are given based on the problem identified by the user, and classified based on priority. Priority is given to certain problems based on impact, frequency of occurrence and persistence of an error. Some TP’s suggestions for improvements are also included in the recommendations.

2.1.3.1 ALFRED Sensor T-shirt

The first part of the test consisted of tasks about the ALFRED sensor T-shirt. This component is focused on health data and therefore, for the Cycle 2, the early smart t-shirt prototype was only tested in Germany (at CHA) where medical staff was present. The users had to perform 2 tasks with 9 questions to answer. They needed to try on the T-shirt and to activate the sensors. Users had the choice between two sizes of the T-shirt.

In order to gain a more detailed insight on how satisfied the users were with the sensor T-shirt, they were instructed to wear it for a period of one day (12 hours); or as long as they could tolerate to wear the shirt in their daily environment while performing everyday activities. After sending the material (Sensor T-shirt and questionnaire) back to the end users office, they participate a short telephone interview to gather additional comments.

In the current test 3 users were tested, 2 of them normally wear T-shirts in size small and 1 in a size medium. The users had little trouble putting on the T-shirt despite the necessary tight fit of the shirt. This was an initial concern of the seniors and also the end-user organisation, since putting on a compression type shirt like the ALFRED sensor T-shirt requires a great amount of flexibility in the shoulder joints.

The activation switch of the sensor unit of the T-shirt is not suitable for seniors due to its small size and poor accessibility. The transmission of the signal from T-shirt to the Nexus 5 smart phone worked without any problems during the test session. However, the position of the sensor unit on the T-shirt has to be changed for comfort and aesthetic reasons. Future sensor T-shirts should be made out of thinner and cooler fabric with better air permeability. Alternatively, the T-shirt could be substituted by a chest belt (which can be made from the same material as the T-shirt) which only covers the area where the sensors need to have contact with the skin of the user. This could reduce the heat which gets

D8.2.2 Piloting & Validation I: Individual Usability	Document Version: 1.0	Date: 2016-09-23	Status: For Approval	Page: 13 / 43
http://www.alfred.eu/		Copyright © ALFRED Project Consortium. All Rights Reserved. Grant Agreement No.: 611218		

trapped under the current T-shirt and avoid issues with putting on and off the T-shirt in users with reduced flexibility of the shoulder joints. A sleeveless shirt would be another option to further reduce excessive sweating experienced by the users

The transmission of the vital parameters and body position from the T-shirt to the Nexus smartphone worked equally well for both participants regardless of the tested size of the T-shirt. With the currently described issues the users would not want to wear the sensor T-shirt for a prolonged period of time.

2.1.3.2 ALFREDO Marketplace

The users had to perform 4 tasks with 12 questions to answer. They needed to change their username, check the apps that were already downloaded, check the new apps are available in the Marketplace as well as try to update current apps.

The concept of a marketplace was unfamiliar to most users and had to be explained before the test by the test administrators. When starting with the first task, a lot of participants were nervous, as they were not sure what to expect of an online marketplace. However, most participants found the tasks not as hard as expected. Unfortunately, there were technological issues which made it impossible for the participant to complete the first task by him/herself; in all cases the test administrator had to intervene and point out the way the name could be changed. Regarding the second, third and fourth task, some participants experienced difficulties understanding the concept of an application. For some participants, the questions itself were difficult and so the test administrator had to explain the meaning of the question beforehand. For a lot of participants (especially those who did not have a smartphone), the idea of 'my apps', 'newest apps' and 'update apps' was difficult to understand, as they did not know what an app is.

Except for the previous comments, users were in general quite positive about the size of the buttons and the icons. The buttons were big enough and the icons were clear. Participants especially liked that there was a description below every icon, even though they did not know what they all meant (e.g. my apps). In general, participants liked the design of the application, the colours, the size of the buttons and the pictograms used – even if they could not extract the exact meaning from them right away.

2.1.3.3 Dance with ALFRED

The users had to perform two tasks and answer 6 open questions in order to get important feedback for the Dance with ALFRED game: the first was to just play the game on medium level as an introduction. The second was to play the game again, on the same level or on easy or hard mode, depending on the participants' performance before. The participants chose an instrument and started playing autonomously; the test administrator monitored the experiment.

The final number of participants for this part of the test was 11, because one of the participants left after finishing the tasks for the ALFREDO Marketplace. Most remaining participants had fun trying the game and were enthusiastic about this concept, as some of them did not expect it to be for older people. They all tried the game in medium modus first. In the second try, eight participants played the game in hard mode, as they wanted a bigger challenge. The other three participants played the game in medium mode for a second time.

D8.2.2 Piloting & Validation I: Individual Usability	Document Version: 1.0	Date: 2016-09-23	Status: For Approval	Page: 14 / 43
http://www.alfred.eu/		Copyright © ALFRED Project Consortium. All Rights Reserved. Grant Agreement No.: 611218		

When participants were asked what they liked about the game, almost everyone answered that they liked to play it together with someone else. Others liked the aspect of exercising: “*I have a home trainer but this is way more fun!*” In order to improve the accessibility of the game, a short introduction on how to play the game should be included.

2.1.4 ASQ Ratings

First and foremost, ASQ ratings show that users in general were satisfied with *the ease of completing the tasks* 2, 3, and 4, time it took them complete it and the information they received from the screen. All mean values were below 3, with the highest mean value being 2.92 and the lowest 1.83. Considering that the ASQ has a rating from 1-7 where 1 means a high satisfaction and 7 means a low satisfaction, this rating can be considered to be a high general satisfaction value. Lowest values were found for *the ease of task completion* ($M = 1.83$, $SD = 0.94$) and ($M = 2.25$, $SD = 1.91$) of the second on fourth task. In general, the participants found it easy to find the place to download new apps, even if they did not succeed immediately. Task 1 only showed a medium satisfaction ranging from 3.33 to 3.75. By looking at the values it can be seen that the German users were particularly dissatisfied with this task by rating it either with a score of 6 or 7. The reasons for this were mainly the small size of the keyboard and the wrong ordering of the buttons in order to change the user name.

Additional reasons are probably also that these users all had little ($N = 2$) or medium ($N = 1$) and only used ICT tools rare ($N = 2$) or once a week ($N = 1$) technological experience and did own basic mobile phones. The Dutch and French users had fewer problems with this task and three of the Dutch users even were very satisfied with this task, which can be explained with the fact that in these countries the testers were more familiar with the usage of ICT-tools. Dutch and French users also used ICT tools more often and 6 users use it daily and 3 weekly. The difference in the satisfaction in these tasks between German and French users can also be seen in the large variance in task 1, which ranges from 4.06 on the amount of time that it took users to complete the task to 4.57 to the amount of supporting information that was provided.

The mean values for supportive information while completing the tasks, were higher for task 2 ($M = 2.92$, $SD = 3.72$), task 3 ($M = 2.75$, $SD = 1.84$) and task 4 ($M = 2.50$, $SD = 1.55$). The marketplace doesn't provide any support information in case the users had navigational problems. The fact that for some users the keys of the Nexus phone were not clear confirms this. The mean ASQ score of task 4 (2.25) was lower than in task 3 (2.7) and task 2 (2.53). The users were probably more familiar with the app in the last task and therefore the lowest values were experienced here. This corresponds with the finding that it was not difficult to find the right place for app updates but there were no applications that needed to be renewed. Participants did not understand this from the instructions given on the screen and often thought they were in the wrong place.

The following table summarizes these findings.

Table 2: ASQ Ratings in Cycle 2

ASQ Ratings (1=high satisfaction, 7=low satisfaction)	
Parameters	M (mean)

Ease of task completion	
Task 1	3.53
Task 2	1.83
Task 3	2.7
Task 4	2.25
Ease of supporting information while completing	
Task 1	2.92
Task 2	2.75
Task 4	2.50
ASQ Score	
Task 2	2.53
Task 3	2.7
Task 4	2.25

2.1.5 PSSUQ Ratings

The results of the PSSUQ can be divided in four subsections, of which the first entails the overall satisfaction of the user with the system and is calculated by the mean of all 19 questions. The rating (1 – 7) is the same as for the ASQ. The overall satisfaction was rather high ($M = 2.72$, $SD = 0.48$) which indicates that users were in general very satisfied with the ALFREDO marketplace.

The second subsection, system quality, is derived from question 1 – 8. It measures whether the system is efficient and whether users feel comfortable using it. Values were low ($M = 2.52$, $SD = 0.28$), which means users were satisfied. Especially questions regarding the feeling on how comfortable the system can be used ($M = 2.17$, $SD = 2.15$) and one's believe of being able to learn how to use the system in a quick and easy way ($M = 2.08$, $SD = 2.08$) had lower values.

The third subsection regarding information quality is being calculated by the answers to question 9 – 15 and shows how well a system is capable of providing the right kind and amount of information to participants. Average values showed a positive feeling ($M = 3.0$, $SD = 0.49$), however the highest values (lowest satisfaction) were found in this section. Information about clear error messages ($M = 3.75$, $SD = 0.93$) and help from the system when a mistake was made ($M = 3.75$, $SD = 1.66$) were rated higher. Ratings were still in the positive aspect of satisfaction (below a value of 4) but rated less positive than the other questions. There are two explanations for this. First of all, a lot of users were confused when they could not find anything (e.g. in the last task) and they found the information (“no updates found”) not sufficient enough, as they were not aware that they did not have any applications downloaded at all. A second explanation is that a lot of users were unsure of

having seen any error messages (*"I don't think I have seen any error messages so I don't know how to answer this question"*). Therefore, these participants rated the question with the middle, neutral answer, which is rated a 4 in the statistical procedure.

The final subsection of the PSSUQ is about the internal quality of the system. Does the system have all the right capabilities and does it have a good interface? The average value showed that users were in general satisfied with the internal quality of the system ($M = 2.6$, $SD = 0.47$). They were especially positive about the interface ($M = 2.08$, $SD = 1.72$), which corresponds with findings of users who liked the colours and thought the pictograms with words under them were very clear and understandable.

The following table summarizes the findings based on the PSSUQ.

Table 3: PSSUQ Ratings Cycle 2

PSSUQ Ratings (1=high satisfaction, 7=low satisfaction)	
Parameters	M (mean)
Overall satisfaction (Q1-Q19)	2.72
System quality (Q1-Q8)	
Total average value	2.52
How comfortable can the system be used	2.17
Belief to become quickly productive with the system	2.08
Information quality (Q9-Q15)	
Total average value	3.0
Clear error messages	3.75
Support for mistakes	3.75
Internal quality of the system (Q16-Q19)	
Total average value	2.6
Interface	2.08

2.2 Iterative Evaluations Cycle 3

The aim is to give the key findings of the user studies in Cycle 3, organised in fall 2015. The details of the used methodology can be found in D8.1.2 and the Cycle 3 workbook (see Annex 3) that gives the predefined tasks performed by the users. The results of the study were collected and analysed for description in the following section.

2.2.1 Summary

The Iterative Evaluation Cycle 3 was done in France and in the Netherlands in November 2015 and in Germany in December 2015. This section contains the results of the third prototype evaluation, including the User Profile Editor, the ALFRED T-shirt and the connected health app, the Dancicians game and the ALFREDO marketplace. The Post-Study System Usability Questionnaire (PSSUQ) was used to evaluate the usefulness and usability of the currently available prototypes with older adults. Altogether 25 (22 females and 3 males) participated in the testing in the Netherlands, France and Germany.

In the User Profile Editor (developed by TIE/WP5) users needed to create, edit and update his/her own user profile with a visual user-interface. The ALFRED T-shirt (developed by AITEX/WP6) was worn by users to measure vital data and to track on his/her health status with the connected mobile app. The Dancicians game (developed by TUDA/WP7) challenged the users to move with a sound stimulating body and mind and the ALFREDO Marketplace (developed by WORLD/WP3) users could check the availability of new apps. The recommendations of all the tested applications are sorted in tables and prioritized by how many times a recommendation was mentioned by the participants.

2.2.2 Participants

Representatives from the primary target groups were involved in the testing as demonstrated by the following table:

Table 4: Test Participants in Cycle 3

	Netherlands (NFE)	France (ESE)	Germany (CHA)	Total
Number of participants	19	3	3	25
Number of Female and Male participants	17(F) 2(M)	2(F) 1(M)	3(F) 0(M)	22(F) 3(M)
Average age of participants	75	68	73	72

Five of the participants had a smartphone (four Android, one iOS). 18 Participants had a normal mobile phone and 2 seniors only had a home phone. From the total of participants only 2 people said to have a negative attitude towards technology (technology makes place of social inclusion), 14 people had a positive attitude towards technology, 9 said to be neutral. There were huge differences in the technological experiences of the participants. Only 3 rated themselves very high experienced, 9 high experienced, 7 medium experienced, 3 low experienced and 3 very low experienced. Most of the participants make daily use (16) of different ICT tools and although the amount and type of activities differ, most of the participants use ICT for e-mail and/or playing games and/or in general web searching.

The initial purpose (as planned in D8.1.2) of involving more participants in the Dutch evaluation was to have a more reliable test case on the speech interaction within the different applications and draw conclusion and develop guidelines on development of speech interaction. Due to delay of the technical delivery, the testing of speech interaction is postponed, and an additional iterative testing cycle of ALFRED Alpha version will be implemented just before the pilot in France and the Netherlands.

It was decided to maintain the initial plan of the bigger test pool in the Netherlands, as an added value compared to the earlier Cycle 1 and Cycle 2. The large test pool in Cycle 3, made it possible to test the ALFRED components with different ALFRED target groups (see the group definitions in D2.3).

When we look at the three defined target groups, 8 participants belong to the Primary Target Group (PTG)¹ PTG1² and 10 participants to the PTG2³ or PTG3⁴. In France all the three participants belong to the PTG1 but one of the testers' is also Secondary Target Group (STG)⁵ as she is taking care of a family member. In Germany three participants were part of the predefined PTG1 group and one user was taking care of both her parents so he can be considered a user of the STG1 group as well. 24 of the participants live independently at their home, one of them in a home for older people. Annex 4 (Overview of test participants Cycle 3) gives further information on the TP's characteristics.

2.2.3 Evaluation Tasks and Results

This section contains the results of the third prototype evaluation, including the User Profile Editor, the ALFRED Sensor T-shirt, the game 'Dancicians' and the ALFREDO Marketplace. The Post-Study System Usability Questionnaire (PSSUQ) was used to evaluate the usefulness and usability of the currently available prototypes with older adults.

In the result analysis focuses on providing testers' recommendations for the technical partners and not on defining task success rates. Moreover, this Evaluation Cycle is focused on different components of the ALFRED system and therefore there is no added value in measuring success rates. In the final Iterative Evaluation success rates will be included on each of the tasks, taking also up the comments of the EC reviewers.

The evaluation was performed with two test administrators guiding the session and one participant. One person guided the user through the questionnaires and questions, another observed, helped and monitored the different tasks that needed to be done.

2.2.3.1 User Profile Editor

This application was tested the first time with users in this cycle. The users had to perform 5 tasks with 7 questions to answer with the User Profile Editor. They needed to register and log in, to edit and use their own profile, to add a new contact, to look for their own health statistics and to log off (see Annex 3 for further task definitions).

¹ See the specific definitions in D2.3 User Stories Report.

² Independently living and autonomous older adults 60+

³ Independently living older adults 60+ having informal caregiver's support

⁴ Independently living older adults 60+ having formal caregiver's support

⁵ Informal caregiver

D8.2.2 Piloting & Validation I: Individual Usability	Document Version: 1.0	Date: 2016-09-23	Status: For Approval	Page: 19 / 43
http://www.alfred.eu/		Copyright © ALFRED Project Consortium. All Rights Reserved. Grant Agreement No.: 611218		

In general, the users considered that the user profile was relatively easy to create, but the majority of the users experience inconsistencies in the user interface: the way they needed to save changes and to add information in an easier way. Also the language (only an English version) was experienced by most participants as a problem. Moreover, some suggestions were given to improve the usability. For the User Profile Editor, recommendations are the following:

2.2.3.2 Sensor T-shirt

The ALFRED T-shirt had already been tested in Germany during the iterative testing cycle 2 but this was the first time in France and in the Netherlands. All the three countries tested the linked health data app for the first time.

The users had to perform 3 tasks with 11 questions to answer regarding the comfort and usability of the Sensor T-shirt (see the tasks in the workbook in Annex 3). They needed to try on the T-shirt, to insert and activate the sensors and to connect the T-shirt with the ALFRED application. Although the all testers agreed to try on the T-shirt, the results do not confirm that everybody is ready to wear it on a more regular basis to monitor their health. An often mentioned sentence was “*if I get more complaints*”, “*if my health gets worse*” or “*if it is necessary*”. In consequence, the most testers would agree to wear a smart T-shirt only under certain conditions, notably if they have health problems.

A special attention must be paid to t-shirt sizes as those are sized very small. Moreover, the testers liked the app related to the t-shirt and the idea to be able to check on their health status and be informed about potential changes in their health but they want to be able to decide with whom this data is shared.

2.2.3.3 Dancicians

The previous version of this game was called “Dance with ALFRED” and it was tested in the three countries in the iterative testing cycle 2, and in the cycle 3 a new version of the game “Dancicians” was tested in these three countries.

The users had to perform 3 tasks with 6 questions to answer about the experiences of the exergame. The users were asked to play the game in the offline mode and to choose a (one of the five options) sound. The second task was to restart the game and to choose another instrument and the third task was to play the game in the online mode, with another player (the researcher) (see Annex 4 for the task definitions).

Not all participants performed the three different tasks, because they experienced motivational difficulties. The main problem was that the music stopped when participants made a mistake in the game, as some kind of playful punishment. It would be highly suggested to keep the music playing even if an incorrect move of the player is detected and just give visual feedback in case an error was detected by the game. On the other hand, the participants also liked the game, because of the challenge in making the right move (with another person), it was comparable with the gym class some participants were in.

Even though improvements have been made since the very first version of the game (i.e. Dance with ALFRED), this game (Dancicians) remains difficult for many players. The game gives instructions to do movements but it is often not sensing correctly when the

D8.2.2 Piloting & Validation I: Individual Usability	Document Version: 1.0	Date: 2016-09-23	Status: For Approval	Page: 20 / 43
http://www.alfred.eu/		Copyright © ALFRED Project Consortium. All Rights Reserved. Grant Agreement No.: 611218		

player is moving, which can easily decrease the player's motivation. However, the effort made in the game layout and musical options were appreciated by the testers.

The following tables lists the users' recommendations for Dancicians game:

2.2.3.4 The ALFREDO Marketplace

This app was only tested in Germany due to timing issues. The following only gives the results from Germany. The ALFREDO marketplace for older adults was tested with the representatives of primary target group's users, whereas the ALFREDO marketplace for caregivers was tested by the researchers from the end-user partner organizations.

The users had to perform 4 tasks and to answer 12 questions. They were asked to change their username, to find the list with the apps already downloaded, to see possible new apps available for download and to find the possibility to update apps (into newer versions).

Furthermore, the end-user partner researchers tested the second version of the ALFREDO marketplace designed for the caregivers that are the STG of the project. The following table lists the recommendations from the researches. Moreover, this application will be tested by the actual STG during the forthcoming ALFRED pilot phase aiming to validate the integrated ALFRED system.

2.2.4 PSSUQ Rating

A description on how the PSSUQ ratings can be interpreted was already previously given in section 2.1.5.2.2. The testers were asked to fill in the PSSUQ questionnaire in the very end of the testing session. In this cycle we tested the overall satisfaction of the user with the system which. This showed similar values ($M=2.9$, $SD=1.51$) compared to the last testing cycle. The second subsection is calculated by the mean values from questions 1-8 measures whether the system is efficient and whether users feel comfortable using it. This cycle has slightly worsened the values ($M=2.9$, $SD=1.5$) compared to the previous iteration ($M=2.52$, $SD=0.28$). The third subsection of the PSSUQ questionnaire regards information quality and is being calculated by the answers to question 9 – 15 and shows how well a system is capable of providing the right kind and amount of information to participants. Values were on a similar level ($M=2.9$, $SD=1.53$) compared to the last test with ALFRED ($M=3.0$, $SD=0.49$). The final subsection of the PSSUQ is about the internal quality of the system. Does the system have all the right capabilities and does it have a good interface? Also in this section of the PSSUQ the values of the previous ($M=2.6$, $SD=0.47$) tests are fairly close to the observations which were made in the last test session ($M=2.8$, $SD=1.48$).

The user ratings of ALFRED remain on a high level and confirm the results from the previous tests that were conducted and show that the different ALFRED components that were tested already can provide a high user satisfaction besides the relative immaturity of some of the functions which were tested. Despite the fact that no major changes, neither on the negative nor the positive side could be observed, the results are still promising and show that even in a broader population of seniors, the users still did not experience any significant problems using ALFRED apps. Future tests in the ALFRED pilots will evaluate an integrated version of ALFRED which is likely to open these users a broader range of potential benefits.

D8.2.2 Piloting & Validation I: Individual Usability	Document Version: 1.0	Date: 2016-09-23	Status: For Approval	Page: 21 / 43
http://www.alfred.eu/		Copyright © ALFRED Project Consortium. All Rights Reserved. Grant Agreement No.: 611218		

Table 5: PSSUQ Ratings Cycle 3

PSSUQ Ratings (1=high satisfaction, 7=low satisfaction)	
Parameters	M (mean)
Overall satisfaction (Q1-Q19)	2.9
System quality (Q1-Q8)	
Total average value	2.9
Information quality (Q9-Q15)	
Total average value	2.9
Internal quality of the system (Q16-Q19)	
Total average value	2.6

3 Key findings from Iterative Evaluations on the Integrated Prototype

The aim of this chapter is to give the key findings of the user studies performed on two integrated versions of ALFRED prototypes. The chapter will first present what the prototypes entailed. The methodology will be shortly explained and the results of the Fourth (May 2016) and Fifth (June 2016) Evaluation Cycles will be presented.

3.1 Integrated version of ALFRED

In the Integrated ALFRED prototype of May 2016, a total of 14 apps were available. These apps were accessible through voice commands through the Personal Assistant and were developed within the four ALFRED pillars.

gives an overview of the apps.

Table 6: Apps of the Integrated ALFRED Prototype

App		Main functionality
1	Tutorial app	ALFRED introduces himself and gives basic instructions on how to be used
2	Alarm clock	ALFRED asks how the user is feeling today and based on the answer makes a suggestion for an event or to call someone.
3	Location app	The user can ask ALFRED where he is, who then explains the exact location.
4	Reminder app	The user can ask ALFRED to set a reminder for a certain period of time.
5	Posture app	ALFRED gives suggestions for good postures
6	Battery app	ALFRED indicates the battery status
7	Agenda app	User can set an appointment in his agenda
8	Microphone app	User can set ALFRED to his preferences by changing the microphone colour
9	Navigation app	User can ask directions to a certain place
10	Contact app	User can ask ALFRED to make a call to a contact
11	Help call	User can ask for help

12	Meeting app	User can organize a meeting with friends
13	Group app	User can set up a group discussion
14	Health monitor app	User can ask what his vitals are.

The Personal Assistant and the 14 apps were previously installed by the researchers on the Smartphone Google Nexus 5X from the ALFREDO marketplace. For each of these apps a set of commands was available in the pilot country languages. The TP's started the tests directly in the ALFRED Personal Assistant as shown in Figure 1. So they did not have to launch ALFRED on the smartphone.

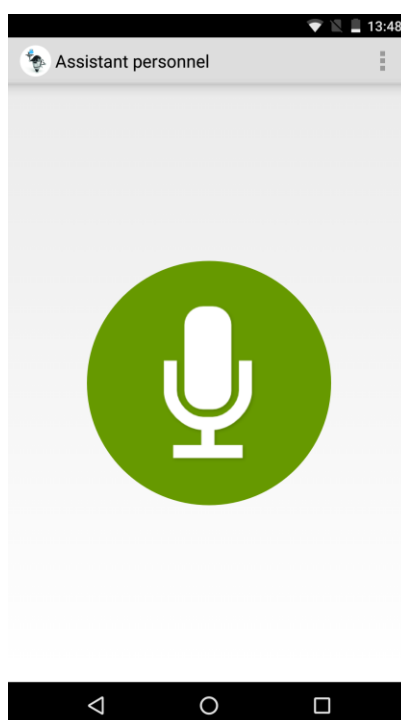


Figure 1: ALFRED Personal Assistant

As the ALFREDO Marketplace, Serious Games and ALFRED T-shirt were already extensively tested in previous evaluation cycles (see section 2) and full integration with voice interaction was still missing, these functionalities were not included in Evaluation Cycles 4 and 5, nor the pilot (see D8.4).

Unfortunately not all apps were fully functional at the time of the Evaluation Cycle 4. Due to technical problems at the time of the Evaluation Cycle 4 in May, it was decided to implement a final and Fifth Evaluation Cycle on an improved Integrated ALFRED prototype. This Evaluation Cycle 5 was implemented as a session with the TPs of the pilot in June. The same prototype was used by the TP's during the pilot independently at home.

3.2 Methodology

3.2.1 Overall methodology

The Evaluation Cycles 4 and 5 have been implemented in order to test the integrated ALFRED prototypes and define the success rate of the tasks performed by TPs. The methodology that was implemented is similar to the methodology for the other Evaluation Cycles with some specifics for these two cycles.

TPs were recruited from the Primary Target Group (See D2.3) through national projects and services of the Dutch and French end user organizations.

As in the previous sessions, the users performed the tests in a controlled environment. The users received a list of standardized tasks in a workbook (see Annex 5). As in the previous sessions the users filled in an intake questionnaire on demographics. They performed the tasks, where the researcher observed the voice interaction with the ALFRED system without interfering. After performing each task, the users commented on the interaction and their findings (replay-the-test, see D8.1). Finally the TP filled in the PSSUQ questionnaire on the overall system usability.

3.2.2 Success Rate

In addition to the previously mentioned tools, these two cycles also included the definition of a success rate. After each task the researcher indicated a success rate on the performance. A success rate is a simple way to indicate the usability of a system in numbers. It is the percentage of tasks that users complete correctly⁶.

There are three rates that can be given to a task. First of all a task is either completed and given a 'success' (S) or a task cannot be completed and is given a 'failure' (F).

In two specific occasions a Partial Success (P) was given to the tasks:

- When someone finished the task with some specific help from the Researchers.
- As the prototype was not always working correctly and would get stuck, the Researchers sometimes had to reset the system, close all apps and start again. As these were technical issues and not so much part of the interaction itself, it was decided to rate these tasks with a P, once the task was completed.
- When someone finishes the task after trying a few times (max. 3).
- As the ALFRED system works with voice interaction, it was sometimes complicated for users to find the initial command to set off an app and start the interaction. It was decided that users can try out several commands and still have a partial success if they find the right command and complete the task.

These simplified success rates provide a general picture of how the ALFRED system and different apps are working. As we are working with a small number of observations and rough estimate of partial success scores they do not give a complete picture, but an indication on the improvement needed to make the ALFRED system work properly. According to some literature a success rate of less than 50% is not uncommon for website scores when using the first time [2001, Nielsen].

⁶ <https://www.nggroup.com/articles/success-rate-the-simplest-usability-metric/>

D8.2.2 Piloting & Validation I: Individual Usability	Document Version: 1.0	Date: 2016-09-23	Status: For Approval	Page: 25 / 43
http://www.alfred.eu/		Copyright © ALFRED Project Consortium. All Rights Reserved. Grant Agreement No.: 611218		

3.2.3 Analysis

The data that was gathered during the Evaluation Cycles 4 and 5 consisted of the demographic information of the TPs, Task Rate Scores and PSSUQ scores that will be presented in the following sections.

Additionally there were many observational remarks from researchers and end users themselves (replay-the-test). Instead of reporting the observational remarks into a spreadsheet with findings and recommendations as was done in previous cycles, it was decided to directly report the findings as issues into Gitlab. Due to the short period for improvement and technical updates and the many issues that rose, it was considered a more effective way of collaborating between the end user and technical partners. This way the responsible partners for the different apps could directly resolve technical and usability issues. A total of 135 issues were created with an indication on priority (Low, Medium, High) for improvement (see Figure 2).

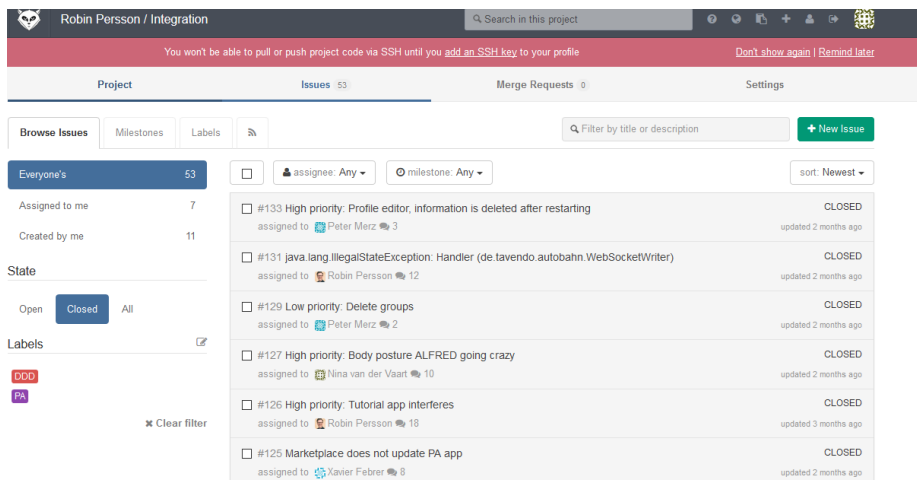


Figure 2: Gitlab Findings and Recommendations

3.3 Iterative Evaluation Cycle 4

3.3.1 Summary

The Iterative Evaluation Cycle 4 was performed in the Netherlands and in France in May 2016 on the first integrated ALFRED prototype. As described in section 3.1 the evaluation entailed the User-Driven Interaction Assistant with 14 apps from the four ALFRED pillars that were already installed and presented on the Google Nexus 5x phone.

The TP’s received the workbook and a set of 14 tasks. Examples of tasks are: Please ask ALFRED to introduce himself. Please set a reminder to take your medication. Please check on your battery, etc. For each app a task was defined. It must be noted that during the Iterative Evaluation Cycle 4 different technical difficulties rose in the integration of the different components and that some apps were not working properly at the time of the tests.

3.3.2 Participants

A total of 5 Dutch and 5 French TP's were involved in this Iterative Evaluation Cycle (see Annex 3). Only 1 male and 9 were female. The average age of all participants was 73.7 years old. The average age of the French participants was 69 and of the Dutch participants was 78,4 (see Table 7)

Table 7: TPs Evaluation Cycle 4

	Netherlands (NFE)	France (ESE)	Total
Number of participants	5	5	10
Number of Female and Male participants	4(F), 1(M)	5 (F)	9(F), 1(M)
Average age of participants	78.4	69	73.7

Six TP's use a normal mobile phone. Two persons use an Apple smartphone and three persons use an Android mobile phone. One person indicated to have a negative position towards new technology, three were neutral and the rest considered themselves to be positive. The experience in the use of technology varied very much per user as can be seen in **Fout! Verwijzingsbron niet gevonden..**

Table 8: TP Experience in Use of Technology

	Netherlands (NFE)	France (ESE)	Total
Very high	0	3	3
High	1	0	1
Medium	3	0	3
Low	0	2	2
Very low	1	0	1

As in previous cycles we see that most of the users make daily or weekly use of different ICT tools, such as e-mailing, web searching and playing games. Only one Dutch user indicated that she never uses any ICT solution. In terms of educational level, we see that in France 4 TP's are higher educated and one TP has a lower education. In the Netherlands 3 persons are intermediate educated, one is higher educated and one is lower educated.

3.3.3 Evaluation Tasks and Results

3.3.3.1 Findings and recommendations

The findings and recommendations for improvement with a priority between low, medium and high were reported in Gitlab. The findings ranged from changes in the sequence of commands, additional commands that were required for a more intuitive interaction, adjustments in the pace of the interaction, etc. One of the main points in the interface is that each app added a menu button, so that users could go back to the main menu in each moment. Additionally it was suggested to lower the pace of the interaction with ALFRED, making it easier to use for especially the older end users.

3.3.3.2 PSSUQ ratings

The testers filled in the PSSUQ rating after finishing all the tasks at the very end of the session. If we look first at the overall satisfaction of the user with the system (average of 1 to 19) we see an overall satisfaction rate that is considerably lower (so with a higher score from 1 to 7) than in any of the other PSSUQ ratings up to this moment ($M=4.49$, $SD=1.7$).

If we then divide this overall satisfaction rate into the corresponding subsections, we see that the system usefulness is rated highest by the TPs compared to the other subsections ($M=3.63$, $SD=1.63$), although still quite low it is still mildly positive (above 4 is positive, below 4 is negative). The quality of the information provided by the system has been rated lowest of all subsections ($M=4.69$, $SD=1.64$). This corresponds also to the previous PSSUQ ratings in the iterative evaluations, where the quality of the information always received the lowest rating. Finally the interface quality of the system and its capabilities were rated also slightly positive ($M=3.70$, $SD=1.60$). The following table summarizes these results.

Table 9: PSSUQ ratings

PSSUQ Ratings (1=high satisfaction, 7=low satisfaction)	
Parameters	M (mean)
Overall satisfaction (Q1-Q19)	4,49
System quality (Q1-Q8)	3,63
Information quality (Q9-Q15)	4.69
Interface quality (Q16-Q19)	3.70

These negative results of the PSSUQ can be directly linked to the low success rates of the task performance. In overall there were still too many technical issues and inconsistencies in the use of this ALFRED prototype to make a smooth interaction possible. The very low rate on the information quality can also be linked to the fact that voice interaction provides for relatively little information related to other interaction modalities. Due to these low ratings and the struggles on the technical side, it was decided to implement a final and fifth User Evaluation Cycle on a further improved integrated ALFRED prototype.

3.4 Iterative Evaluations Cycle 5

3.4.1 Summary

The Iterative Evaluation Cycle 5 was performed in the Netherlands and in France in June and July 2016 on the second integrated ALFRED prototype. This prototype was similar as the prototype tested during Iterative Evaluation Cycle 4, with technical and usability improvements implemented, as reflected in Gitlab, in a short cycle. Additional commands were added to make the interaction more natural and easier for the end users.

The Cycle was implemented as a session previous to the pilot with the same TP's from the pilot. The pilot results will be reported in D8.4. This section will highlight the results for the usability session previous to the pilot.

The TP's received the same workbook and tasks as in the previous Cycle 4, making the results highly comparable. The same methodology was applied as in the previous session.

3.4.2 Participants

A total of 21 TPs participated in Iterative Cycle 5, 11 Dutch and 10 French (see Annex 7). Four were male and 16 were female. The average age of the TP's was 77,5 years old. In the Netherlands the TP's were in general a bit older with 81.2 than in France with an average age of 74 (see **Fout! Verwijzingsbron niet gevonden.**). As mentioned the TPs are the same as in the pilot, with the exception of one Dutch TP who could not be scheduled for the previous session to the pilot.

Table 10: TPs Evaluation Cycle

	Netherlands (NFE)	France (ESE)	Total
Number of participants	11	10	21
Number of Female and Male participants	9 (F), 2(M)	8 (F), 2 (M)	17(F), 4(M)
Average age of participants	81.2	74	77.5

In France all TP's were higher educated (Higher Secondary Education 1, and University, 9). In the Netherlands the education level was more evenly divided with 1 higher educated TP, 7 TP's at intermediate education level and 3 were lower educated.

Almost all TP's, except for two in the Netherlands indicate that they online activities, such as e-mailing or searching on the internet.

gives a short overview on the technological experience indicated by themselves.

Table 11: TP Experience in Use of Technology

	Netherlands (NFE)	France (ESE)	Total
Very high	1	6	7
High	5	3	8
Medium	3	0	3
Low	2	0	2
Very low	1	1	1

The use of phone devices varied a great deal among the participants. A total of 11 persons has experience with a smartphone (either Apple or Android) and 5 persons didn't have any experience with a mobile device. **Fout! Verwijzingsbron niet gevonden.** is a short overview of the types of smart phone devices in use.

Table 12: Use of Mobile Device

	Netherlands (NFE)	France (ESE)	Total
No mobile device	4	1	5
Normal mobile phone	2	2	4
Smart phone (apple)	3	4	7
Smart phone (Android)	2	2	4
Senior phone	1	0	1

3.4.3 Drop out

During the Evaluation Sessions two TP's dropped out of the research. Although they had been informed extensively on the research and signed the Informed Consent, they decided to stop halfway between the tasks. They did not feel like using ALFRED independently at home for two weeks and felt a bit overwhelmed. Besides that they dropped out of the pilot, this meant that they dropped out of the Evaluation session.

3.4.4 Evaluation Tasks and Results

3.4.4.1 Findings and Recommendations

Also during this evaluation cycle the findings and recommendations were reported through Gitlab with the intention to make further improvements after the pilot possible.

3.4.4.2 PSSUQ Rating

When looked at the PSSUQ rating for Iterative Cycle 5, an overall satisfaction rate of ($M=3.96$, $SD=1.86$) can be seen. Although slightly tipping over to the positive side (4 is neutral) it does not seem to be totally in line with the success rates previously presented. Considering the much higher success rate, one would expect also a considerably higher level of satisfaction. This rating in Cycle 5 is only slightly higher (so a lower number) than in the previous Iterative Cycle 4. In **Fout! Verwijzingsbron niet gevonden.** Cycle 4 and Cycle 5 are shown to easily compare the results between the two sessions.

The difference between the improvement in the success rate and the less improved PSSUQ is probably explained by the fact that there were still some technical difficulties in the prototype that could easily be solved by the researcher during the session and did not affect the success rates, but gave the TP a negative impression on the quality of the prototype.

Table 13: PSSUQ Ratings Cycle 4 and 5

PSSUQ Ratings (1=high satisfaction, 7=low satisfaction)		
Parameters	M (mean) Cycle 4	M (mean) Cycle 5
Overall satisfaction (Q1-Q19)	4,49	3.96
System quality (Q1-Q8)	3,63	3.93
Information quality (Q9-Q15)	4.69	4.04
Interface quality (Q16-Q19)	3.70	3.91

The usefulness of the system in Cycle 5 is still mildly positive ($M=3.93$, $SD=1.9$), but compared to Cycle 4 slightly lower (higher number). The information quality was rated much better in Cycle 5 ($M=4.04$, $SD 1,86$) although still slightly negative and the Interface quality changes only little ($M=3.91$, $SD=1.77$) to neutral.

4 Results from the Usability Perspective

The aim of this chapter is to draw some parallels throughout the usability perspective as a whole, taking into account all iterative evaluation cycles.

4.1 Overall results on the Iterative Cycles

The aim of the Iterative Cycles is to identify user demands, needs and behaviours of potential users at an early stage of the project so that they can be taken up and provided back to the technical partners. The research establishes and validates user performance measures and identifies potential concerns to be addressed in order to improve efficiency, productivity and end-user satisfaction [D8.1 Piloting Definitions].

Finally a total of 5 iterative cycles have been implemented. In each cycle different aspects of the ALFRED system have been tested (see Figure 3).

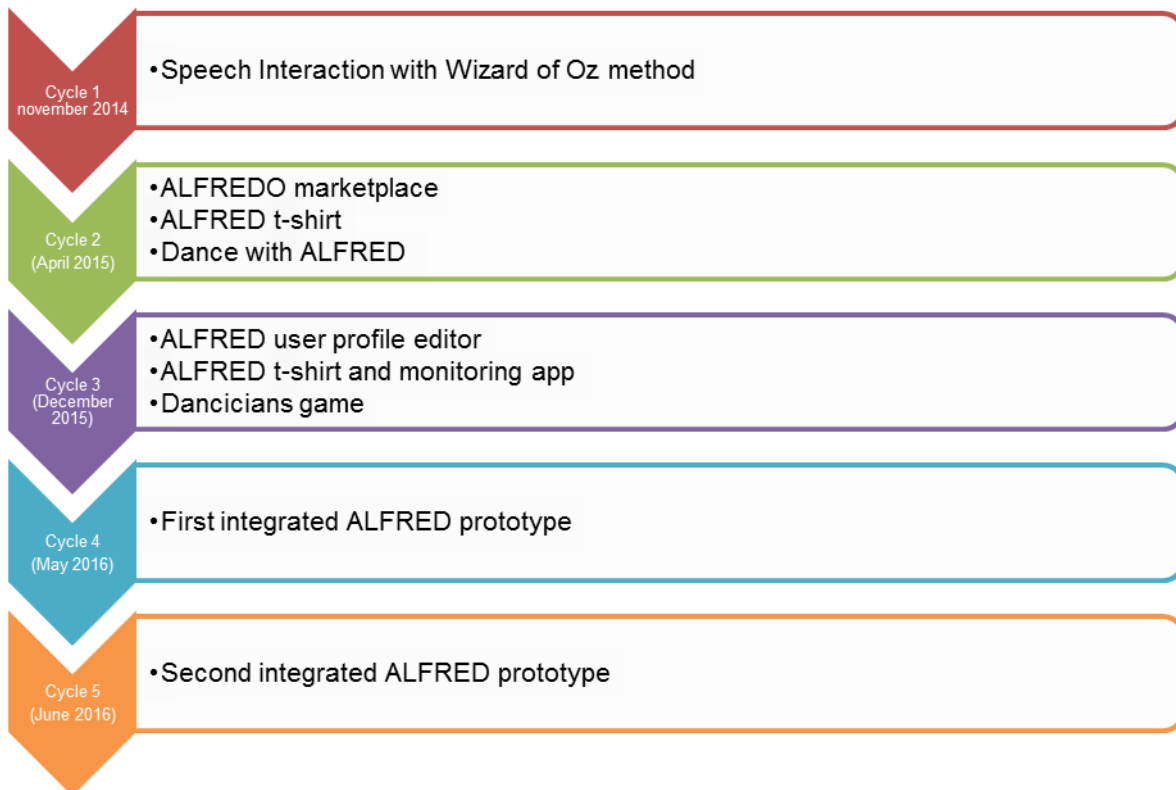


Figure 3: Iterative Evaluation Cycles

Each cycle was very different in terms of the tested prototypes, ranging from prototypes of apps with a GUI, a t-shirt, up to an integrated app driven by voice command. Although each cycle was very different in terms of the tested prototypes, this section will try to give some overall conclusions that can be made throughout the different Iterative Evaluation Cycles.

4.1.1 PSSUQ Rating

Although each cycle was very different in terms of the tested prototypes, the PSSUQ questionnaire was used in all cycles to define the satisfaction of the users. Due to the

D8.2.2 Piloting & Validation I: Individual Usability	Document Version: 1.0	Date: 2016-09-23	Status: For Approval	Page: 32 / 43
http://www.alfred.eu/		Copyright © ALFRED Project Consortium. All Rights Reserved. Grant Agreement No.: 611218		

varying nature of the tests, the results of the PSSUQ are not so much comparable. However when we put all the PSSUQ results in line, it provides a nice summary of the different Evaluation Cycles (see Figure 4). Also here it is important to take into account that 1 is the highest level of satisfaction and 7 the lowest level of satisfaction. If we consider this, we can see that the first session was evaluated very positive. This is logical if we consider that in Iterative Cycle 1 the tests took place in a Wizard of Oz setting, where voice interaction was simulated and users could fairly easy complete the tasks. These sessions gave important insight in the use of voice interaction.

After this a very small decline in level of satisfaction can be seen in Iterative Evaluation Cycle 2. This was expected as here actual working prototypes were tested. Between Cycle 2 and 3 one would have expected a slight improvement, as some apps were tested for the second time. Although there is a slight improvement visible in the Information Quality, the overall score is a bit lower. The difference between Cycle 3 and Cycle 4 is explained by the fact that Cycle 4 consisted of an integrated version of ALFRED, where still work was needed on further improvement. Based on the experience between Cycle 2 and 3, the end user partners decided to improve the uptake of feedback from end user partners to technical partners by working directly in Gitlab. This approach seemed to work better if we look at the improvement between Cycle 4 and 5, where continuous feedback could be given and the link between feedback and uptake was much faster. Also considering the short time frame this is can be considered as good progress on the integrated ALFRED system from a user perspective.

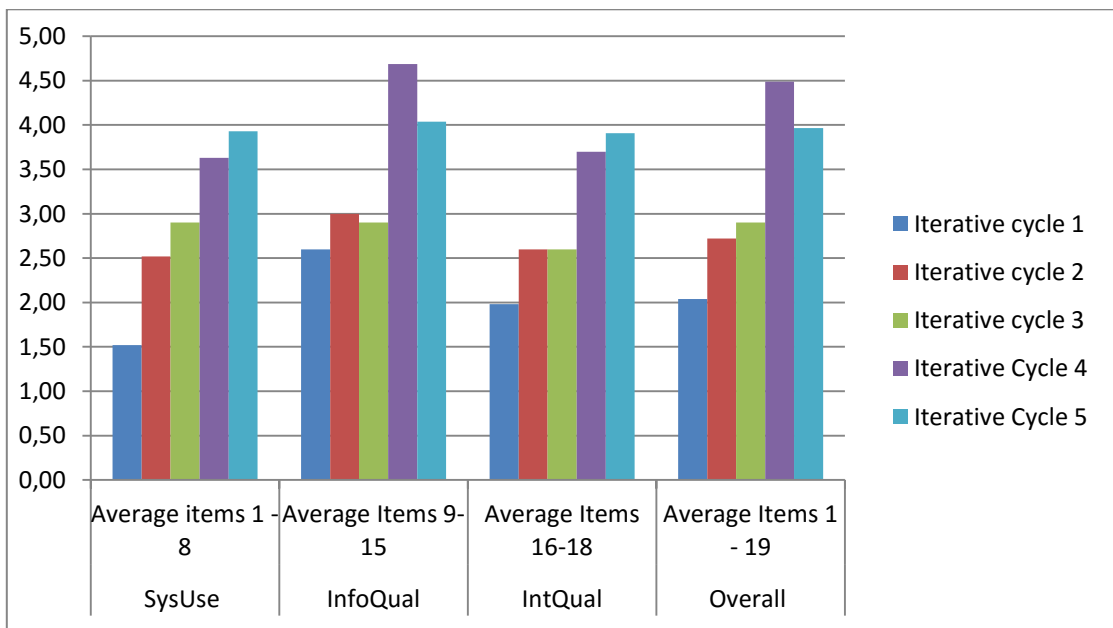


Figure 4: Overall PSSUQ Ratings

4.2 KPI's of the Usability Perspective

Finally this section goes into the KPIs of the Usability Perspective. The first two KPI's were defined as part of the DoW. The second two KPI's were additionally defined in D8.1.2 and adapted in D8.2.2 to reflect better the realities of the pilot and the methods of measurement (see Table 14).

Table 14: KPIs of the Usability Perspective

	Concept	KPI	Strategic Goal
1	Obtain high end user acceptance by a fully voice controlled and non-technical environment	Positive feedback in the validation groups after the last prototype of ALFRED	More than 75% rated positively.
2	Obtain high end user acceptance by a fully voice controlled and non-technical environment	Improvement of the positive rating in each feedback loop.	Overall improvement, throughout the cycles.
3	Adaptation of the needs of older end users and uptake of end user input.	Rise of PSSUQ satisfaction levels towards the final usability cycle.	Increase of satisfaction levels (excluding the WoZ tests).
4	Ease of use of the final ALFRED system.	Success rate based on the failure, partial successes and successes.	A success rate of 80%

4.2.1 KPI 1: Positive Feedback in the Validation Groups

The aim of this KPI was to make sure that the ALFRED solution would meet a high end user acceptance, using fully voice controlled environments. To this end the Iterative Evaluations were implemented already on an early stage of the project from M12. As seen in section 3.4.4.2 on the Iterative Evaluation Cycle 5, the level of satisfaction on the PSSUQ rating was ($M=3.96$, $SD=1.86$). Taking this point for measurement of the strategic goal of the KPI it can be concluded that the feedback of the last prototype was slightly positive. However, considering the average rating of each participant individually and look whether their overall satisfaction was positive or negative (where positive is below 4 and negative is above 4), it is revealed that 58% of the users rated the services positive. This would mean that the strategic goal of 75% has not been reached. As mentioned previously, it is considered that this is mainly due to the fact that there were still some technical problems during the Evaluations that gave a less positive impression despite the high success rates.

4.2.2 KPI 2: Improvement of Positive Rating in Each Feedback Loop

Also this KPI is meant to give an indication on the end user acceptance. Section 4.1.1 already explained with Figure 4, that between Evaluation Cycle 4 and 5 an improvement took place in the satisfaction rates. If we take the Success rates as a starting point for this KPI, it indicates that between Cycle 4 and Cycle 5 a considerably improvement took place. The conclusion must be made that his Strategic Goal has partly met.

4.2.3 KPI 3: Rise of PSSUQ Satisfaction Levels

This KPI is meant to give an indication on whether the end user needs were implemented. As mentioned in section 4.1.1 it is difficult to use the PSSUQ rates as a measurement for

this, as there were many different prototypes tested. Considering the increase between Evaluation Cycle 2 and 3, the Strategic Goal has not been met. However, as communication and feedback implementation was improved, thereby the Strategic Goals has been met between Cycle 4 and 5.

4.2.4 KPI 4: Success Rates

This KPI gives an indication on the ease of use of the final system. The final success rate after Evaluation Cycle 5 was 72.9% (see section **Fout! Verwijzingsbron niet gevonden.**). The Strategic Goal of 80% was therefore not met in the pilot. It must be said that a success of 80% is very ambitious (see section 3.2.2) and a success rate of 72.9% can be considered quite satisfactory for this specific target group.

D8.2.2 Piloting & Validation I: Individual Usability	Document Version: 1.0	Date: 2016-09-23	Status: For Approval	Page: 35 / 43
http://www.alfred.eu/	Copyright © ALFRED Project Consortium. All Rights Reserved. Grant Agreement No.: 611218			

5 Conclusion

The end-user partners and their research work with the end users are essential for the viability of the developed ALFRED solution. End-users were continuously involved to validate the usability and the usefulness of the system. A total of 5 Iterative Evaluation Cycles were implemented in the project, where different prototypes and parts of the system were tested with the involvement of the target groups. The last two Cycles focused on the usability and satisfaction of the users with an integrated prototype of the ALFRED system. In these sessions TP's fulfilled set of tasks in a controlled environment, where different methods were implemented to analyse the results of the tasks. The results of the sessions were communicated back to the technical partners to implement improvements to adjust to the end user needs.

This early testing also helped the end-user partners to plan and define the ideal final validation (pilot) test settings, reported on in D8.4. The researchers from end user partners are more involved in the development process and have seen the earlier versions of the system components. The pilots have, as the technical system, been developed in an iterative process adjusting to the reality of each moment.

D8.2.2 Piloting & Validation I: Individual Usability	Document Version: 1.0	Date: 2016-09-23	Status: For Approval	Page: 36 / 43
http://www.alfred.eu/	Copyright © ALFRED Project Consortium. All Rights Reserved. Grant Agreement No.: 611218			

Annex 1: Overview Test Participants Cycle 2

TP	Year of Birth	Gender	Nationality	Household income	Marital status	Living situation	Education	Employment	Residential situation	Self rated health
				0 = no comment 1 = 500-1000 € 2 = 1000-2000 € 3 = 2000-3000 € 4 = > 3000 €	1 = Single 2 = Married 3 = Divorced 4 = Widowed	1 = Living alone 2 = Living with Children 3 = Living with partner, no children 4 = Living with partner and children	1 = Primary education 2 = Lower vocational education 3 = Intermediate vocational education 4 = Intermediate secondary education 5 = Higher secondary education 6 = Higher vocational education 7 = University	1 = Retired 2 = Employed 3 = Unemployed 4 = Voluntary work 5 = Other	1 = Living independently at home 2 = Independent planned housing, congregate housing 3 = Home for older people	1 = Poor 2 = Not so good 3 = Fair 4 = Good 5 = Very good
1	1943	F	DEU	3	2	3	7	1	1	4
2	1943	F	DEU	4	2	3	7	1	1	4
3	1940	F	DEU	4	2	3	5	1	1	4
4	1935	M	NLD	3	2	3	3	1	1	4
5		F	NLD	1	1	1	2	4	1	4
6	1952	F	NLD	0	2	3	2	5	1	5
7	1947	F	NLD	3	2	1	3	1 & 4	1	4
8	1941	F	DEU	0	4	1	3	1 & 4	1	4
9	1932	M	DEU	0	4	1	3	5	1	2
10	1941	M	ITA	0	4	1	5	5	1	4
11	1941	F	FRA	0	2	3	7	5	1	4
12	1952	F	FRA	0	1	3	7	4	1	4

TP	Vision	Hearing	Motor control A	Motor control B	Self rated mobility level	Technological Experience	Attitude on Technology	Use of ICT tools	Type of phone
	1 = Yes, without difficulty 2 = Yes, with minor difficulty 3 = Yes, with major difficulty 4 = No, not able to	1 = Yes, without difficulty 2 = Yes, with minor difficulty 3 = Yes, with major difficulty 4 = No, not able to	1 = Yes, without difficulty 2 = Yes, with minor difficulty 3 = Yes, with major difficulty 4 = No, not able to	1 = Yes, without difficulty 2 = Yes, with minor difficulty 3 = Yes, but I need some help 4 = No, I'm not able to	1 = Poor 2 = Not so good 3 = Fair 4 = Good 5 = Very good	1 = Very high 2 = High 3 = Medium 4 = Low 5 = Very low	1 = Positive 2 = Neutral 3 = Negative	1 = Daily 2 = Weekly 3 = Monthly 4 = Rarely 5 = Never	1 = Normal mobile phone 2 = Senior mobile phone 3 = Smartphone (Android) 4 = Smartphone (Apple) 5 = Other
1	1	1	1	1	5	3	2	2	1
2	1	2	1	1	4	4	2	4	1
3	1	2	2	1	5	4	2	4	1
4	1	1	1	1	4	4	1	2	1
5	1	1	1	1	4	3	1	1	1
6	1	2	1	1	4	4	1	2	2
7	1	1	1	1	4	1	1	1	3
8	1	1	1	1	4	3	1	1	1 & 3
9	1	1	1	1	4	3	2	2	1
10	1	2	2	1	4	2	1	1	3
11	1	1	1	1	4	2	1	1	1
12	1	1	1	1	5	2	1	1	3

Annex 2: Overview Test Participants Cycle 3

TP	Year of Birth	Gender	Nationality	Household income	Marital status	Living situation	Education	Employment	Residential situation	Self rated health
				0 = no comment 1 = 500-1000 € 2 = 1000-2000 € 3 = 2000-3000 € 4 = > 3000 €	1 = Single 2 = Married 3 = Divorced 4 = Widowed	1 = Living alone 2 = Living with Children 3 = Living with partner, no children 4 = Living with partner and children	1 = Primary education 2 = Lower vocational education 3 = Intermediate vocational education 4 = Intermediate secondary education 5 = Higher secondary education 6 = Higher vocational education 7 = University	1 = Retired 2 = Employed 3 = Unemployed 4 = Voluntary work 5 = Other	1 = Living independently at home 2 = Independent planned housing, congregate housing 3 = Home for older people	1 = Poor 2 = Not so good 3 = Fair 4 = Good 5 = Very good
1	1947	F	FRA	0	2	3	7	1	1	3
2	1941	M	ITA	0	4	1	5	1	1	4
3	1952	F	FRA	0	1	3	7	4	1	4
4	1943	F	DEU	3	2	3	7	1	1	3
5	1943	F	DEU	4	2	3	7	1	1	4
6	1940	F	DEU	4	2	3	5	1	1	4
7	1927	M	NLD	3	4	4	6	1	2	4
8	1943	M	NLD	3	2	4	1	1	1	4
9	1933	F	NLD	0	4	4	2	1	2	3
10	1944	F	NLD	0	2	4	3	1 & 4	1	3
11	2015	F	NLD	2	2	4	2	1	2	2
12	1933	F	NLD	0	4	4	5	1 & 4	1	3
13	1950	F	NLD	0	2	4	2	1 & 4	1	3
14	1937	F	NLD	0	4	4	2	2	1	5
15	1936	F	NLD	2	4	4	2	1	3	4
16	1939	F	NLD	0	4	4	5	1	2	4
17	1945	F	NLD	2	2	4	2	4	1	5
18	1953	F	NLD	2	4	1	5	2	1	4
19	1948	F	NLD	3	2	4	3	1	1	4
20	1950	F	NLD	0	2	4	2	1 & 4	1	4
21	1944	F	NLD	0	2	4	5	5	1	4
22	1926	F	NLD	2	4	4	1	1	1	4
23	1923	F	NLD	0	4	4	5	1	2	3
24	1942	F	NLD	1	4	4	2	1 & 4	2	3
25	1944	F	NLD	0	1	2	2	1	1	5

TP	Vision	Hearing	Motor control A	Motor control B	Self rated mobility level	Technological Experience	Attitude on Technology	Use of ICT tools	Type of phone	Involved in testcycle 1 & 2
	1 = Yes, without difficulty	1 = Yes, without difficulty	1 = Yes, without difficulty	1 = Yes, without difficulty	1 = Poor	1 = Very high	1 = Positive	1 = Daily	1 = Normal mobile phone	1 = yes
	2 = Yes, with minor difficulty	2 = Yes, with minor difficulty	2 = Yes, with minor difficulty	2 = Yes, with minor difficulty	2 = Not so good	2 = High	2 = Neutral	2 = Weekly	2 = Senior mobile phone	2 = no
	3 = Yes, with major difficulty	3 = Yes, with major difficulty	3 = Yes, with major difficulty	3 = Yes, but I need some help	3 = Fair	3 = Medium	3 = Negative	3 = Monthly	3 = Smartphone (Android)	
	4 = No, not able to	4 = No, not able to	4 = No, not able to	4 = No, I'm not able to	4 = Good	4 = Low		4 = Rarely	4 = Smartphone (Apple)	
					5 = Very good	5 = Very low		5 = Never	5 = Other	
1	1	1	1	2	5	1	1	1	3	2
2	1	2	2	1	4	2	1	1	3	1
3	1	1	1	1	5	2	1	1	3	1
4	1	1	1	1	5	3	2	2	1	1
5	1	2	1	1	4	4	2	4	1	1
6	1	2	2	1	5	4	2	4	1	1
7	2	2	1	1	3	2	1	1	5	2
8	1	1	1	1	4	2	1	1	1	2
9	1	1	2	1	3	5	2	4	1 & 5	2
10	1	1	1	1	3	3	2	1	1 & 5	2
11	1	1	2	1	2	2	1	1	2	2
12	1	1	1	1	4	3	1	2	1	2
13	2	1	2	1	4	2	3	1	1	2
14	1	1	1	1	4	4	2	1	1	2
15	1	2	1	1	3	5	2	5	1	2
16	1	1	2	1	3	3	2	2	1	2
17	1	2	1	1	5	1	1	1	3	2
18	1	1	1	1	4	3	1	1	1	2
19	1	1	1	1	4	1	1	1	4	2
20	1	2	1	1	5	2	1	1	1	2
21	1	1	1	1	5	2	1	1	1	2
22	1	1	1	1	4	3	1	1	1	2
23	1	2	1	1	3	5	2	5	5	2
24	2	1	1	1	4	2	1	1	1	2
25	1	1	1	1	5	3	3	2	1 & 5	2

Annex 3: Test Persons Evaluation Cycle 4

Participant ID	Birth year	Gender	Nationality	Household income	Marital status	Living situation	Education	Employment	Residential situation	Self rated health	Vision	Hearing	Motor control A	Motor control B	Self rated mobility level	Technological Experience	Attitude on Technology	Use of ICT tools	Type of phone	Online activities
User 1_NL	1933	F	NL	1000-2000	Single	Living alone	Higher Vocational Education	Retired	Living independently at home	Average	Yes without difficulty	Yes without difficulty	Yes without difficulty	Yes with minor difficulty	Fair	Medium	Positive	Daily use	Normal mobile phone	Email, Look up a receipt, Web search, look on a map for directions, news, playing games,
User 2_NL	1935	F	NL	1000-2000	Widow	Living alone	Primary education, lower vocational training	Retired	Living independently at home	Average	Yes without difficulty	Yes without difficulty	Yes without difficulty	Yes without difficulty	Fair	Very little	Neutral	Never	Normal mobile phone	-
User 3_NL	1932	M	NL	1000-2000	Single	Living alone	Intermediate vocational training	Retired	Living independently at home	Not so good	Yes, without difficulty	Yes with minor difficulty	Yes without difficulty	Yes without difficulty	Fair	Medium	Positive	Daily use	Smart phone (Apple)	Email, Chat, web search, playing games
User 4_NL	1947	F	NL	2000-3000	Widow	Living alone	Secondary education	Retired	Living independently at home	Good	Yes, without difficulty	Yes, without difficulty	Yes, without difficulty	Yes, without difficulty	Good	High	Positive	Daily use	Normal mobile phone / Smartphone (Android)	Email, Chat, web search, look for transport information, online banking, online shopping, playing games
User 5_NL	1941	F	NL	2000-3000	Widow	Living alone	Secondary domestic school	Volunteering	Living independently at home	Good	Yes, without difficulty	Yes, without difficulty	Yes, without difficulty	Yes, without difficulty	Good	Medium	Neutral	Daily use	Normal mobile phone / Smartphone (Android)	Chat, playing games
User 1_FA	1934	F	FR		Married	Living with partner and children	University	Retired	Living independently at home	Good	Yes with minor difficulty	Yes without difficulty	Yes without difficulty	Yes without difficulty	Very good	Low	Positive	Daily use	Normal mobile phone	Email, Web search, look for information on a hobby or interests, look on a map for directions, news, weather
User 2_FA	1938	F	FR		Single	Living alone	Primary education, lower vocational training	Retired	Living independently at home	Very good	Yes without difficulty	Yes without difficulty	Yes with minor difficulty	Yes without difficulty	Very good	Low	Negative	Weekly use	Normal mobile phone	Email, Look up a receipt, Web search, look for health/medical information, look for information on a hobby or interests, weather, online banking, playing games, visit a local, state or federal government website
User 3_FA	1959	F	FR		Married	Living alone	University	Unemployed	Living independently at home	Very good	Yes without difficulty	Yes without difficulty	Yes without difficulty	Yes without difficulty	Very good	Very high	Neutral	Daily use	Smartphone (Apple)	Email, Chat, Web search, look for health/medical information, look for information on a hobby or interests, look for transport information, look on a map for directions, job search, visit a local, state or federal government
User 4_FA	1958	F	FR		Married	Living with partner and children	University	Employed	Living independently at home	Very good	Yes without difficulty	Yes without difficulty	Yes without difficulty	Yes without difficulty	Very good	Very high	Positive	Daily use	Smartphone Android	Email, Chat, Web search, look for health/medical information, look for information on a hobby or interests, look for transport information, news, weather, sports, online banking
User 5_FA	1946	F	FR		Single	Living alone	University	Retired	Independent planned housing	Very good	Yes without difficulty	Yes without difficulty	Yes without difficulty	Yes without difficulty	Very good	Very high	Positive	Daily use	Smartphone Android	Email, Chat, Look up a receipt, Web search, look for transport information, look on a map for directions, news, weather, online banking

Annex 4 Test Persons Evaluation Cycle 5

Participant ID	Birth Year	Gender	Nationality	Household Income	Marital status	Living situation	Education	Employment	Residential location	Self-rated health	Vision	Hearing	Motor control A	Motor control B	Self-rated mobility (self-rated)	Technical proficiency	Attitude on technology	Use of ICT tools	Type of phone	Help needed (scale 0-5)	Online activities
NL_001	1938 F	Female	NL	3000	Widowed	Living alone	Intermediate vocational	Retired	Living independently at home	Good	Yes, without difficulty	Yes, without difficulty	Yes, without difficulty	Yes, without difficulty	Good	High	Positive	Daily use	No mobile phone	Rarely	
NL_002	1952 M	Male	NL	€1,000	Widowed	Living alone	Primary education	Retired	Independent planned housing, congregating	Not so good	Yes, without difficulty	Yes, with minor difficulty	Yes, without difficulty	Yes, without difficulty	Fair	Medium	Positive	Daily use	Smart phone (Apple)	Rarely	Email, Chat, web search, news, weather, online banking, visit a local, state or federal government website
NL_003	1931 F	Female	NL		Widowed	Living alone	Primary education	Retired	Independent housing, congregating	Good	Yes, without difficulty	Yes, with major difficulty	Yes, without difficulty	Yes, without difficulty	Fair	Low	Neutral	Monthly use	Normal mobile phone (with big buttons)	Weekly	only to call
NL_004	1929 F	Female	NL		widowed	living alone	lower vocational	Retired	Independent living at home	fair	Yes, without difficulty	yes, without difficulty	Yes, without difficulty	Yes, without difficulty	Good	Medium	Positive	Daily use	Normal mobile phone	Rarely	Email, web search, news, weather, playing games, visit a local, state or federal government website
NL_005	1939 F	Female	NL		Divorced	living alone	Intermediate vocational training	Retired	Home for older people	fair	Yes, without difficulty	yes, without difficulty	Yes, without difficulty	Yes, with minor difficulty	Not so good	High	Positive	Daily use	Smart phone (Apple)	Rarely	Chat, web search, news, weather, online banking, shopping, playing games
NL_006	1940 F	Female	NL		Single	Living alone	Intermediate vocational training	Retired	Living independently	fair	Yes, without difficulty	Yes, without difficulty	Yes, without difficulty	Yes, without difficulty	Not so good	High	Neutral	Daily use	Smart phone (Android)	Rarely	web search, look for transport information, look for a map for directions, news, sports, online banking, shopping, playing games, visit a local, state or federal government website
NL_007	1944 F	Female	NL		Married	Living with partner	Intermediate vocational	Volunteering	Living independently at home	Good	Yes, without difficulty	Yes, without difficulty	Yes, without difficulty	Yes, without difficulty	Good	High	Positive	Daily use	Smart phone (Android)	Rarely	web search, look for transport information, playing games
NL_008	1935 F	Female	NL		Widowed	Living alone	second degree	Volunteering	Living independently at home	Very good	Yes, without difficulty	Yes, without difficulty	Yes, without difficulty	Yes, without difficulty	Good	High	Positive	Daily use	Smart phone (Apple)	Rarely	Playing games, Facebook
NL_009	1927 M	Male	NL		Widowed	Living alone	Higher vocational education	Retired	Independent planned housing, congregating	fair	Yes, with minor difficulty	Yes, with minor difficulty	Yes, without difficulty	Yes, with minor difficulty	Fair	High	Positive	Daily use	Normal mobile phone	weekly	Email, web search, look for health/medical information, look for an information on a hobby or interests, look on a map for directions, news, weather, playing games
NL_010	1933 F	Female	NL		Single	Living alone	school	second job	Independent housing, congregating	fair	Yes, without difficulty	Yes, with minor difficulty	Yes, without difficulty	Yes, with minor difficulty	Not so good	low	Positive	Daily use	No mobile phone	weekly	

User 10 RR	1391 M	RR		Divorced	Living alone	Living with university and colleague	Retired independent living	Very good	Yes, with minor difficulties	Yes, with minor difficulties	Yes, without difficulty	Yes, without difficulty	Yes, without difficulty	Yes, but I need some help, for example with things my services provide on my tasks	Very good	Very high	Positive	Daily use	Smart phone (Android)	Never	Web search, look for health/medical information, look on a map for directions, news, online banking, playing games
User 9 RR	1338 F	RR		Single	Living alone	Higher secondary education	Living independent living at home	Good	Yes, without difficulty	Yes, without difficulty	Yes, without difficulty	Yes, without difficulty	Yes, without difficulty		Very good	Low	Neutral	Weekly use	Normal mobile phone	Never	
User 8 RR	1366 F	RR/Scoti		Divorced	Living alone	Unvers higher education	Living independent living at home	Very good	Yes, without difficulty	Yes, without difficulty	Yes, without difficulty	Yes, without difficulty	Yes, without difficulty		Very good	Very high	Positive	Daily use	Smartphone (Apple)	Rarely	Email, look up a recipe, web search, look for an information on a hobby or interests, look for transport information, look on a map for directions, news, weather, online banking, games
User 7 RR	1366 F	RR		Single	Living alone	Unvers higher education	Living independent living at home	Very good	Yes, without difficulty	Yes, without difficulty	Yes, without difficulty	Yes, without difficulty	Yes, without difficulty		Good	Very high	Positive	Daily use	Smartphone (Android)	Never	Email, chat, web search, look for an information on a hobby or interests, look for transport information, look on a map for directions, news, weather, games, social networks
User 6 RR	1344 F	RR/US		Single	Living alone	Unvers higher education	Living independent living at home	Good	Yes, without difficulty	Yes, without difficulty	Yes, without difficulty	Yes, without difficulty	Yes, without difficulty		Fair	High	Positive	Daily use	Smartphone (Apple and Android)	Once a month	Email, web search, look for an information on a hobby or interests, look for transport information, look on a map for directions, news, weather, online banking, visit a local, state of federal government website, and go about active aging.
User 5 RR	1347 F	RR		Single	Living alone	Unvers higher education	Living independent living at home	Very good	Yes, without difficulty	Yes, without difficulty	Yes, without difficulty	Yes, without difficulty	Yes, without difficulty		Very good	High	Positive	Daily use	Normal mobile phone	Rarely	Email, look up a recipe, web search, look for an information on a hobby or interests, look for transport information, look on a map for directions, news, weather, online banking, visit a local, state of federal government website
User 4 RR	1322 M	RR		Widowed	Living with children	Unvers higher education	Living independent living at home	Good	Yes, with minor difficulties	Yes, with minor difficulties	Yes, without difficulty	Yes, without difficulty	Yes, without difficulty		Fair	Very high	Positive	Daily use	Normal mobile phone	Rarely	Health/medical information, look for an information on a hobby or interests, look for transport information, look on a map for directions, news, weather, online banking, visit a local, state of federal government website
User 3 RR	1337 F	RR		Widowed	Living alone	Unvers higher education	Living independent living at home	Good	Yes, with minor difficulties	Yes, with minor difficulties	Yes, without difficulty	Yes, without difficulty	Yes, without difficulty		Good	High	Positive	Daily use	None but own iPad	Rarely	Health/medical information, look for an information on a hobby or interests, look for transport information, look on a map for directions, news, weather, online banking, visit a local, state of federal government website
User 2 RR	1347 F	RR		Divorced	Living alone	Unvers higher education	Living independent living at home	Very good	Yes, without difficulty	Yes, without difficulty	Yes, without difficulty	Yes, without difficulty	Yes, without difficulty		Very good	Very high	Positive	Daily use	Smartphone (Apple)	Rarely	Health/medical information, look for an information on a hobby or interests, look for transport information, look on a map for directions, news, weather, online banking, playing games, visit a local, state of federal government website
User 1 RR	1361 F	RR		Widowed	Living alone	Unvers higher education	Living independent living at home	Very good	Yes, without difficulty	Yes, without difficulty	Yes, without difficulty	Yes, without difficulty	Yes, without difficulty		Very good	Very high	Positive	Daily use	Smartphone (Apple)	Rarely	Health/medical information, look for an information on a hobby or interests, look for transport information, look on a map for directions, news, weather, online banking, playing games, visit a local, state of federal government website
NL 012	1355 F	NL		Widowed	Living alone	second school	Living independent living at home	Fair	Yes, without difficulty	Yes, without difficulty	Yes, without difficulty	Yes, without difficulty	Yes, without difficulty		Fair	Very low	Positive	Never	No mobile phone	Never	Email, look up a recipe, web search, look for an information on a hobby or interests, look for transport information, weather, playing games, visit a local, state of federal government website
NL 011	1331 F	NL		Widowed	Living alone	second school	Living independent living at home	Poor	Yes, with minor difficulties	Yes, without difficulty	Yes, without difficulty	Yes, with major difficulties	Yes, without difficulty		Poor	Medium	Positive	Weekly use	No mobile phone	Rarely	Web search, look for health/medical information, look on a map for directions, news, online banking, playing games

